Imperial College London

Department of Bioengineering

A Retinal Vasculature Tracking System Guided by A Deep Architecture

Author
Fatmatulzehra Uslu

Supervisor
Dr. Anil Anthony Bharath

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy in Bioengineering of Imperial College London
and the Diploma of Imperial College London

November 22, 2017

I hereby declare that the work submitted in this thesis is my own and all else is appropriately referenced.

'The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work'

Acknowledgement

Firstly, I would like to express my sincere gratitude to my supervisor, Dr. Anil Anthony Bharath, for being always supportive and encouraging over the course of my PhD journey. Being a part of his group gave me a chance to experience/observe different methods/approaches, which expanded my horizons regarding knowledge and how to do research. His attention to detail and his search for perfection inspired me to improve my technical skills and knowledge. In particular, I can not emphasise enough the benefits of his quick and useful feedback during the writing of my thesis.

I am utterly grateful to Prof. Dr. Kim Parker for his invaluable discussions regarding retinal vasculature segmentation in the early years of my PhD. These discussions helped me to further understand my research question.

I am also grateful to Dr. Claudia Clopath and Dr. Tobias Reichenbach for their invaluable feedback during my PhD examinations, Early Stage Assessment and Late Stage Review. It was a great opportunity to discuss my research with other researchers, which helped me to see my own work from different point of views.

My special thanks go to the members of my research group, Dr Jose Rivera-Rubio, Christoforos Charalambous, Kai Arulkumaran, Cher Bachar and Antonia Creswell for being always supportive and motivating.

I am much obliged to my sponsor, Republic Of Turkey Ministry Of National Education. Without their financial support, I would not be experiencing such a fulfilling research experience in Imperial College.

I would like to thank Prof. Dr. Mehmet Emin Yuksel for motivating me to apply for a PhD program in Imperial College.

Last but not the least, I would like to thank my family for being supportive at every stage of my research and the writing of this thesis. Without their encouragement, I would not progress in my academic life that much.

Abstract

Many diseases such as diabetic retinopathy (DR) and cardiovascular diseases show their early signs on retinal vasculature. Analysing the vasculature in fundus images may provide a tool for ophthalmologists to diagnose eye-related diseases and to monitor their progression. These analyses may also facilitate the discovery of new relations between changes on retinal vasculature and the existence or progression of related diseases or to validate present relations.

In this thesis, a data driven method, namely a Translational Deep Belief Net (a TDBN) [1], is adapted to vasculature segmentation. The segmentation performance of the TDBN on low resolution images was found to be comparable to that of the best-performing methods. Later, this network is used for the implementation of super-resolution for the segmentation of high resolution images. This approach provided an acceleration during segmentation, which relates to down-sampling ratio of an input fundus image. Finally, the TDBN is extended for the generation of probability maps for the existence of vessel parts, namely vessel interior, centreline, boundary and crossing/bifurcation patterns in centrelines. These probability maps are used to guide a probabilistic vasculature tracking system.

Although segmentation can provide vasculature existence in a fundus image, it does not give quantifiable measures for vasculature. The latter has more practical value in medical clinics. In the second half of the thesis, a retinal vasculature tracking system is presented. This system uses Particle Filters to describe vessel morphology and topology. Apart from previous studies, the guidance for tracking is provided with the combination of probability maps generated by the TDBN. The experiments on a publicly available dataset, REVIEW, showed that the consistency of vessel widths predicted by the proposed method was better than that obtained from observers. Moreover, very noisy and low contrast vessel boundaries, which were hardly identifiable to the naked eye, were accurately estimated by the proposed tracking system. Also, bifurcation/crossing locations during the course of tracking were detected almost completely. Considering these promising initial results, future work involves analysing the performance of the tracking system on automatic detection of complete vessel networks in fundus images.

Contents

1 Introduction **20**

1.1 Introduction . . . 20
1.2 Fundus Cameras and Colour Fundus Photos . . . 21
1.3 Retinal Fundus Image Analysis . . . 22
1.3.1 Computer-Assisted Diagnosis . . . 22
1.3.2 Biomarkers . . . 23
1.4 Challenges Associated With Retinal Vasculature Analysis . . . 24
1.4.1 Imaging . . . 24
1.4.2 Anatomy . . . 24
1.4.3 The Designs of Methods . . . 25
1.4.4 Availability and Reliability of Manually Labelled Datasets . . . 25
1.5 The Aim and The Scope of This Thesis . . . 26
1.6 Contributions . . . 27
1.6.1 Vasculature Segmentation . . . 27
1.6.2 Quantitative Vasculature Analysis . . . 27
1.7 The Organisation Of The Thesis . . . 28

2 Literature Review **29**

2.1 Introduction . . . 29
2.2 Publicly Available Datasets . . . 29
2.2.1 The DRIVE dataset [2] . . . 29
2.2.2 The STARE dataset [3] . . . 30
2.2.3 HRF dataset [4] . . . 31
2.2.4 CHASE_DB1 dataset [5] . . . 31
2.2.5 REVIEW dataset [6] . . . 32
HRIS dataset . . . 32
VDIS dataset . . . 33
CLRIS dataset . . . 33
KPIS dataset . . . 33

2.3 Retinal Vessel Segmentation Methods . . . 33
2.3.1 Supervised Methods . . . 34
Deep Learning Methods . . . 35
Other Supervised Methods . . . 38
2.3.2 Unsupervised Methods . . . 39
Shape Based Methods . . . 40
Intensity Based Methods . . . 41
Spatial Connectivity . . . 42
Both Shape and Intensity . . . 43
2.4 Quantitative Vasculature Analysis . . . 43
2.4.1 Model Based Methods . . . 43
2.4.2 Tracking Methods . . . 48
Deterministic Tracking . . . 48
Probabilistic Tracking . . . 49
2.4.3 Landmark location detection . . . 50

3 Retinal Vasculature Segmentation 53
3.1 Introduction . . . 53
3.1.1 Introduction to Deep Learning for Segmentation . . . 53
3.1.2 Segmentation: Cross-Modality vs Classification . . . 54
3.1.3 A Proposed Method for Retinal Vessel Segmentation . . . 55
3.1.4 The Remainder of this Chapter . . . 56
3.2 Autoencoders . . . 56
3.2.1 Autoencoders & Their Applications . . . 56
3.3 RBMs . . . 58
3.3.1 A Traditional RBM vs A Traditional Autoencoder . . . 58
3.3.2 The Traditional RBM from A Generative Learning Perspective . . . 59
3.3.3 Training RBMs . . . 60
3.3.4 Gibbs Sampling & Contrastive Divergence . . . 62
3.4 Denoising To Improve Hidden Representation . . . 63
3.5 Stacking Representations . . . 64
3.5.1 Stacking AEs : Deep Autoencoders . . . 64
3.5.2 Stacking RBMs: DBNs . . . 65
3.6 Cross Modality Learning . . . 66
3.6.1 Translational Deep Belief Networks . . . 66
3.6.2 Denoising in Cross-Modality Learning . . . 67
3.6.3 An Improved Denoising Strategy for CM Learning . . . 68
3.6.4 The Proposed Network For Vessel Segmentation . . . 69
3.7 Evaluating the Effects of Denoising . . . 71
3.7.1 Experiment 1: The Effects of Denoising on The Generated Features 73
3.7.2 Experiment 2: Evaluation of Generated Vessel Masks . . . 75

3.8 Conclusion 75

4 Enhanced TDBNs For Vasculature Segmentation 79

4.1 Introduction 79
4.2 Probability Maps from Enhanced TDBNs 80
4.3 Evaluation Criteria For Segmentation Performance 80
4.4 The Segmentation of Vasculature on Low Resolution Fundus Image Datasets 81
4.4.1 Transforming Patch Based Vessel Mask Generation To Vesselness Probability Maps of Fundus Images 82
4.4.2 Experimental Settings 83
Hyper-parameters of The Network 83
Data Preparation 84
4.4.3 Preliminary Results: Vessel Masks and Features 85
4.4.4 Segmentation Results on the DRIVE dataset 85
4.4.5 Segmentation Results on the STARE dataset 88
4.5 Super-resolution Segmentation With Enhanced TDBNs 90
4.6 The Segmentation of Vasculature in High Resolution Fundus Images 91
4.6.1 From Super-resolution Vessel Masks to Probability Maps 92
4.6.2 Experimental Settings 92
4.6.3 Preliminary Results : Vessel Masks and Features 93
4.6.4 Segmentation Results on the CHASE_ DB1 dataset 93
4.6.5 Segmentation Results on the HRF dataset 96
4.6.6 Evaluation of Accelerated Segmentation With Super-resolution . . . 97
4.7 Multiple Vasculature Part Segmentation 99
4.8 Vessel Interior, Centreline and Boundary Probability Maps 101
4.8.1 From Vessel Masks to Probability Maps 101
4.8.2 Experimental Settings and The Preparation of Training Dataset . . 101
4.8.3 Preliminary Results : Vessel Masks and Features 102
4.8.4 Segmentation Results 105
4.9 Conclusion 108

5 Quantitative Vasculature Analysis 110

5.1 Introduction 110
5.1.1 Motivations For The Proposed Vessel Tracking System 111
5.2 Tracking as a Dynamic System 112
5.2.1 Bayesian Solution to The State-Space Model 113
5.3 Particle Filters 115

5.4 Proposed Tracking System For Quantitative Analysis of Retinal Vasculature 118
5.4.1 *Mode 1*: Tracking Vessel Segments . . . 121
Motion Model . . . 122
Observation Model . . . 123
The Proposed Models for *Mode 1* . . . 125
Final Estimation Of Centreline and Boundary Locations . . . 128
5.4.2 *Mode 2*: Detecting Landmark Locations and Collecting Information for the Initialisation of Future Tracking of Connected Vessels . . . 129
1- Landmark Location Detection . . . 129
2- Daughter Vessel Identification . . . 130
5.5 Conclusion . . . 134

6 Tracking Performance Evaluation 137
6.1 Introduction . . . 137
6.2 Demonstration of Vessel Tree Tracking . . . 137
6.2.1 Experimental Settings . . . 137
6.2.2 *Mode 1*: Regular Vessel Segment Tracking . . . 138
6.2.3 *Mode 2*: Landmark Location Detection . . . 139
6.2.4 Detecting Daughter Vessels . . . 139
6.3 Quantitative Vasculature Analysis . . . 141
6.3.1 Introduction . . . 141
6.3.2 Experimental Settings . . . 144
For the DRIVE dataset . . . 144
For the REVIEW dataset . . . 144
6.3.3 Evaluation Criteria . . . 145
For Vessel Width Estimation . . . 145
For Vessel Tree Extraction and Landmark Location Estimation . . . 145
6.3.4 Vessel Width Estimating in the REVIEW dataset . . . 146
The First Experiment . . . 146
The Second Experiment . . . 147
Performance Comparison With Previous Studies . . . 155
6.3.5 The Tracking of A Vessel Tree Fraction . . . 156
6.4 Conclusion . . . 156
6.4.1 The Performance Evaluation . . . 159
6.4.2 Vessel Cross-Section Modelling . . . 162

7 Discussion and Conclusion 164
7.1 Introduction . . . 164
7.2 Contributions . . . 165
7.2.1 Vasculature Segmentation . . . 165
7.2.2 Quantitative Vasculature Analysis . . . 165

7.3 Limitations 167
7.3.1 Vasculature Segmentation 167
7.3.2 Quantitative Vasculature Analysis 168
7.4 Future Work 170

8 Appendix **187**

List of Tables

4.1 The evaluation of the effect of the depth of the proposed network on the segmentation performance. 84
4.2 Average and the best and the worst segmentation performances of the proposed method on the DRIVE dataset. The order of performances obtained from individual images is based on accuracy. 87
4.3 The comparison of segmentation performance of the proposed method with the performances of the art methods regarding the DRIVE dataset 87
4.4 Average and the best and the worst segmentation performances of the proposed method on the STARE dataset. The order of performances obtained from individual images is based on accuracy. 89
4.5 The comparison of segmentation performance of the proposed method with the performances of the art methods regarding the STARE dataset. 89
4.6 The comparison of segmentation performance of the proposed method with the performances of the art methods regarding the pathological images of the STARE dataset. 89
4.7 Average and the best and the worst segmentation performances of the proposed method on CHASE_DB1 dataset. The order of performances obtained from individual images is based on accuracy and down-sampling rate for fundus image patches is 2. 95
4.8 The comparison of segmentation performance of the proposed method with the performances of state of the art methods regarding CHASE_ DB1 dataset. 96
4.9 Average and the best and the worst segmentation performances of the proposed method on HRF dataset. The order of performances obtained from individual images is based on accuracy and down-sampling rates for fundus image patches are 2 and 4. 97
4.10 The comparison of segmentation performance of the proposed method with the performances of the art methods regarding HRF dataset 98
4.11 Comparison of segmentation times of super-resolution TDBN on high resolution fundus image datasets with those of enhanced TDBN on low resolution fundus image datasets. 99

4.12 The average segmentation performances of the proposed method on the DRIVE dataset. 106
4.13 The average segmentation performances of the proposed method on the STARE dataset. 107

6.1 The performance of the proposed method on vessel width estimation when it follows centreline location given in ground truth data 147
6.2 The performance of the proposed method on vessel width estimation when it is used as an automatic tracker. 151
6.3 The comparision of the performance of the proposed method with that of previous studies regarding the accuracy of vessel width estimates in subsets of the REVIEW dataset 157

List of Figures

1.1 A fundus image with marked the optic disc, the fovea, an artery and a vein (An image from publicly available CLRIS dataset [6]) 21

2.1 Some proposed models for the estimation of vessel widths. Blue lines in (a)-(d) are identical and show a $1D$ vessel cross-section intensity distribution. Red lines in (a) show the locations of related intensities (e.g. left max) while those in (b)-(d) demonstrate the approximated representations of the intensity distribution. The intensity distribution is approximated in (d) by adding two Gaussian-like curves in green in order to model the central light reflex. Red surfaces in (e)-(f) approximate a $2D$ vessel cross-section intensity distribution. 46

3.1 (a) An application of classification based retinal segmentation approach with a shallow Convolutional Neural Network (This network can be converted to be a deeper one by adding extra layers of convolutional and pooling before fully connected layer. The number of fully connected layers can be also increased) : The class (belong to a vessel or not) of pixel shown with x is evaluated through the network (b) An application of cross-modality learning based retinal segmentation approach with a fully connected network. 54

3.2 A fundus image patch in the left and its vessel map in the right 55

3.3 (a) An autoencoder (b) An unfolded autoencoder by showing encoder $f(\mathbf{x})$ and decoder $g(f(\mathbf{x}))$ functions. 57

3.4 An RBM architecture, where $\mathbf{v}$ represents visible units and $\mathbf{h}$ hidden units 58

3.5 The training of an RBM with Gibbs Sampling, where an equilibrium distribution of the model is supposed to be reached after n steps 62

3.6 The training of a denoising autoencoder as described by Vincent *et al.* [7], where binary masking noise (which replaces randomly selected pixels in training sample with zero) is applied to the training data and the original training data is approximated. 64

3.7 The architecture of a DBN. 65

3.8 The training of a DBN with the stack of RBMs. 66

3.9 The training of a T-RBM: initially, a traditional RBM in the right side is trained on multiple data modalities, where purple circles and yellow ones belong to different data modalities. After the training of this RBM, the training of a T-RBM (shown in the left side) is performed based on the model learned in the RBM. The weight updates in the TRBM are realised for only data modality ($data'$) represented with purple circles. 67
3.10 The combined training of an RBM with a special denoising procedure on the couple of an fundus image patch and (a) single type of segmentation mask or on the couple of an fundus image patch and (b) multiple types of segmentation masks. 69
3.11 The training of an RBM with I_x and I_y in order to learn a joint representation of them. Note that this joint representation is directly learned from the samples of data modalities. In the other words, I_x and I_y are connected to form an input vector in this layer. 69
3.12 Denoising procedure implemented with a block masking noise (a) when the second data modality I_y is missing (d) when the first data modality I_x is missing. The reconstruction of both modalities given the hidden representation generated by only one modality is detailed in (b)-(c) and (e)-(f); (b) The generation of I_x given I_x (c) The generation of I_y given I_x (e) The generation of I_y given I_y (f) The generation of I_x given I_y. 70
3.13 Training of the second layer of a DBN with binary masking noise. 71
3.14 (a) The weights of a trained DBN with a training dataset containing two data modalities I_x and I_y are unfolded. The resulting network can be viewed as a deep autoencoder with the symmetric weights of encoder and decoder (b) The adjustment of pretrained weights to cross-modality learning: The weights connecting the input layer and the first hidden layer of this autoencoder are changed with the weights learned during the training of a T-RBM. The units and weights associated with I_x in the output layer of this autoencoder are also removed. This final network can be optimised with a stochastic gradient descent algorithm. 72
3.15 The features learned for the first layer of a TDBN in the generative training stage without and with the application of denoising (the noise type is block masking noise) and the evolution of these features during the finetuning. . 76
3.16 Details of features generated during the pretraining, demonstrated in Figure 3.15. In each sub-figure, the top layer is associated with fundus image patches and the bottom layer with binary labels. Some selected features (a) from the training of a TDBN without denoising, (b) from the training of a TDBN with denoising 77

3.17 Fundus Image Patches and Generated Probability Maps for 'Vesselness': The first row shows the fundus image patches, the second one probability maps generated by the TDBN without denoising, the third one with denoising and the last one represents the ground truth image patch. The first column belongs to an example of single vessel, the second one to that of parallel single vessels and the third one to that of crossing vessels. 78

4.1 The segmentation of vasculature in the DRIVE dataset: the rows from top to bottom show images with the best and the worst segmentation performances based on accuracy. The columns from left to right belong to green channels of fundus images, ground truth images, generated probability maps with the proposed method and their segmented versions respectively. The red disc shows a pathological region. 87

4.2 The segmentation of blood vessels in the STARE dataset: the rows from top to bottom show images with the best and the worst segmentation performances based on accuracy. The columns from left to right belong to green channels of fundus images, ground truth images, generated probability maps with the proposed method and their segmented versions respectively. The green and purple discs show vessels with almost the same features. Altough vessels in the purple disc are labelled in the ground truth data, those in the green one are missing. 90

4.3 (a) A pair of a fundus image patch and its vessel mask for super-resolution application, where the scale of the fundus image patch is 0.5 (b) The training of an RBM with denoising in order to learn the correlation between a low resolution image patch and corresponding high resolution vessel mask given in (a). 91

4.4 A trained enhanced TDBN for super-resolution segmentation task: $\mathbf{I}_x$ is an input vector transformed from a low resolution image patch and $\hat{\mathbf{I}}_y$ is an output vector to be transformed to a high resolution vessel mask. 92

4.5 Some of generated features learned during the fine-tuning stage of the training of the proposed network for super-resolution segmentation in HRF dataset (the down-scaling factor for fundus image patches was 4) . (a) The first layer features (b) The last layer features 94

4.6 (a) Randomly sampled fundus image patches from HRF test set, where the down-scaling factor for fundus image patches was 4 (b) Generated vessel masks corresponding to these image patches by the proposed method (c) Ground truth image patches 94

4.7 The segmentation of blood vessels in CHASE_DB1 dataset: the rows from top to bottom show images with the best and the worst segmentation performances based on accuracy. The columns from left to right belong to green channels of fundus images, ground truth images, generated probability maps with the proposed method and their segmented versions respectively. (Both images were thresholded at 0.15). 95
4.8 The segmentation of blood vessels in HRF dataset in the scale of 0.5: the rows from top to bottom show images with the best and the worst segmentation performances based on accuracy. The columns from left to right belong to green channels of fundus images, ground truth images, generated probability maps with the proposed method and their segmented versions respectively. (Both images were thresholded at 0.15). 97
4.9 The segmentation of blood vessels in HRF dataset in scale 0.25: the rows from top to bottom show images with the best and the worst segmentation performances based on accuracy. The coloumns from left to right belong to green channels of fundus images, ground truth images, generated probability maps with the proposed method and their segmented versions respectively. (Both images were thresholded at 0.15). 98
4.10 (a)The training of an RBM with denoising in order to learn the correlation between a fundus image patch and a group of label patches. 100
4.11 A trained enhanced TDBN for multiple types of label patch generation: $\mathbf{I}_x$ is an input vector transformed from an image patch and $\hat{\mathbf{I}}_y$ is an output vector to be transformed to multiple vessel masks, namely vessel interior, centreline, boundary and bifurcation/crossing pattern in centreline. 100
4.12 Learned features during the training of an enhanced TDBN for multiple types of mask generation. Some examples from features connected to (a) fundus image patches in the input layer of the final network weights from the training of a TDBN (b) vessel interior (c) vessel centerline (d) vessel boundary (e) bifurcation/crossing pattern label patches in the output layer of this network 103
4.13 Randomly sampled fundus image patches from the DRIVE test set and generated labels with the proposed method 104
4.14 (a) A bifurcation vessel pattern in a fundus image patch in the DRIVE dataset, (b) The profile of a generated bifurcation pattern for image patch in (a), (c) A crossing vessel pattern in a fundus image patch in the DRIVE dataset (b) The profile of a generated crossing pattern for image patch in (c) 105

4.15 The segmentation of vasculature in the DRIVE dataset regarding vessel interior, centreline and boundary segmentation: the rows from top to bottom show images with the best and worst segmentation performances based on accuracy. The columns from left to right belong to green channels of fundus images, probability maps of vessel interior, centreline and boundary respectively. 106
4.16 Bifurcation/crossing pattern detection in the DRIVE dataset: Probability maps belong to (a) 1^{st} image (b)12^{th} image(c) 13^{th} image (d) 20^{th} image. (The contrast of images is inverted for better visualisation) 107
4.17 The segmentation of vasculature in the STARE dataset regarding vessel interior, centreline and boundary segmentation: the rows from top to bottom show images with the best and worst segmentation performances based on accuracy. The columns from left to right belong to green channels of fundus images, probability maps of vessel interior, centreline and boundary respectively. (The contrast of images is inverted for better visualisation) . 107

5.1 A state space model of a first order Markovian system: z^k shows the state vector at iteration k and y^k observations at this iteration. 113
5.2 The sequential estimation of the posterior probability distribution of state vectors by using a Particle Filter, which samples from SIR and uses State Transition Probability Distribution as the proposal distribution. In the diagram, the weights of particles are visualised with their sizes. The particles with the same state vector (identical particles) are located top of another (adapted from [8]). 120
5.3 The proposed vessel tracking system; an example of (a) observation model and (b) motion model 121
5.4 The difference between the responses of the proposed network to thinner and thicker vessels. 124
5.5 The calculation of likelihood for *Model 1* for (a),(c) a strong hypothesis (b),(d) a weak hypothesis, where E_1 and E_2 denote estimated edge locations. In (a) and (b), vesselness probability profile is shown with red dashed line and samples from this profile with red arrows. Similarly, edge profile with blue dashed line and blue arrows and samples from the complement of edge profile with green arrows. (c) and (d) shows samples from these distributions. 128
5.6 (a) A bifurcation map belongs to the first image in VDIS dataset. (The contrast of images is inverted for better visualisation.) (b)-(c) $3D$ visualisation of some selected regions from (a), where bifurcating/crossing locations appear peaks. (Please, find matching colours of frames in (a) and (b)-(c) for region identification.) 131

5.7 Detecting the mode of a set of data points with the Mean Shift algorithm. x^1 represents the first estimate of a mode location. 133

6.1 (a)A fundus image from VDIS dataset: Green arrow shows the starting location of the tracking, which is shown with a green dot (b) Zoomed version of (a) 138

6.2 Visualisation for the tracking of a regular vessel segment: (a)-(c) Estimates of each particle regarding centreline and edge locations are respectively shown with red and green dots. The expectations of centreline and edge posterior probability distributions are illustrated with a black cross. The expectation for the direction of tracking is illustrated with a big black arrow and the direction of each particle is demonstrated with a blue small arrow connected to its centreline location. (b)-(d) Each subfigure shows both vesselness (curves appear as peaks) and edge (curves appear as valleys) profiles for 10 particles. The first and second row shows the first 100 particles with highest likelihood while the third and the last row belong to the last 100 particles in the same order. The number on the top of each subfigure shows the average of the likelihoods of 10 particles in this subfigure. 140

6.3 An estimation for bifurcation location with the Landmark Detection Function: The estimated landmark location is visualised with a yellow disc and other possible candidates with red discs (a) on the fundus image (b) on the landmark location probability map from $3D$ view. The estimation time for the yellow disc was $t = 1$ and for others $t = 2 : 4$ (the increase in the time is illustrated with the darker tones of red). The landmark location probabilities at these locations are respectively 0.799,0.796, 0.768 and 0.742 from $t = 1$ to $t = 4$. 141

6.4 (a)Prior Probability Distribution of daughter vessel existence: (b) Zoomed version of (a); where blue arrows show the direction vector for each particle and green dots centreline locations of possible daughter vessels. 142

6.5 (a) Posterior probability distribution of being a part of the centreline of a daughter vessel: which is the result of one step of Particle Filtering ($P(z^{k+1}|y^{k+1})$). Each particle is represented with its direction vector and centreline location. (b) Estimated particle sets corresponding to possible daughter vessels by using *Mean Shift Clustering Function*, where each particle set is shown with a black circle and a direction arrow. 142

6.6 Starting the tracking of estimated daughter vessels (Refer back to Figure 6.2). 143

6.7 The segmentation of blood vessels in VDIS dataset (8^{th} image): The images from left to right are respectively a fundus image and the probability maps of its vessel interior pixel, centerline and boundary segmentation 145

6.8 The precison distribution for individual vessel segments: vessel widths are estimated given centreline locations in the ground truth data (Whiskers represent $+/-2.7\sigma$ in (a)-(c). '*' indicates the result of a individual image in (d)). 148
6.9 Vessel segments with the worst average precision from HRIS and VDIS. Estimated locations are shown with green dots and reference locations with red dots. The black ellipse in (a) shows a mispredicted vessel boundary in the reference data. The black circles in (b) and (c) encircle vessel beading. 149
6.10 The precison distribution for individual vessel segments: vessel widths are autonomously estimated by the tracking system. (Whiskers represent $+/-2.7\sigma$ in (a)-(c). '*' indicates the result of a individual image in (d)). 151
6.11 Performance comparison of the three models of vessel centreline and boundary locations (vessel segments with the worst precision). The top row belongs to the CLRIS dataset, the second one to the HRIS dataset, the third one to the VDIS dataset and the bottom one to the KPIS dataset; while the first column to *Model* 1, the second column to *Model* 2 and the third column to *Model* 3. Estimated locations are marked with green and reference locations with red. 152
6.12 Performance comparison of the three models of vessel centreline and boundary locations (vessel segments with the best precision). The top row belongs to the CLRIS dataset, the second one to the HRIS dataset, the third one to the VDIS dataset and the bottom one to the KPIS dataset; while the first column to *Model* 1, the second column to *Model* 2 and the third column to *Model* 3. Estimated locations are marked with green and reference locations with red. 153
6.13 Some examples of predicted boundaries by *Model* 4 and reference boundaries (a)-(d) from CLRIS (e)-(h)from HRIS, (i) from KPIS and (j)-(l) from VDIS (green lines show predicted profiles and red lines reference profiles). 154
6.14 The tracking of a part of vasculature in 10^{th} test image in the DRIVE dataset: (a) The traced path (the starting point for tracking is shown with the pink arrow.) (b) Estimated landmark locations during the tracking (c)-(d) Estimates of vessel centreline and edge locations, where estimates of boundary locations are marked with blue and those of centreline locations are with red. (c) shows the bottom left corner of the image and (d) shows its top right corner (Some missed branches during tracking are illustrated with circles). 158
6.15 Approaches to vessel tracking regarding their observation spaces: the observation space is the image space in approach 1, the binary image in approach 2 and a set of probability maps in approach 3. 161

Chapter 1

Introduction

1.1 Introduction

The retina is the inner layer of the eye, which roughly resembles a sphere with a diameter of 2.2 cm (in human adults) [9]. Main landmarks in the retina are the optic disc, fovea, macula and retinal vasculature. The optic disc is located in the center of the retina. The human optic disc has an approximate ellipsoid shape, with a major axis of 0.2 cm and a minor axis of 0.15 cm [10]. The fovea, which is positioned in the center of the macula, in a fundus image is located on the left or right side of the optic disc depending on which eye is captured [10]. The fovea appears 'reddish' in colour and is close to an ellipse in shape. This region has a lower density of visible blood vessels.

The optic disc accomodates optic nerves leaving the eye and the central region artery entering the retina [10]. The blood is distributed to the inner layer of the retina with arterioles bifurcated from the central region artery and blood is collected back with various branches of venules into the central retinal vein to be returned to the heart [9]. The branches from arterioles can cross with those of venules, while branches belonging to the same vessel group seem almost parallel [9].

There are several reasons for monitoring the retina [9]. This process may be performed in a non-invasive way by simply transmitting white light through the pupil and collecting it back with a simple device. During this imaging operation, the only intervention to the human body is the usage of pupil dilating agents, if necessary. Therefore, the comfort of a patient is maintained during imaging, and the side effects of this imaging are almost none. Secondly, being easily monitored makes the retina the best organ to detect early signs of chronic diseases such as diabetes, high blood pressure, which systemically affects the vasculature. Also, some morphological changes of the retina have been found to be related to the existence or the severity of some diseases. These changes can be used as biomarkers to facilitate early diagnosis of these diseases or to follow their progress or to change their course with preventative therapies [11]. Fundus photographs are currently the accepted standard media for screening diabetic retinopathy in England [12]. Moreover,

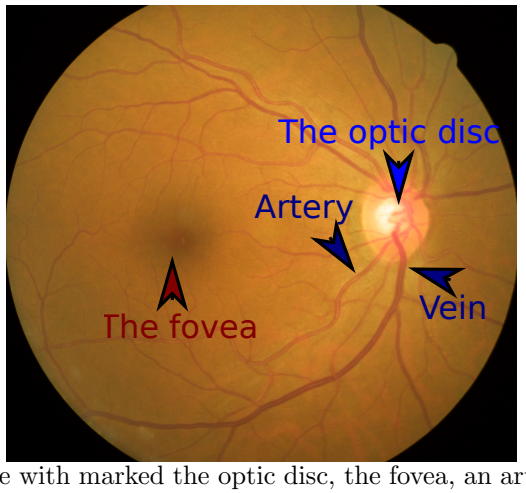


Figure 1.1: A fundus image with marked the optic disc, the fovea, an artery and a vein (An image from publicly available CLRIS dataset [6])

the vasculature of the retina and that of the brain share some properties as a result of being located in a close distance and give the similar responses to diseases and stress. This has attracted the attention of neurologists to indirectly monitor brain vasculature via monitoring retinal vasculature in an easy way [13].

Although, to date, there have been many studies devoted to fundus image analysis, these studies have mostly focused on the segmentation of vasculature in low resolution and healthy fundus images. However, analysing/quantifying vasculature in the presence of pathologies and in high resolution fundus images is still an open area and it may be more useful for medical research. This chapter will start with a discussion about how fundus image analysis can be used in medical practice and will continue with another discussion regarding challenges related to image analysis in fundus photos. These discussions will lead to the aims of this thesis. The chapter will finish with the contributions of this research to the literature and the organisation of this thesis.

1.2 Fundus Cameras and Colour Fundus Photos

There are various imaging modalities used for the visualisation of the retina for various purposes. Main retinal imaging modalities are fundus cameras, scanning laser ophthalmoscopes (SLO), angiographies and optical coherence tomographies (OCT) [9]. Because the images processed in this thesis were fundus photographs, only fundus cameras will be covered.

Fundus photos are $2D$ projections of the inner layers of $3D$ retinas. These photos show the optic disc, macula, retinal vasculature and pathologies and can be used for the diagnosis of eye-related diseases or for routine monitoring of their prognosis or their responses to treatment. The photos have the advantage of being electronically transferable to another center by allowing them to be examined by medical experts in a different location than the place they are captured. Also, the ability to be stored in a digital environment permits these photos to be re-examined for follow-up examinations [14]. Finally, the evolution

of fundus cameras from the traditional bulky systems to portable fundus cameras makes them generally more accessible [15]. More technical details about fundus cameras can be found in [15].

1.3 Retinal Fundus Image Analysis

There are several aims for performing retinal image analysis. Trucco *et al.* [16] divided methods related to retinal image analysis into three main groups depending on their functions:

1. Identification of diseases: These methods classify each fundus image in databases collected during screening/monitoring programs according to the presence of a disease of interest such as diabetic retinopathy, with the aim of referring them to a medical expert in the presence of this disease.

2. Computer-assisted diagnosis: These methods calculate the likelihood of a disease to be exist in a fundus image depending on the existence of pathologies or morphological changes in retinal vasculature.

3. Biomarkers: These methods evaluate the relation between characteristics of retinal vasculature with the presence or the severity of a disease.

The second and the third group may require a quantitative analysis of retinal vasculature. Mostly used morphological parameters of the retina include vessel calibres, arteriovenous ratio (AVR), length-to-diameter ratio, branching angles, tortuosity, fractal dimensions [17]. These two groups will be explained further as follows.

1.3.1 Computer-Assisted Diagnosis

Analysing fundus photos can facilitate the diagnosis of diseases causing visual dysfunction or even blindness when they are left without treatment. Related diseases include diabetic retinopathy (DR), glaucoma, age-related macular degeneration (AMD) [18]. The functionality of the eye can be maintained if these diseases can be recognised in early stages, and their progression is annually checked [19]. Before the introduction of screening programs in the UK, treatable eye-diseases were the most common diseases leading to blindness. However, after screening programs, hereditary eye diseases became the most prevalent ones. [20].

Among these diseases, DR and AMD are related to the dysfunction of retinal vasculature while glaucoma to the neurological structure of the retina. Diabetic retinonpathy (DR) can be caused by diabetes mellitus. The symptoms of DR are not visible in early stages but DR can cause blindness in severe stages. DR shows different symptoms during its progression such as microaneurysms, hemorrhages, venous beading, intraretinal microvascular abnormalities, neovascularization, vitreous/preretinal hemorrhage [21].

AMD has two versions [19]: dry AMD and wet AMD (known as choroidal neovascularization (CNV)). Among them, CNV is more dangerous. It progresses faster and can cause blindness; though, both are associated with the loss of visual acuity. The main characteristics of CNV are : (i) choroidal vasculature may grow into the macula, (ii) vascular permeability may rise by fluid collection in the retina or below the retina, leading to a deterioration in visual function.

Glaucoma is described with 'a gradual damage to the optic nerve' [19]. Regarding the examination of this disease in fundus images, cup-to-disc ratio, defined as 'the ratio of the optic disc cup to neuroretinal rim surface areas', is an important indicator [19].

Apart from these commonly encountered diseases, birdshot chorioretinopathy (BCR) is an autoimmune disease, which can also lead to blindness and shows its sign in retinal and choroidal vasculature. Agrawal *et al.* found that venuler caliber was statistically significant between patients with birdshot chorioretinopathy (BCR) and a control group in a small scale population analysis [22]. According to their evaluation, patients with BCR had smaller venuler calibres and their 'arteriole-to-venule ratio' was larger than that of the control group. Agrawal *et al.* suggested that venular calibre can be used as a biomarker after validating its power in large scale analysis.

1.3.2 Biomarkers

A biomarker can be defined as an 'objectively measurable characteristic', which is an indicator of 'normal biological processes, pathogenic processes, or pharmacological responses to a therapeutic intervention' [23].

Monitoring the retina is rather easier (non-invasive) and less costly than monitoring other organs, such as the brain, and retinal vasculature is affected by systematic vascular diseases in a manner similar to other organs. Therefore, many studies have explored the relation between morphological changes in retina vasculature as an indicator of the incidence or progression of related diseases [11]. The morphological characteristics of the retina listed in Section 1.3 can be used as biomarkers. For example, the change of vessel calibres was associated with the incidence or the progression of diabetes or the incidence of diabetic nephropathy [24], and the change of oxidative stress regulation and inflammation in cardiovascular system [25]. Similarly, both increases in venule calibres and decreases in arteriole calibres were related to the increased risk of hyper-tension [26] and 'more diffuse and severe ' coronary artery diseases in a female population [27]. In a recent study, McGowan found a significant link between hypertension and reduced central retinal arteriolar equivalent (CRAE) [28].

Apart from systematic diseases, neurological diseases have been found to be related to changes in the morphology or functionality of the retina. For instance, narrower retinal venules were linked with the incidence of Alzhemir's diasease [29, 30]. Cheung *et al.* observed that vessel tortuosity increased for patients with Alzhemir's diasease, whereas branching angles at bifurcation locations maintained. Schizophrenia patients were also

found to have wider retinal venules [11]. Also, the incidence of stroke was associated with decreased artery-vein ratio (AVR), narrower arterioles, wider venules, increased tortuousity and reduced complexity of branching [11]. In addition to these changes, reduced density of retinal vasculature in fundus images and reduced branching angles were associated with degraded cognitive functions [11]. According to a meta-analysis performed by Wu *et al.* [31], the narrowing of retinal arteriole, 'retinal arteriovenous nicking', the existence of hemorrhage and microaneurysm in the retina and the reduction in fractal dimension of retinal vasculature were found to be significantly correlated with stroke.

1.4 Challenges Associated With Retinal Vasculature Analysis

Because the iris has a very small diameter, in a range of 2 to 8 mm, illuminating and imaging the retina in the same optic path poses a big challenge [19]. Some challenges regarding fundus image analysis are directly related to how well a fundus image is captured and others may be associated with the design of the analysis/segmentation methods.

1.4.1 Imaging

Image acquisition may be a challenge in screening programs because poor quality fundus images can be captured with non-mydriatic cameras or by non-professionals. For example, if the camera is not well focused, it can cause blurry images [32] or if the size of dilated iris is not big enough, it can degrade the quality of fundus images taken. Also, some occlusions such as eyelashes and cataract, or artefacts caused by the camera can obstruct a clear view of the retina [16]. Using a different camera such as mydriatic, non-mydriatic or hand-held also may generate different fundus images from the same eye [33]. The distance between a camera and the retina, or using a different camera lens, can change vessel widths appeared in fundus images [34]. On the other hand, 'adequate' quality images may be acceptable for a clinical examination but not be suitable for automatic image analyses [16].

1.4.2 Anatomy

The projection of a three dimensional structure to two dimensional images can affect the appearance of the retina in photos. This can cause an appearance of overlapping structures such as the overlapping of retinal vasculature and choronoidal vessels [33]. Similarly, the magnification rate of retinal structures may be affected by the dimension of the retina, such as axial length or depth or by the degeneration on the refractive surface of the eye such as myopia [35]. Also, the sizes of vascular arterial structures may vary in sequentially captured images from the same eye due to the cardiac cycle [35]. Uneven illumination of the retina, which may be due to its spherical shape or less dilation of the pupil, may cause various contrast levels for similar parts of vasculature in the same photo.

1.4.3 The Designs of Methods

In addition to the intrinsic characteristics of fundus photos, there are other challenges, which may be more associated with their analyses. Firstly, a range of diameters of retinal vessels may make vasculature segmentation/analyses more difficult. This thickness variation is also mostly accompanied with a variation of contrast levels against the detection of thinner vessels, in addition to uneven illumination of fundus images. Secondly, the optic disc or pathological lesions - for instance hemorrages, cotton wool spots, microaneurysm- may exacerbate vasculature segmentation further because they may be in close neighbourhood to vasculature or may appear in a similar contrast to vasculature contrast. Thirdly, the reflections along arterioles, called the central light reflex and mostly seen in the retinas of young people, may deceive algorithms as if they represent vessel edges. Finally, the topological characteristics of vasculature such as having branches or crossing of arterioles with venules or closely located or curvy vessels may pose major challenges for segmentation, where the tubular appearance of vessels disappears at overlapping regions or the identification of vessels may become harder.

To date, segmentation methods for retinal vasculature have shown big improvements on the segmentation of large vessels in healthy images. However, the segmentation of thin vessels with low contrast levels, segmentation in pathological images and segmentation in bifurcation/crossing locations have remained problematic [33]. Also, the identification of pixels near the field of view (FOV) stops of fundus images can be highly confusing for automatic methods [36].

1.4.4 Availability and Reliability of Manually Labelled Datasets

Sufficiently large amounts of labelled datasets are crucial for the training of supervised methods and important factors in the performance evaluation of proposed methods. Although there are a couple of image datasets with manual labels, the variety of photos in these datasets is far from representing images encountered in clinics or in screening programs. Firstly, their age and ethnicity representative capacity is limited [2–6]. Secondly, the proportion of images with any symptoms of eye related diseases is much lower than those encountered in real situations [2]. Also, 'not gradeable' or 'poor' images are not or rarely represented in publicly available datasets, which may be an important reason behind reduced performance of automatic retinal analysis methods in real situations [37]. Given the variability of pathologies and other artefacts in fundus images, a larger number of representative images is needed to deal with these problems.

The reliability of labelling provided for publicly available fundus image datasets have been questioned by some researchers [38]. These datasets were accompanied by at least two sets of ground truth from independent observers in order to eliminate the subjectivity on manually traced vessel maps. According to Sofka *et al.* [38], inter-observer variability in manually segmented vessel maps of the DRIVE and the STARE datasets was so large (larger for thinner and low contrast vessels) that it can also introduce subjectivity to the

performance evaluation of proposed methods. The difference of labelling by independent experts may be due to different interpretation of vessels. Some observers traced vasculature more cautiously by giving more chance for the labelling of larger vessels, whereas some observers assigned larger numbers of pixels for thinner vessels with low contrast [2,2,4,5]. In addition to the existence of a vessel, variations can be observed in the determination of pixels representing vessel boundaries. There was no numerical analysis performed to show the accuracy of manual labelling, as far as I am aware.

1.5 The Aim and The Scope of This Thesis

Recently, data-driven methods have been shown to obtain better accuracy for image analysis. Although many methods have been devoted to the segmentation of retinal vasculature to date, increased amounts of available data may offer new opportunities to improve segmentation methods. Given variations on fundus images due to various ethnicities, age groups, patient groups; the design/training of retinal vessel analysis methods appear to benefit from a data-rich evaluation. Existing methods for retinal analysis are mostly focused on the segmentation of interior vessel pixels as an individual task. However, describing the vasculature with various structured features such as vessel centreline and boundary may open new doors to retinal analysis applications. In this thesis, the simultaneous detection of vessel interior, centreline and boundary locations stretched the boundaries of segmentation to tracking of retinal vasculature for the purpose of obtaining quantitative estimates in pixel coordinates.

The main aims of this thesis are listed below:

1. To benefit from data in the design/training of the method
2. To shorten segmentation time for high resolution images
3. To provide/improve the accuracy of vasculature analysis

The scope of this thesis is limited to the analysis of retinal vasculature in fundus photos from commonly used publicly available datasets. These images are used only for the segmentation of vasculature and for the identification of the topology of vessel tree fractions. However, the detection of other landmarks such as the location of fovea or that of the optic disc or their characteristics are not among the aims of the thesis; though, the performance of the proposed tracking method can be improved by using the location of the optic disc to pick seed locations. In addition, there is no intention to give a diagnostic decision related to any eye-related or other diseases nor to detect/identify/characterise pathological lesions in fundus images; though, the results of this thesis such as binary vessel maps, the properties of vessel topology or vessel morphology can be used for diagnosis or research purposes, image registration or follow-up checks.

Among several morphological properties of vasculature, only vessel width is estimated in this thesis, because the change of vessel calibre has been mostly related to eye/systematic

/neurological diseases in the literature. Also, examples of vessel boundary localisation and traced vessel tree fractions are visually demonstrated. Although the quantitative measurements of this thesis are presented in terms of pixels, some of them can be dependent on camera or image acquisition conditions. However, dimensionless measurements such as tortuosity, length-diameter-ratio and branching angles can be calculated with the information provided with the tracker.

1.6 Contributions

The contributions of this thesis will be given separately for vasculature segmentation and detailed analysis of vasculature as follows.

1.6.1 Vasculature Segmentation

1. The training of the Translational Deep Belief Net (TDBN) is improved for vessel segmentation.

2. The applicability of enhanced TDBNs to various segmentation applications is shown: conventional segmentation, super-resolution segmentation, multi-labelling segmentation.

3. The super-resolution concept is combined with segmentation in the training of enhanced TDBNs with the aim of reducing the segmentation time of high resolution images.

4. The labelling of multiple vessel parts in the vasculature is realised: vessel interior, centreline and boundary pixels and bifurcation/crossing patterns in vessel centrelines.

1.6.2 Quantitative Vasculature Analysis

1. A probabilistic system is introduced to estimate some morphological properties of retinal vasculature. This system uses multiple probability maps generated with the multi-labelling TDBN to guide a particle filter.

2. This system is used to estimate morphological properties of retinal vasculature, namely centreline/boundary locations and vessel widths, and to predict the topology of vessel tree fractions with the detection of bifurcation/crossing locations between identified vessel segments.

3. New observation models for vessel tracking are introduced. These models combine the probability maps generated by the multi-labelling TDBN and calculate the presence of vessel likelihood, where vessel is described with its centreline and width.

4. A method is proposed to identify connected vessels at bifurcation/crossing locations; this is augmented with another new method used for the detection of these locations.

5. The usage of these probability maps for tracking is observed to improve the detection of thin vessels.

1.7 The Organisation Of The Thesis

Following this **Introduction Chapter**, the organisation of the thesis is given as follows

- Chapter 2 is dedicated to a literature review for both vasculature segmentation and quantitative vasculature analysis and gives details of fundus image datasets used for the analysis.

- Chapter 3 introduces a retinal vasculature segmentation method (an improved training of TDBNs) after firstly giving background information about translational deep belief nets and cross-modality learning. The preliminary results of this training are also presented, with generated features and vessel masks.

- Chapter 4 proposes a segmentation approach to reduce the segmentation time and a training of TDBN for this aim (super-resolution segmentation) and another training of TDBN to generate multiple label patches for each input patch. Both approaches are realised by training networks, whose architectures are derived from the architecture of the first network introduced in Chapter 3. After giving preliminary results, the networks are utilised to segment fundus images. The segmentation performance of the enhanced TDBN and that of the multi-labelling TDBN are evaluated on the DRIVE and the STARE datasets, that of super-resolution TDBN on CHASE_DB1 and HRF datasets. Also, the performance of the networks is compared with that of recent and best-performing methods.

- Chapter 5 introduces a tracking system to obtain quantitative information about retinal vasculature by incorporating the output probability maps of a multi-labelling TDBN in the execution of Particle Filters. After giving the background information for Particle Filters, the proposed tracking system is explained.

- Chapter 6 begins with a short simulation of the tracking described in Chapter 5 and continues with the performance evaluation of the tracking system on vessel width estimation and with the demonstration of traced vessel tree fractions.

- Chapter 7 summarises and discusses the contribution of this thesis to the literature.

Chapter 2

Literature Review

2.1 Introduction

This chapter will summarise the literature related to retinal vasculature analysis and provide an overview of the datasets used for the analysis of this thesis. Retinal vasculature analysis in fundus images can be examined in two groups according to the information to be extracted from the images: the segmentation of the vasculature and the quantitative analysis of the vasculature, including vasculature topology and morphology. Because it is not possible to review all studies in this thesis, only relevant and recent studies will be summarised. Readers can refer to literature reviews of Kirbas and Quek [39], Fraz *et al.*, [40] and Srinidhi *et al.* [33] for more comprehensive information.

2.2 Publicly Available Datasets

In the performance evaluation of the proposed methods, publicly available fundus image datasets were used in order to enable comparisons with the performance of other studies. These datasets are the DRIVE, the STARE, High Resolution Fundus Images (HRF), CHASE_DB1 and REVIEW datasets. Among these datasets, the resolutions of the first two are much lower than those of the second two, which are all provided for vasculature segmentation. On the other hand, REVIEW dataset consists of several sub-datasets with different resolutions and characteristics, and presents quantitative information (e.g. vessel boundary locations) about vasculature.

2.2.1 The DRIVE dataset [2]

Image Collection Process: This dataset contains 40 fundus images, which were selected from a larger set collected during a screening program in Netherlands. The images were captured with a Canon CR5 nonmydriatic 3CCD camera at 45^o of field of view (FOV) with the resolution of 768×584. Images in this dataset were divided into two groups of 20 images to maintain a standard in the evaluation of supervised methods. 3 out of 20 images

in the training set were captured from diabetic retinopathy patients, while the number of pathological images in the test set is 4.

Vessel Labelling Process: The images were manually labelled by three independent experts. One of these experts was an ophthalmologist, who trained the other two experts. Two sets of ground truth vessel maps were generated for the test set, while one set was produced for the training set. This is the only dataset among other datasets reviewed here, which divided images into training and test groups. Due to subjectivity involved in the human decision, the existence of some thinner vessels may vary in the sets of ground truth vessel maps so the number of vessel pixels do. The percentage of vessel pixels labelled in the first set was reported to be 12.7%, while that in the second set to be 12.3%. In order to resolve any conflict that may occur due to this different labelling, the first expert's ground truth images have been accepted as the reference in many studies. Apart from vessel maps provided with this dataset, Azzopardi and Petkov labelled vessel bifurcation and crossing locations for 40 fundus images [41]. In addition to ground truth images, the DRIVE dataset includes FOV masks.

2.2.2 The STARE dataset [3]

Image Collection Process: Images in this dataset were captured with a TopCon TRV-50 fundus camera with 35^0 of FOV. The size of images is 605 by 700 pixels. This dataset contains 20 images with a half belonging to healthy subjects and the other half to pathological cases, where pathological changes may obscure the detection of blood vessels. The collection of these images was made with two main aims. One was to provide a dataset that consists of both pathological images and healthy images in order to evaluate the performance of any segmentation method on both types of images and to compare the performance of each group with that of the other one. Another aim was to encourage researchers to consider pathological images in the design of segmentation methods because these images have more chance to be observed in clinics and they have more significant value in terms of grading of diseases or discovering relations between vessel characteristics and potential diseases.

Vessel Labelling Process: The images were manually segmented by two experts in image processing and retinal vasculature analysis, resulting in two independent ground truth vessel maps. Similar to the DRIVE dataset, one of the experts (the first one) estimated less number of pixels for vessels. To be more explicit, $32,200$ pixels were labelled by the first experts as vessel pixels and $46,100$ pixels by the second one. These extra pixels are mostly due to labelling thinner vessels ignored by the first expert or tracing vessels larger than those traced by the first expert.

Because this dataset does not contain FOV masks, they were generated by using the technique explained in [42]. Also, there are two accepted approaches for how to divide the datasets for training and evaluation in the literature. The first aproach is *leave-one-out*

training, where an image is selected in turn for the performance assesment of a method trained on the other 19 images. This approach requires repeating the training for 20 times to completely test all images in the dataset. The second approach is based on evaluating the performance of a method on the entire dataset despite of the collection of training samples (pixels) randomly over all images. The performance of studies depending on this approach might be, somehow, overoptimistic due to not using separate sets for training and testing [43]; though, the authors of these studies justified the approach by stating the usage of ignorable percentage of pixels used for training [5, 44, 45]. For instance, Cheng *et al* [45] reported the usage of 6% of pixels for the training.

2.2.3 HRF dataset [4]

Image Collection Process: Images in this dataset were collected in an ophthalmology clinic in Czech Republic. During the collection of these 45 images, a mydriatic fundus camera CANON CF–60 UVi, which had CANON EOS– 20D digital camera with a 60^o of FOV, was used for capturing the images. The size of the images is 3504 by 2336 pixels.

The image set consists of 3 groups of fundus images. Each group has 15 images, which were compressed to JPEG format. The first group contains images from healthy people, the second one from Diabetic Retinopathy patients and the third one from Glaucoma patients in advanced stage. The second and the third groups show the sign of related diseases. For example, images in the second group have spots that occurred due to laser treatment, haemorrhages, neovascular nets and bright lesions. Also, those in the third group bear symptoms related to nerve fibre layer loss.

Vessel Labelling Process: Manual labelling was performed by 3 experts trained by ophthalmologists. FOV masks of all images were provided.

2.2.4 CHASE_DB1 dataset [5]

Image Collection Process: These images were collected from both eyes of 14 school children in the UK from various ethnicities so there are 28 images in the dataset. The images were captured with a Nidek NM-200-D fundus camera with 30^o of FOV. The resolution of the images is 1280 by 960 pixels. This dataset consists of fundus images with central light reflex, non-even background illumination and poor contrast of blood vessels.

Vessel Labelling Process: Images in this dataset have two sets of ground truth vessel maps, where each set was labelled by an independent expert. The set labelled with the first expert is accepted as the reference. This dataset does not include FOV masks. In this research, FOV masks were generated by using the technique in [42].

2.2.5 REVIEW dataset [6]

Images in this dataset were taken from a fundus image dataset, which was collected in the diabetic retinopathy clinic at Sunderland Eye Infirmary during clinical routine. This dataset has 4 sub-datasets according to the existence of pathology, central light reflex or the aim of providing a comparison at subpixel accuracy level. These groups are a set of high resolution images, the high resolution image set (HRIS), that of images affected from vascular diseases, the vascular disease image set (VDIS), that of images bearing central light reflex, the central light reflex image set (CLRIS), and that of images labelled at kick points, the kick point image set (KPIS).

Vessel Labelling Process: Initially, three experts independently located vessel edges by selecting sample locations from the images. Two of the experts had an experience in retinal vessel analysis and the other one took a training to locate vessel edges. After obtaining this initial edge information from the experts, this information was edited by an algorithm to ensure even spaces between consecutive profiles along vessels [6]. Although three versions of an edge location estimated by three experts are available, the average of these estimates is used as reference for this location to provide a standard in performance evaluation and to alleviate inter-subject variability in the estimates of vessel boundaries. The total number of manually located pairs of vessel edge locations, which were obtained in sub-pixel accuracy, was 5066. From these edge locations, one can estimate vessel centerlines and vessel widths based on the decision of each expert. Also, this dataset contains grades of any diseases, whose symptoms appear in these images.

HRIS dataset

This set was aimed to provide a comparison with subpixel accuracy for the evaluation of related methods. Subpixel accuracy means that one can estimate vessel widths smaller than a pixel. To do that, reference vessel widths are initially measured on high resolution images. Then, the images are downsampled for the purpose of performance evaluation and initial reference vessel widths are also divided by the down-sampling factor. The images in HRIS were down-sampled by a factor of 4 after measuring their widths. As a result of that down-sampling, the accuracy of widths was limited with an error of ± 0.25 pixels.

Image Collection Process: There are 4 RGB images with a resolution of 3584 by 2438 pixels. These images were captured by a Cannon 60 UV film camera with 60^o of FOV. The number of vessel segments evaluated in this image set was 90 and that of manually detected pairs of edge locations was 2368. 2 of these images were graded with severe non-proliferative retinopathy and 1 of them with moderate and the other one minimal.

VDIS dataset

This dataset serves the purpose of evaluating the accuracy of vessel width estimation on images with pathologies and higher noise. These images were associated with a larger variance of expert-based labelling.

Image Collection Process: The images were captured with the Zeiss fundus camera and JVC 3CCD with 50^o of FOV. The resolution of images is 1360 by 1024. This image set consists of 8 images, where only 2 of them was reported to belong to healthy people and the rest to patients with various types of Diabetic Retinopathy. 79 vessel segments were assessed from these images and 2249 boundary locations were detected.

CLRIS dataset

This set was established to evaluate vessel width estimation on images showing signs of atherosclerosis. The symptoms of atherosclerosis can be counted as the exaggeration of central light reflex and changes in vessel walls.

Image Collection Process: These images were obtained with a Zeiss FF 450 fundus camera equipped with 3-CCD JVC camera. Images were captured with 50^o of FOV after centring the macula. This set consists of only 2 images with 2160 by 1440. The number of vessel segments is 21 and that of edge location pairs is 285.

KPIS dataset

This set provides an alternative method to determine edge locations, which is the detection of kick points.

Image Collection Process: These images were captured with a Canon $60uv$ fundus camera with a 60^o of FOV. This set contains 2 images with a resolution of 3300 by 2600 pixels, which is almost 5 times of the resolution of fundus images commonly used for performance assessment, such as those in the DRIVE. The number of vessel segments is 3 and the number of vessel profiles is 164. These profiles are located on vessel segments between bifurcation locations. These segments have larger calibres and are not tortuous.

Vessel Labelling Process: After identifying vessel centrelines with 'tramline' algorithm [46], the observers located kick points. The images were down-sampled to provide a sub-pixel accuracy.

2.3 Retinal Vessel Segmentation Methods

These methods aim to identify pixels belonging to vasculature in fundus images. The majority of past studies related to retinal vasculature can be placed in this group [14,33,40].

The aims of these studies include removal of vasculature in fundus images so that the clutter/distraction in these images can be reduced to detect other pathologies or other structures in them for diagnosing eye-related diseases. Also, vessel maps can be used for the registration of fundus images from different imaging modalities. Moreover, these maps can be processed with morphological structures to obtain vessel centrelines or boundaries for quantitative vasculature analysis [47–50].

Retinal vessel segmentation methods can be categorised into supervised and unsupervised methods based on the use of ground truth data in the decision process of a pixel to belong to vasculature. In contrast to unsupervised methods, supervised methods use ground truth data in the design, which was acknowledged with a 'training' process. Despite being located in separate groups, many unsupervised methods can be used in conjuction with supervised methods, when they are combined with a trainable decision maker. For example, line detectors were used in an unsupervised manner by Nguyen *et al.* [51]; though, their original introduction to retinal vessel segmentation was made in conjunction with an SVM classifier by Ricci and Perfetti [52].

2.3.1 Supervised Methods

These methods are generally based on minimising a cost function by maximising the similarity between the estimates of vessels by the model/models and binary vessel masks given in ground truth images. The designs of the models and ways to minimise their cost functions can vary. Generally, trainable classifiers were employed to evaluate individual pixels in fundus images to calculate their membership to vasculature. These methods include shallow and deep neural networks [36, 44, 53–58], support vector machines (SVMs) [52], fuzzy logic [59], decision trees and random forests [5, 45]. Also, regression [60] and graph methods such as conditional random fields (CRFs) [61] were also used for vessel segmentation.

When studies in this group carefully are examined, a recent tendency can be seen towards the usage of more powerful classifiers such as random forests and deep learning architectures. Due to the close relevance to the present study, deep learning methods will be explained in more details in a separate subgroup, whereas other supervised methods will be mentioned in another one. The other reasons to put deep learning methods to a subgroup can be due to their higher performance than that of other supervised methods [33] and their ability to automatically generate features in a hierarchical way, in contrast, other methods (both supervised and unsupervised) whose success heavily depends on the design of discriminative features or on the careful selection of existing ones [52, 62, 63]. When the features generated at each layer are examined, it can be observed that the lower layers of deep networks provide more local features such as edges, while the higher layers more abstract features, such as belonging to a class of objects [64]. In addition to generating data-driven features automatically, deep learning based methods can be utilised as standalone systems for a specific task such as classification or regression without needing

to be combined with a separate method [36, 43]. In this case, the training is called 'end-to-end training'.

Deep Learning Methods

Studies using deep architectures were dominated by pixel-wise segmentation approach, where the output layer of the network contains usually one unit to output the probability of a pixel of interest [36, 44, 54–58]. After evaluating each pixel through the network in a fundus image, the output probability map is thresholded to obtain a binary vessel map. Because this approach is based on accepting vessel pixels to be sampled from a normal distribution, the studies adopting this approach may not explicitly consider the spatial dependencies between pixels in fundus images. In other words, these methods may ignore the fact that a vessel pixel/a background pixel is highly probable to be an immediate neighbour of the pixel from the same class. On the other hand, some recent studies appeared to be more willing to consider the spatial dependencies between pixels. These studies will be summarised under spatial connectivity methods.

Pixel-wise Segmentation Methods Among these studies, Wang *et al.* [44], Maji *et al.* [55] utilised deep architectures for feature extraction then trained a powerful classifier such as Random Forests or another network on the output of the architectures to obtain the probability of a pixel belonging to vasculature.

In Maji *et al.*'s study, a fully connected deep network was trained in an unsupervised way as a stack of sparse denoising auto-encoders (not 'unfolded') and the output of this network (100 units for each image patch) became an input to a random forest classifier to classify the center pixel of an input image patch fed into the network [55]. On the other hand, Wang *et al.* trained a feature extractor, which was a CNN, as a classifier in a supervised way [44]. Then, they used the features learned in various levels of this network as the input to the ensemble RFs classifier, which adopted the *winners -take-all* technique as the ensemble model. This model was used to determine the final decision about segmentation according to the decisions returned from three classifiers fed with the features from each layer independently.

Apart from Maji *et al.* and Wang *et al.*, many deep networks for retinal vessel segmentations have been trained in an 'end-to-end training' fashion. These methods include the studies of Lahiri *et al.* [54], Maji *et al.* [56], Tan *et al.* [58], Liskowski and Krawiec [36]. Also, Maninis *et al.* 's study [60] can be put into this group; though, the output of the network was based on regression. These methods will be explained as follows.

Maji *et al.* [56] and Lahiri *et al.* [54] used ensembles of networks for vessel pixel classification. Maji *et al.* [56] used an ensemble of 12 CNNs designed for pixel classification. Each network was independently trained with such a dataset whose samples were randomly picked from a common training set. The output of this ensemble was calculated by averaging the probabilities generated by all these networks. On the other hand, Lahiri *et al.* classified vessel pixels with an ensemble of fully connected networks [54], where each

network was trained in an unsupervised manner as the stack of two sparse denosing auto-encoders in a bootstrapped training dataset similar to the approach of Maji *et al.* [55]. Lahiri *et al.* claimed that training the networks on such training datasets provided more diversity in learned features with these networks. This unsupervised training was followed by a supervised training after adding a SoftMax classifier on top of the last hidden layer of each stack of auto-encoders. In this study, Lahiri *et al.* reported the usage of two levels of ensemble. The first level was described with the training of the same architecture on bootstrapped training dataset. The second level was defined with combining of two collections of networks, where each collection had different architecture of networks. The outputs of networks in each collection were weighted averaged to obtain a joint decision in the first level and the output of each collection was averaged to reach a final probability of being vesselness in the second level.

Liskowski and Krawiec [36] trained a CNN architecture for both one pixel classification and structured prediction (which will be explained later) to segment retinal vasculature. In order to avoid over-fitting of the network to the training dataset, they increased the number of selected samples for training by using data augmentation, which increased the initial number of samples by a factor of 10. This data augmentation stage included scaling of intensities in an interval, rotating patches in various angles, flipping them in vertical or horizontal directions and Gamma correction in HSV space. Their performance outperformed that of many deep learning approaches in retinal vessel segmentation.

Apart from aforementioned studies, Maninis *et al.* [60] and Tan *et al.* [58] dealt with both segmentation of vasculature and that of other organs in fundus images. Maninis *et al.* proposed the training of a CNN to concurrent segmentation of both retinal vasculature and the optic disc. Rather than working on the classification of vessel pixels, they trained the network as a regressor. They modified a trained VGG network [65] for their method. They used finer features generated in the first layers of the network for vessel segmentation and coarser features generated in the last layers of the network for the segmentation of the optic disc.

Similarly, Tan *et al.* classified pixels in fundus images according to their likelihood to be a part of vasculature, that of the optic disc, that of fovea and that of the background image by using a CNN [58]. They used a multi-scale approach in the input layer of the network, where 3 channels were dedicated to up-scaled, original-scale and down-scaled image patches centred in the same pixel in respective channels. However, they found their specificity lower than that of previous work. They stated that a possible reason behind this issue may be the network confusing the pixels in the optic disc with those in the retinal vasculature during multiple classification.

Spatial Connectivity Methods In contrast to pixel-wise segmentation, spatial connectivity methods predict the label of a pixel in a fundus image not independent of close neighbour pixels but somehow considering them in the label prediction of the pixel either by simultaneously generating possible labels for these neighbour pixels with the pixel of

interest or by specifically using the correlation between the labels of neighbour pixels in the label estimation of the pixel of interest (as in CRF). Studies using cross-modality learning [43, 66], structured predictions [36] or CRF in segmentation [57] can be put into this category. Also, the studies of Ganin and Lempitsky [67] and Wu *et al.* [68] can be included to this category, where they queried a binary vessel map by a reduced dimension of a fundus image patch.

Cross-modality learning was introduced to retinal vessel segmentation by Li *et al.* [43,66]. In this approach, fundus images and their labels were assumed to be from different data modalities. Li *et al.* presented a CNN generating a label matrix for an image patch [66]. In another study, Li *et al.* trained another type of architectures, a fully connected network, for this aim [43]. The weights of the first layer of the network were initialised with those generated by training a de-noising autoencoder. This autoencoder was trained with the the concatenated inputs of fundus image patches and their labels. Because the aim of the authors was to generate a function to facilitate the conversion of a fundus image patch to its labels, only the weights connecting fundus image patches to the hidden layer of this trained auto-encoder were utilised by ignoring those connecting labels to this hidden layer. The weights of the other layers of the deep network were initialised randomly. After this initialisation stage, the network was optimised by a gradient descend algorithm. Although the performance of both networks was found almost the same on the DRIVE dataset; the training time required for the latter network was almost the third of that required for the former one.

A similar approach to cross-modality learning, which was 'structured prediction', was adopted for vessel segmentation by Liskowski and Krawiec [36]. In this approach, the labels of pixels only in the central region of an image patch was estimated collectively. Liskowski and Krawiec compared the performance of the network trained with structured prediction and that with pixel-wise classification (which was mentioned before). Apart from the output layer, the networks had the same architecture. Liskowski and Krawiec experimented on several sizes of the output matrix and observed that when the output of the network was a matrix of $5x5$, its performance was better than that obtained with $3x3$ or $7x7$ or pixel based classification. They also reported that structured prediction made the network insensitive to small vessels but maintained vessel boundaries better.

Apart from the aforementioned studies, Ganin and Lempitsky queried label matrices by feature vectors (which corresponded to dimension reduction applied input image patches) of the input image patches by using the nearest neighbor search [67]. In order to produce feature vectors, a CNN was trained to perform dimension reduction on input patches. Then, a dictionary was established by coupling the outputs of CNNs for selected input image patches with their binary vessel maps. Ganin and Lempitsky called the augmentation of a CNN with the nearest neighbour search 'N^4 fields'. They observed a slight increase in AUC when the CNN was coupled with nearest neighbour search, when compared with the mere use of a CNN for generating output vessel maps.

Based on Ganin and Lempitsky's work [67], Wu *et al.* proposed a tracking method,

which incorporated generalised particle filtering [69] with 'N^4 fields' for exploring the connectivity of retinal vasculature [68]. With this approach, they aimed to reduce the segmentation time of retinal vasculature by arguing that the number of vessel pixels were far less than that of non-vessel pixels in the same image so only estimating vessel pixels by using the connectivity of the vasculature could reduce the computation complexity of N4 fields in vessel segmentation. In this study, another network was trained to classify image patches to be tracked, according to containing a new vessel branch (bifurcating or crossing region), an existing vessel segment and background image.

Conditional Random Fields (CRFs) [70] were also introduced in conjunction with a deep architecture for retinal vessel segmentation [57]. In [57], a CRF was modelled as a Recurrent Neural Network (an RNN) on the top of a CNN to train an end-to-end network. Fu *et al* [57] called their final architecture 'DeepVessel' .

Other Supervised Methods

Regarding supervised methods, most of early work on retinal vessel segmentation reviewed here was based on simpler classifiers such as SVMs, Gaussian mixture classifiers and shallow/narrow networks.

Ricci and Perfetti introduced line detectors for the segmentation of retinal vessels. Line detectors were described with fixed length and their centers were located at the pixel to be evaluated. These detectors were rotated through certain number of equal distant angles to capture vesselness information in various orientations. The line strength returned for each pixel was used as a measure of vesselness, whose larger values indicated higher probabilities of vesselness. They trained an SVM for the classification of vessel pixels by using feature vectors obtained from the outputs of these detectors [52].

Soares *et al.* classified vessel pixels by using a Gaussian mixture classifier trained with feature sets consisting of pixel intensity, $2D$ Gabor wavelet responses in multiple scales [71]. Marín *et al.* trained an ANN with 3 hidden layers of 15 units with $7D$ feature vectors to classify vessel pixels [53]. Their feature vectors were obtained from pre-processed fundus images and contained features generated from intensities and those based on moment invariants.

In one of the recent work, Kaur and Mittal trained an ANN classifier to improve the connectivity of predicted vasculature and to reduce the false positive pixels in binary vessel maps, generated from the responses of Gaussian matched filters in multiple-scales [32]. They utilised this segmentation method on images which suffer from various grades of pathologies from none to severe. They used both shape and intensity based features for training. Shape based features were extracted from connected objects (predicted vessel segments) appearing in initial vessel maps generated from the responses of matched filters. These features provided information about the areas and the lengths of detected vessel segments and their distances to the optic disc. On the other hand, the latter features utilised intensities from fundus images such as mean intensity for each detected vessel

segment in each channel of RGB.

On the other hand, most recent work focused more on the usage of stronger classifiers or the combination of weak classifiers. Fraz *et al.* introduced an ensemble of bagged and boosted decision trees to classify vessel pixels by using $9D$ feature vectors [5]. These feature vectors included features to discriminate lesions from vessels in various contrasts and to highlight vessels in fundus images, such as the orientation analysis of gradient vector field, multi-scale Gabor filter responses and morphological transformations, line strength measures and intensities in the green channel of RGB fundus images.

Cheng *et al.* classifed vessel pixels by using a random forest which was trained with a combination of context-aware features such as Weber's local descriptors (which measures the intensity difference between neigbor pixels), the stroke width transform (which gives estimates of widths for each pixel), intensities, Frangi's vesselness measures, Gabor responses and location information [45]. Initially, they calculated a local context for each pixel to be classified, from where features mentioned above were produced. A local context was defined with uniformly sampled points in an oriented and fixed size of rectangular region whose center was located at the pixel of interest. In order to generate rotation invariant features, this region was rotated through several angles. Ultimately, the generated features (280 to 344 features for each pixel depending on the dataset) were fed into a Random Forest classifier for the evaluation of the membership of this pixel to vasculature.

Barkana *et al.* used fuzzy logic, an ANN and a SVM for vessel pixel classification [59]. Apart from previous studies designing features with expertise, they employed descriptive statistics as features: 4 features were designed as the mean of intensities in four directions (horizontal, vertical, each diagonal) and another 4 as their median corresponding in the same directions.

Apart from methods based on classification, Orlando and Blaschko introduced the usage of conditional random field (CRF) for the automatic segmentation of retinal vasculature [61]. They extended a CRF to a fully connected mode, where the range of connections were extended to span all pixels so that each pixel in an image was connected with remaining pixels. In order to deal with increased computational cost during inference, they learned the parameters of this method with a structured output generating SVM.

2.3.2 Unsupervised Methods

Until recently, unsupervised methods were considered superior to supervised methods, being faster and less computationally expensive [33]. Unsupervised methods mainly consider expertise as the main source to define techniques to distinguish vessels from other parts of fundus images, without heavily relying on an estimator trained with manually annotated data in the decision of a pixel/pixels to be a part of the vasculature. Despite this clear definition, the border between unsupervised and supervised methods appears more vague in practice. For example, some unsupervised methods can provide features for a trainable estimator to obtain retinal vasculature by resulting in being grouped as super-

vised methods. In the same way, some methods can benefit from training in any part of their segmentation process; though, they mostly rely on expertise-motivated techniques. Despite including some training, these methods will be reviewed as unsupervised methods because the main focus/motivation of these methods seems to be based on expertise on vessel segmentation.

Unsupervised methods can be roughly grouped into four categories according to which properties of vasculature were exploited: shape based methods, intensity based methods, both shape and intensity based methods and spatial connectivity based methods.

Shape Based Methods

These methods solely describe vasculature with its shape-based properties such as its elongation or width. For example, matched filters model the cross-section intensity profiles of vessels by considering their tubular shape, Frangi's filter, Gabor wavelets and line detectors use the elongation of vessels as the main discriminator for segmentation and multi-scale approaches deal with various widths of vessels.

The cross-section profiles of vessels have been usually assumed to resemble Gaussian functions whose standard deviations vary depending on the diameters of vessels. Chaudhuri introduced Gaussian matched filters to improve the appearance of vessels by using this property of vessels [72]. Sofka and Stewart extended the usage of matched filter responses to multi-scales [38]. They also proposed an overall vesselness measure, called 'Likelihood Ratio Vesselness (LRV)', by combining matched filter responses obtained in multi-scales with 'confidence' measures generated for both vesselness and vessel boundaries. Confidence measures indicated the similarity between the estimated model and the appearance of a vessel in a fundus image. The overall vesselness measure was calculated by dividing a conditional likelihood given a pixel to be a vessel pixel by a conditional likelihood given the pixel to be a non-vessel pixel. These likelihoods were learned with a histogram-based method [73] by using the features obtained from vesselness and edge confidence measures in addition to vesselness and edge responses obtained from matched filters.

Nguyen *et al.* combined line detectors (which was explained in Section 2.3.1.) with a multi-scale approach [51]. The multi-scale apparoch was realised by varying the length of an edge detector for each scale. Vesselness maps were obtained by averaging the responses of line detectors at various scales after adding original image intensities to them. Zhang *et al.* calculated edge location probabilities for a probabilistic tracking method from vessel cross-section profiles obtained with multi-scale line detection [74].

Eigenvectors of Hessian matrices have been usually associated with the elongation of a tubular structure in images. One of the most used filters in retinal vessel segmentation, Frangi's filter [75], used Hessian matrix analysis to enhance the appearance of vessels by emphasising their elongation. According to [75], the orientations of blood vessels can be estimated by considering the eigenvectors with the smallest eigenvalues based on the observation that these eigenvectors reflect the minimal variations in the directions of blood

vessels in fundus images. On the other hand, Bekkers *et al.* adopted orientation analysis inspired by 'the cortical orientation columns in the primary visual cortex' for retinal vasculature analysis [76]. An orientation score was described as a function, which returns a score given both a specific location and a specific orientation. This function was obtained by convolving an image with various anisotropic wavelets. Bekkers *et al.* experimented on both cake wavelets and Gabor wavelets, and both invertible and non-invertible orientation scores. The orientation score was used in two new tracking methods, which will be explained in detail later.

Oliveira *et al.* augmented the responses of matched filters [72], those of Frangi's filter [75], those of Gabor wavelet filters [71] and contrast enhanced images to emphasise vessel locations in fundus images in order to use the specific strength of each filter. The reason behind this approach was explained with improving the performance of individual methods by combining their strengths such as noise insensitivity of Frangi's filter and *sensitivity to small vessels* of both matched filters and Gabor filters [62]. They used both median ranking and weighted mean for response combination. When median ranking was used for the combination, final segmentation masks were obtained with a global thresholding. On the other hand, in the case of using weighted mean, the weight for each filter response was learned with a Genetic Algorithm [77]. Then, the output of this combination was segmented with either fuzzy C-means or deformable models.

Intensity Based Methods

One of the most striking characteristics of blood vessel appearances in fundus images is having a lower contrast than that of image background. This property provides a motivation to utilise thresholding techniques for vasculature segmentation. However, the discriminative ability of this property can be weak to separate vessel pixels from non-vessel pixels due to non-homogeneous background illumination, changing contrast of vessels depending on their thickness and high levels of noise in fundus images. Due to these reasons, either thresholding techniques include some local information to detect vessels [3, 4, 78–80] or they can be applied to responses of appearance enhancing filters such as matched filters or Gaussian filters [81, 82].

Hoover *et al.* utilised an iterative thresholding method to segment vessels [3]. This method was applied to matched filter responses and incorporated both global and local properties. Budai *et al.* used hysteresis thresholding, implemented with Canny edge detector [83], to the outputs of Frangi's filter applied in multiple resolutions [78]. The resolution of Frangi's filter response in each scale was brought to the original one before applying thresholding. The final segmented image was the result of combining each thresholded image with an 'OR' operation. Odstrcilik *et al.* used a minimum error thresholding method [84], which assumed that an image can be represented with probability distributions of object pixels and background pixels, to the responses of matched filters [4].

Annunziata *et al.* used an image based threshold to remove pathological regions in a

fundus image, then retinal vasculature was detected by using a 'percentile-based' thresholding over the responses of Frangi's filter in multiple-scales after filling removed regions with inpainting [81]. The percentile-based thresholding was based on the assumption that the ratio of the number of vessel pixels to that of non-vessel pixels in a fundus image is similar in various images and a threshold, which was estimated by considering the percentage of vessel pixels in ground truth images, can be used to segment vessels. Similar to Annunziata *et al.*, Imani *et al.* applied adaptive thresholding to the responses of Morlet Wavelet Transform after removing pathologies in fundus images with the use of morphological component analysis (MCA) [82].

Roychowdhury *et al.* used global thresholding to detect larger vessels, then applied an adaptive thresholding to residual images obtained after removing estimated large vessels from related fundus images, to identify smaller vessels [79]. Neto *et al.* segmented vessels in two steps with a 'coarse-to-fine' fashion, where they roughly detected vessels in the first step by using an adaptive thresholding and then refined/improved this segmentation by using morphological reconstruction and curvature analysis [80].

Spatial Connectivity

Region growing and tracking methods mainly exploit the spatial connectivity of vessel pixels for vessel segmentation. These methods iteratively detect vessel pixels by starting from seed locations until stopping criteria are fulfilled. Among other segmentation methods, tracking methods can provide more detailed information about the topology and the geometry of retinal vasculature, such as vessel width and vessel orientation. However, tracking methods can be more prone to missing branches or false detection due to their dependence to the accuracy of previous steps of tracking when compared with segmentation methods based on pixel evaluation [74]. Because of the relevance of tracking methods to the proposed method in this thesis, they will be reviewed exclusively in Section 2.4.2.

Martinez-Perez *et al.* combined multi-scale approaches with a region growing algorithm [85]. In multi-scale approach, the geometrical properties of a vessel segment such as vessel width and orientation were estimated from the magnitude of gradient and maximum principal curvature (maximum eigen-value), which were calculated from Hessian matrices of intensities. Local maxima across the scales were used to select the best fitting scale. The discrimination of vessel pixels for region growing was based on these two features (width and orientation) of the selected scales. Martinez-Perez *et al.* applied two stages of region growing. In the first step, pixels with lower gradient magnitude for both vessels and background were grown. For vessels, these regions correspond to vessel interiors in longitudinal direction. In the second stage, the growing started for pixels located near the edge of vessels.

In a recent study, Zhao *et al.* proposed to detect thinner vessels by using a region growing algorithm [86]. In this study, a level set algorithm [87] was applied to detect larger vessels and remaining thinner vessels after this segmentation were located with the

region growing algorithm.

Both Shape and Intensity

The studies in this group can include model based approaches. In 2009, Al-Diri *et al.* proposed an active contour model called 'Ribbon of Twins (ROT) to detect vessel boundaries by enforcing 'width consistency' [88] whose details will be explained in Section 2.4.1. Zhao *et al.* proposed an IPACHI model, which was an extension of an active contour model (infinite perimeter active contour model (IPAC) [89], which was recently introduced for the detection of structures with irregular edges) by combining it with multiple sources of image information [90]. Apart from other active contour models, IPAC used infinite perimeter regularisation instead of the constraint of the shortest length, which was counted as one of the reason why the authors selected this model for vessel segmentation. Regarding the other contribution of this research, the usage of multiple sources of information for segmentation, Zhao *et al.* used vesselness information coming from local phase-based filter (LP) [91] and intensity values as two sources of image information.

2.4 Quantitative Vasculature Analysis

This type of analysis aims to extract quantifiable information about retinal vasculature such as vasculature morphology (e.g. vessel width, tortuousity) and vasculature topology (vein-artery identification, branching locations, branching angles). The extracted information may be directly used for clinical applications. These analyses usually perform iterative estimations of vessel centreline, boundary locations and landmark locations by starting from seed locations. In this section, vessel width estimation and bifurcation/crossing location detection methods will be reviewed.

Methods related to vessel width estimation may be examined in two groups; model-based and tracking-based.

2.4.1 Model Based Methods

Although the estimation of vessel width can be simply described with the distance between edge locations on both sides of a vessel, it poses big challenges due to the presence of central light reflex and hardly visible vessel edges due to noise, very low contrast and non-even illuminated background intensity levels. Chapman [92] showed that the estimation of vessel width based on the detection of vessel edge locations with an edge detector produced the least reliable results due to its similar response to central light reflex as vessel edges. On the other hand, the intensity profiles of vessel cross-sections have been used as the most reliable tool to estimate vessel widths. These methods can be categorised into two groups. The first group includes methods based on the direct usage of vessel cross-section profile for width estimation. The second group contains methods indirectly estimating vessel width from vessel profile.

The first group: One of the oldest methods in the first group is the full width half maximum (FWHM) proposed by Brinchmann-Hansen *et al.* for the estimation of retinal vessel width [93, 94]. FWHM considered the possibility of different intensity levels for either side of vessel and estimated vessel width from the horizontal distance between the locations with the average intensity levels of cross-section profile for each vessel side. These average intensity levels were independently calculated by considering the intensity range of profile for each vessel side. Later, Rassam *et al.* showed that the 'kick points' of a vessel could provide closer estimates of vessel widths to ground truth widths than those obtained with FWHM in fundus images [95]. The kick locations occur at the locations where blood columns meet with vessel walls, and these locations were observed with 'skew locations' in the slopes of a vessel profile. The reason why these locations appear in images was explained with the different light absorption coefficient of blood column from that of vessel wall during imaging. However, this method was stated to heavily rely on good focusing during imaging and to demand high resolutions of images; otherwise, kick points may not be visible in images with lower resolutions [6].

Apart from methods based on FWHM and kick points, Gregson *et al.* introduced a rectangular profile to describe a vessel cross-section profile [96]. This rectangular profile had a fixed height which was equal to the difference between the minimum and maximum intensities of the original vessel profile, while the width of this rectangular profile was adjusted in such a way that the area under the proposal profile was equal to the area under original vessel profile. Chapman *et al.* detected vessel edge locations based on the maximum variation of the slopes of intensity profiles [92]. They fit lines with linear regression to vessel cross-section profiles in a sliding window across the profile. The locations where the steepest lines were obtained were used for the identification of edge locations. This method was called 'the sliding linear regression filter (SLRF)'. Chapman *et al.* showed that this method estimated more consistent vessel width than the estimations made with fitting a double Gaussian to the vessel profile and with using Sobel edge detection algorithm. Also, they reported that predicting edge locations by using Sobel algorithm was not reliable due to its responses to centreline light reflexes as if they represented vessel edges.

On the other hand, Chutatape *et al* modelled a vessel cross-section with a negative step gate function and calculated vessel width from zero crossing locations of the convolution of this function with a second order derivative Gaussian matched filter [72], which was followed by other studies using the similarity of vessel cross-section to Gaussian functions and their derivatives such as modelling vessel profiles with $1D$ Gaussian functions [97] or $2D$ [34] and modelling of central light reflex with 'twin' Gaussians [98]. They convolved cross-section profiles with a second order Gaussian matched filter for the estimation of vessel edge locations, where zero-crossing locations of generated responses were assigned as vessel edge locations. Later, they used this modelling in a probabilistic vessel tracking application [99].

Zhou *et al.* estimated vessel edge locations by fitting a Gaussian function to vessel profile

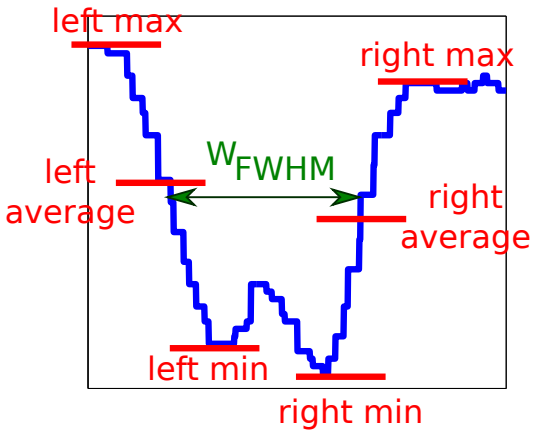
by using matched filters [97]. The distance between an estimated centreline location and either of edge locations was assumed to be 1.96σ of the Gaussian function. This profile model was used for vessel tracking, where a better accuracy for vessel width estimation was reported to be obtained with the proposed model than to be predicted with FWHM. The first time, Gao *et al* [98] modelled the cross-section profiles of vessels with central light reflexes by using 'twin' Gaussian functions. In this model, a regular vessel profile and the reflection were modelled with separate Gaussian functions. The difference between these functions was utilised to represent the vessel cross-section with the reflection.

Lowell *et al.* used a $2D$ Gaussian function to model the cross-section of a $2D$ vessel segment [34]. They also modelled the central light reflex with $2D$ 'Difference of Gaussians', where a Gaussian with a smaller magnitude was extracted from one with a larger magnitude. When compared with the methods based on $1D$ vessel profiles, namely FWHM [93, 94], Gregson [96] and a 1-D Gaussian [97], Lowell *et al.* reported more consistent width estimation. Despite the success of their method to deal with tortuous vessels and central light reflex, Lowell *et al.* also reported that the method could fail in the case of identification of crossing vessels or very closely located parallel vessels.

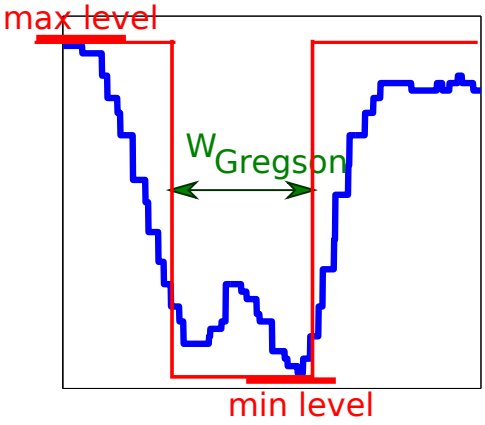
Figure 2.1 demonstrates some models describing vessel cross-section profiles in a direct way for the estimation of vessel width.

The second group: In 2009, Al-Diri *et al.* introduced 'the extraction of segment profiles (ESP)' algorithm which combined vessel segmentation and vessel width estimation [88]. This algorithm contained tramline algorithm [46] to estimate vessel pixels, an active contour model called 'Ribbon of Twins (ROT)', which was introduced to detect vessel boundaries by enforcing 'width consistency', and an algorithm for the estimation of vessel connectivity. After estimating vessel centrelines with the tramline algorithm, a segment growing algorithm was applied to iterate and to connect boundary locations detected by ROT over vessel segments. Another algorithm [100] was used to connect these segments. Al-Diri *et al.* observed that their method can deal with the central light reflex and noisy and blurry boundaries. However, they also observed that their method can fail when vessel boundaries were not able to be detected due to parallel running vessels in a close distance or overlapping vessels, sharp changes on the gradient of background intensity or vessels being close to FOV stops. Moreover, vessels whose diameters were less than 3 pixels were detected with a bigger error. The reason behind this increased error was explained with inadequacy of tramline algorithm. They also stated that the performance of the method was sensitive to some of its parameters.

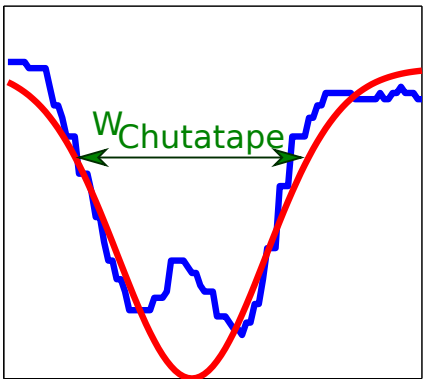
Xu *et al.* proposed an algorithm to estimate vessel width based on the simultaneous estimation of vessel edges by using a graph search method [47]. In this study, each boundary of a vessel was modelled with a surface slice in a $3D$ graph. In order to generate a cost image, they convolved one dimensional first order derivative of Gaussian functions parallel to vessel cross-sections with the green channel of fundus images. They estimated vessel boundary locations by using centreline locations obtained from segmented fundus images.



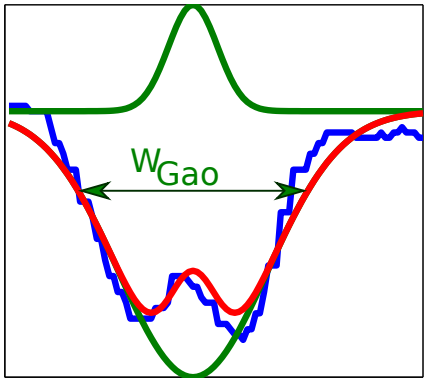
(a) *FWHM [93]*



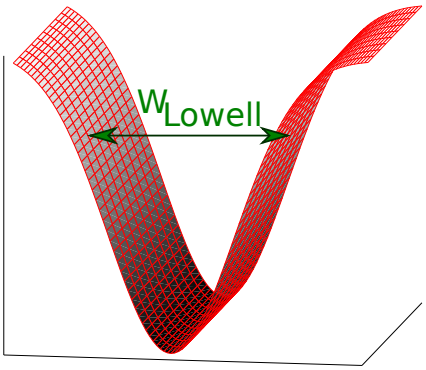
(b) *Gregson's rectangular profile [96]*



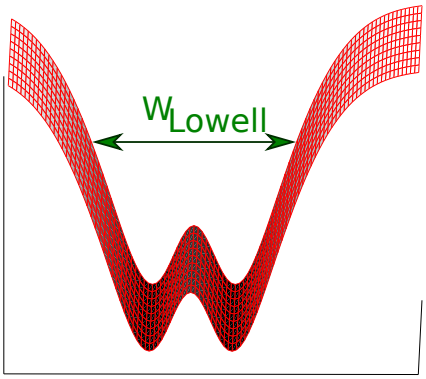
(c) *Chutatape et al. [99]*



(d) *Gao et al. [98]*



(e) *Lowell et al. [34]*



(f) *Lowell et al. [34]*

Figure 2.1: Some proposed models for the estimation of vessel widths. Blue lines in (a)-(d) are identical and show a $1D$ vessel cross-section intensity distribution. Red lines in (a) show the locations of related intensities (e.g. left max) while those in (b)-(d) demonstrate the approximated representations of the intensity distribution. The intensity distribution is approximated in (d) by adding two Gaussian-like curves in green in order to model the central light reflex. Red surfaces in (e)-(f) approximate a $2D$ vessel cross-section intensity distribution.

As a result of the dependency to initial binary vessel maps, the accuracy of vessel boundary estimation was stated to vary. Also, the performance of the method was reported to depend on the resolution of the image, where the performance was stated to be degraded in low resolution images. On the other hand, they argued the method to be robust in the case of edges to be blurry and not showing high contrast. Also, they considered the unequal distances of edge locations to estimate vessel centrelines.

Apart from the aforementioned studies, Lupaşcu *et al.* used a Hermite model instead of Gaussian functions by discussing that Gaussian functions can cause under-estimation of vessel widths for large vessels [17]. Lupaşcu *et al.* introduced a $3D$ parametric surface model for the representation of vessel cross-section profiles of retinal vessels. This surface model was based on the multi-resolution Hermite model proposed by Wang *et al.* [101] for the modelling of $1D$ cross-section profiles of retinal vessels. Lupaşcu *et al.* extended Wang *et al.*'s model by using multiple cross-section profiles then generated cross-section vessel surfaces by combining these $1D$ profiles. They also modified Wang *et al.*'s Hermite model for non-symmetric boundaries around centreline. After generating cross-section surfaces, Lupaşcu *et al.* used a training-based method to estimate vessel widths. Ensembles of bagged decision trees were trained for regression to predict widths from the parameters of the cross-section surfaces which were found to be the best fit to local vessel region. However, they observed that estimated widths for large vessels were not very close to reference widths. The reason behind this failure was explained with not having a balanced training dataset regarding vessel widths. The existing dataset was found to have less number of large vessels, when compared with the number of thin vessels.

Similar to Lupaşcu *et al.*'s approach, Araújoa *et al.* used $2D$ cross-section surfaces for model fitting to vessel segment, which considered the asymmetry of vessel boundaries and the existence of central light reflex, and estimated vessel width with a training based method [50]. The model used by Araújoa *et al.* was a $3D$ parametric model, 'DoG-L7' and the training based method for width estimation was Random Forests using ensembles of bagged decision trees. They reported better performance than Lupaşcu *et al.* regarding the usage of a multi-resolution Hermite model. Araújoa *et al.* explained the superiority of their performance with both better modelling of the asymmetry of vessel cross-section profiles and pre-processing the image prior to model fitting. They also claimed that the performance of their method did not depend on the vessel width and the method was able to generate reliable estimates in noisy and pathological regions. On the other hand, because of being a supervised method, the performance of their method was related to the uniform distribution of vessel widths in the training dataset.

Apart from previous studies requiring the detection of vessel edges or the segmentation of vessels, Aliahmad and Kumar estimated vessel width by associating it with 'self-similarity' characteristics of vessel cross-section described by Hicohus 's dimension [102]. Aliahmad and Kumar found that the relation between Hicohus 's dimension and vessel width sensitive to background noise level. In order to reduce this sensitivity, they built a $3D$ model, which related vessel width to noise level in addition to Hicohus 's dimension. They trained their

model on synthetic vessel images with known vessel widths and with various noise levels.

2.4.2 Tracking Methods

When compared with pixel based segmentation, tracking provides the connectivity of the vasculature in addition to giving extra information related to topology, such as centreline, boundary and bifurcation/crossing locations, and geometrical features, such as the orientation and the width of vessels at particular locations. Tracking can be realised with deterministic and probabilistic methods depending on the usage of probability in the estimation of next step parameters, regardless of the use of probabilities in other stages of methods.

Deterministic Tracking

Some examples of non-probabilistic tracking applications on retinal images were proposed by Zhou *et al.* [97], Sofka and Stewart [38] and Bekkers *et al.* [76]. Zhou *et al.* tracked vessels by using Gaussian matched filters [97]. This method predicted vessel boundaries based on the standard deviation of a Gaussian function showing the best fit to the vessel. They used an adaptive step size to accommodate the change on the direction of a vessel. The larger the change was, the smaller the step size was. This method was observed to be prone to failure in the tracking of crossing vessels, if both had similar brightness levels. This failure was caused by not having any intensity difference to identify edge locations.

As explained in Section 2.3.2, Sofka and Stewart proposed a vesselness measure called 'Likelihood Ratio Vesselness (LRV)' and used it for the segmentation of vessel centrelines. Also, they evaluated the success of this measure for the tracking of vessel centrelines. For this application, initially, they used tracking with matched filters to generate seed locations prior to the tracking with LRV. For the former tracking, an initial tracking direction was estimated as the eigenvector with the smallest eigenvalue calculated from Hessian matrice. An initial vessel width at this location was predicted from the scale of matched filter that generated the maximum response. The size of the search space of tracking direction, vessel width and step size was fixed for all locations. After this step was completed, the tracking with LRV started and eliminated locations whose LRVs were not above a threshold. When compared with another algorithm based on tracing parallel edges [103], Sofka and Stewart observed that their method was better on the detection of low contrast vessels and very close parallel vessels and also dealing with pathologies.

As explained in Section 2.3.2, Bekkers *et al.* introduced two tracking methods based on orientation analysis for retinal vasculature analysis [76]. The methods were claimed to tackle main concerns of tracking of retinal vasculature such as overlapping vessels, parallel vessels located in a very close distant, a range of vessel widths and high curvature. The first algorithm for tracking, called 'edge tracking in orientation scores (ETOS)', was about the concurrent detection of both vessel edges by using orientation scores, which were obtained by using cake wavelets. In this algorithm, they also incorporated the smooth

change of vessel width along a vessel with an edge location estimation process with 'a probability envelope' demonstrating the most likely locations of both edges. After the detection of edge locations, they estimated the current orientation as the one maintaining the highest orientation score for both edges and iterated the algorithm accordingly. The second algorithm introduced was 'multi-scale vessel center-line tracking in orientation score (CTOS)'. This algorithm included Gabor wavelet responses in multiple-scales to mitigate the effect of the central light reflex on the estimation of vessel centrelines. CTOS was stated to be quicker than ETOS but less stable at crossing locations or in the detection of parallel running vessels due to using multi-scale approach.

Prior to the tracking, they manually selected seed locations at bifurcation locations and at the starting points of vessels. They found ETOS better at dealing with challenges aforementioned and at estimating vessel widths than CTOS. They also stated that ETOS overestimated widths of vessels smaller than 7 pixels and underestimated those larger than 7 pixels. Regarding the topology of retinal vasculature, they observed that their method was reliable with rare estimation of false vessels.

Probabilistic Tracking

Chutatape *et al* introduced extended Kalman filters for retinal vessel segmentation [99]. Kalman filters initially estimated next values of centreline locations, vessel widths and tracking directions in a probabilistic way by using both the current values of these variables and their all previous values. These initial estimates were later updated by estimating vessel widths from the convolution of matched filters (second order derivatives of Gaussian functions) with cross-section intensity profiles of the vessels. In order to predict bifurcation/crossing locations, they also added a simple branch detection method to this tracking scheme. This detection method was based on estimating the presence of branches depending on the responses of the same kind of matched filters on both sides of the estimated vessel through a predetermined range of angles.

Yin *et al.* introduced a probabilistic tracking method to segment retinal vasculature and to analyse it by estimating vessel edges and centrelines in addition to bifurcation/crossing locations [104]. They assumed a vessel cross-section to be Gaussian shaped and background intensity to be constant. During the probabilistic tracking, they used a dynamic search window in the shape of a semi-ellipse, whose major axis length was adapted to the previous estimate of vessel width, and whose minor axis length was to the previous estimate of vessel curvature. They sampled candidate boundary locations on this window. They estimated the type of a vessel part regarding connectivity by maximising the criterion of maximum a posteriori (MAP). Evaluated connection types were single vessel, bifurcating vessels and crossing vessels. The number of vessel edge locations sampled on the dynamic search window depended on the connection types: 2 for single vessel, 4 for bifurcating ones and 6 for crossing vessels. During tracking, the probability of each vessel connection type was calculated over various combination of candidate edge locations and

the connection type with the maximum likelihood was selected as the one best explaining connection type of the current tracking. According to the best hypothesis, the probabilities of candidate edge locations were calculated both by evaluating sampled locations between these edge locations regarding their belongness to a vessel and by evaluating remained sampled locations, outside of these edge locations, with respect to their belongness to background image by considering their intensity levels. The candidate edge locations with the highest probabilities were selected as estimates of edge locations at the current step of tracking. Later, Zhang *et al.* [74] improved Yin *et al.* [104]'s method by combining vessel longitudinal profiles, obtained with multi-scale line detection, with vessel cross-section profiles for the calculation of edge probabilities.

Nayebifar and Moghaddam intoduced Particle Filters to estimate the trajectory of retinal vasculature [105]. Their probability density distribution was generated by multiplying the green and blue channels of an RGB fundus image pixel by pixel and then by applying median filter. Their proposal distribution was a uniform distribution along a ring. The initial radius of the ring was adapted to the vessel under tracking by examining intensity profiles, obtained along the vectors connecting the center of the ring to particles on the ring.The standard deviations of the intensity profiles were used to estimate thresholds to determine locations of particles (e.g. inside or outside vessel or, crossing a vessel). If any particles cross the vessel, their intensity profiles were used to locate vessel wall locations. These vessel wall locations were eventually utilised to update the radius of the ring. After this update, particles were propagated on the new ring and those with larger intensity values than the local threshold were evaluated by using their intensity profiles (obtained as before) in terms of being located inside the vessel or in a crossing vessel or in a parallel one. The particles found inside vessels were clustered with the quality threshold clustering (QT-clustering) method [106]. The center of this cluster appeared to be used for the next center location of the ring. The vectors between previous and current estimates of center locations were stored as the traced path. The existence of a bifurcating vessel was estimated from having two clusters. They reported that the method missed some small vessels during tracking.

2.4.3 Landmark location detection

The detection of landmark locations such as bifurcation and crossing locations is an important task regarding facilitating better segmentation of vessels in these regions and also providing information for building the connectivity of vasculature [33]. Especially tracking algorithms may explicitly need to deal with these locations by detecting them during tracking [97,99,104,105] or prior to tracking with external methods [41,48,107–110]. Possible reasons for detecting these locations may be providing seed locations to trace new branches or a criterion to stop tracking or avoiding the disruption of tracking due to the distortion of vessel shape and the complexity of intensity distribution of vessel cross-sections at these locations. On the other hand, the detection of these landmark locations

may be challenging because of the variations on their patterns and the involvement of a range of vessel widths at these locations.

Some landmark detection algorithms have searched bifurcation/crossing locations on vasculature skeleton images in order to simplify the problem by removing the effects of vessel widths on bifurcation and crossing patterns and those of imaging modalities [48, 107, 110]. Bhuiyan *et al.* [48] and Calvo *et al.* [107] explored these locations in retinal images after applying the thinning operation to the binary vasculature obtained with a segmentation method. On the other hand, Morales *et al.* directly generated the skeleton of the vasculature from a retinal image in order to remove the noise, such as missing or false vessels, inherited from an additional step of segmentation [110].

Bhuiyan *et al.* utilised a rotational invariant mask to detect candidate landmark locations [48]. They estimated validity of candidate locations, the accurate locations of correctly detected landmark locations and their types (bifurcation or crossing) by examining geometrical and topological characteristics of these locations such as the connectivity of pixels in the centreline or the number of vessels passing from these locations. Calvo *et al.* used 'intersection number' as a measure to identify possible bifurcation and crossing locations in the vessel skeleton [107]. This measure was related to the number of neighbour pixels of a pixel, whose connectivity type was on examination, in the vasculature skeleton. If a pixel had more than two neighbours, it was accepted to be a candidate landmark location. In order to examine the validity of these candidate locations, they performed tracking from an end point of a vessel to the other end or to a candidate landmark location and removed those mistakenly detected. However, the performance of their method was reported to depend on the parameter selection. Morales *et al.* applied 'the hit-or-miss transformation (HMT)' to vasculature skeletons by using 'composite structuring element', which resembled various patterns of bifurcation and crossing [110]. They reported that their performance on the detection of landmark locations (both bifurcation and crossing) was much better than that of Calvo *et al.* [107].

Apart from the approaches using vasculature skeleton as the search space, there are other methods performing directly on binary retinal vasculature image [41, 108] or processed vesselness probability maps [109]. These methods were found to be more robust because they avoided additional noise occurring due to thinning process. Azzopardi and Petkov used trainable filters to detect only vessel bifurcation locations, which were called 'Combination Of Shifted Filter Responses (COSFIRE)' [41]. Filters defined in this method can automatically learn complicated properties of connectivity patterns at bifurcation locations by combining simpler features such as orientations of vessels and their widths. The authors stated that the more examples from various vessel connections were observed during training, the more types of bifurcation locations can be detected in the validation stage. They also observed that trained filters could detect similar patterns to some extent. In a follow up study [108], the authors also reported robustness to missing links at bifurcation locations that occurred as a result of segmentation.

Fang *et al.* used the determinant of Hessian matrices calculated in multi-scales to

detect landmark locations, where vessel connection patterns were described with higher curvature, on vesselness probability maps generated by a deep network designed for the cross-modality learning approach [109]. They argued that using vessel probability maps reduced the effect of pathology and noise in the detection of these landmark locations. The local maxima locations on final processed probability maps were assigned as landmark locations. They observed that their method detected more locations with more reliability than those estimated with the COSFIRE algorithm [108].

Chapter 3

Retinal Vasculature Segmentation

3.1 Introduction

3.1.1 Introduction to Deep Learning for Segmentation

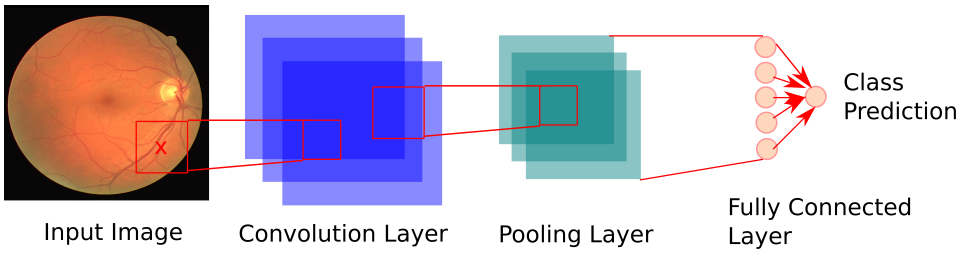
Artificial neural networks with more than two hidden layers are called *deep neural networks* [111]. Deep neural networks have various architectures depending on their types of connections between layers or operations performed in a layer or unit types in a layer. For example, a multi-layer perceptron has feed forward connections while a Recurrent Neural Network (RNN) has recurrent connections which provide previous signals to be processed along with the current signal during the training. A Convolutional Neural Network (CNN) has convolution layers, performing convolution between input data and a series of feature detectors. On the other hand, a Deep Belief Net (DBN) has stochastic units and connections between layers are directed from the top layer to the bottom layer.

These networks have been applied to a range of problems from image analysis to language processing [112]. Deep learning has been found very successful at image segmentation, with very extensive examples of object, human or semantic segmentation in natural images [113, 114] To date, there have been limited attempts at organ segmentation in medical images, such as brain part segmentation from MRI images and cell segmentation from microscopy [113, 115]. Applications of deep networks to retinal vessel segmentation have also started to appear in recent years [36, 43–45, 55, 57].

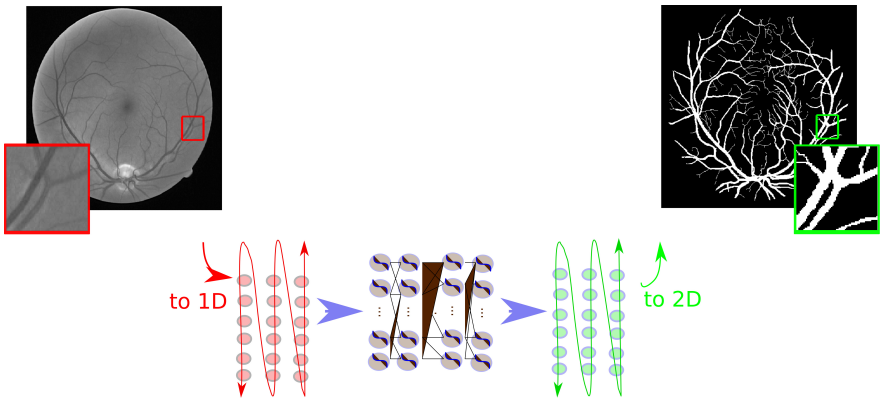
One key advantage of using a deep network in medical image segmentation can be the adaptation of the method to segment new data, acquired by a different acquisition system, by only retraining the network. In contrast, traditional methods require the adaptation for the segmentation of new data, often entailing the redesign of features according to the new dataset or searching for optimum parameters. On the other hand, the training of a deep network can be challenging in terms of collecting large amounts of labelled data, and this can be viewed as the biggest disadvantage of this method.

3.1.2 Segmentation: Cross-Modality vs Classification

Deep learning approaches for retinal vessel segmentation were grouped in two categories in Chapter 2: pixel-wise segmentation methods and spatial connectivity evaluation based methods. Of these, the most commonly used are the pixel-wise segmentation methods. Such methods usually individually classify each pixel in a fundus image. In the case of using a CNN, a region whose center is the pixel to be classified is evaluated. Despite using supportive information coming from neighbouring pixels, the classification of a pixel does not depend on labels assigned to surrounding pixels. Figure 3.1a shows an example application of this approach by using a CNN.



(a)



(b)

Figure 3.1: (a) An application of classification based retinal segmentation approach with a shallow Convolutional Neural Network (This network can be converted to be a deeper one by adding extra layers of convolutional and pooling before fully connected layer. The number of fully connected layers can be also increased) : The class (belong to a vessel or not) of pixel shown with x is evaluated through the network (b) An application of cross-modality learning based retinal segmentation approach with a fully connected network.

On the other hand, methods based on spatial connectivity of label pixels consider the similarity of pixels in the same neighbourhood regarding their memberships to the same object. A subgroup of these methods, those based on cross-modality learning, treat the input image and its segmentation mask as if they are from two different data modalities [43] and aim to transform one data modality to another one through a shared representation. The introduction of cross-modality learning to fundus image segmentation was made by Li *et al.* [43, 66]. In this approach, vessel segmentation involves generating a label vector/matrix of the same size as the input image. Regarding vasculature segmentation, the

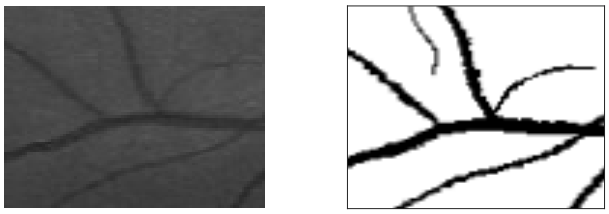


Figure 3.2: A fundus image patch in the left and its vessel map in the right

network performs a complete transformation from fundus image patches to their corresponding vessel maps.

Regarding cross-modality learning, auto-encoders or similar networks, such as Restricted Boltzmann Machines (RBMs), are commonly used to learn a shared representation of two data modalities [43, 116, 117]. Figure 3.1b shows an example application of this approach by using a deep fully connected network. When compared with the classification based approach, cross-modality learning may consider the consistency of pixel labels to a greater extent and eventually may lead to better segmentation performance. This claim can be supported easily with the results of Liskowski and Krawiec [36], where they showed that generating labels for a group of neighbour pixels, instead of only one pixel, significantly improved segmentation performance.

3.1.3 A Proposed Method for Retinal Vessel Segmentation

In this current work, the cross-modality learning approach is explored for vessel segmentation because it considers the solidarity of label pixels from the same class during the segmentation. Also, this approach suits the nature of the problem as explained below.

Fundus image patches can be viewed as noisy versions of vessel masks. Figure 3.2 shows a fundus image patch and its possible vessel mask. As seen from the figure, the relation between these fundus image patches and their vessel masks is not so complicated, when compared with the relation between the samples of audio and video data in previous applications of cross-modality learning [117]. In an unrealistic case, even a linear mapping between fundus images and their vessel maps can be possible if the noise level is really low, virtually zero, for fundus images. Because of this similarity, a shared representation learned between fundus images and their vessel masks can reflect the characteristics of both data modalities by highlighting the main structures of interest, blood vessels, at the same time.

The implementation of this approach can be achieved with a generative learning method, such as through using a DBN. Why a generative learning method is selected can be explained by two reasons. The first is because both generative learning and cross-modality learning require a good representation of the input data. The second is that the features learned during the generative training of a DBN, which can also be called *pretraining*, can be manipulated to obtain useful features for cross-modality learning.

3.1.4 The Remainder of this Chapter

Initially, auto-encoders will be reviewed as a way of learning hidden representation from data from the perspective of cross-modality learning. There are various techniques to improve the representation learned during the training of autoencoders. Among them, three techniques will be explained. The first is denoising, which underlies the principle of denoising autoencoders. The second is using stochastic units in training. One class of networks with stochastic units is known as *RBMs*. The third one is using deeper representations, which is possible through stacking autoencoders or RBMs.

The main contributions of this chapter are a denoising technique, which will be introduced for cross-modality learning, and an implementation of the cross modality learning approach by using a DBN improved with this denoising technique for the segmentation of retinal vasculature. The improved hidden representation obtained through usage of this denoising method and the proposed network will be investigated with experiments on the DRIVE dataset, a widely used fundus dataset for the evaluation of vessel segmentation.

3.2 Autoencoders

3.2.1 Autoencoders & Their Applications

Traditional autoencoders can be defined as networks of two layers. Figure 3.3a demonstrates the architecture of an autoencoder. The top layer is called the hidden layer while the bottom layer is called the input layer. Autoencoders have directed connections between these layers. Depending on the size of the hidden layer relative to that of the input layer, these hidden representations take different names. The representation is called undercomplete if the number of hidden units is smaller than that of input units and overcomplete if larger.

The units in these layers are deterministic, which means that the activations of the units are directly used as the outputs. Depending on the activation function, these units can be called *sigmoid units* if a sigmoid function is used to squash the weighted addition of signals coming from connected units to the range of $[0.5, 1]$. To be more precise, the activation of a sigmoid unit is calculated from $a(h_j) = \sigma(\sum_i W_{ij}x_i + b_j)$, where $a(h_j)$ is the activation of a unit h_j and b_j is the bias of this unit. x_i is an input unit connected to the hidden unit h_j with the weight W_{ij}. $\sigma(\cdot)$ is the sigmoid function, $\sigma(x) = (1 + \exp(-x))^{-1}$.

An autoencoder is trained to learn a hidden/latent representation of the input data in the hidden layer. The training of an autoencoder is usually performed by minimising the error between the input data and its reconstruction with backpropagation. During this training, two functions are learned: encoder and decoder. The encoding function $f(\cdot)$ transforms the input data x to a representation h in the hidden layer and the decoding function $g(\cdot)$ converts this representation to the approximation of the input data x'. These functions respectively correspond to affine mappings $\sigma(Wx + b)$, $\sigma(W^T h + b)$, when both input and hidden layer units are binary and weights in the encoder and decoder are tied,

symmetric. These functions are shown with an unfolded autoencoder in Figure 3.3a.

Autoencoders can be used for dimensionality reduction and information retrieval tasks by using hidden representations as descriptors or features of the input data [118]. These hidden representations can provide concise and discriminative information for further analysis such as classification and regression tasks.

The representation capability of an autoencoder can be improved by preferring undercomplete representation in its architecture or with some regularization on its training such as using binary units in the hidden layer or limiting the activation of hidden units with sparsity [111]. Using undercomplete representation prevents the network from learning an identity matrix. The undercomplete representation is expected to force the network to reconstruct the input data by using a limited number of hidden units, which may correspond to distinctive features of the input data. Similarly, binary hidden units limit the capacity of the model learned during the training, which encourages the network to learn more descriptive features of the input data to regenerate it.

Recently, denoising was introduced by Vincent et al. to improve the performances of autoencoders [7]. Their way of denoising randomly suppresses some information in the input data temporarily at each iteration, by adding Gaussian noise to it or replacing a proportion of the data with zero; however, the network is encouraged to reconstruct the original version of the input data [7]. The denoising forces the network to reconstruct the input data by using higher level interactions between the input units. This method has been shown to generate global features, which can compensate the missing information in the noisy input data. When denoising is used in the training of autoencoders, they are called denoising autoencoders. Now, they mostly replace autoencoders due to their better performances.

Also, sparsity can be used to improve data representation in autoencoders, where one limits the the number of hidden units activated in each training epoch to a pre-determined number [119]. Because a few hidden units are allowed to be activated, sparsity mostly reduces the cooperation between hidden units when reconstructing the input data, by leading to each hidden unit to be specialised on a specific feature [111].



Figure 3.3: (a) An autoencoder (b) An unfolded autoencoder by showing encoder $f(\mathbf{x})$ and decoder $g(f(\mathbf{x}))$ functions.

3.3 RBMs

3.3.1 A Traditional RBM vs A Traditional Autoencoder

Similar to traditional autoencoders, traditional RBMs have two layers. However, an RBM [120–122] is a generative model and it learns the statistical distribution of the input data, while a traditional autoencoder learns only a deterministic mapping between the input data and a latent space. Because of it being a generative model, an RBM can generate plausible samples from this distribution when it is activated from its hidden layer. This function of an RBM does not limit it from being used in applications where autoencoders are employed to generate latent representations.

Figure 3.4 demonstrates the architecture of an RBM. The bottom layer is called the *visible layer*, because the units in this layer are observed. The second layer is called *hidden layer*, because units here are not observed. The connections between these layers are bi-directional. As a result of this, the weights of these connections, which were described with encoder and decoder functions for autoencoders, are symmetric for RBMs.

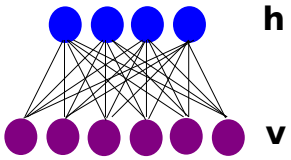


Figure 3.4: An RBM architecture, where **v** represents visible units and **h** hidden units

Despite having undirected (bi-directional) connections, RBMs can learn the dependency between a pair of visible units through two connections: the first one is between either of the visible units and a hidden unit and, the second one is between this hidden unit and the other visible unit [111]. The undirected connections between visible units allow learning of dependencies between units regardless of their location in the input vector.

The ability of RBMs to learn statistical distribution is partly due to them having stochastic units. Both autoencoders and RBMs aim to increase the representation of input data during training but training in RBMs is made in a stochastic way, usually with the Contrastive Divergence (CD) [111], which will be explained in Section 3.3.4. The activations of hidden units in autoencoders, calculated with $a(h_j) = \sigma(\sum_i W_{ij}x_i + b_j)$, correspond to the probabilities of hidden units to be active in RBMs, where hidden units are binary. Then, the states of hidden units in RBMs are determined after sampling from these probabilities. This sampling and using CD training are the most discriminative characteristics of RBMs from autoencoders.

Traditional RBMs only consist of binary units in both visible and hidden layer. These RBMs are called *Bernoulli-Bernoulli RBMs (BB-RBMs)* because the probability distribution learned with binary units represents a Bernoulli probability distribution. However, BB-RBMs have limited representation capacity for real valued data. The representation of this type of data can be realised through using specialised units such as *Gaussian Units* [123] and *Rectified Linear Units* [124], which provide better approximation to Gaus-

sian distribution, instead of binary units. However, the training of RBMs when Gaussian or rectified linear units are used is not easy because these units make the training very sensitive to learning rate and can be saturated during training, despite the existence of some remedies [125, 126, 126]. Binary units provide a more stable training of an RBM because using a squashing function (sigmoid) in the calculation of probabilities keeps the learning signal limited [125].

3.3.2 The Traditional RBM from A Generative Learning Perspective

An RBM is an energy based generative model. Because of an RBM being a generative model, one wants to learn a model over visible and hidden units representing the statistical distribution of data in an RBM, which is stored in the weight matrix of the connections between the visible and the hidden layers. With this model from a trained RBM, one can generate new plausible data samples from this distribution by activating units in the hidden layer and sending this activation to the visible layer. In order to learn this probability distribution of data, the training of an RBM aims to maximise the probability of observed data given in (3.1).

$$P(\mathbf{v}) = \frac{P(\mathbf{v}, \mathbf{h})}{P(\mathbf{h}|\mathbf{v})} \tag{3.1}$$

Because of it being an energy based model, the probability function of an RBM is related to an energy function which is an input to an exponential function as given in (3.2).

$$P(\mathbf{v}, \mathbf{h}) = \frac{1}{Z} e^{-E(\mathbf{v},\mathbf{h})} \tag{3.2}$$

where $P(\mathbf{v}, \mathbf{h})$ is the probability of the model to be in both a particular state vector of visible units $\mathbf{v}$ and that of hidden units $\mathbf{h}$.

The energy function $E(\mathbf{v}, \mathbf{h})$ in (3.3) includes the state vector of visible units, $\mathbf{v}$, and that of hidden units, $\mathbf{h}$, in addition to trainable parameters, which are weights between visible and hidden units and bias values of visible and hidden units.

$$E(\mathbf{v}, \mathbf{h}) = -\mathbf{v}^T \mathbf{W}\mathbf{h} - \mathbf{b}^T \mathbf{v} - \mathbf{c}^T \mathbf{h} \tag{3.3}$$

where $\mathbf{W}$ is the weight matrix between visible and hidden units.

Describing the probability function in this way guarantees that the probability of any possible combination of $\mathbf{v}$ and $\mathbf{h}$ is larger than zero because this exponential function does not produce zero or negative values for real valued inputs (of energy function), even though, any related probability value can be virtually zero [111]. In order to calculate normalised probability values, the output of this exponential function must be divided by a constant value Z given in (3.4), which is the sum of the exponential negative energy of all possible state vectors of $\mathbf{v}$ and $\mathbf{h}$.

$$Z = \sum_{\mathbf{v},\mathbf{h}} e^{-E(\mathbf{v},\mathbf{h})} \tag{3.4}$$

Because it is not possible to calculate this large number of the state vector combination, this probability function is intractable and cannot be calculated analytically. However, Monte Carlo based sampling methods can be used to approximate this probability calculation. The most common one used in the training of an RBM is Gibbs sampling [111], which will be explained later.

Despite the joint probability of the states of visible and hidden units being intractable, the conditional probabilites of hidden units given visible units can be calculated analytically. The conditional probabilities of hidden units given visible units and the conditional probabilities of visible units given hidden units are independent because of the lack of connections between units in the same layer. Therefore, these conditional probabilities may be factored as in (3.5) and (3.6). This property permits the update of the conditional probabilities of visible or hidden units as a block.

$$p(\mathbf{h}|\mathbf{v}) = \prod_i p(h_i = 1|\mathbf{v}) \tag{3.5}$$

$$p(\mathbf{v}|\mathbf{h}) = \prod_j p(v_j = 1|\mathbf{h}) \tag{3.6}$$

The conditional probability of a hidden unit given the states of visible units can be calculated from (3.7) for a *Bernoulli-Bernoulli RBM*. In the same way, the conditional probability of a visible unit given the states of hidden units can be calculated from (3.8) for the same RBM.

$$p(h_i = 1|\mathbf{v}) = \sigma\left(\sum_{j=1}^{m} w_{ji} v_j + c_i\right) \tag{3.7}$$

$$p(v_j = 1|\mathbf{h}) = \sigma\left(\sum_{i=1}^{n} w_{ji} h_i + b_j\right) \tag{3.8}$$

where $\sigma(x)$ is the sigmoid function. w_{ji} represents the weight of the connection between visible unit v_j and hidden unit h_i. b_j and c_i respectively denote the bias of visible unit v_j and that of hidden unit h_i.

3.3.3 Training RBMs

As suggested by (3.5) and (3.6), the calculation of probability values requires many multiplications. Addition is generally preferred over multiplication because it gives more accurate and faster results. So, log probability (the logarithm of the probability) is usually preferred over probability in training algorithms because it turns multiplications in a probability calculation to sum operations in a log probability calculation. The log probability of the pair of visible and hidden state vectors is given in (3.9) for the probability

in (3.2). Referring to (3.9), the maximization of log probability of the pair of visible and hidden state vectors means the minimization of the energy function calculated from these state vectors. When the energy function given in (3.3) is differentiated with respect to weights, (3.10) is obtained, where the derivative is simply the multiplication of the states of related visible and hidden units, because the energy function is a linear combination of weights and bias terms.

$$log(P(\mathbf{v},\mathbf{h})) = -E(\mathbf{v},\mathbf{h}) \tag{3.9}$$

$$\frac{\partial E(\mathbf{v},\mathbf{h})}{\partial w_{j,i}} = -v_j h_i \tag{3.10}$$

The main aim of the training of an RBM is maximising the probability of the input data (3.1). One can also write this probability in terms of log probability as given (3.11).

$$\log(P(\mathbf{v})) = \log(P(\mathbf{v},\mathbf{h})) - \log(P(\mathbf{h}|\mathbf{v})) \tag{3.11}$$

The derivative of this log probability of data (visible state vectors) with respect to weights can be calculated as (3.12).

$$\frac{\partial \log(P(\mathbf{v}))}{\partial w_{j,i}} = \frac{\partial \log(P(\mathbf{v},\mathbf{h}))}{\partial w_{j,i}} - \frac{\partial \log(P(\mathbf{h}|\mathbf{v}))}{\partial w_{j,i}} \tag{3.12}$$

(3.12) has two terms: the first one belongs to the model when no data is observed and the second one when data vectors are clamped to visible units. Due to being intractable, the derivative of the probability of model requires an approximation method, such as Gibbs sampling. The second term can be calculated analytically, because the calculation of conditional probabilities of hidden state vectors is possible in RBMs.

When the derivative calculated in (3.10) is substituted into (3.12), (3.13) is obtained:

$$\frac{\partial \log(P(\mathbf{v}))}{\partial w_{j,i}} = \left\langle v_j h_i \right\rangle_{data} - \left\langle v_j h_i \right\rangle_{model} \tag{3.13}$$

where $\langle\rangle_{data}$ represents the expectation under the distribution of data and $\langle\rangle_{model}$ denotes the expectation under the distribution of the model.

The expectations under these distributions can be calculated by using a Monte Carlo method, in which many samples are drawn from the related distribution and the mean of these samples approximates the expectation under this distribution. In order to draw samples from the data distribution, a sample vector of the data is clamped to the visible units and this vector corresponds to the state vector of visible units. Then, the state vector of hidden units is sampled from the conditional probabilities of hidden units given visible units as given in (3.7). At this point, we can calculate the expectation of data distribution $\left\langle v_j h_i \right\rangle_{data}$ by averaging the multiplication of the states of a visible and hidden units in a pair. Then, we can start to draw samples from the model distribution by using Gibbs

sampling.

3.3.4 Gibbs Sampling & Contrastive Divergence

Gibbs Sampling is the most popular Markov Chain Monte Carlo method in the training of RBMs [111]. When it is not possible to directly draw samples from an intractable multivariable distribution, Gibbs Sampling can approximate this distribution by starting from the state of a variable and by drawing a sample for the second variable from the conditional distribution given the state of the first variable. Similarly, the states of the third and the other variables are sampled conditioned on the state of the previous variable. After sampling all variables, the state of the first variable is sampled conditioned on the last variable and the sampling of other variables is also performed in the same order until reaching an equilibrium distribution. In RBM training, Gibbs Sampling is performed for sampling from model distribution $P(\mathbf{v}, \mathbf{h})$. Because the states of hidden units are previously sampled conditioned on data, this state vector is used to start Gibbs Sampling. Then, the states of visible units is sampled from $P(\mathbf{v}|\mathbf{h})$ in (3.8) and the states of hidden units is sampled from $P(\mathbf{h}|\mathbf{v})$ in (3.7). This cycle of sequential sampling of the visible and hidden state vectors is called *one step* in Gibbs Sampling, and the sequential sampling procedure continues until it starts to sample independent and identically distributed samples from a stable distribution. After this point, the sampled state vectors of visible and hidden units can be used to estimate the expectation under model distribution in (3.13). Figure 3.5 shows the training of an RBM with Gibbs Sampling.

Although Gibbs Sampling is useful to obtain an unbiased estimate of the expectation of model distribution, it takes a long time to reach an equilibrium distribution. Moreover, it is usually accepted to reach this distribution after a number of steps because there is no exact method to measure it except evaluating generated samples or finding correlations between successive samples [111]. As a practical alternative to Gibbs Sampling, Hinton introduced *Contrastive Divergence (CD)* [122]. In CD training, one does not need to wait until reaching this equilibrium distribution of the model. This training corresponds to a limited number of steps of Gibbs Sampling. Hinton showed that even 1 step of CD can be enough to train a Bernoulli-Bernoulli RBM in practice.

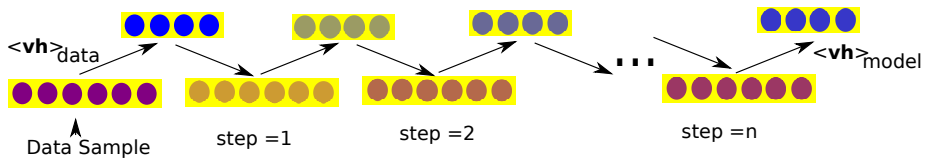


Figure 3.5: The training of an RBM with Gibbs Sampling, where an equilibrium distribution of the model is supposed to be reached after n steps

In BB-RBM, the states of visible unit vector $\mathbf{v}$ and hidden unit vector $\mathbf{h}$ consist of binary values of 0 and 1. However, because a BB-RBM can also model real valued training data in the range of $[0, 1]$ by accepting the data as the probabilities of the activations of corresponding visible units [125]. In this case, the condition that a binary visible unit can

have binary values (0 or 1) is relaxed and the state of binary visible units is allowed to have real values in the range of $[0, 1]$. This relaxation is only applicable to visible units so, the states of binary hidden units are strictly maintained as binary states during training. This situation changes the derivation of log probability of the data to

$$\frac{\partial \log(P(\mathbf{v}))}{\partial w_{j,i}} = \langle p_j h_i \rangle_{data} - \langle p_j h_i \rangle_{model} \tag{3.14}$$

where p_j is the probability of the visible unit j being active. h_i is the state of hidden unit i.

In order to maximise the probability of data, the change in the weights of the connections between visible and hidden units and biases of visible and hidden units for each training iteration can be calculated from (3.15-3.16-3.17).

$$\Delta w_{ji}(t) = \alpha \cdot \Delta w_{ji}(t-1) + \epsilon \left[\left\langle v_j^o h_i^o \right\rangle - \left\langle v_j^N h_i^N \right\rangle \right] \tag{3.15}$$

$$\Delta c_i(t) = \alpha \cdot \Delta c_i(t-1) + \epsilon \left[\left\langle h_i^o \right\rangle - \left\langle h_i^N \right\rangle \right] \tag{3.16}$$

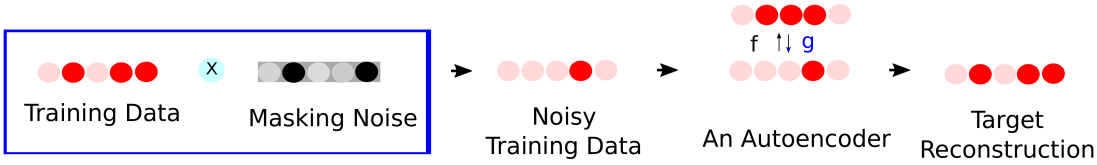
$$\Delta b_j(t) = \alpha \cdot \Delta b_j(t-1) + \epsilon \left[\left\langle v_j^o \right\rangle - \left\langle v_j^N \right\rangle \right] \tag{3.17}$$

where ϵ is a learning rate and α corresponds to the momentum variable. t represents the current iteration number during training and $t-1$ denotes the previous iteration number.

3.4 Denoising To Improve Hidden Representation

Vincent *et al.* introduced a denosing procedure to the training of autoencoders or RBMs to improve the features learned with these networks [7]. In this denoising training, the activation of hidden units is activated by noisy samples of the training dataset; however, the noise-free training dataset is used for the minimization of the error between training data and reconstructed data. In the preparation of this noisy dataset, random noise is applied to the original training dataset at each training iteration so, the network observes a slight change in the training data in each iteration but the original training dataset is retained as the target for reconstruction at each iteration.

The noise types recommended for this application by Vincent *et al.* are additive isotropic Gaussian noise, masking noise and, salt and pepper noise [7]. Among these noise types, additive isotropic Gaussian noise was especially suggested for natural images due to their real valued pixels and salt and pepper noise for binary or binary like data (e.g. the output of a sigmoid function). In the former type, the noise value for a pixel is sampled from a Gaussian distribution, whose mean is equal to the original pixel value and whose sigma determines the noise level (higher sigma means higher noise level). In the latter type,



(a)

Figure 3.6: The training of a denoising autoencoder as described by Vincent *et al.* [7], where binary masking noise (which replaces randomly selected pixels in training sample with zero) is applied to the training data and the original training data is approximated.

a pre-determined fraction of pixels in a training sample is replaced either with 0 or 1 with equal probability (0.5). Apart from these two noise types, masking noise can also be applied to both real valued and binary/ binary-like data. This noise type replaces randomly selected pixels in a predetermined fraction of an image with zero values. In salt and pepper, and masking noise, some pixels are unchanged, while the Gaussian noise affects overall data. Figure 3.6 demonstrates the training of a denoising autoencoder with the masking noise.

Vincent *et al.* found that this denoising regularisation is superior than other well-known regularisation approaches, such as using undercomplete representation or L_2 regularization in the loss function and generating better descriptive features [7]. The reason why denoising leads to better feature learning was suggested to be because the network needs to learn stable, higher order dependencies between input units in order to reconstruct the original data by ignoring the noise. The experiments of Vincent *et al.* showed that larger amounts of noise produce more global features, which eventually leads to better classification performance [7].

3.5 Stacking Representations

3.5.1 Stacking AEs : Deep Autoencoders

Autoencoders can be stacked regardless of their type such, as traditional or denoising, by forming a network with more than one hidden layer. This stack of auteoncoders is called a deep autoencoder. The training of a deep autoencoder is performed by individually training each hidden layer with the immediate lower layer as an autoencoder, starting from the first hidden layer. Let's examine the training of a deep autoencoder of 3 hidden layers. The first two bottom layers of the network become the first autoencoder in the training. This first autoencoder takes training samples as input and learns hidden representations for these samples. Then, the hidden layer of the first autoencoder and the second hidden layer of the deep network constitute the second autoencoder. The data representation generated with the hidden layer of the first autoencoder becomes the training data for the second autoencoder. The training of the third hidden layer of the deep network is also repeated in the same way.

3.5.2 Stacking RBMs: DBNs

RBMs can also be trained as a stack in the same way as autoencoders. The training of the stack of RBMs was proposed by Hinton *et al.* as an approximation to the trainining of a well-known generative model called *Deep Belief Networks (DBNs)* [127]. A deep belief network (DBN) is a deep generative model which consists of many layers of hidden variables. The top two layers of a DBN are linked with undirected connections and the connections in the rest of the network are directed from top to bottom, representing a directed graph (see Figure 3.7).

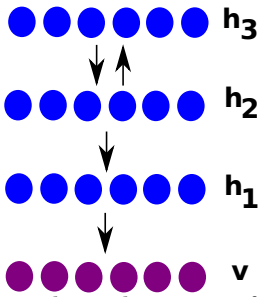


Figure 3.7: The architecture of a DBN.

According to Hinton *et al.*, such a pairwise training is performed for the training of a DBN so that the pair of each immediate neighbour layers of a DBN is trained as an RBM by ignoring the other connections coming to or going from these layers [127]. Figure 3.8 demonstrates this training of a DBN. Similar to the training of a stack of autoencoders, the training of a DBN starts from the bottom layer and moves to upper layers by evaluating each layer pair. As an example, for the training of the first hidden layer of the DBN, the input layer of the DBN (where training data is fed into the network) becomes the visible layer of an RBM and the first hidden layer of the DBN becomes the hidden layer of this RBM. After training this RBM with CD, a hidden representation of training data is generated by passing the data up until the second hidden layer of the DBN and this hidden representation is used to train the second hidden layer of the DBN. Similarly, the first hidden layer of the DBN and the second hidden layer of the DBN is treated as an RBM, and the training of the DBN continues until completing the training of all layers. Note that a layer which behaved as a hidden layer in the training of an RBM previously becomes the visible layer for the next layer.

The main aim behind the training of a DBN is to generate new samples from the model distribution by activating the top layer and passing this activation down to the bottom layer. On the other hand, a DBN can be used as a deep autoencoder, too. In this case, the bottom-up connections of a DBN correspond to the encoder part of the autoencoder and the top-down connections of the DBN to the decoder part of the autoencoder. Deterministic training is performed to minimize the loss function between the input data and the reconstructed data. In some applications, the multilayer connections of the DBN can be unfolded by generating a feed forward network, where the output layer corresponds to the reconstruction of the input data. Apart from being exploited as an autoencoder,

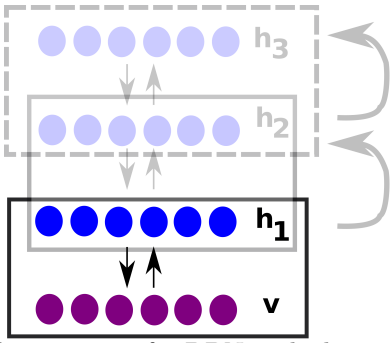


Figure 3.8: The training of a DBN with the stack of RBMs.

a DBN can be used as a feature extractor and can be combined with external classifiers and regressors such as SVMs, or directly be trained as a classifier or regressor by training another layer with deterministic units on the top of the DBN.

3.6 Cross Modality Learning

3.6.1 Translational Deep Belief Networks

A specialised type of a DBN, known as a TDBN, was proposed by Fasel and Berry [1] to realise cross-modality learning between two data modalities through a single deep network. The main motivation behind a T-DBN is to predict one of the data modalities, which is missing at run time, given the other one. A TDBN has two generative and one deterministic training stage. In the first generative training stage, a TDBN is similar to an unfolded DBN (which can also be viewed as an unfolded deep autoencoder) which has been trained with a dataset of two data modalities. Each sample of this dataset contains an example from each modality. In this training stage of a TDBN, the network is expected to regenerate the input data containing examples from these two modalities in the output layer.

In the second generative training stage, a Translational version of the Restricted Boltzmann Machine (a T-RBM) [1] is trained. The T-RBM was proposed to maintain a joint latent representation of two input data modalities (that was learned through training, when both of them are available) when one of them is missing at run time. The training of a T-RBM is similar to the training of an RBM regarding using CD training and almost the same update rules. However, a T-RBM does not aim to reconstruct the training data (of two modalities) as it is, but only partially (only available data modality at run time), and its training is based on a previously trained RBM using the original input data. The training of a T-RBM is demonstrated in Figure 3.9.

In the training of a T-RBM, the calculation of conditional probabilities of hidden units when conditioned on data is performed by using the weights of a previously trained RBM with the same training dataset. However, when calculating the conditional probabilities of visible units in the first step of CD training, the dimension of visible units is reduced to the size of a sample from the data modality present at run time. Later, the conditional

probabilities of hidden units in this step of CD training are also calculated from this modified visible layer. In other words, the samples of only one data modality are reconstructed during CD training, though the hidden activation conditioned on data was generated by using both data modalities before CD training. Although there is no change on the update rules, it should be noted that only weights and biases connected to the reconstructed data modality are updated during this training, which is given in (3.18),

$$\frac{\partial \log(P(\mathbf{v}))}{\partial w_{j,i}} = \langle v_j h_i \rangle_{data'} - \langle v_j h_i \rangle_{model'} \tag{3.18}$$

where $data'$ represents the data modality reconstructed during the training of a T-RBM and $model'$ shows new model learned during the training of the T-RBM.

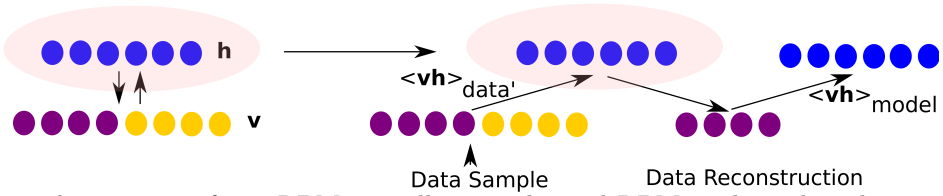


Figure 3.9: The training of a T-RBM: initially, a traditional RBM in the right side is trained on multiple data modalities, where purple circles and yellow ones belong to different data modalities. After the training of this RBM, the training of a T-RBM (shown in the left side) is performed based on the model learned in the RBM. The weight updates in the TRBM are realised for only data modality ($data'$) represented with purple circles.

After completing the training of the T-RBM at this second stage of training the DBN, some of the weights of the unfolded DBN obtained in the first generative stage are either removed or replaced with those of the T-RBM. Because one wants to feed the network with the data modality present at run time and to generate plausible values for the other one, the input layer weights of the unfolded DBN are replaced with the weights of the T-RBM. With the same aim, only the weights associated with the data modality to be predicted are maintained in the output layer of this unfolded DBN by removing the rest of the weights in this layer.

The third training stage of the TDBN might be seen as a deterministic stage, because the stochastic activations of the DBN are replaced with deterministic units. In this stage, the TDBN is similar to a fully connected network whose input and output layers have the same dimension (by assuming the samples of these two modalities have the same size) but are related to different data modalities. A loss function related to this cross-modality learning problem is minimised by using stochastic gradient descent to optimise the network.

3.6.2 Denoising in Cross-Modality Learning

Vincent *et al.*'s denoising procedure to obtain better features from image data inspired Ngiam *et al.* [117] to learn better features in the training of an autoencoder which is fed with training samples from multiple data modalities. Ngiam *et al.* used a different type of noise, absence of a data modality in a training sample, because their aim was to learn a

shared representation which is robust to the absence of some data modalities in the input data.

Ngiam *et al.* [117] used audio and video data as two modalities. The correlation between these two data modalities is revealed by reconstructing a sample pair (each from a modality) during CD training, while one sample in this pair is replaced with zeros in the visible layer (unobserved) when the data is observed. Therefore, the probabilities of hidden units conditioned on data depend on only the informative sample from this sample pair. This may also suggest that the model learned during CD training depends on the data sample that was present. Because Vincent *et al.* applied *masking noise* to randomly selected elements of a data vector while Ngiam *et al.*'s noise replaced a complete sample of a related data modality with a zero vector, Ngiam *et al.*'s noise will be called *block masking noise* to differentiate it from the other.

In order to implement this denoising, Ngiam *et al.* used data augmentation [117]. In the augmented dataset, two different sets of noisy version of the original dataset were added to the training dataset. The first noisy set had sample pairs in which the first sample in each pair was replaced with zero. Similarly, the second noisy set had zero vectors in the place of samples from the second modality.

3.6.3 An Improved Denoising Strategy for CM Learning

Similar to Ngiam *et al.*' work [117], block masking noise is used in this research to reveal the correlation between fundus images and their label masks. On the other hand, the block masking noise is applied differently from Ngiam *et al.*'s denoising implementation in two aspects.

Firstly, the noise is applied to all samples of the training dataset without leaving any data pair untouched. This means that either the sample from the first modality, or that from the second one, is missing in sample pairs in the noisy dataset. In other words, there is no sample pair which contains informative samples from both data modalities, as was the case in Ngiam *et al.*'s work [117]. The proposed preparation of a noisy dataset provides stronger regularisation for the training, when compared with Ngiam *et al.*'s approach.

Secondly, the generation of this noisy dataset is performed independently for each training iteration. Because training is realised with the usage of mini-batches from the training data, this type of noisy data generation allows one to adjust the amount of noise for each training iteration by encouraging each mini-batch to equally contribute to weight updates associated with each data modality, or to produce larger gradients for weights linked to a selected data modality. In this study, because learning the representation of both data modalities is equally important, the sample pairs in each mini-batch are randomly divided into two sets, and each set suffers from either missing the first data modality or the second one. Figure 3.10a shows the preparation of noisy training data for this training of the RBM with denoising.

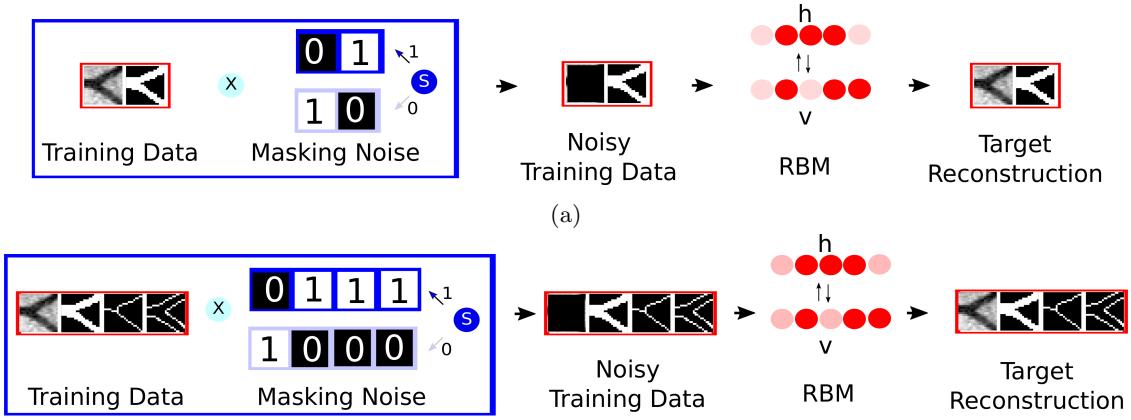


Figure 3.10: The combined training of an RBM with a special denoising procedure on the couple of an fundus image patch and (a) single type of segmentation mask or on the couple of an fundus image patch and (b) multiple types of segmentation masks.

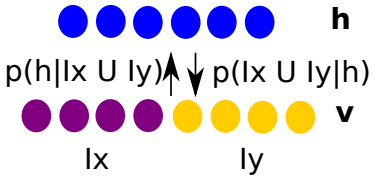


Figure 3.11: The training of an RBM with I_x and I_y in order to learn a joint representation of them. Note that this joint representation is directly learned from the samples of data modalities. In the other words, I_x and I_y are connected to form an input vector in this layer.

3.6.4 The Proposed Network For Vessel Segmentation

In this thesis, the training of the TDBN [1] is combined with the improved denoising procedure. Also, its architecture is slightly changed to obtain better segmentation performance, inspired from the network of Li *et al.* [43]. In order to discriminate the new network from the one proposed by Fasel and Berry [1], the former one will be called the enhanced TDBN. The details of the training of the enhanced TDBN are given below.

Initially, the first hidden layer of the enhanced TDBN and its input layer are trained as an RBM by combining the CD training with the improved denoising. A sample pair of training data is a one dimensional row vector which contains vectors of both fundus image patches and their vessel maps. These one dimensional vectors are obtained after individually cropping two dimensional image patches and their vessel maps from related data matrices and by converting them to one dimensional vectors. A mini-batch of the training dataset contains a group of these sample pairs.

Before training an RBM with this mini-batch, half of the sample pairs is randomly selected to replace the samples from fundus image patches with zero, and samples from the vessel maps are replaced with zero in the rest of the sample pairs. When training with a sample pair whose vessel map I_y is replaced with zero (see Figure 3.12a), the conditional probabilities of hidden units are based only on the information from the fundus image patch. However, the RBM is supposed to reconstruct both present data (I_x in Figure

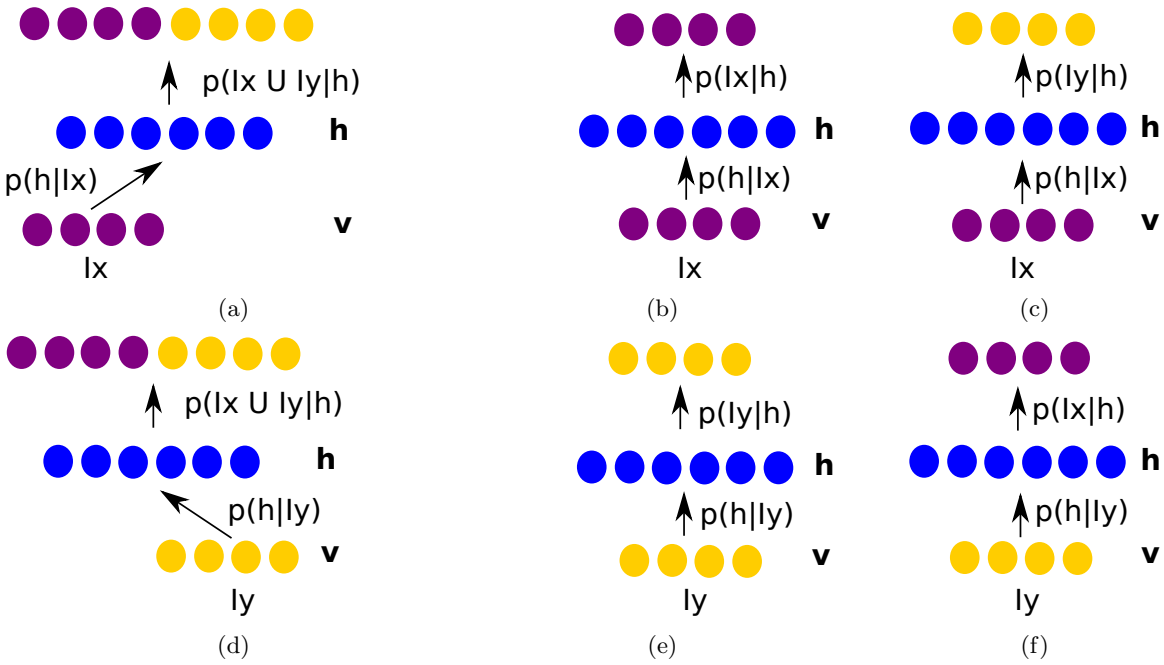
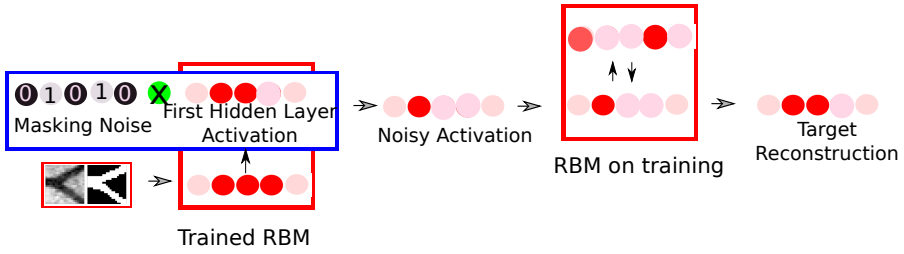


Figure 3.12: Denoising procedure implemented with a block masking noise (a) when the second data modality I_y is missing (d) when the first data modality I_x is missing. The reconstruction of both modalities given the hidden representation generated by only one modality is detailed in (b)-(c) and (e)-(f); (b) The generation of I_x given I_x (c) The generation of I_y given I_x (e) The generation of I_y given I_y (f) The generation of I_x given I_y.

3.12b) and absent data (I_y in Figure 3.12c) together in its visible layer during CD training due to the denoising procedure.

Let's factorise the training of an RBM when the proposed denoising is applied to either data modalities. Figure 3.12a shows the training when the second data modality (I_y) is missing. This training can be factorised into the simultaneous training of two networks: the first network given in Figure 3.12b reconstructs the data modality I_x (a fundus image patch) from the hidden representation learned from I_x and the second network given in Figure 3.12c predicts the absent data modality I_y (a vessel map) from the same hidden representation. In the same way, Figure 3.12d shows the training of the network when fundus image patches are missing. In this way, the activations of hidden units depend on the information coming from a vessel map. This hidden activation is responsible for both the reconstruction of the vessel map given as input in Figure 3.12e, and the prediction of a possible fundus image patch in Figure 3.12f. As can be concluded from these figures, the hidden representation takes a very important role in the inference of the data modality that is present or the prediction of the absent data modality. In order to perform these tasks, this hidden representation should include modality specific features of both fundus image patches and their vessel maps. Also, it should contain cross-modality features, which can transform a fundus image patch to its vessel map, or vice versa.

After learning modality-specific and cross-modality features in the first hidden layer of a DBN, the joint representations of both fundus image patches and their vessel maps are



(a)

Figure 3.13: Training of the second layer of a DBN with binary masking noise.

fed into its upper layers. In order to improve feature learning (hidden representations) in the training of its upper layers (see Figure 3.8 for the training of a DBN with the consecutive trainings of RBMs), the denoising procedure described in [7] with masking noise is combined with the training of RBMs. As an example, the training of the second hidden layer of the DBN with this latter denoising is demonstrated in Figure 3.13. This time, the masking noise replaces the randomly selected unit activations of the first hidden layer of the DBN (which becomes the training dataset for the training of the second hidden layer of the DBN) with zero at each training iteration. Although the conditional probabilities of hidden units in the training of the second hidden layer of the DBN depend on the noisy data vectors, their noise-free versions are aimed to be reconstructed during CD training.

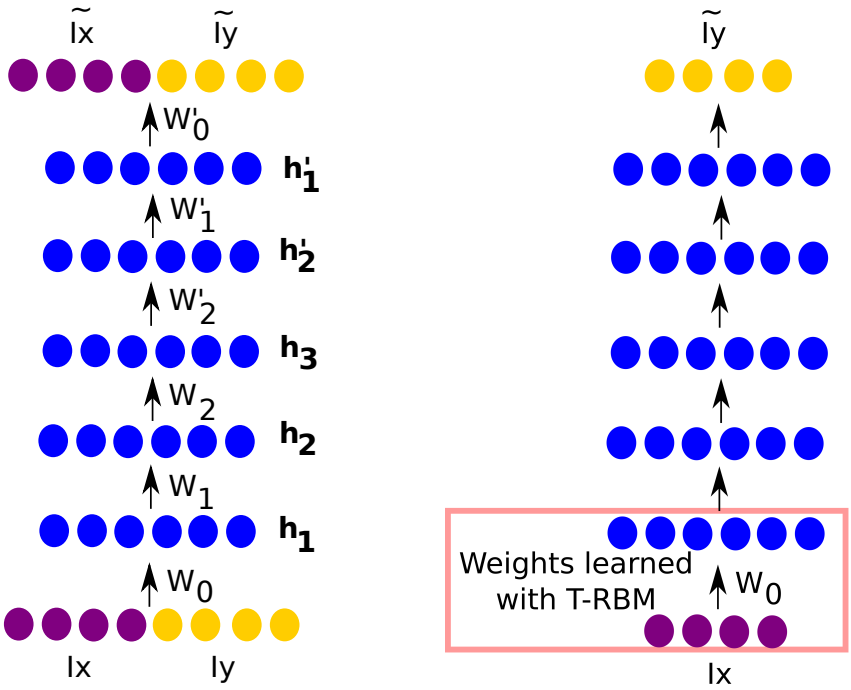
The proposed denoising with block masking noise can also be applied to the training of a T-RBM. In this case, the hidden representation generated by a sample of either data modality is responsible for the reconstruction of a fundus image patch. After completing the generative trainings of the DBN and T-RBM, the connections of the DBN are unfolded (see Figure 3.14a for the unfolded version of a DBN with 3 hidden layers) and the necessary modifications are made to weights in order to convert the unfolded DBN to a TDBN (see Figure 3.14b) as explained in Section 3.6.1. The final network (the TDBN) has directed connections from the input layer to the output layer and has sigmoid units including input, hidden and output layers. This network is optimised by using stochastic gradient descent. The optimisation of this network was realised by minimising the L_2 loss function given in (3.19). This loss function is also called Euclidean distance loss,

$$L_2 = \sum (I_y - \hat{I}_y)^2 \tag{3.19}$$

where L_2 is the loss function. $\hat{I}_y$ is the predicted vessel mask in 1D and I_y is the ground truth vessel mask.

3.7 Evaluating the Effects of Denoising

I conducted two pilot experiments to assess the effects of the proposed approach to denoising on generated features and vessel maps with TDBNs by training one TDBN with denoising and another one without it. The first experiment evaluated the effects of using



(a) (b)

Figure 3.14: (a) The weights of a trained DBN with a training dataset containing two data modalities I_x and I_y are unfolded. The resulting network can be viewed as a deep autoencoder with the symmetric weights of encoder and decoder (b) The adjustment of pretrained weights to cross-modality learning: The weights connecting the input layer and the first hidden layer of this autoencoder are changed with the weights learned during the training of a T-RBM. The units and weights associated with I_x in the output layer of this autoencoder are also removed. This final network can be optimised with a stochastic gradient descent algorithm.

the proposed denoising during generative training on features obtained after pretraining and on those evolved during finetuning whereas the second one examined generated vessel maps given fundus image patches with the finetuned TDBN.

3.7.1 Experiment 1: The Effects of Denoising on The Generated Features

A way to scrutinise the performance of generative networks such as RBMs and DBNs is to visualise the features, in other words weights of the network, learned with training and to evaluate them regarding their descriptive ability [111]. In order to evaluate if there is any improvement in the features when the training of a TDBN is combined with and without denoising, two TDBNs were trained with the same training dataset and with the same parameter set. Also, the effect of training a T-RBM on the features, initially obtained from the training of an RBM, is investigated.

The detail of the setting of the experiment will be given in Section 4.4.4 of the next Chapter. Figure 3.15 compares features learned with the training of a TDBN and those generated with the training of an enhanced TDBN, at different steps of pre-training (e.g. RBM and T-RBM trainings) and, finetuning. Because only features connected to input image couples can be directly visualised, these features are illustrated in the figure. The features associated with fundus image patches and those binary vessel maps are examined separately in two groups because the features in each group have distinctive characteristics.

Feature sets in the top of the figure belong to RBM training and are associated with fundus image patches. When denoising is not combined in the training of RBM, the network generates features for fundus image patches, which mostly look like parts of black discs with brighter strips on their borders. These features seem far from representing any vasculature structures but rather the general contrast difference between pixels inside FOV and those outside FOV. On the other hand, when the training of RBM is combined with denoising, the features demonstrate various relations between vessel segments in the vasculature tree, with combinations of line detectors-kind patterns such as bifurcating or crossing vessels.

The second row in the figure shows features generated with the T-RBM for fundus image patches. When compared with the first row features, the second row shows big improvement on features, generated with the RBM training without denoising. However, there is almost no change for those produced with the RBM with denosing. Also, the patterns of features in the second row can be matched with those of the features in the forth row. The features in the forth row belong to binary vessel maps. Regarding matching patterns of features for fundus image patches and binary maps, one can observe the influence of binary vessel maps on the generated features for fundus image patches. This influence starts earlier, at the training of RBM, when denoising is combined with the network training; in contrast, this influence begins later at the training of T-RBM when denoising is not a part of the training.

Although there are visible differences between the characteristics of features for both fundus image patches and vessel maps depending on whether denoising takes part in the training of RBMs and T-RBMs or not, the differences become smaller during finetuning. The third and bottom rows in the figure show the features generated during fine-tuning for fundus image patches and binary vessel maps respectively. It seems that features in the right column in these rows, which are related to training with denoising, are more various and similar to line detectors.

There are three striking differences regarding the features in the figure.The first difference is that the features learned in the pretraining of the latter network are more representative of vessel tree patterns, though both networks have the same architecture. As seen in Figure 3.15, the features learned during the training of an RBM when the training is combined with denoising are much more descriptive for both fundus image patches and their vessel maps. In particular, the features associated with vessel maps are much richer and more visually interpretable. One can see various patterns of vessel combinations in a vessel tree such as different orientations and number of vessels in the same neighbourhood or various connection types (e.g. crossing, parallel). Also, nearly all of the hidden units corresponds to a different pattern of vessel tree, which makes the representation richer.

Clearly, one can differentiate the features related to vessel connection patterns in fundus images such as single vessels, bifurcatings, crossings or multiple vessels in both feature groups (one for fundus image patches and the other for vessel maps) learned with this training. On the other hand, the features generated without using denoising in the training of the RBM do not represent either the structure of a vessel nor the patterns associated with vessel tree fractions. This may show that the proposed denoising makes an RBM focus on vessel structures rather than low level interactions between pixels.

The second difference is that matching features for both fundus image patches and their labels appear in different times for both training. When denoising is not added to the training, these features appear after training the T-RBM while they appear after training the RBM when denoising is used for training. This may suggest that the proposed denoising can be an alternative to a T-RBM for cross-modality learning approaches. As seen from the figure, features generated during T-RBM training combined with denoising are very close to those generated in the previous training of the RBM. This may suggest that the training of a T-RBM for this type of application is necessary when denoising is not combined with the training of an RBM; otherwise, it is optional.

The third difference is that features learned for fundus image patches and those for their labels in the training of an RBM are similar in their overall shape, when denoising is involved during the training. In other words, both feature groups are in line with each other in the structures they describe. This is better demonstrated in Figure 3.16. One can say that features learned from fundus image patches are being guided by features learned for vessel maps. This is partly due to the stronger activation of label inputs in the hidden layer than that of fundus image patches. This is a desired situation because one wants to detect such structures in fundus images patches that match with the structures in vessel

maps. This suggests that learning better features for binary vessel maps encourages the network to learn useful features for the detection of vessels in fundus images.

3.7.2 Experiment 2: Evaluation of Generated Vessel Masks

In this experiment, the inferred vessel masks from fundus images will be examined. This will give readers a chance to compare vessel masks inferred by a TDBN without denoising and those inferred by a TDBN with denoising before moving to the segmentation an entire fundus image in the next Chapter.

The most important benefit of using a segmentation method based on cross-modality learning is to be able to infer a vessel mask (not a pixel label) for a given fundus image patch. This immediate evaluation of a fundus image patch can be used for online applications, where only the segmentation of a local region is necessary/sufficient for the time being such as tracking. However, only offline tracking will be performed in the rest of the thesis and tracking with online segmentation will be left for future work.

A TDBN generates a probability map when it is fed with a fundus image patch. This probability map is expected to have larger values for vessel pixels and lower values for non-vessel pixels. The aim of the training of a TDBN is to increase the difference between these probability values so that pixels in a fundus image can be represented with saturated two label values 0 and 1. Figure 3.17 demonstrates three examples of image patches with the inferred probability maps by a TDBN without denoising and a TDBN with denoising. According to this figure, a TDBN with denoising infers more flat probability surfaces for both vessels and vessel backgrounds in all image patches.

3.8 Conclusion

The TDBN was originally proposed for the estimation of tongue contours on ultrasound images. In this work, the training and the architecture of the TDBN were adapted for patch-based retinal vasculature segmentation. As far as I am aware, it is the first application of the TDBN for retinal vasculature segmentation. According to the results of preliminary experiments, augmenting denoising with the training of RBMs was found to encourage features which were more associated with intrinsic topology of retinal vasculature. This accords with the results of Vincent *et al.* [7].

An expectation from a cross-modality learning based segmentation method is to maintain the consistency between label pixels representing the same object. Given the results of the preliminary experiments, it can be concluded that probabilities of vessel pixels in generated probability maps for input fundus image patches were mostly consistent with each other, regardless of whether denoising was used in the training. However, in some cases, using denoising was found to increase this consistency of vesselness probabilities of vessel pixels and to significantly decrease the vesselness probability of non-vessel pixels. These small improvements on probabilities may not be considered to be valuable with

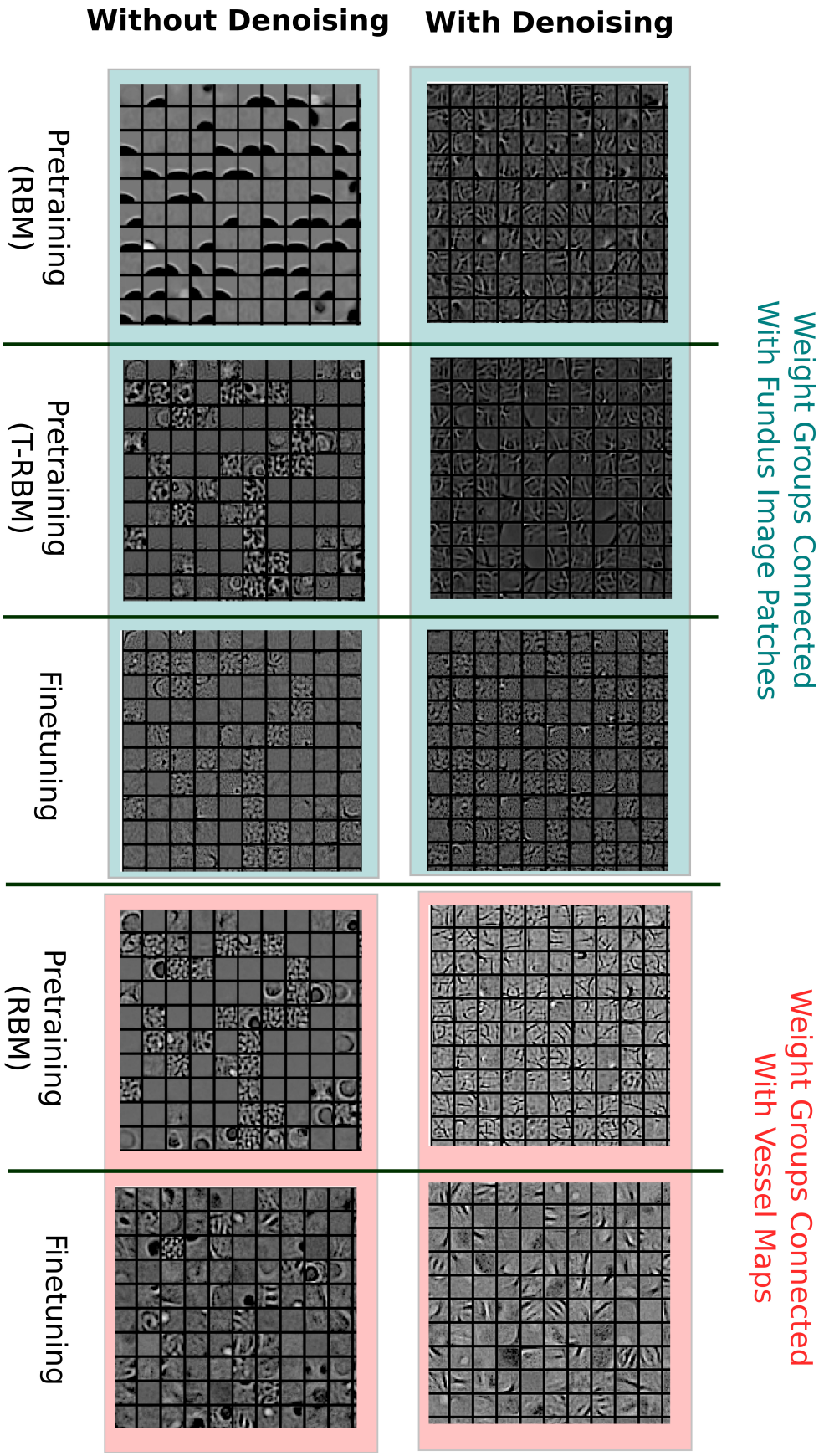
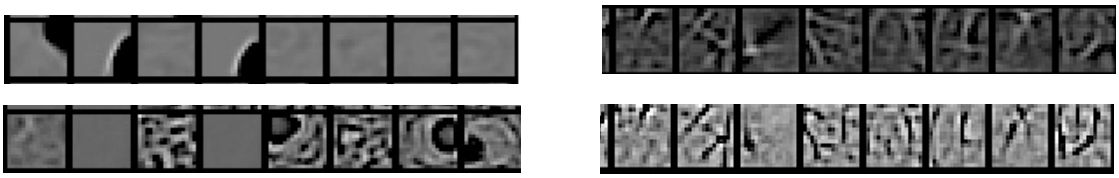


Figure 3.15: The features learned for the first layer of a TDBN in the generative training stage without and with the application of denoising (the noise type is block masking noise) and the evolution of these features during the finetuning.



(a) (b)

Figure 3.16: Details of features generated during the pretraining, demonstrated in Figure 3.15. In each sub-figure, the top layer is associated with fundus image patches and the bottom layer with binary labels. Some selected features (a) from the training of a TDBN without denoising, (b) from the training of a TDBN with denoising

many segmentation methods because their segmentation maps are usually the result of using binary thresholding, which is oblivious to small variations of vesselness probabilities unless the variations change the labels of related pixels. However, the improvements, emerged from using denoising, in probability maps may be much useful for a tracking system, which will be introduced in Chapter 5, because the system directly uses probability maps, not binary maps, for the estimates of vessel centreline and boundaries.

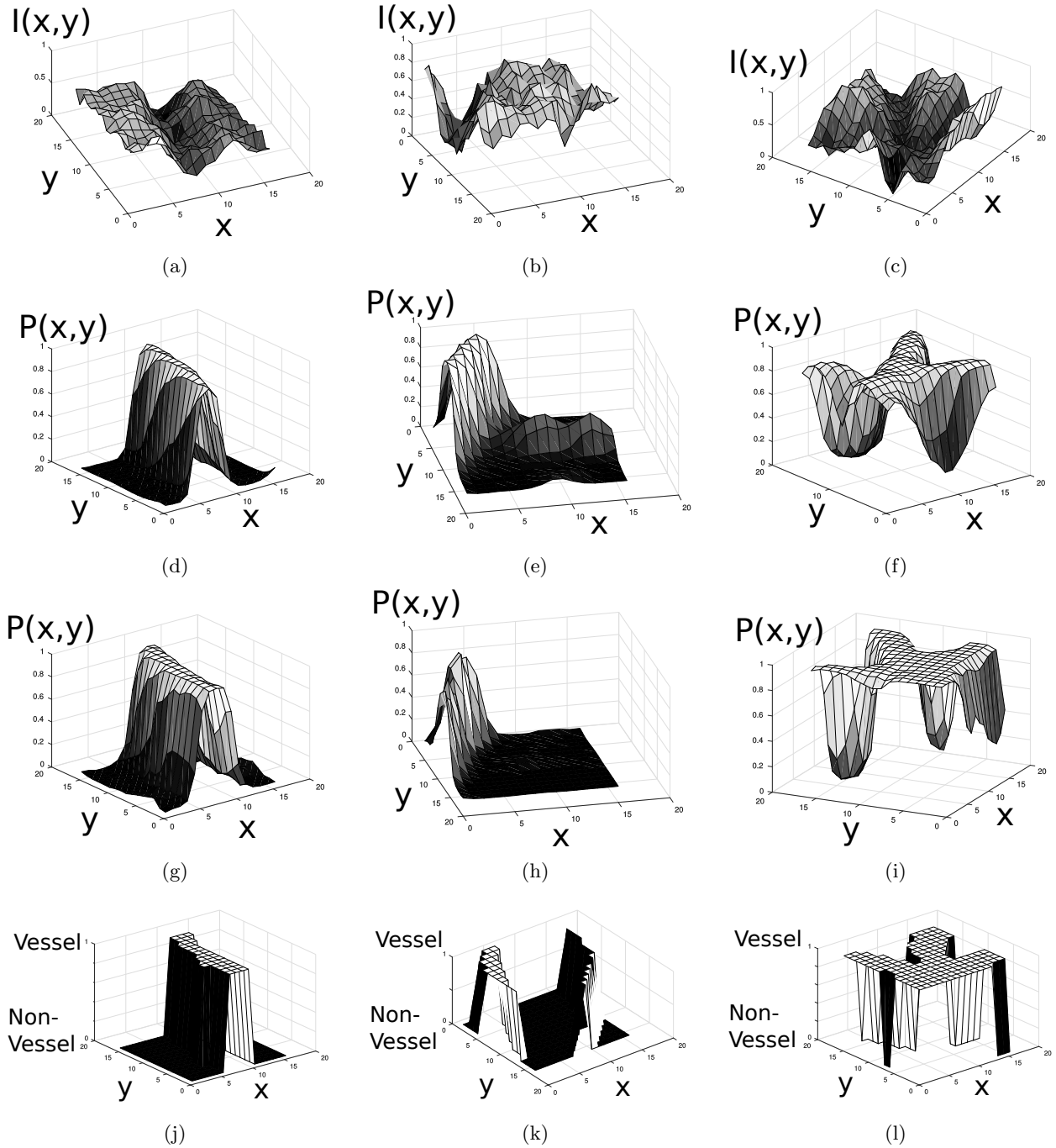


Figure 3.17: Fundus Image Patches and Generated Probability Maps for 'Vesselness': The first row shows the fundus image patches, the second one probability maps generated by the TDBN without denoising, the third one with denoising and the last one represents the ground truth image patch. The first column belongs to an example of single vessel, the second one to that of parallel single vessels and the third one to that of crossing vessels.

Chapter 4

Enhanced TDBNs For Vasculature Segmentation

4.1 Introduction

In the previous chapter, the TDBN was adapted for the segmentation of fundus images in patch level and named the enhanced TDBN. This adaptation was realised in two ways (i) by using an architecture inspired by another application of cross-modality learning on vessel segmentation [43] and (ii) by utilising a denoising procedure to improve representation learned by a TDBN. As an extension to the previous chapter, this chapter introduces three applications for vasculature segmentation which can be realised by the enhanced TDBN.

The first application is traditional segmentation, where vessel interior pixels are identified in fundus photos. When the network is trained for this purpose, fundus image patches and their vessel masks are of the same size. The performance of the network will be evaluated on low resolution fundus images from the DRIVE and the STARE.

The second application is related to accelerating the segmentation of vessels in high resolution images and was inspired by super-resolution [128]. In this application, the restriction about generating vessel masks of the same size as its corresponding fundus image patches is relaxed and the size of vessel masks is made deliberately larger than that of fundus image patches. Therefore, this approach to vessel segmentation can reduce segmentation time for high resolution images significantly. The application of this will be performed on high resolution fundus image datasets, namely CHASE_DB1 and HRF, where the segmentation of an entire fundus image can take much longer than that of a low resolution counterpart.

The third application aims to segment multiple parts of vasculature with the training of a single network. These vessel parts include vessel centreline, boundaries, vessel bifurcation and crossing patterns in vessel centrelines in addition to vessel interior pixels. Generated probability maps as a result of these segmentations will be integrated to a vasculature

tracking method, which will be introduced in Chapter 5.

4.2 Probability Maps from Enhanced TDBNs

An enhanced TDBN can be used to transform an input vector to an output vector of the same size as done for traditional retinal vasculature segmentation applications. This method will be applied to the segmentation of retinal vasculature in low resolution fundus image datasets: the DRIVE and STARE. These datasets are publicly available and they are widely used to compare the performances of different segmentation methods.

4.3 Evaluation Criteria For Segmentation Performance

The probability maps produced by a trained network are able to assign higher probability values to the pixels belonging to retinal vasculature and, lower probability values for other pixels. Binary vessel maps can be produced by applying a threshold to these probability maps. This thresholding can be viewed as a binary classification task, where pixels belonging to vasculature are expected to be assigned to the same class (vessel class) and the other pixels to the other class (non-vessel class).

In these binary vessel maps, the pixels correctly classified as vessel pixels are called *true positives (TP)* and the number of vessel pixels wrongly classified is called *false negatives (FN)*. The number of non-vessel pixels correctly classified as non-vessel pixels represents the *true negatives (TN)* and the number of non-vessel pixels mistakenly classified as vessel pixels represents *false positives (FP)*. Using these measures, *accuracy (Acc)*, *sensitivity (Sens)* and *specificity (Spec)* can be calculated as given in 4.1.

$$Acc = \frac{TN+TP}{TP+TN+FN+FP} \ , \ Sens = \frac{TP}{TP+FN} \ \ Spec = \frac{TN}{TN+FP} \tag{4.1}$$

Accuracy is a measure of the ability of the classifier to correctly label vessel pixels as vessel pixels and non-vessel pixels as non-vessel pixels. Sensitivity captures the ability of the classifier to correctly label vessel pixels, and specificity to correctly label non-vessel pixels.

Another measure to evaluate the performance of a classifier is to draw *receive operating characteristic curves (ROCs)*. These curves have FP on the x axes and TP on the y axes, which are calculated at a range of thresholds. In other words, x axes show $(1-specificity)$ and y axes *(sensitivity)* at various thresholds. *The area under ROC curve (AUC)* gives information about the certainty of the classifier, which relates to the segmentation method that generates probability maps. This measure, AUC, has a maximum value of 1. AUC to be closer to 1 means better performance of the classifier. When compared with aforementioned measures, the AUC metric is a better indication of how well a classifier performs according to Fawcett [129].

The performance of the proposed segmentation method is evaluated by using these

aforementioned four measures: *accuracy*, *sensitivity*, *specificity* and AUC. The first three measures are obtained at the threshold where maximum accuracy is achieved on the test datasets [5, 44, 53]. From the calculation of these four measures, only pixels inside FOV masks were used. When evaluating the performance of the proposed segmentation method on a particular dataset, these measures were calculated for each image in this dataset, then average measures were reported as the overall performance on the dataset.

Although these criteria have been widely used for performance evaluation of segmentation methods, they have some limitations. Firstly, the calculation of accuracy, sensitivity and specificity requires binary vessel maps, which can be obtained by thresholding the output probability maps. Threshold is usually selected a value, where the largest accuracy is reached. The same threshold is used for the calculation of sensitivity and specificity. Comparing the performance of methods according to accuracy, sensitivity and specificity can be viewed as evaluating the performance at a certain operation point, without considering overall performance of the methods. Secondly, the performance evaluation measures described above do not consider the intactness of the retinal vasculature in generated binary vessel maps because these measures implicitly assume pixel labels to be spatially independent. Therefore, these measures may not identify methods with better representation of vasculature. For example, the higher performance of a method may be due to accurate labelling of stray pixels, whose neighbour pixels may be from a different class even though they are located in the same object (e.g. vessel). Finally, the number of vessel pixels in a fundus image is far smaller than that of non-vessel pixels. Therefore, a method with larger accuracy may not be guaranteed to identify larger part of the vasculature. Regarding this issue, sensitivity is viewed to be better at identifying vessel pixels than specificity. However, high sensitivity may be also deceiving because it may be achievable by using lower thresholds, which may ensure larger overlapping between actual vessel pixels and estimated ones; however, it may result in a mis-representation of the vasculature, with thicker vessels.

4.4 The Segmentation of Vasculature on Low Resolution Fundus Image Datasets

Fasel and Berry [1] proposed Translational Deep Belief Nets (TDBNs) to transform ultrasound tongue images to tongue contours. In this work, they downsampled ultrasound images to be able to train the network with complete images (not patch based). These ultrasound images had very simple geometry of tongue curves with moderate noise. The output of the network was supposed to only facilitate the detection of contours, which may mean that not exact regenerations of these counters in the output layer of the network were so important but they should imply or somehow give information about the locations of tongue contours for a further calculation.

When compared with the prediction of counters in ultrasound images [1], vessel segmen-

tation is harder because of the range of variation in vessel thickness and curvature and the complicated patterns of vessel connections including bifurcating and crossing vessels. Furthermore, a fundus image presents a complicated network of a large number of vessels, when compared with the simplicity of tongue contours in ultrasound images. Regarding this issue, it does not seem possible to learn the internal representation of fundus images in this scale. In order to simplify the problem, one can benefit from patch based approach. In this approach, an Enhanced TDBN aims to learn the transformation of *a fundus image patch to a vessel mask*. Later, these vessel masks can be combined in such a way that the final stitched label image represents the vasculature of an entire fundus image.

4.4.1 Transforming Patch Based Vessel Mask Generation To Vesselness Probability Maps of Fundus Images

The easiest way to combine vessel masks generated as outputs of the network can be to put these vessel masks at the same positions, where input image patches were cropped. This approach can be called *a non-overlapping fashion* and assumes that each fundus image patch is independent. Because the length of the time to complete the segmentation of a fundus image is related to the number of image patches cropped from this image, it can be expected that the segmentation time can be shorter than when each pixel in this image is evaluated for the segmentation. Regarding spatial dependencies among the pixels, this approach ignores the relationship between neighbour image patches, which may cover different parts of the same vasculature fraction.

In order to consider this dependency between fundus image patches for the generation of the complete vessel map, one can obtain the location of next image patch by changing the location of the current one with the distance of one pixel. Therefore, the probability of each pixel in the entire fundus image will be evaluated multiple times and with various neighbourhoods. This allows one to capture the variation on the probabilities of a particular pixel, when it interacts with different neighbours. The usage of the average of these probabilities that a particular pixel can be assigned in an entire fundus image, can give a better estimate of its probability of belonging to a vessel structure. How to calculate the probability of each pixel in an entire fundus image by considering its interaction with other pixels in close neighbourhoods is given in (4.2) and (4.3).

$$\hat{P}_i = \frac{1}{W_p \cdot H_p} \sum_{n=1}^{W_p \cdot H_p} P_i^n, \qquad i = 1 : W_f \cdot H_f = 1 : W_c \cdot H_c \tag{4.2}$$

$$P_i^n \in g\big(\mathbf{I}_p^n\big), \qquad \mathbf{I}_p^n \subset \mathbf{I}_f, \qquad n = 1 : W_p \cdot H_p \tag{4.3}$$

where W_f,W_c and W_p respectively correspond to the width of the entire fundus image, that of the complete vessel map and that of a fundus image patch; while, H_f, H_c and H_p represent the height of the entire fundus image, that of the complete vessel map and that of a fundus image patch respectively. $\mathbf{I}_p^n$ represents n^{th} image patch sampled from a

fundus image $\mathbf{I}_f$. $g(\cdot)$ shows the transformation function which was learned during the training of the Enhanced TDBN. This function transforms a one dimensional vector of a fundus image patch $\mathbf{I}_p^n$ to a vector containing a probability map. The lengths of the input vector and the output vector are both equal to $W_p \times H_p$ while total number of pixels in the entire fundus image $\mathbf{I}_f$ is equal to $W_f \times H_f$. The total number of pixels in the complete vessel map is $W_c \times H_c$. Because the number of probabilities that a pixel i can take in fundus image $\mathbf{I}_f$ is limited with the number of image patches this pixel can be part of, the average probability $\hat{P}_i$ for this pixel i is calculated by only considering its probability values in these neighbour image patches. P_i^n represents the probability of the pixel i in an image patch I_p^n .

Li *et al.* and Liskowski and Krawiec [36, 43, 66] used the same method to generate the probability map for an entire fundus image. This method, as will be shown later, generates a smoother probability map than the one generated with *non-overlapping fashion* for an entire fundus image by bringing the probabilities of pixels belonging to the same vasculature to almost the same level, and this agreement of pixel probabilities over a complete probability map allows a single threshold to be enough to binarise the probability map. Examples of these probability maps and those of binary vessel maps can be found in Figure 4.1 and Figure 4.2.

4.4.2 Experimental Settings

Hyper-parameters of The Network

The architecture of the proposed network was inspired by the work of Li *et al.* in [43]. Li *et al.*'s method is the earliest work using cross-modality learning for vasculature segmentation as far as I am aware. Their network was faster and less data hungry than CNNs [36]. Also, the performance of their network was comparable or better than that of state of the art methods (see Table 4.3). Li *et al* used a deep fully connected network of 3 hidden layers with 400 sigmoid units in each hidden layer. Both the input and the output layers had 256 linear units. They decided on the size of a hidden layer according to the experiment they conducted. The experiment compared the performance of the network given the size of input and output layers (256 units) and a range of hidden layer size (from 200 to 700 units).

I used the results of their experiment to decide on the size of input, output and hidden layers. Then, I evaluated the performance of my network by changing the number of hidden layers. Table 4.1 shows the measured performances of the network when varying its depth regarding AUC, accuracy, sensitivity and specificity. The training was performed with the DRIVE dataset. Layer numbers in the table show the total number of layers including input, hidden and output layers in the final version of the network. As seen from this table, increasing the number of layers in this DBN from 3 layers to 7 layers generated better segmentation performance; however, a further increase in the depth of the network may increase computation cost without yielding any remarkable improvement

Table 4.1: The evaluation of the effect of the depth of the proposed network on the segmentation performance.

Layer No	AUC	Accuracy	Sensitivity	Specificity
3	0.9694	0.9513	0.7491	0.9806
5	0.9735	0.9519	0.7593	0.9799
7	0.9752	0.9535	0.7662	0.9808

in the performance. Therefore, I preferred to use 5 hidden layers.

A TDBN has three different training stages, where the number of hidden layers and the number of units in a layer can change between these stages. In the first of these training stages, a DBN of 3 hidden layers was trained, where each hidden layer had 400 binary units and its visible layer had 512 binary units. At the second training stage, a T-RBM whose hidden layer had 400 binary units and its visible layer had 256 binary units was trained.

In these two stages of generative learning, the initialisation of weights were realised by sampling from a normal distribution with zero mean and standard deviation of 0.001 $N(0, 0.001)$. Then, these weights were updated during the training with a learning rate of 0.005. The iteration number for both training stages was 50. In the first 5 epochs, a momentum of 0.5 was used. This momentum increased to 0.9 after this point.

In the third training stage, the final network contained 5 hidden layers of 400 sigmoid units due to unfolding this DBN [123]. Also, both the input and the output layers had 256 sigmoid units because of replacing the weights in the input layer and deleting unwanted weights in the output layer. The learning rate in this stage was 0.3 for 120 epochs for both experiments.

Data Preparation

The green channel of RGB fundus images was used for all training and experiments. This channel was used in many previous studies due to its better contrast of vessel appearances facilitating the detection of vessels [40]. Fundus image patches and their vessel masks were normalised in patch based manner in the range of $[0, 1]$ because image patches were assumed to be independent. This was observed to improve segmentation performance of the method on less illuminated image regions, where the contrast of vessels is very low. In the preparation of training datasets, the inclusion of many examples of vessel patterns was given importance in order to show the network a range of vessel patterns by aiming to increase the generalisation capacity of the network. With this aim, pixels outside FOV masks, which did not give much information about the correlation between fundus image patches and their vessel masks, were multiplied with zero, and image patches which only included these zero valued pixels and their corresponding vessel masks were removed from the training dataset.

Despite uniformly sampling over entire fundus images/binary vessel maps, training samples happened to be densely sampled from inside FOV masks due to removing these image

patches from the training dataset later. The numbers of training samples (pairs of a fundus image patch and its vessel mask) in both experiments were approximately $2,000,000$. These training samples were separated to mini-batches of 100 training samples. The complete training set was used during the generative training whereas a quarter of this training dataset was used for validation during the fine-tuning stage and the rest of it for the training. Because the time spent for pretraining is minute when compared with that for finetuning (e.g. 20 minutes to 5 hours) and the locations of some weights learned at pretraining are changed in the network prior to finetuning, using complete training dataset for pretraining was not observed to cause any overfitting during fine-tuning, which was checked by plotting training and validation errors after finetuning. The same architecture and the training parameters were used for both the DRIVE and STARE datasets, which were trained independently.

4.4.3 Preliminary Results: Vessel Masks and Features

In an enhanced TDBN, the weights in the input and the output layer of the final network can be directly visualised due to the similarity of dimensions of these weights to image patches. To be more precise, each vector of weights $[w_{1,i}, w_{2,i}, ..., w_{i,i}, ...w_{256,i}]$ linking a unit h_i^1 in the first hidden layer to units $[v_1, v_2, ..., v_i, ...v_{256}]$ in the input layer can be visualised after transforming this vector to a $2D$ array, which has the same size as the input image patches of 16 by 16. The vectors of weights can be thought of as *vessel detection features* because they represent the properties of vessel patterns in fundus image patches, as shown in Figure 3.15. When a vessel pattern described with these weights is detected in an image patch, the related hidden units get activated and the collection of these hidden activations is processed through the network.

In the same way, each vector of weights $[w_{j,1}, w_{j,2}, ..., w_{j,j}, ...w_{j,256}]$ connecting a unit h_j^5 in the last hidden layer to units $[o_1, o_2, ..., o_i, ...o_{256}]$ in the output layer can also be visualised. Because the vectors of these weights transfer hidden representations to the output units, they can be viewed as *vessel labelling features*. When the features demonstrated in Figure 3.15 in Section 3.7.1 in Chapter 3 are closely examined, it is obvious that both vessel detection and vessel labelling features are shaped to generate a specific vessel pattern when any hidden unit is activated.

Examples of generated vessel maps can be found as follows for both the DRIVE and STARE datasets.

4.4.4 Segmentation Results on the DRIVE dataset

In this section, the performance of the proposed network is evaluated on the DRIVE dataset. After examining the segmentation performance of the proposed method on the best and the worst case images, the performance will be compared with that of state of the art methods.

Table 4.2 tabulates the overall performance of the proposed network on the DRIVE

dataset with respect to the evaluation criteria reviewed in Section 4.3. In addition, the highest and lowest performances based on the maximum and minimum accuracies are shown. As understood from this Table, there is not too much difference between the performance metrics of the best and the worst cases based on accuracy. The proposed network obtained its best and worst performances respectively on the 19^{th} and the 3^{rd} images. These images were also reported as having the best and the worst performances in recent studies using supervised methods [5,36,43,44]. In the binarisation of these vessel probability maps, the average threshold value was found to be 0.1305 ± 0.0432 (average $\pm$ standard deviation).

Generated vessel probability maps and binary vessel maps for these images can be found by visual examination of Figure 4.1. In this Figure, the discrimination of the optic disc from blood vessels in both binary maps (best and worst cases) is performed well, despite the similarity of its border to blood vessels regarding contrast levels. Also, inhomogeneous illumination over fundus images and poor contrast of blood vessels do not seem to cause any disruption in the detection of even tiny blood vessels. On the other hand, the proposed network seems to be sometimes misled by pathologies in fundus images and can sometimes respond to these pathologies as if they are a part of blood vessels. This can be observed in the red circular region in the binary vessel map corresponding to the worst case in the same figure. The proposed network was also seen to mistakenly respond to a fraction of cotton wool spots. Although these responses seem weaker than or almost the same as those of neighbourhood capillaries in the related probability map, some pathological responses appear in the final binary map because of very low threshold. Also, readers should be aware that the DRIVE dataset mostly contains healthy fundus images, so it can also be a factor affecting the performance of the network on pathological images.

When compared with the previous studies in Table 4.3, the performance of the proposed network produced larger AUC value, surpassing the performance of other state of the art deep networks proposed by Liskowski and Krawiec [36], Li *et al.* [43], Wang *et al.* [44] and Maji *et al.* [55]. Regarding other evaluation metrics, the performance of the proposed method is comparable with the performances of the previous methods. Among them, Li *et al.*'s approach [43] is the closest to the proposed one: both used a cross-modality learning approach for the vasculature segmentation, and both used fully connected networks. However, the ways the methods are trained vary. My network is a TDBN, where the weights in each layer is initialised with weights learned with the probabilistic training of RBMs. However, Li *et al.*'s network is a traditional fully connected network with a modification, where the first layer of their network is initialised by the weights learned by training a denoising autoencoder. The other two layers were initialised by sampling from a normal distribution. Also, the number of hidden layers used in the networks is different. While my network has 5 hidden layers, Li *et al.*'s network has 3 hidden layers.

Table 4.2: Average and the best and the worst segmentation performances of the proposed method on the DRIVE dataset. The order of performances obtained from individual images is based on accuracy.

	AUC	Accuracy	Sensitivity	Specificity
Mean	0.9752	0.9535	0.7662	0.9808
Max	0.9846	0.9646	0.7955	0.9878
Min	0.9706	0.9429	0.7435	0.9769

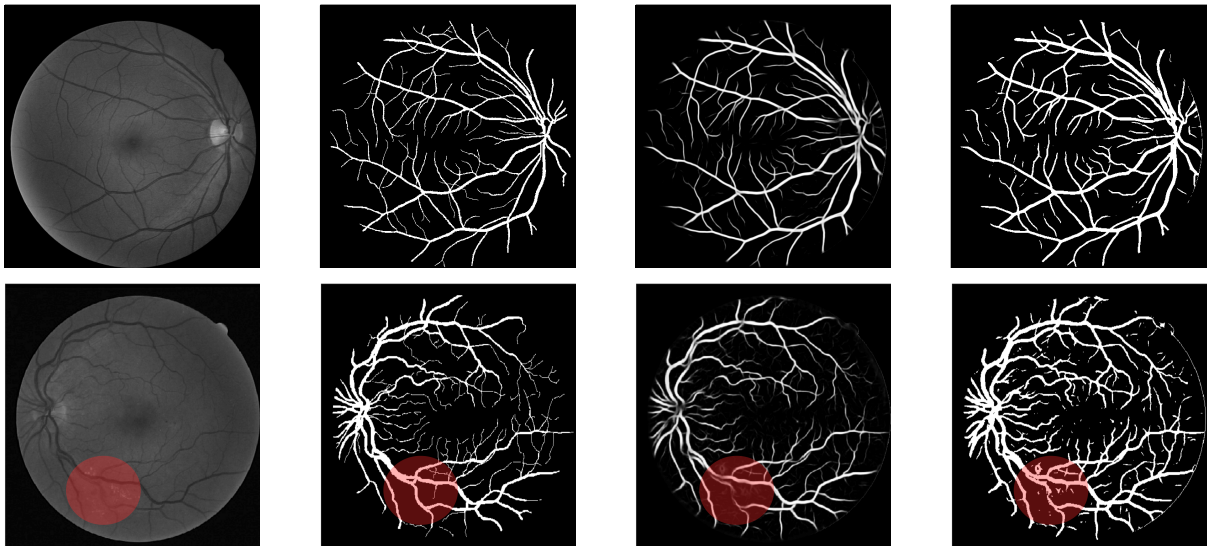


Figure 4.1: The segmentation of vasculature in the DRIVE dataset: the rows from top to bottom show images with the best and the worst segmentation performances based on accuracy. The columns from left to right belong to green channels of fundus images, ground truth images, generated probability maps with the proposed method and their segmented versions respectively. The red disc shows a pathological region.

Table 4.3: The comparison of segmentation performance of the proposed method with the performances of the art methods regarding the DRIVE dataset

–	Year	Method	AUC	Accuracy	Sensitivity	Specificity
Supervised M.	2017	The Proposed Method	**0.9752**	0.9535	0.7662	0.9808
	2016	Liskowski and Krawiec [36]	0.9710	0.9515	0.7520	0.9806
	2015	Wang *et al.* [44]	0.9475	**0.9767**	0.8173	0.9733
	2015	Li *et al.* [43]	0.9738	0.9527	0.7569	0.9816
	2015	Li *et al.* [66]	-	0.9522	0.7659	0.9797
	2015	Maji *et al.* [55]	0.9195	0.9327	-	-
	2014	Gonzalez *et al.* [130]	-	0.9412	-	-
	2014	Cheng *et al.* [45]	0.9648	0.9474	0.7252	0.9798
	2012	Fraz *et al.* [5]	0.9747	0.9480	0.7406	0.9807
	2011	Marin *et al.* [53]	0.9526	0.9452	0.7067	0.9801
Unupervised M.	2016	Oliveira *et al.* [62]	0.9513	0.9464	**0.8644**	0.9556
	2014	Zhao *et al.* [86]	-	0.9477	0.7354	0.9789
	2013	Nguyen *et al.* [51]	-	0.9407	-	-
	2013	Fraz *et al.* [131]	-	0.9422	0.7302	0.9742
	2013	Budai *et al.* [78]	-	0.9572	0.644	**0.987**
	2012	Fraz *et al.* [132]	-	0.9430	0.7152	0. 9759

4.4.5 Segmentation Results on the STARE dataset

In this section, the performance of the proposed network is evaluated on the STARE dataset. Table 4.4 compares overall, the best and the worst performances of the proposed method in the STARE dataset. Similar to the DRIVE Dataset, the differences between performance metrics of the best and the worst cases based on accuracy are trivial. Image *im*0291 and image *im*0240 respectively yielded the maximum and minimum accuracies in the STARE dataset. The average of thresholds used for the binarisation of each probability map was 0.1386 ± 0.0745 (average $\pm$ standard deviation). This average threshold was very close to that calculated for the DRIVE dataset, which was 0.1305 ± 0.0432.

When considering the threshold values for the DRIVE and STARE dataset, it is obvious that the proposed method generated almost the same threshold value, which showed consistency of the method regarding how to react vessel pixels. On the contrary to Marín *et al.* 's results, where their threshold was 0.63 for the DRIVE and 0.91 for the STARE dataset. The threshold values for Fraz *et al.* were 0.55 and 0.64 for the DRIVE and the STARE respectively [5]. Cheng found the threshold of 0.84 for both the DRIVE and STARE datasets [45].

Figure 4.2 shows the probability and binary maps of the best and worst case images. As seen from this Figure, the performance of the proposed network does not depend on the contrast of a fundus image, and the network can deal with highly uneven illumination. The best case image has poorer contrast and worse inhomogeneous illumination than the worst case image. Another interesting observation from this figure is that the ground truth image provided for image *im*0240 is not consistent regarding the labelling of very small vessels. Only labelling of the capillaries in the purple circular region is apparent though capillaries in this region and those in the green circular regions have almost the same visibility. The inconsistency in vessel labelling affected the performance of the network because both capillary groups were detected by the proposed network. Some of the capillaries detected by the proposed method seems acceptable even though they are not labelled in the ground truth. These factors might be the main reasons why this image yields the lowest accuracy despite its lack of response to the background noise as in the best case image and its high responses to the vessels. The main downside of the method visible in both images, is the false detection of the borders of the optic discs, which was not the case for the best and worst case images in the DRIVE dataset. This can be due to their thicker and sharper appearances in the STARE dataset.

Table 4.5 compares the performance of the proposed method with the performance of previously reported methods. Although the performance of the proposed method is not at the best level regarding any particular performance metrics in the STARE dataset, overall performance is comparable with the state of the art methods. On the other hand, readers should be aware that some pixels in training datasets used by Wang *et al.* [44] and Fraz *et al.* [5] overlapped with those in their test datasets. Because of using some common pixels for the training and testing, their results may lead to slightly higher expectation.

Table 4.4: Average and the best and the worst segmentation performances of the proposed method on the STARE dataset. The order of performances obtained from individual images is based on accuracy.

	AUC	Accuracy	Sensitivity	Specificity
Mean	0.9859	0.9661	0.7914	0.9858
Max	0.9881	0.9771	0.7787	0.9920
Min	0.9756	0.9507	0.7908	0.9768

Table 4.5: The comparison of segmentation performance of the proposed method with the performances of the art methods regarding the STARE dataset.

	Year	Method	AUC	Accuracy	Sensitivity	Specificity
Supervised M.	2017	The Proposed Method	0.9859	0.9661	0.7914	0.9858
	2016	Liskowski and Krawiec [36]	**0.9880**	0.9696	0.8145	**0.9866**
	2015	Li *et al.* [43]	0.9879	0.9628	0.7726	0.9844
	2015	Wang *et al.* [44]	0.9751	**0.9813**	0.8104	0.9791
	2014	Cheng *et al.* [45]	0.9844	0.9633	0.7813	0.9843
	2012	Fraz *et al.* [5]	0.9768	0.9534	0.7548	0.9763
	2011	Marin *et al.* [53]	0.9769	0.9526	0.6944	0.9819
Unsupervised M.	2016	Oliveira *et al.* [62]	0.9544	0.9532	**0.8254**	0.9647
	2015	Annunziata [81]	0.9655	0.9562	0.7128	0.9836
	2014	Zhao *et al.* [86]	-	0.9509	0.7187	0.9767
	2013	Nguyen *et al.* [51]	-	0.9324 -	-	-
	2013	Fraz *et al.* [131]	-	0.9423	0.7318	0.9660
	2013	Budai *et al.* [78]	-	0.9386	0.58	0.982

Among some state of the art methods, the worst accuracy was related to image $im0004$, which was not the case for the proposed method. This image is striking with its very low contrast for vessels and higher level of inhomogeneous background illumination. A very interesting finding regarding this image is that some very successful methods produced very low sensitivity values because they could not detect some vessels properly. For example, Fraz *et al.* [5] reported a sensitivity of 0.4360, Wang *et al.* [44] that of 0.5483 and Li *et al.* [43] that of 0.5759. However, the sensitivity produced by the proposed method was 0.7221, which was not very different from the average sensitivity.

The STARE dataset contains equal numbers of pathological and healthy fundus images. This dataset can be used to evaluate the performance of the segmentation methods on pathological images. Table 4.6 compares the performance of the proposed method with the state of the art methods regarding only pathological images. The proposed method has the second best AUC after that of Liskowski and Krawiec [36], who used CNNs for retinal vasculature segmentation.

Table 4.6: The comparison of segmentation performance of the proposed method with the performances of the art methods regarding the pathological images of the STARE dataset.

	Year	Method	AUC	Accuracy	Sensitivity	Specificity
Supervised	2017	The Proposed Method	0.9819	0.9641	0.7484	0.9860
	2015	Li *et al.* [43]	-	0.9672	**0.7800**	0.9805
	2015	Wang *et al.* [44]	0.9670	**0.9773**	0.7473	0.9750
	2016	Liskowski and Krawiec [36] *	**0.9846**	0.9691	0.7779	**0.9884**

* These results were calculated over the probability maps provided by the authors in their website

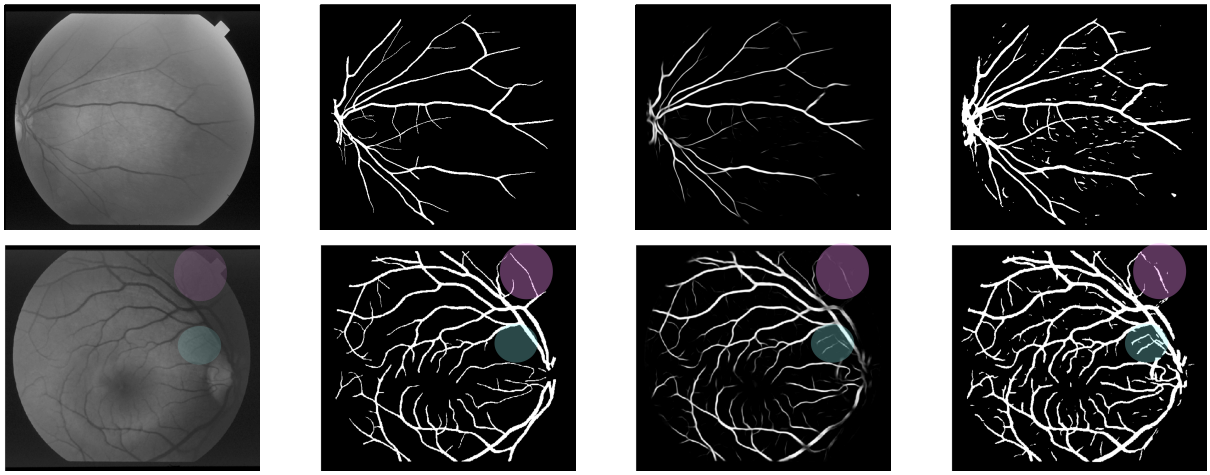


Figure 4.2: The segmentation of blood vessels in the STARE dataset: the rows from top to bottom show images with the best and the worst segmentation performances based on accuracy. The columns from left to right belong to green channels of fundus images, ground truth images, generated probability maps with the proposed method and their segmented versions respectively. The green and purple discs show vessels with almost the same features. Altough vessels in the purple disc are labelled in the ground truth data, those in the green one are missing.

4.5 Super-resolution Segmentation With Enhanced TDBNs

The aim of the super resolution applications is to transform low resolution input images to higher resolution images by improving the appearances of objects with increased sharpness of detail. There are various approaches to super resolution such as reconstruction based, learning based and interpolation based [128, 133, 134]. Among learning based approaches, RBM training was also used for super resolution applications by Gao *et al.* [135]. Gao *et al.*'s approach was based on the generative training of an RBM to learn a joint representation of low resolution natural images and their higher resolution counterparts, which were provided as pairs in the visible layer. When high resolution estimates of low resolution images were required at run time, initial estimates of high resolution images were paired with their low resolution images. These initial estimates of high resolution images were generated with a basic super-resolution method. The evolution of these initial estimates to examples of high resolution images with a trained RBM were realised by performing several steps of CD training.

In this study, the generation of vessel masks given fundus image patches (the segmentation of fundus image patches with cross-modality learning) is combined with a super-resolution approach. This new segmentation approach will be called *super-resolution segmentation*. This approach is proposed to detect vasculature in high resolution fundus images in shorter segmentation time. The implementation of this segmentation approach will be performed with an enhanced TDBN that will be used to learn a mapping between low resolution fundus image patches and high resolution vessel masks. Although the candidate applications of this super-resolution segmentation approach seem limited with the segmentation based on the generation a vessel mask given an image patch, this

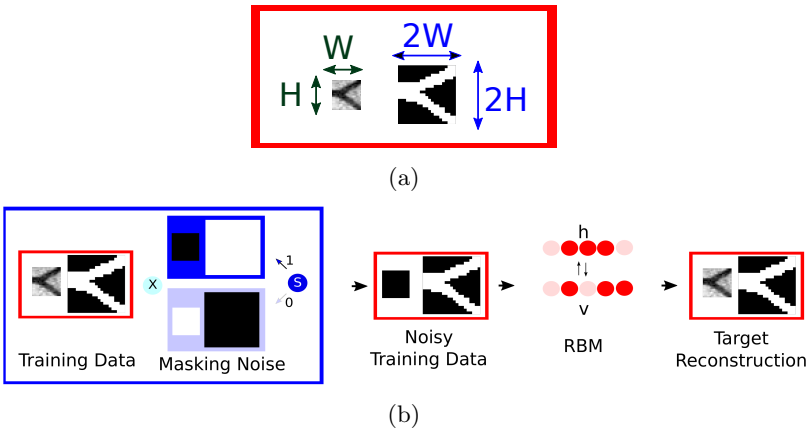


Figure 4.3: (a) A pair of a fundus image patch and its vessel mask for super-resolution application, where the scale of the fundus image patch is 0.5 (b) The training of an RBM with denoising in order to learn the correlation between a low resolution image patch and corresponding high resolution vessel mask given in (a).

approach can be implemented with different network architectures such as a CNN, when such architectures are used to label multiple pixels given a fundus image patch [36, 66].

In the implementation of this super-resolution approach, a fundus image patch of low resolution is associated through an enhanced TDBN with a vessel mask of high resolution in the training dataset. In the preparation of this training dataset, initially both fundus image patches and their vessel masks are sampled from high resolution images and their vessel maps respectively. Then, the resolution of fundus image patches is reduced by down-sampling these image patches by the factor of an arbitrary number d. d can be 2 or 4 as preferred for the following applications. *Scale* will also be used in the place of inverse of down-sampling rates in order to inform readers about the reduced resolution of fundus image patches.

There is no significant change on the training of an enhanced TDBN for this type of segmentation. Figure 4.3b shows the adaptation of the training of an RBM for a super-resolution segmentation task that learns the correlation between low resolution fundus image patches and their high resolution vessel masks. Similarly, Figure 4.4 demonstrates the optimisation of the pre-trained network for super-resolution vessel mask generation. The same architecture of the enhanced TDBN, introduced for the segmentation of low resolution fundus image datasets, can be used in super-resolution segmentation applications by adapting only unit numbers in the input and output layers in the final network demonstrated in Figure 4.4.

4.6 The Segmentation of Vasculature in High Resolution Fundus Images

In this section, CHASE_DB1 and HRF datasets will be used to evaluate the performance of an enhanced TDBN in the super-resolution vessel segmentation task.

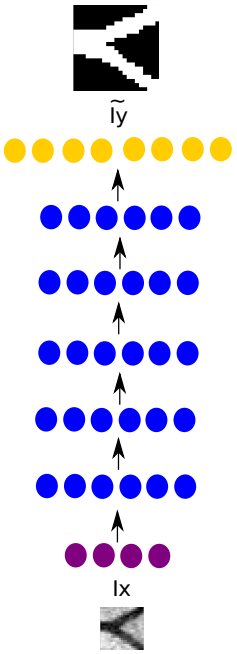


Figure 4.4: A trained enhanced TDBN for super-resolution segmentation task: $\mathbf{I}_x$ is an input vector transformed from a low resolution image patch and $\hat{\mathbf{I}}_y$ is an output vector to be transformed to a high resolution vessel mask.

4.6.1 From Super-resolution Vessel Masks to Probability Maps

When the size of a fundus image patch was the same as the size of a generated vessel mask, the size of the probability map for an entire fundus image was equal to the size of this fundus image. This probability map was calculated as explained in Section 4.4.1. However, in the present application, the size of a generated vessel mask is naturally larger than that of an image patch so the size of the probability map for the entire fundus image is also larger than the size of this fundus image. Therefore, (4.2) is replaced with (4.4) in the calculation of the super-resolution probability map.

$$\hat{P}_i = \frac{1}{W_p \cdot H_p} \sum_{n=1}^{W_p \cdot H_p} P_i^n, \qquad i = 1 : W_c \cdot H_c \tag{4.4}$$

where $W_c = a \cdot W_f$ and $H_c = a \cdot H_f$; a corresponds to magnification factor.

4.6.2 Experimental Settings

In the following experiments, CHASE_DB1 and HRF datasets were used for the evaluation of the performance of the enhanced TDBN for the super-resolution segmentation. For this task, fundus image patches sampled from either dataset were down-sampled by a factor of 2. Also, HRF dataset was used for another experiment, where downsampling rate was 4 because of its higher resolution.

The size of a fundus image patch used in Section 4.4.2 was maintained for this ap-

plication, which was 16 by 16. However, the size of a vessel mask varied depending on down-sampling factor. When this factor was 2, the size of a vessel mask became 32 by 32 and when it was 4, the size of labelling mask became 64 by 64.

The training dataset of CHASE_ DB1 was prepared by using the first 20 images in this dataset. The number of samples in this training dataset was $1,349,693$. The remaining 8 images were used to evaluate the performance of the network. The same division was also used in [43].

The training dataset of HRF used the first 24 images of the dataset. This kind of division of HRF dataset provided equal numbers of examples from healthy, glaucoma and diabetic retinopathy subsets. The number of training samples was $2,502,706$ at the scale of 0.5 and $1,176,700$ at the scale of 0.25. Dense sampling was used inside FOV masks for both datasets as explained in Section 4.4.2.

The training parameters of the proposed network were the same as those detailed in Section 4.4.2. The only difference was that the learning rate during the finetuning was changed to 0.1 and the training lasted for 70 epochs, when the down-sampling factor was 2. When this factor was 4, the learning rate during the finetuning was reduced further to 0.01, and training was completed at 70 epochs.

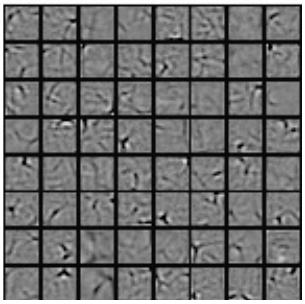
4.6.3 Preliminary Results : Vessel Masks and Features

Figure 4.5 shows obtained features in the fine-tuning stage of training of the proposed network with HRF dataset. The scale of fundus image patches in the training dataset was 0.25. As seen from the figure, both the first layer features and the last layer features are well described. Figure 4.6 demonstrates some examples of low resolution fundus image patches from the training dataset and the high resolution vessel masks generated by the enhanced TDBN for them. In many cases, larger vessels are correctly regenerated despite the resolution difference between images represented in the input and output layers.

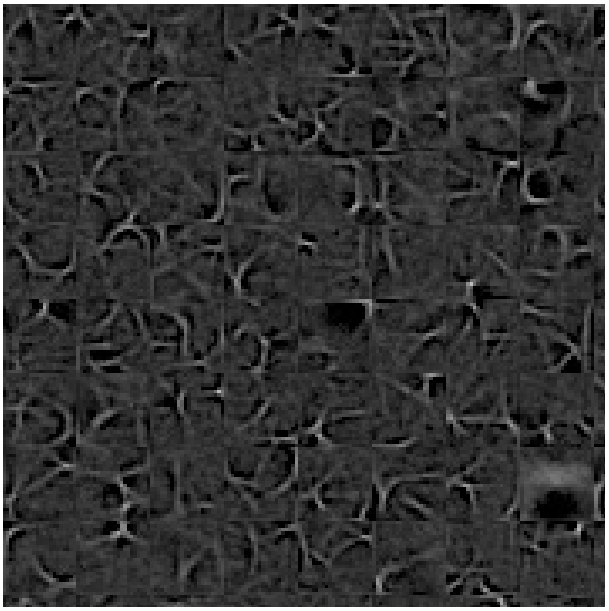
Examples of generated vessel maps for super-resolution segmentation can be found in the following sections for both HRF and CHADE_DB1 datasets.

4.6.4 Segmentation Results on the CHASE_ DB1 dataset

In this section, the performance of the proposed network will be evaluated on the vasculature segmentation of CHASE_DB1 dataset. Table 4.7 tabulates the overall performance of the proposed method on the segmentation of retinal vasculature in CHASE_ DB1 dataset. The maximum and minimum accuracy levels achieved with this method on this dataset are associated with image Image_11R and image Image_12L. The average threshold level for binarising images in this dataset was found 0.1393. Figure 4.7 demonstrates the best and worst case segmentation results. As seen from this Figure, both fundus images suffer from very bad imaging conditions in addition to pathologies (white lesions). The contrast of blood vessels is very poor in both images, especially in the worst accurately segmented image. Despite not having clearly visible blood vessels to the naked eye, the proposed

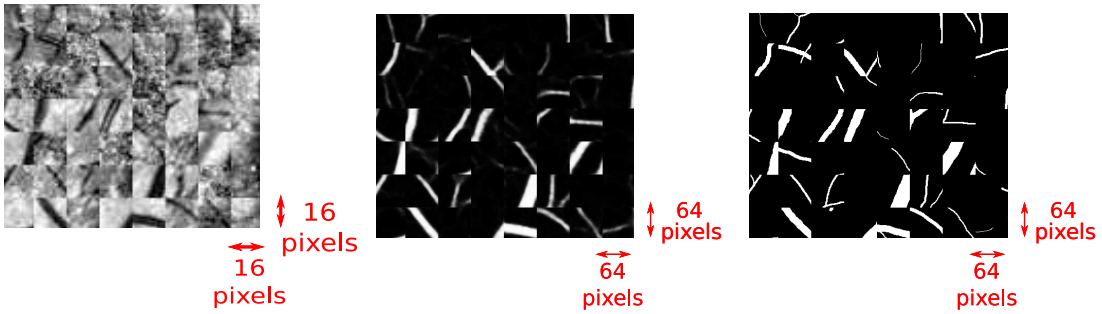


(a)



(b)

Figure 4.5: Some of generated features learned during the fine-tuning stage of the training of the proposed network for super-resolution segmentation in HRF dataset (the down-scaling factor for fundus image patches was 4) . (a) The first layer features (b) The last layer features



(a) (b) (c)

Figure 4.6: (a) Randomly sampled fundus image patches from HRF test set, where the down-scaling factor for fundus image patches was 4 (b) Generated vessel masks corresponding to these image patches by the proposed method (c) Ground truth image patches

Table 4.7: Average and the best and the worst segmentation performances of the proposed method on CHASE_DB1 dataset. The order of performances obtained from individual images is based on accuracy and down-sampling rate for fundus image patches is 2.

Scale		AUC	Accuracy	Sensitivity	Specificity
	Mean	0.9751	0.9589	0.7230	0.9824
0.5	Max	0.9808	0.9686	0.7579	0.9861
	Min	0.9734	0.9513	0.7383	0.9760

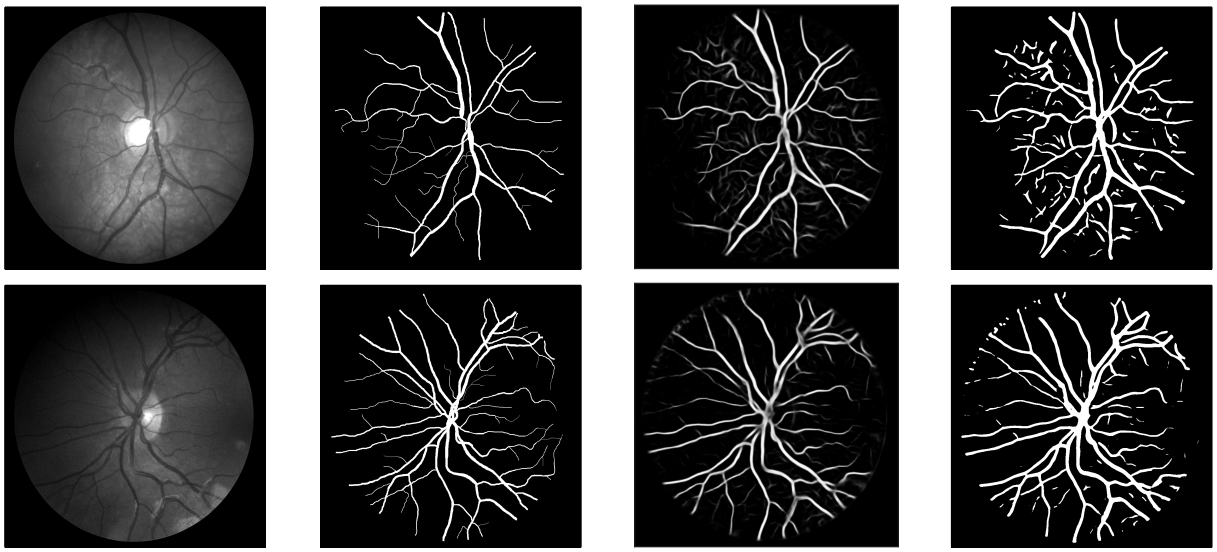


Figure 4.7: The segmentation of blood vessels in CHASE_DB1 dataset: the rows from top to bottom show images with the best and the worst segmentation performances based on accuracy. The columns from left to right belong to green channels of fundus images, ground truth images, generated probability maps with the proposed method and their segmented versions respectively. (Both images were thresholded at 0.15).

method detected the thicker blood vessels of poor contrast, and even many of the capillaries. There is almost no response to the white lesions in both images. However, the proposed method seems very sensitive to vessel-like structures. This can result in some ghost vessels in binary vessel maps as happened in Image_11R.

Table 4.8 compares the performance of the proposed method for super-resolution segmentation with the performances of the state of the art methods applying the segmentation in the original resolution. Despite the resolution difference of the images where vessels are detected and labelled, the proposed method has the best AUC and the best specificity values among supervised and unsupervised methods. It also has the largest sensitivity among supervised methods. Interestingly, the proposed method surpassed the performance of a similar network proposed by Li *et al.* [43] in all four metrics, which performed vessel segmentation in the original resolution of the images. Because Fraz *et al.* [5] used a different split of images for training and testing dataset and Roychowdhury *et al.* [79] used all images in this dataset for the evaluation of their method, it is not possible to make a direct comparison with Fraz *et al.*'s or Roychowdhury *et al.*'s results.

Table 4.8: The comparison of segmentation performance of the proposed method with the performances of state of the art methods regarding CHASE_ DB1 dataset.

	Year	Method	AUC	Accuracy	Sensitivity	Specificity
Supervised	2017	Proposed Method	**0.9751**	0.9589	0.7230	**0.9824**
	2015	Li *et al.* [43]	0.9628	0.9429	0.7118	0.9791
	2012	Fraz *et al.* [5]	0.9712	**0.9649**	0.7224	0.9711
Unsupervised	2015	Roychowdhury *et al.* [79]	0.9623	0.9467	**0.7615**	0.9575

4.6.5 Segmentation Results on the HRF dataset

In this section, the performance of the proposed network will be evaluated on the vasculature segmentation of HRF dataset. Aside from the CHASE_DB1 dataset, the HRF dataset was used in two experiments. A factor of 2 was used to reduce the resolution in the first one and the factor of 4 was used in the second one. The average, the best and the worst performance (based on accuracy) of the proposed method on the HRF dataset are given in Table 4.9 for both experiments. Image 15_h and image 14_dr are respectively the most and the least accurately segmented images in the first experiment and image 15_h and image 10_dr in the second experiment.

Figure 4.8 illustrates the best and the worst case images when fundus image patches were in the scale of 0.5, when compared to their label patches. Images in this dataset were also observed to have poor contrast of vessels. Pathological lesions seems very spread in the worst case image. Despite the poor contrast, the best case image yields almost the same result as the ground truth image. In the worst case image, the vasculature is almost completely detected. However, there are some falsely detected structures. Under close examination, they seem to correspond to dark patterns inside lesions.

Figure 4.9 demonstrates the best and worst case images when fundus image patches are at the scale of 0.25. Despite an increased resolution difference between input image patch and generated vessel map, nearly complete vessel networks are obtained in both cases.

The performance of the proposed method is compared with the performance of state of the art methods in Table 4.10. Because the performance of the proposed method was evaluated in a small subset selected from the entire dataset, it is not possible to make any direct comparison of this perfomance with the segmentation perfomances reported by Cheng *et al.* [45], Yu *et al.* [136] and Annunziata *et al.* [81], which provided the segmentation performance calculated over all images in HRF dataset. However, their results are presented to give a general view to readers. Also, it was possible to recalculate the segmentation performance of the method proposed by Odstrcilik *et al.* [4] for the test subdataset used in this research, because of reported image based segmentation performance. This new performance is given with the reference of Odstrcilik *et al.** .

When compared with the performance metrics obtained from Odstrcilik *et al.**, the proposed method has better AUC, accuracy and specificity levels in both scales (0.5 and 0.25) in all subsets of diabetic retinopathy, glaucoma and heathy and over the entire dataset. Although Odstrcilik *et al.** provided better sensitivities for subsets of diabetic retinopathy and glaucoma, the sensitivity of the proposed method surpassed their sensitivity in

Table 4.9: Average and the best and the worst segmentation performances of the proposed method on HRF dataset. The order of performances obtained from individual images is based on accuracy and down-sampling rates for fundus image patches are 2 and 4.

Scale		AUC	Accuracy	Sensitivity	Specificity
0.5	Mean	0.9803	0.9645	0.7428	0.9864
	Max	0.9903	0.9732	0.8258	0.9883
	Min	0.9682	0.9540	0.6638	0.9829
0.25	Mean	0.9772	0.9622	0.7314	0.9850
	Max	0.9888	0.9708	0.8248	0.9857
	Min	0.9749	0.9523	0.7199	0.9791

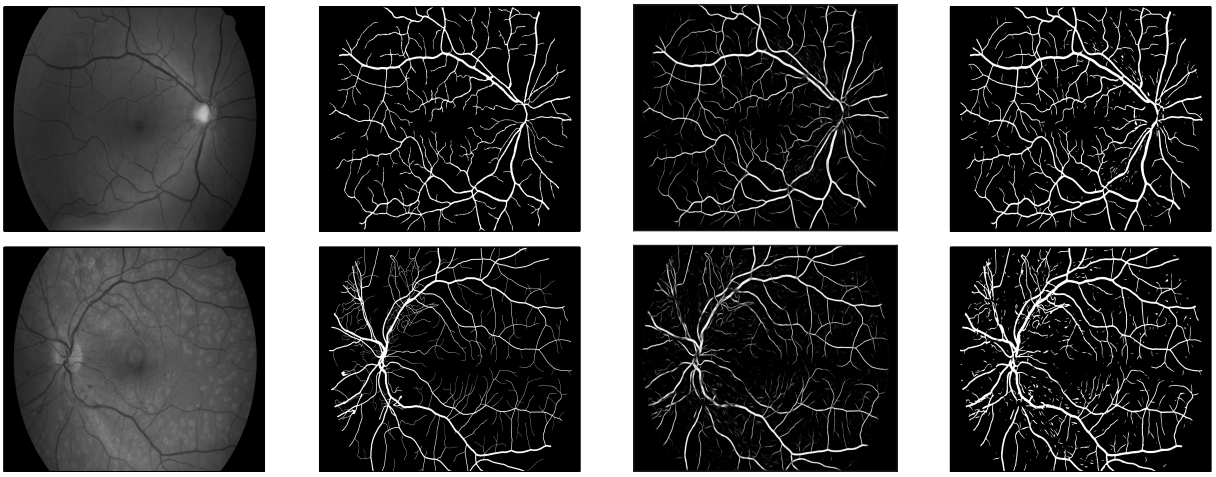


Figure 4.8: The segmentation of blood vessels in HRF dataset in the scale of 0.5: the rows from top to bottom show images with the best and the worst segmentation performances based on accuracy. The columns from left to right belong to green channels of fundus images, ground truth images, generated probability maps with the proposed method and their segmented versions respectively. (Both images were thresholded at 0.15).

healthy subset for both resolution levels.

Regarding the proposed method, increased resolution difference between fundus image patches and their vessel masks in the training dataset made a trivial difference on the performance of the network in terms of AUC, accuracy and specificity metrics. Larger down-sampling rate in this super-resolution segmentation seems to cause only small reductions in sensitivities calculated from individual subsets.

4.6.6 Evaluation of Accelerated Segmentation With Super-resolution

The segmentation of a fundus image is usually performed in the original resolution of this image. In the other words, both input images and their output labels are in the same resolution and in the same size. However, the proposed super-resolution segmentation approach allows to perform the detection of structures of interest in a reduced resolution of fundus images, while generating output labels for these structures at the original resolution of the same images. It can be claimed that super-resolution segmentation can reduce the segmentation time roughly a square of the amount of the reduction in the resolution of the input image patches.

In order to evaluate this claim, Table 4.11 compares segmentation time per pixel of high

Table 4.10: The comparison of segmentation performance of the proposed method with the performances of the art methods regarding HRF dataset

Dataset	Method	AUC	Accuracy	Sensitivity	Specificity
Entire	Proposed Method(Scale 0.5)	**0.9803**	**0.9645**	0.7428	**0.9864**
	Proposed Method(Scale 0.25)	0.9772	0.9622	0.7314	0.9850
	Odstrcilik *et al.**[4]	0.9685	0.9479	**0.7753**	0.9653
	Annunziata *et al.* [81]	-	0.9581	0.7128	0.9836
	Cheng *et al.* [45]	-	0.9614	0.7041	0.9864
	Budai *et al.* [78]	-	0.961	0.669	0.985
	Odstrcilik *et al.* [4]	0.9678	0.9494	0.7741	0.9669
	Yu *et al.* [136]	-	0.9515	0.7811	0.9685
Diabetic Retinopathy	Proposed Method (Scale 0.5)	**0.9736**	**0.9590**	0.6911	**0.9843**
	Proposed Method(Scale 0.25)	0.9707	0.9574	0.6812	0.9836
	Odstrcilik *et al.**[4]	0.9614	0.9429	**0.7550**	0.9606
	Annunziata *et al.* [81]	-	0.9554	0.6997	0.9787
	Cheng *et al.* [45]	-	0.9561	0.6313	0.9851
	Budai *et al.* [78]	-	0.955	0.658	0.977
	Odstrcilik *et al.* [4]	0.9589	0.9445	0.7463	0.9619
	Yu *et al.* [136]	-	0.9460	0.7604	0.9625
Glaucoma	Proposed Method (Scale 0.5)	**0.9796**	**0.9644**	0.7216	**0.9869**
	Proposed Method (Scale 0.25)	0.9752	0.9618	0.7057	0.9856
	Odstrcilik *et al.* *[4]	0.9680	0.9473	**0.7804**	0.9626
	Annunziata *et al.* [81]	-	0.9603	0.7566	0.9785
	Cheng *et al.* [45]	-	0.9634	0.6921	0.9875
	Budai *et al.* [78]	-	0.965	0.687	0.986
	Odstrcilik *et al.* [4]	0.9704	0.9497	0.7900	0.9638
	Yu *et al.* [136]	-	0.9518	0.7890	0.9662
Healthy	Proposed Method (Scale 0.5)	**0.9879**	**0.9702**	**0.8156**	**0.9879**
	Proposed Method (Scale 0.25)	0.9858	0.9675	0.8071	0.9859
	Odstrcilik *et al.* *[4]	0.9762	0.9535	0.7904	0.9728
	Annunziata *et al.* [81]	-	0.9587	0.6820	0.9935
	Cheng *et al.* [45]	-	0.9647	0.7889	0.9865
	Budai *et al.* [78]	-	0.961	0.662	0.992
	Odstrcilik *et al.* [4]	0.9742	0.9539	0.7861	0.9750
	Yu *et al.* [136]	-	0.9566	0.7938	0.9767

* These performance metrics were calculated from the related paper and only include the performance evaluation regarding the last 21 images.

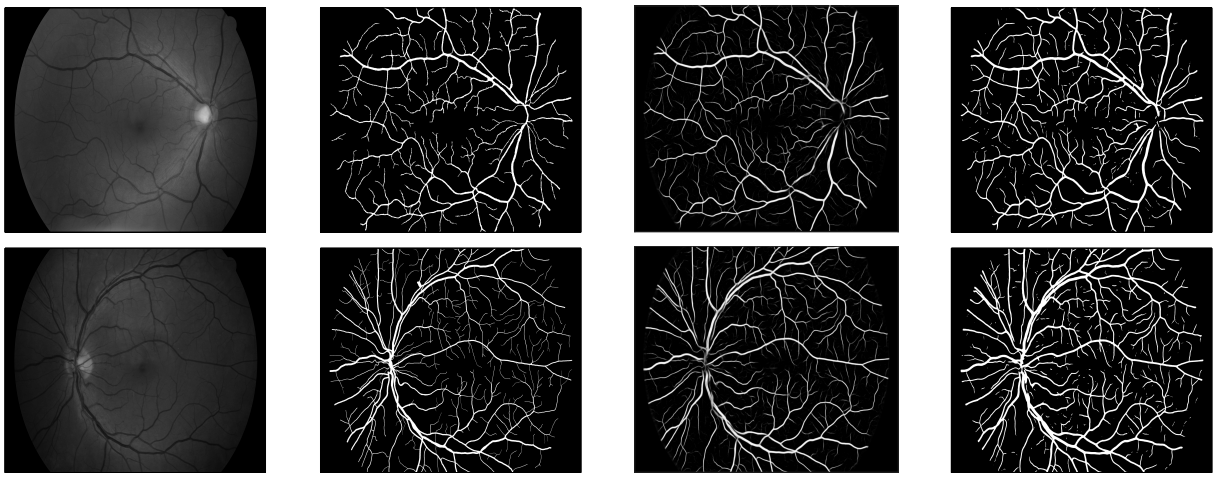


Figure 4.9: The segmentation of blood vessels in HRF dataset in scale 0.25: the rows from top to bottom show images with the best and the worst segmentation performances based on accuracy. The coloumns from left to right belong to green channels of fundus images, ground truth images, generated probability maps with the proposed method and their segmented versions respectively. (Both images were thresholded at 0.15).

Table 4.11: Comparison of segmentation times of super-resolution TDBN on high resolution fundus image datasets with those of enhanced TDBN on low resolution fundus image datasets.

Method	Dataset	Scale of Fundus Image Patches	Segmentation Time (second per pixel)
Super-Resolution Segmentation	CHASE_DB1	0.5	1.2660 e^{-5}
	HRF	0.5	1.1685 e^{-5}
	HRF	0.25	0.6528 e^{-5}
Traditional Segmentation	DRIVE	1	3.0334 e^{-5}
	STARE	1	3.0337 e^{-5}

resolution images obtained with super-resolution segmentation and that of low resolution images generated with traditional segmentation. In this comparison, enhanced TDBNs were used for the implementation of both segmentation types. The calculation of segmentation time per pixel provides a fair comparison for different sizes of image datasets. In this table, the scale of fundus image patches shows the relative resolution of fundus image patches to that of generated vessel masks. For example, a scale of 1 means that the resolution of a fundus image patch and that of a vessel mask is same. A scale of 0.5 means that the resolution of a fundus image patch is half of the resolution of the generated vessel mask so, the size of the vessel mask is quadruple of the size of an image patch.

As seen in this table, there is almost a linear relation between the segmentation time and the resolution of input images. In the original resolutions, the DRIVE and STARE datasets are segmented almost in the same pace (second per pixel). On the other hand, reducing the resolution of the input image with a down-sampling rate of 2 or 4 accelerates the segmentation with the same amount of the down-sampling rate. This is due to reducing the number of pixels evaluated during the segmentation with the super-resolution approach. When compared with the segmentation in the original resolution, a super-resolution segmentation can lead to an acceleration of almost 5 times, without degenerating the accuracy of the segmentation in a remarkable amount (refer to Table 4.10).

4.7 Multiple Vasculature Part Segmentation

Vessel segmentation has been usually viewed as the initial step to calculate the geometrical properties of retinal vasulature. Binary vessel maps created as the result of this segmentation can be further processed to generate vessel centerlines, required to calculate turtousity [137] or can be used as landmarks with bifurcation/crossing points for image registration. Also, vessel centerlines are sometimes seen as a precursor to boundary detection, from which vessel widths [47, 138] may be estimated.

When the segmentation maps are available from ground truth data or as a result of segmentation method [47,48], the detection of vessel centerlines or boundaries has been made through consecutive steps of various segmentation tasks [47–49, 88]. However, this process may make the final step of segmentation/detection/calculation prone to errors which come from earlier steps [47, 48]. However, a method capable of simultaneous segmentation of

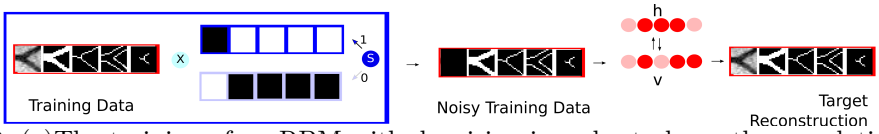


Figure 4.10: (a)The training of an RBM with denoising in order to learn the correlation between a fundus image patch and a group of label patches.

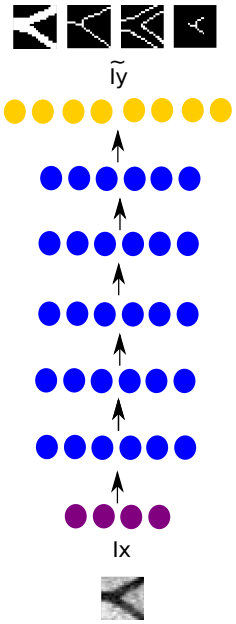


Figure 4.11: A trained enhanced TDBN for multiple types of label patch generation: $\mathbf{I}_x$ is an input vector transformed from an image patch and $\hat{\mathbf{I}}_y$ is an output vector to be transformed to multiple vessel masks, namely vessel interior, centreline, boundary and bifurcation/crossing pattern in centreline.

vessel interior, centreline and boundary may generate more reliable performance by providing more consistent vessel descriptions. To date, deep networks have been suggested for multiple location detection in fundus images such the optic disc or fovea in addition to vasculature [58, 60]. However, the segmentation of vasculature parts with a deep network was not introduced.

The proposed architecture for multiple vasculature part segmentation is demonstrated in Figure 4.11. As seen in this figure, the only difference between the architecture of the enhanced TDBN and that of the proposed network here is the number of units in the output layer. In the new network, the unit number in this layer is scaled up with the number of label vector (e.g centreline, boundary). This difference does not prevent one from using the training scheme of the enhanced TDBN for the training of the multi-labelling network. However, a moderate change is necessary in denoising procedure applied for the training of RBMs/TRBM in the first layer of the network, in order to motivate the network to simultaneously recognise similar structures over all labels/vasculature parts. The new denoising is shown in Figure 4.10.

4.8 Vessel Interior, Centreline and Boundary Probability Maps

4.8.1 From Vessel Masks to Probability Maps

In this application, the final network generates an output vector which contains four types of label: vessel interior, vessel centreline, vessel boundary and bifurcation/crossing patterns. Because the size of each mask type is equal to the size of the input image patch, this output vector can be split to four equal length of label vectors. After converting these label vectors to label patches, the probability maps for an entire fundus image can be generated as explained in Section 4.4.1.

4.8.2 Experimental Settings and The Preparation of Training Dataset

An architecture similar to the enhanced TDBN was used for this application. Apart from unit numbers corresponding to output vessel maps, the other parts of the architecture is adopted completely. The size of vessel masks for each type was the same as the size of fundus image patches, which was 16 by 16. Therefore, total unit number in the output layer of the final network was $256 \times 4 = 1024$, whereas the number of units in the input layer was 256. Only the learning rate was changed to 0.08 for 120 epochs in this application and other training parameters detailed in Section 4.4.2 were maintained.

In this application, the ground truth images of vessel centrelines, boundaries and bifurcation/crossing patterns were necessary in addition to those of conventional vesselness used in aforementioned applications. In order to avoid any confusion, vessel maps or vessel masks will be called *vessel interior* maps or *vessel interior* masks after this point. The generation of ground truth images of vessel centreline and boundaries was performed respectively by applying a thinning algorithm, and by applying an edge detection algorithm to given ground truth vessel interior images. In order to produce ground truth images for bifurcation and crossing patterns, the locations of bifurcation and crossing points, detected by Azzopardi and Petkov [41], were utilised to determine locations of bifurcation and crossing patterns at vessel centreline ground truth images. Ground truth images of these patterns were obtained by maintaining the centreline information inside matrices of 5 by 5, which were centred at the locations of bifurcation and crossing points, and by removing the centreline information outside these matrices. In order to evaluate the performance of the application, the DRIVE and STARE datasets were used.

Bifurcation detection was only applied to the DRIVE dataset because the STARE dataset does not have bifurcating/crossing location detection ground truth images. The number of training samples was approximately $2,400,000$ for the DRIVE dataset and $1,800,000$ for the STARE dataset for this application. Examples of training samples can be found in Figure 4.13. Similar to previous applications, denser sampling was performed inside FOV masks in order to give image patches with vessel pixels greater chance to be represented.

4.8.3 Preliminary Results : Vessel Masks and Features

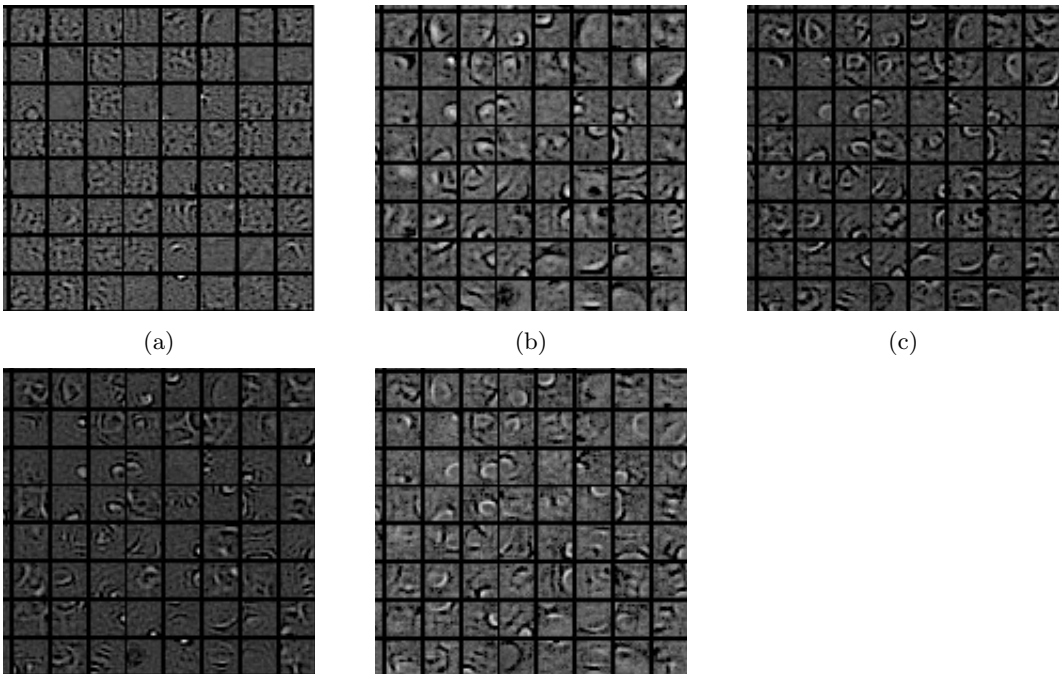
Figure 4.12 visualises weights (features) connecting the input layer and the first hidden layer, and those linking the last hidden layer and the output layer in the final network. In this figure, features belong to four types of label patches (vessel interior, centreline, bifurcation and bifurcation/crossing location) are borrowed from the same unit group for each label type in the output layer; for example, from the first unit to $25th$ unit for each label type. Because of visualising the same unit group in each label type, a similarity between these feature groups is noticeable. For example, the features for centreline labels and those for vessel interior labels look very similar in terms of representing *line detectors*. On the other hand, the features associated with boundary labels are mostly *edge detectors* for the same patterns of vessel tree fractions. Because bifurcation/crossing patterns are cropped from centreline labels and resemble these labels in terms of structure, the features related to these patterns are similar to features of centreline labels

Figure 4.13 shows examples of fundus image patches and generated label patches regarding vessel interior, centerline, boundary and bifurcation/crossing patterns. As demonstrated in Figure 4.13c, hypothetical vessel segments are generated for the vessel segments in given fundus image patches. Also, the effect of background noise and the variation on intensity levels of a vessel segment are not reflected in these hypothetical vessel segments but only the confidence of the network regarding a possible structure in the evaluated patch. The confidence of the network for vessel segments seems not to vary depending on the thickness of the segments but rather on their possibility to be vessels. Because of stretching the contrast in image patches to the range of $[0, 1]$ by applying patch based normalisation, even noise components that appear as if they are very faded vessel segments to the naked eye can be assigned low level of probabilities of being vessel. This may one of the reasons why the network is very sensitive to possible vessel segments even if they are not traced by experts in ground truth images.

Figure 4.13e shows hypothetical vessel centrelines for given fundus image patches. These centrelines accord with the existence of vessel segments detected in Figure 4.13c. The generated boundary label patches demonstrated in Figure 4.13g are also in harmony with generated vessel segments in Figure 4.13c. These boundary label patches have higher probabilities for larger vessels and lower probabilities for tiny vessels. The detection of boundaries also seems to be independent of vessel intensities in fundus image patches, but related to the geometrical properties of vessel patterns.

A further investigation of hypothetical label patches is demonstrated in Figure 5.4. Figure 5.4a shows fundus image patches sampled from a fundus image. Although these image patches are in different contrast ranges to the original fundus image, the contrast levels of vessel segments in these patches are brought to almost the same range by using patch-based normalisation.

Figure 4.13i shows generated bifurcation/crossing patterns for given two fundus image patches. Because the preparation of training samples from related ground truth images was



(a) (b) (c)

(d) (e)

Figure 4.12: Learned features during the training of an enhanced TDBN for multiple types of mask generation. Some examples from features connected to (a) fundus image patches in the input layer of the final network weights from the training of a TDBN (b) vessel interior (c) vessel centerline (d) vessel boundary (e) bifurcation/crossing pattern label patches in the output layer of this network

based on random sampling, not many locations of bifurcation/crossing points were located at the centres of fundus image patches. Also, this random sampling can cause the ground truth patches to include parts of vessel centreline in addition to bifurcation/crossing patterns in some examples. Even some examples may only contain parts of vessel centreline. These factors may prevent the network from learning more discriminative features for the regeneration of bifurcation/crossing patterns. As seen in Figure 4.14, the probabilities of pixels belonging to these patterns increase towards the centres of bifurcation/crossing patterns. This response to the crossing location in Figure 4.14d can be characterised with the sum of two almost perpendicularly elongated Gaussian-like functions with the similar mean values. The response to the bifurcation location in Figure 4.14b can also be described with the sum of Gaussian-like functions but there is a half of one of the Gaussian-like functions missing due to not having 4 vessel branches connected at this location. The angles between these Gaussian-like functions mainly depend on the angles between daughter vessels entering these regions.

Examples of generated probability maps for multiple vasculature part segmentation can be found in the following sections for both the DRIVE and the STARE datasets.



(a) *Fundus Image Patches*

(b) *Ground Truth Vessel Masks*

(c) *Generated Vessel Masks*

(d) *Ground Truth Vessel Centreline Masks*

(e) *Generated Vessel Centreline Masks*

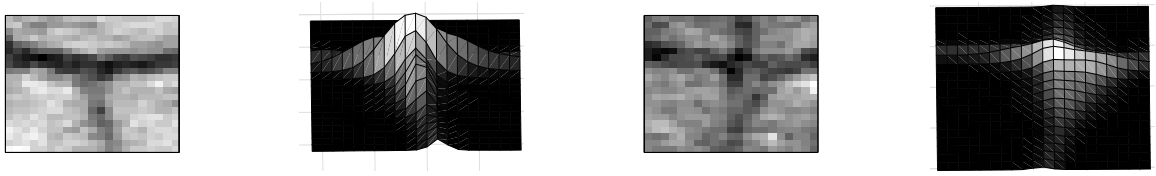
(f) *Ground Truth Vessel Edge Masks*

(g) *Generated Vessel Edge Masks*

(h) *Ground Truth Vessel Bifurcation/Crossing Pattern Masks*

(i) *Generated Vessel Bifurcation/Crossing Pattern Masks*

Figure 4.13: Randomly sampled fundus image patches from the DRIVE test set and generated labels with the proposed method



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

Figure 4.14: (a) A bifurcation vessel pattern in a fundus image patch in the DRIVE dataset, (b) The profile of a generated bifurcation pattern for image patch in (a), (c) A crossing vessel pattern in a fundus image patch in the DRIVE dataset (b) The profile of a generated crossing pattern for image patch in (c)

4.8.4 Segmentation Results

In this section, the performance of the proposed network will be evaluated by using the DRIVE and the STARE datasets. Figure 4.15 and Figure 4.17 show two fundus images with the best and the worst segmentation performance based on the accuracy of vessel interior segmentation results and their probability maps for the DRIVE and the STARE datasets respectively.

As seen in these figures, the probability maps of vessel interior, centreline and boundary point to the same vessels but their related parts. The pixels with high probabilities in these probability maps mostly overlap with pixels representing vessel trees in ground truth images depending on the segmentation performance of the network on vessel interior pixels. Also, there is almost no extra noise/false positives emerged in centreline and boundary probability maps apart from those inherited from vessel interior probability maps.

Such a consistency across probability maps for different vessel parts may be due to the usage of special denoising in the training of RBMs/TRBMs. This consistency across probability maps may bring a big advantage to a vasculature tracking system, which will be introduced later, in terms of maintaining vessel parts even for vessels with lower probabilities; otherwise, whose presences may be lost if they are obtained from binary vessel maps.

Although the aim of this experiment was not to evaluate the performance of the network on segmentation, such an assessment may be useful to compare the performance of the network with that of the enhanced TDBN on vessel interior segmentation, whose accuracy may highly affect the detection of vessel centreline and boundaries. In addition to the performance evaluation on vessel interior segmentation, those on centreline and boundary segmentations are tabulated in Table 4.12 for the DRIVE dataset and Table 4.13 for the STARE dataset. When compared with the performance of the enhanced TDBN in Table 4.2 for the DRIVE and in Table 4.4 for the STARE, the segmentation performances of the multi-labelling network on vessel interior segmentation on both datasets are slightly better in three criteria, AUC, accuracy and sensitivity.

This may indicate that down-sides of training a network with a very high dimensional output layer may be reduced by generating highly related multiple labels and by using a structured pretraining to reveal the relation between the labels; otherwise, the network

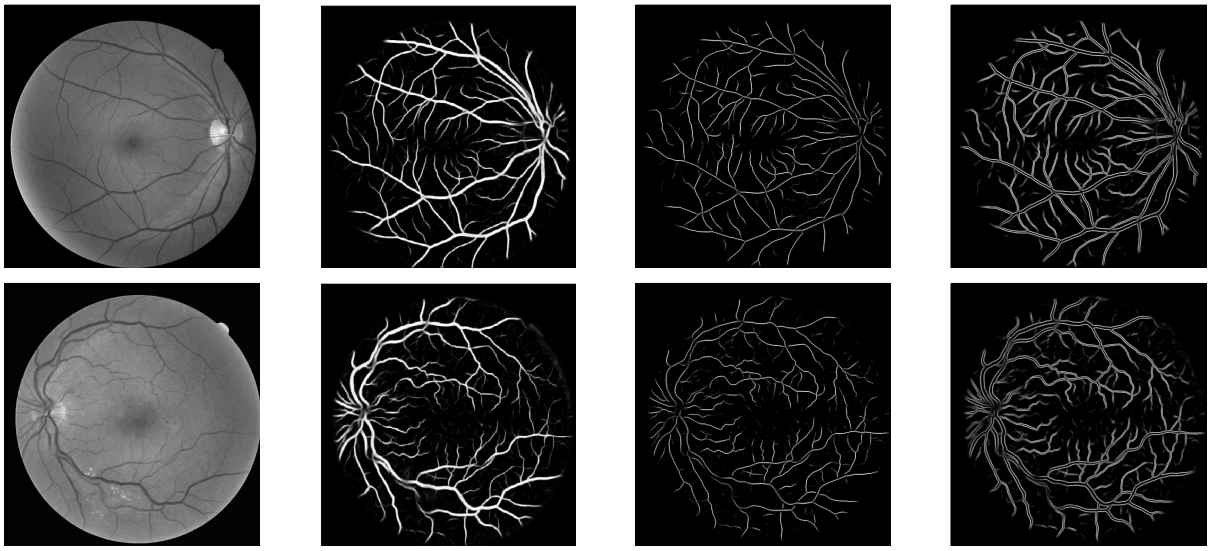


Figure 4.15: The segmentation of vasculature in the DRIVE dataset regarding vessel interior, centreline and boundary segmentation: the rows from top to bottom show images with the best and worst segmentation performances based on accuracy. The columns from left to right belong to green channels of fundus images, probability maps of vessel interior, centreline and boundary respectively.

Table 4.12: The average segmentation performances of the proposed method on the DRIVE dataset.

Segmentation Type	AUC	Accuracy	Sensitivity	Specificity
Vessel Interior	0.9761	0.9542	0.7752	0.980
Centerline	0.9485	0.9611	0.3945	0.9894
Boundary	0.9131	0.9317	0.1902	0.9894

would behave in favour of the training of label-dense image patches (vessel interior) by mostly ignoring the training of label-scarce ones (centreline and boundary). Also, when compared with the performance of the multi-labelling network on the DRIVE dataset regarding vessel interior, centreline and boundaries, its performance on the STARE dataset is better, which may be due to better identification of vasculature on the STARE dataset due to higher resolution of fundus photos.

In addition to vessel part detection, the multi-labelling network was trained for the generation of bifurcation/crossing patterns with the DRIVE dataset. Figure 4.16 demonstrates examples of generated probability maps for these patterns. In these maps, these patterns appear much brighter than other structures by indicating higher probabilities, which also be used to guide the tracker.

Regarding the training time, there was almost no considerable increase in the training time for multiple types of label generation, when compared with single type of label generation. The approximate time for the training of 19 images of the STARE dataset was 5 hours for the single label generation task and was 6 hours for the multiple type label generation. The training was performed by a Linux operation system installed computer with 64 GB RAM.

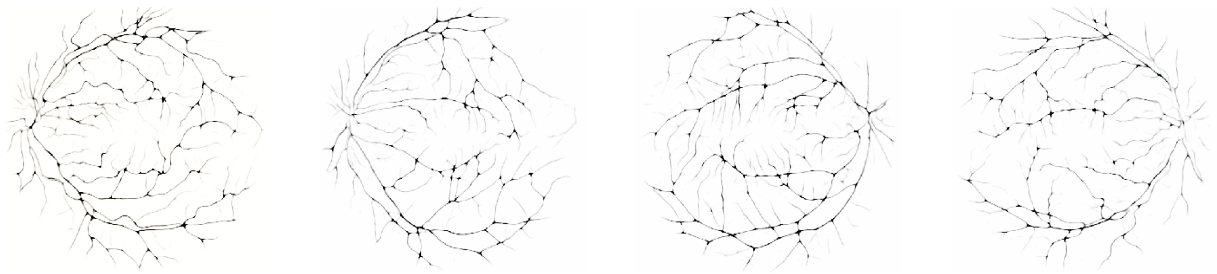


Figure 4.16: Bifurcation/crossing pattern detection in the DRIVE dataset: Probability maps belong to (a) 1^{st} image (b)12^{th} image(c) 13^{th} image (d) 20^{th} image. (The contrast of images is inverted for better visualisation)

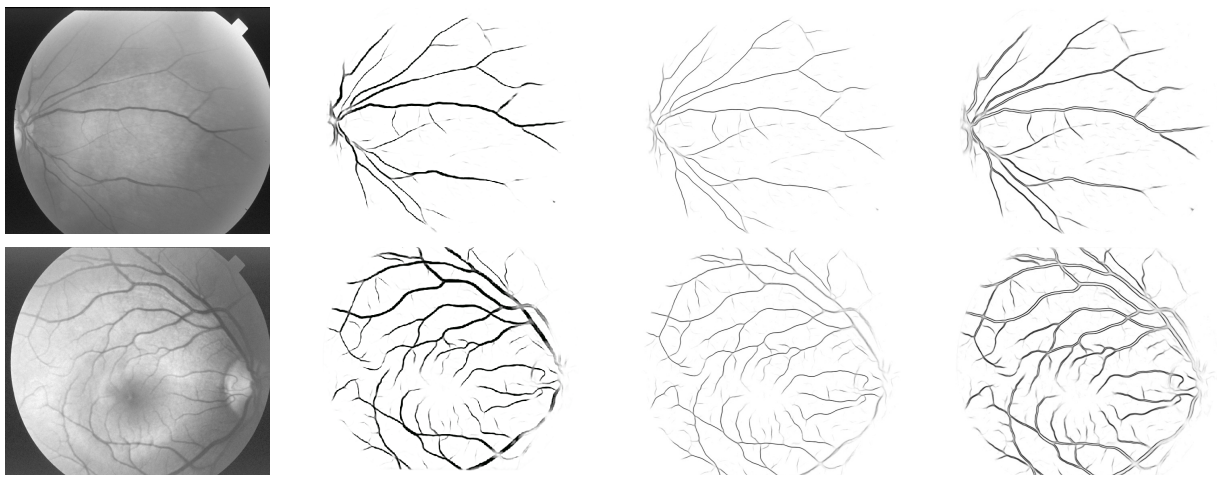


Figure 4.17: The segmentation of vasculature in the STARE dataset regarding vessel interior, centreline and boundary segmentation: the rows from top to bottom show images with the best and worst segmentation performances based on accuracy. The columns from left to right belong to green channels of fundus images, probability maps of vessel interior, centreline and boundary respectively. (The contrast of images is inverted for better visualisation)

Table 4.13: The average segmentation performances of the proposed method on the STARE dataset.

Segmentation Type	AUC	Accuracy	Sensitivity	Specificity
Vessel Interior	0.9860	0.9664	0.7978	0.9855
Centerline	0.9749	0.9794	0.3675	0.9948
Boundary	0.9654	0.9559	0.4605	0.9857

4.9 Conclusion

In this chapter, three networks were introduced for the segmentation of retinal images; an enhanced TDBN, a super-resolution TDBN and a multi-labelling TDBN. The first two networks serve the same aim of segmentation but in different image resolutions; low and high resolutions. The distinction of the second network from the first one was due to combining super-resolution approach with cross-modality learning based segmentation. The introduction of the second network presented some advantages over conventionally segmenting networks regarding segmentation time. On the other hand, the multi-labelling network provided probability maps of various vessel parts with a consistency.

The three networks are initialised from generative training of a DBN. When considering the depth and width of the network, this initial training facilitated the training of the networks. Despite the popularity of another type of network in image segmentation, CNNs, the proposed networks have the advantage of shorter training times, which permits the re-training of the proposed network in feasible times for new datasets. Also, the ability to produce a label patch allows one to check segmentation performance of the network during the training by observing the quality of generated labels.

When compared with that of low resolution images, the segmentation of high resolution images have more importance. Firstly, many images collected in ophthalmology clinics in recent years have high resolutions [9, 139]. Also, the available datasets of high resolution such as CHASE_DB1 and HRF pose more challenges, which may reflect the problems faced in the clinics better. For example, CHASE_DB1 contains images from a multinational ethnicity and young age group, which is mostly described with central light reflex and uneven illumination. The majority of images (two thirds) in HRF dataset is from subjects with either diabetic retinopathy (DR) or glaucoma patients. On the other hand, half of the STARE dataset consists of DR patients, while a fifth of test images in the DRIVE dataset belongs to DR patients. Therefore, the performance evaluation on CHASE_DB1 and HRF datasets may present more meaningful results, which are useful for clinical applications.

Applying a segmentation method to various resolutions of images may pose challenges because the same vessel may be represented with a range of sizes in terms of pixel number. Although less resolution difference between target dataset and the dataset considered for the designs of methods may be tolerable for some methods, they may still need to adjust their design parameters in the case of large differences. This may explain that the number of studies using high resolution fundus image datasets is virtually zero, when compared with the number of methods using low resolution images; though, publicly available datasets for high resolution were publicised approximately 5 years ago (CHASE_DB1 in 2012 and HRF in 2013) [4,5]. To date, the studies on high resolution images were generally either unsupervised [4, 81] or using easily trainable classifier in terms of demanding less number of training samples such as random forests [5, 45].

The application of deep learning to segmentation of high resolution retinal images has been rarely realised [43], despite of the success of deep learning methods on image seg-

mentation and the increased interest of researchers working on vasculature segmentation with these methods in recent years. The difficulty of training a deep network on high resolution images may be due to increased amount of storage for training for CNN examples [140]. Even though a network is successfully trained, the segmentation time may be scaled up with the size of images. On the other hand, the proposed super-resolution TDBN achieved reduction in the segmentation time of high resolution images efficiently while providing better or comparable performance than/with that of previous studies. When compared with the performance of a similar network of Li *et al.* [43], the proposed network outperformed significantly the performance of Li *et al.*'s network on CHASE_DB1 dataset; though, the proposed network generated slightly better or comparable results in low resolution images with those reported by Li *et al.* [43]. This may indicate the necessity of a special treatment for high resolution images, such as larger receptive fields. In the proposed method, an effective increase in the receptive field size was realised by down-sampling of input image patch. Reducing the resolution of input images was also implemented by Budai *et al.* [78] to decrease computation time.

The examples of studies related to multiple location detection in fundus images by using a deep network were limited by the identification of vasculature, the optic disc and fovea [58, 60]. The detection of multiple vessel parts was not realised with deep networks, which may facilitate the quantitative analysis of retinal vasculature. This analysis may provide an important tool to identify changes in the vasculature, which may help to diagnose eye-related diseases such as DR, or can be used as bio-markers to anticipate the incidence of systemic or brain related diseases. In contrast to defining vessel morphology by using binary vessel maps, whose accuracies may vary depending on the performance of segmentation methods, the probability maps for different vessel parts may allow us to analyse even thinner vessels with low contrast, which may be ignored with many segmentation methods.

Chapter 5

Quantitative Vasculature Analysis

5.1 Introduction

In Chapter 3, and Chapter 4, three different architectures of an enhanced TDBN were introduced for traditional segmentation, super-resolution segmentation and multi-labelling segmentation. These architectures were able to generate probability maps, which can be thresholded to give binary vessel maps. Binary vessel maps may not give sufficient quantitative information for research or clinical applications. Quantitative information such as relative vessel widths and branching angles, may be obtained by the use of tracking methods, which might include parameter estimation.

In this chapter, a method for vasculature tracking based on Particle Filters will be introduced. The required information to guide the tracker will be sustained from the multi-labelling network. This network is capable of generating multiple probability maps of retinal vasculature corresponding to vessel interior (vesselness), vessel centreline, vessel boundary and bifurcating/crossing locations. The latter, we refer to as *landmark locations* in the rest of this Chapter. Probability profiles obtained from these probability maps will be utilised as noisy measurements (observations) for the tracker and will be used in the likelihood calculation of vessel description hypothesises such as vessel centreline locations, vessel widths and vessel connection locations (landmark locations).

This tracking system has two modes to deal with (i) regular tracking and (ii) the detection of landmark locations and the identification of daughter vessels. The first, the regular vessel tracking mode, is designed to incorporate various observation types in order to have a better estimation of related structural properties. The second mode, the detection of landmark locations, provides a new approach to estimate landmark locations. This mode also includes a new approach for the identification of all vessels connected to landmark locations.

5.1.1 Motivations For The Proposed Vessel Tracking System

Model based approaches approximate the vasculature trees as if they are built up with the connection of simple structures. For example, vasculature in $3D$ data was modelled with cylinders [141]. Similarly, a vasculature tree in $2D$ images can be simplified with connected vessel segments. Also, the cross section of a vessel in a $2D$ image is usually assumed to be a Gaussian function [34, 72, 97, 98]. This simplifications can be used to detect vessel width for a given centreline location [34, 72, 92, 94, 96, 97, 97, 98, 98].

However, the appearance of a vessel in an image can vary due to the pathologies, noise, such as central light reflex or poor imaging conditions, which can cause the vessel to have lower contrast in addition to be seen more blurred and corrupted. On the other hand, probability based tracking methods do not consider only one assumption to describe a vessel segment. However, they can evaluate multiple assumptions by combining prior beliefs about these assumptions with supports (observations) extracted from the image. Therefore, these methods can model vessels with more complicated descriptions [142]. Examples of the most famous probabilistic tracking methods are Kalman filters and Particle Filters. When compared with Kalman filters, Particle Filters can deal with non-linear, non-Gaussian models. This makes Particle Filters an efficient tool for vasculature tracking, where abrupt changes to vessel shape or contrast level exist. Another superiority of Particle filter to Kalman filters is that it can handle bifurcation locations by incorporating a strategy to detect them during the tracking [143].

Usually, the methods estimating the morphological or the topological characteristics of vasculature may benefit from binary vessel maps obtained with a segmentation algorithm [48, 107, 110]. Therefore, the accuracy of estimated vasculature properties may depend on the accuracy of these segmentation algorithms, which may cause disconnected vessel tree or the absence of tiny vessels with low contrast or false vessels such as perceiving the noise emerged from the segmentation of the borders of the Optic Disc or FOV stops as a vessel. On the other hand, direct usage of probability maps without thresholding may bring benefits due to considering the uncertainity of vessel existence in vessel character estimation. In a recent study, Fang *et al.* used vesselness probability maps to detect bifurcation/ crossing locations [109]. They argued that using probability maps in the place of binary vessel maps reduced "unpredictable" influence of noise and pathology in the detection of these landmark locations.

In this thesis, a tracking method based on Particle Filters is guided with probability maps of vessel parts. The advantages of using probability maps can be summarised as follows. This approach can facilitate the better estimation of vessel characteristics by selecting relatively the best hypothesis, where lower probabilities represent pixels belonging to thinner vessels or any vessels suffering from noise such as uneven illumination of background image or those nearby/in bifurcation and crossing regions. Also, this approach may provide a better tackling for identification of vessel characteristics in pathological regions. As a result, more comprehensive vasculature connectivity maps may be provided.

Such an approach may be superior to methods estimating vasculature properties, such as vessel width, directly from intensity images regarding the accuracy and connectivity of vessel tree because unprocessed fundus images exhibit more noise than vesselness probability maps and may be more troublesome for the estimation of vasculature characteristics; though, they are improved with pre-processing to suppress the noise or highlighting vessels in images with related filters [50, 99].

5.2 Tracking as a Dynamic System

Tracking can be defined as the sequential estimation of locations of an object of interest for a limited number of iterations. Tracking starts from a known location and there is a dependency between subsequent locations of movement. When the location of the object is unknown at a particular point, this location can be estimated given previous locations of the object and observations obtained from noisy measurements can be incorporated as they become available.

Because the locations of an object during tracking are not directly measurable but can be estimated given the observations, the variables defining the locations and other related variables contributing to the estimation of these locations such as velocity are called *latent variables*. The estimated values of these variables at each iteration k form a state vector $\mathbf{z}^k$. $\mathbf{z}^{0:k} = \left\{\mathbf{z}^0, \mathbf{z}^1, \mathbf{z}^2, \mathbf{z}^3, ...\mathbf{z}^k\right\}$ that represents the evolution of a dynamic system with respect to state vectors over iterations. The effects of this evolution can be observed from the information obtained from measurement systems. An observation set taken at iteration k may be represented with $\mathbf{y}^k$, which are the observed variables of this dynamic system. Similarly, the sets of observations related to these estimated locations can be represented with $\mathbf{y}^{0:k} = \left\{\mathbf{y}^0, \mathbf{y}^1, \mathbf{y}^2, \mathbf{y}^3, ...\mathbf{y}^k\right\}$, where $\mathbf{y}^k$ denotes a set of observations at iteration k.

Because the state vector of such a dynamic system for the next iteration $k+1$ depends on the state vectors at previous iterations, this dependency between state vectors over iterations can be shown with a Markov chain [8]. The joint probability distribution of these latent variables over a limited number of iterations can be calculated from the product rule given in (5.1) [144].

$$P(\mathbf{z}^{0:k}) = P(\mathbf{z}^k \mid \mathbf{z}^{0:k-1}) \cdot P(\mathbf{z}^{k-1} \mid \mathbf{z}^{0:k-2}) \cdot P(\mathbf{z}^{k-2} \mid \mathbf{z}^{0:k-3}) \cdot ..... \cdot P(\mathbf{z}^0) \tag{5.1}$$

Regarding tracking problems, knowing only the most recent state vector, $\mathbf{z}^k$, can give enough information for the estimation of the next state vector $\mathbf{z}^{k+1}$. In these cases, a first order Markov chain can be used to model this tracking system [144]. A system whose latent variables modelled with a first order Markov chain is depicted in Figure 5.1. The joint probability distribution of state vectors of a tracking system up to iteration k can be

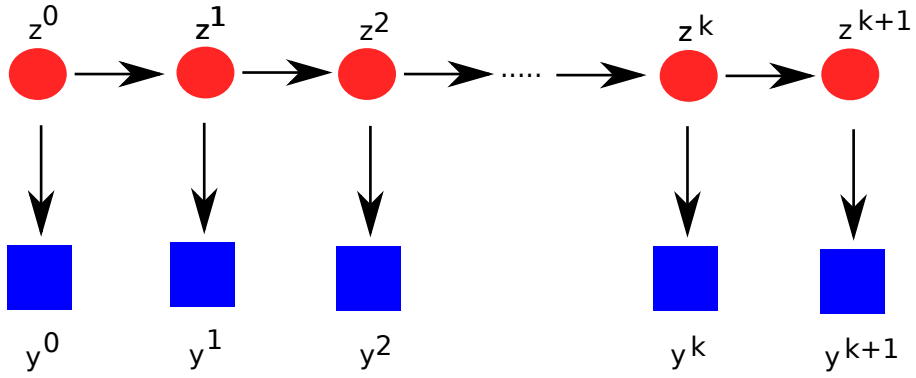


Figure 5.1: A state space model of a first order Markovian system: z^k shows the state vector at iteration k and y^k observations at this iteration.

simply written in (5.2) with a first order Markov chain.

$$P(\mathbf{z}^{0:k}) = P(\mathbf{z}^0) \cdot \prod_{s=1}^{k} P(\mathbf{z}^s \mid \mathbf{z}^{s-1}) \tag{5.2}$$

The evolution of state vectors with corresponding observations of a tracking system can be modelled with a linear state-space model because of having a first-order Markov property. In this case, the evolution of state vectors can be described with (5.3) and the observations affected by these state vectors can be defined with (5.4) [145].

$$\mathbf{z}^{k+1} = f(\mathbf{z}^k, v^k) = \mathbf{z}^k + \epsilon_z^k \tag{5.3}$$

$$\mathbf{y}^{k+1} = g(\mathbf{z}^{k+1}, u^{k+1}) = \mathbf{z}^{k+1} + \epsilon_y^{k+1} \tag{5.4}$$

where $f(\cdot)$ is a state transition function, which models the internal dynamic of the system. The transition of a state vector at k to that at $k+1$ is made through by adding noise ϵ_z^k to the previous state vector at k. This noise can be drawn from a Gaussian distribution [146]. $g(\cdot)$ maps a state vector $\mathbf{z}^{k+1}$ to an observation set $\mathbf{y}^{k+1}$, which is fetched from a noisy measurement system (the tracking system). ϵ_y^{k+1} shows the noise component in the observations. The uncertainty of the tracking system and the noise in the observations are modelled with v and u consecutively.

5.2.1 Bayesian Solution to The State-Space Model

In order to estimate the next state vector of a dynamic system described with (5.3) and (5.4), one can use a probabilistic approach. With this approach, the state transition function $f(\cdot)$ in (5.3) can be modelled with a state transition probability distribution $P(\mathbf{z}^{k+1}|\mathbf{z}^{0:k})$. Similarly, the function $g(\cdot)$ in (5.4), which maps a state vector to an observation set, can be realised with the likelihood $P(\mathbf{y}^{0:k}|\mathbf{z}^{0:k})$.

Because this dynamic system was assumed to have a first order Markov property, the

state transition probability distribution $P(\mathbf{z}^{k+1}|\mathbf{z}^{0:k})$ can be written as $P(\mathbf{z}^{k+1}|\mathbf{z}^{0:k}) = P(\mathbf{z}^{k+1}|\mathbf{z}^{k})$ [144]. Considering the directed graph between latent variables $\mathbf{z}$ and observed variables $\mathbf{y}$ in Figure 5.1, the observation set $\mathbf{y}^k$ at iteration k depends on previous observation sets $\mathbf{y}^{0:k-1}$, which is given in (5.5).

$$P(\mathbf{y}^k|\mathbf{y}^{0:k-1}) \neq P(\mathbf{y}^k|\mathbf{y}^{k-1}) \tag{5.5}$$

However, when the state vector $\mathbf{z}^k$ at iteration k is known, the observation set $\mathbf{y}^k$ becomes independent of other observation sets $\mathbf{y}^{0:k-1}$. The conditional probability of an observation set $\mathbf{y}^k$ given a state vector $\mathbf{z}^k$ to be independent of other observation sets is described in (5.6) and this independence leads to a simpler calculation of likelihood as given in (5.7) [147].

$$P(\mathbf{y}^k) \perp\!\!\!\perp \{P(\mathbf{y}^0), .., P(\mathbf{y}^{k-1})\} \mid P(\mathbf{z}^k) \tag{5.6}$$

$$P(\mathbf{y}^k|\mathbf{z}^{0:k}, \mathbf{y}^{0:k-1}) = P(\mathbf{y}^k|\mathbf{z}^k) \tag{5.7}$$

With respect to a Bayesian approach, one can estimate the next state vector of the tracking system with the posterior probability distribution of this state vector. The posterior probability distribution can be calculated with Bayes's rule given prior probability distribution and likelihood. Because the system is a first order Markovian and an observation set $\mathbf{y}^k$ is independent of previous observation sets $\mathbf{y}^{0:k-1}$ given a state vector $\mathbf{z}^k$, the simplified calculation of likelihood $P(\mathbf{y}^k|\mathbf{z}^k)$ and that of state transition probability distribution $P(\mathbf{z}^{k+1}|\mathbf{z}^k)$ can be used in the calculation of this posterior probability distribution as given in (5.8).

$$P(\mathbf{z}^{k+1}|\mathbf{y}^{0:k+1}) = \frac{P(\mathbf{y}^{k+1}|\mathbf{z}^{k+1})P(\mathbf{z}^{k+1}|\mathbf{y}^{0:k})}{P(\mathbf{y}^{0:k+1}|\mathbf{y}^{0:k})} \tag{5.8}$$

The calculation of the posterior probability distribution in (5.8) involves three terms: prior, likelihood and evidence [148]. $P(\mathbf{z}^{k+1}|\mathbf{y}^{0:k})$ corresponds to the prior term. This probability distribution estimates the next state vector without given next observation sets, which means that the estimates are limited with past experience and prior belief about the system only. This prior can be factorised with Chapman - Kolmogorov equation [149] in (5.9).

$$P(\mathbf{z}^{k+1}|\mathbf{y}^{0:k}) = \int P(\mathbf{z}^{k+1}|\mathbf{z}^k)P(\mathbf{z}^k|\mathbf{y}^{0:k})d\mathbf{z}^k \tag{5.9}$$

where $P(\mathbf{z}^{k+1}|\mathbf{z}^k)$ is the state transition probability at iteration $k+1$ and $P(\mathbf{z}^k|\mathbf{y}^{0:k})$ represents the posterior probability distribution obtained at iteration k (note that, when $k = 0$, $P(\mathbf{z}^0|\mathbf{y}^0) = P(\mathbf{z}^0)$).

$P(\mathbf{y}^{k+1}|\mathbf{z}^{k+1})$ in (5.8) corresponds to the likelihood distribution term given the estimated state vector for iteration $k+1$. The likelihood probability distribution shows how well the predicted state vectors $\mathbf{z}^{k+1}$ with the prior probability distribution can explain the next observation set $\mathbf{y}^{k+1}$.

Finally, $P(\mathbf{y}^{0:k+1}|\mathbf{y}^{0:k})$ in (5.8) represents the evidence term. This shows the evolution of observation sets over iterations. Because the sets of observations are not independent unless they are conditioned on a state vector, the evidence term can not be directly calculated; however, it can be calculated with respect to state vectors of the system by using an integral as given in (5.10).

$$P(\mathbf{y}^{k+1}|\mathbf{y}^{0:k}) = \int P(\mathbf{y}^{k+1}|\mathbf{z}^{k+1})P(\mathbf{z}^{k+1}|\mathbf{y}^{0:k})d\mathbf{z}^{k+1} \tag{5.10}$$

The evolution of the tracking system can be modelled with the recursive updates of (5.9) and (5.8) [149], which are also named prediction and updating steps respectively. In the prediction step, the next state vector z^{k+1} is estimated from the prior probability distribution $P(z^{k+1}|y^{0:k})$, which was the previous estimate of posterior probability distribution, then this estimation is updated by observing the new observation set y^{k+1}.

Many dynamic systems can be modelled with this Bayesian approach: For example, a Kalman filter can analytically calculate the posterior probability distribution $P(\mathbf{z}^{k+1}|\mathbf{y}^{k+1})$, if the system is linear and latent variables z and observed variables y have a Gaussian distribution [149, 150]. However, dynamic systems in the real world are usually non-linear and non-Gaussian. When the system model is non-linear and non-Gaussian, the integrals in (5.9) and (5.10) become intractable, so the state estimation in this type of systems can be performed only with approximation techniques [149].

In order to deal with this type of dynamic systems, linearisation of the system (extended Kalman filter) or representation of posterior probability distribution with a Gaussian mixture distribution (Gaussian sum Kalman filters) were proposed [147,149]. Another method proposed for non-linear and non-Gaussian systems is the point mass filter [147]. This method approximates the posterior probability distribution over a grid that discretizes the state space, while maintaining the original system model. However, when the dimension of the state vector is large, the method suffers from the curse of dimensionality due to having quadratic complexity in the grid size [147]. On the other hand, Particle Filters can approximate the posterior probability distribution with an adaptive stochastic grid [147]. When compared with the point mass filter, the complexity in the number of grid points is linear in Particle Filters [147].

5.3 Particle Filters

In a tracking problem, one wants to estimate the posterior probability distribution of the next state vector $P(\mathbf{z}^{k+1}|\mathbf{y}^{k+1})$. However, the calculation of a posterior probability distribution is often not analytically possible.

In Particle Filters, the probability distributions are not calculated analytically but they are approximated by a set of N particles $\left\{\mathbf{z}^k\right\}_{n=1}^{N} = \left\{\left\{\mathbf{z}^k\right\}^1, \left\{\mathbf{z}^k\right\}^2, \cdots, \left\{\mathbf{z}^k\right\}^N\right\}$ and a set of weights $\left\{W^k\right\}_{n=1}^{N} = \left\{\left\{W^k\right\}^1, \left\{W^k\right\}^2, \cdots, \left\{W^k\right\}^N\right\}$ as given in (5.11). Briefly,

each particle $\left\{\mathbf{z}^k\right\}^n$ in this method corresponds to a possible state vector of the system at a specific iteration k. A particle with a state vector of $\left\{\mathbf{z}^k\right\}^n$ is assigned a weight of $\left\{\mathbf{W}^k\right\}^n$, which shows the probability of this specific state vector in a multivariate distribution $P(\mathbf{z})$.

$$P(\mathbf{z}) \approx \sum_{n=1}^{N} \{\mathbf{W}\}^n \cdot \delta(\mathbf{z} - \{\mathbf{z}\}^n) \tag{5.11}$$

where a continuous probability distribution $P(\mathbf{z})$ is approximated with N particles. $\delta(\mathbf{z} - \{\mathbf{z}\}^n)$ is a dirac delta function, located at the state vector $\{\mathbf{z}\}^n$.

The main power behind Particle Filters to be able to deal with intractable integration is the usage of the sequential Monte Carlo sampling. Monte Carlo sampling is a statistical method for the approximation of a true integral with a discrete sum. For example, the calculation of $\int_{\mathbf{z}} f(\mathbf{z})\, dP(\mathbf{z})$ can be approximated in three steps: (i) by sampling N number of i.i.d samples $\left\{\{\mathbf{z}\}^1, \{\mathbf{z}\}^2, \cdots, \{\mathbf{z}\}^N\right\}$ from the probability distribution $P(\mathbf{z})$ and (ii) by evaluating the function $f(\mathbf{z})$ with these samples and finally (iii) by replacing the integral with sum operation [148]. The expectation $E\left[f(z)\right]$ of the function $f(z)$ under probability distribution $P(z)$ can be approximated by using Monte Carlo sampling with (5.12). The larger N is, the closer $\bar{f}_N$ is to true expectation $E\left[f(z)\right]$ according to the Law of Large Numbers [151]. This means that increasing the number of particles defined for a Particle Filtering application will improve convergence to the true probability distribution.

$$\bar{f}_N = \sum_{n=1}^{N} f(\{\mathbf{z}\}^n)\, \{\mathbf{W}\}^n \tag{5.12}$$

where $\{\mathbf{z}\}^n$ represents the n^{th} sample drawn from $P(z)$ and whose weight is $\{\mathbf{W}\}^n$. $\bar{f}_N$ is the mean of the function values evaluated at these sampled points.

Two main questions have to be addressed when using Monte Carlo sampling: (i) How to sample from this probability distribution? (ii) What to select as proposal distribution?

(i) How to sample from an arbitrary probability distribution? The most popular sampling method for Particle Filtering can be said to be Importance Sampling [149]. In Importance Sampling, one uses a proposal distribution $Q(z)$, which is easier to sample than $P(z)$, in the place of $P(z)$. With respect to $Q(z)$, the expectation of a function under the distribution $P(z)$, $E\left[f(z)\right]$, which was given as an example before, can be written in (5.13) and its approximation $\hat{f}_N$ in (5.14). In this equation, $W(\{\mathbf{z}\}^n)$ represents the importance weight of n^{th} sample and its calculation is explicitly given in (5.15). These weights are normalised in order to make sure $\sum_{n=1}^{N} W(\{\mathbf{z}\}^n) = 1$. Normalised weights $\hat{W}(\{\mathbf{z}\}^n)$ are calculated in (5.16) and these weights are called *normalised importance weights*.

$$\int f(z)P(z)dz = \int f(z)\frac{P(z)}{Q(z)}Q(z)dz \tag{5.13}$$

$$\grave{f}_N = \frac{1}{N}\sum_{n=1}^{N} W(\{\mathbf{z}\}^n)f(\{\mathbf{z}\}^n) \tag{5.14}$$

$$W(\{\mathbf{z}\}^n) = \frac{P(\{\mathbf{z}\}^n)}{Q(\{\mathbf{z}\}^n)} \tag{5.15}$$

$$\hat{W}(\{\mathbf{z}\}^n) = \frac{W(\{\mathbf{z}\}^n)}{\sum_j^N W(\{\mathbf{z}\}^j)} \tag{5.16}$$

$$\bar{f}_N = \frac{\frac{1}{N}\sum_{n=1}^{N} W(\{\mathbf{z}\}^n)f(\{\mathbf{z}\}^n)}{\frac{1}{N}\sum_{j=1}^{N} W(\{\mathbf{z}\}^j)} = \sum_{n}^{N} \hat{W}(\{\mathbf{z}\}^n)f(\{\mathbf{z}\}^n) \tag{5.17}$$

In the equations above $P(z) = P(\mathbf{z}^{k+1}|\mathbf{y}^{0:k+1})$ and $Q(z) = Q(\mathbf{z}^{k+1}|\mathbf{y}^{0:k+1})$. The proposal distribution can be chosen such that $Q(\mathbf{z}^{k+1}|\mathbf{y}^{0:k+1}) = Q(\mathbf{z}^0|\mathbf{y}^0)\prod_{n=1}^{N} Q(\mathbf{z}^k|\mathbf{z}^{k-1}, \mathbf{y}^{0:k})$. Therefore, the weights for the next iteration can be calculated recursively from the previous weights. The calculation of next weights in (5.15) can be related to weights at previous iteration with (5.18) [152].

$$W\left(\left\{\mathbf{z}^{k+1}\right\}^n\right) \approx W\left(\left\{\mathbf{z}^{k}\right\}^n\right)\frac{P\left(y^{k+1}|\left\{\mathbf{z}^{k+1}\right\}^n\right)P\left(\left\{\mathbf{z}^{k+1}\right\}^n|\left\{\mathbf{z}^{k}\right\}^n\right)}{Q\left(\left\{\mathbf{z}^{k+1}\right\}^n|\left\{\mathbf{z}^{k}\right\}^n, y^k\right)} \tag{5.18}$$

The main disadvantage of importance sampling is known as particle depletion. Particle depletion means that only very limited number of particles dominates the approximation of probability distribution because of their weights to get larger as the iteration number increases. This causes computational power to be wasted because of the denser calculation of less likely particles. This situation usually occurs when the proposal distribution $Q(z)$ does not match true probability distribution $P(z)$, especially in high dimensions [148]. In order to avoid the poor representation of true distribution due to particle depletion, the resampling of particles was introduced when it is necessary, or at every update step. When importance sampling is followed by resampling it is called *Sampling-Importance Resampling (SIR)* [153]. SIR method aims to remove the particles with negligible weights while giving more representation chance to particles with large weights. Therefore, the more likely particles can be employed during the posterior estimation.

Apart from Importance Sampling, a new particle set with N particles is drawn from the generated particle set with replacement in SIR after completing the aforementioned steps of importance sampling. The probability of the reselection of a particle depends on its normalised importance weight. After resampling, the weights of particles are made equal which means that $W(\{\mathbf{z}\}^n) = \frac{1}{N}$ for each elements of $\{\{\mathbf{z}\}^1, \{\mathbf{z}\}^2, ..., \{\mathbf{z}\}^N\}$.

(ii) What to select as proposal distribution? In SIR sampling, the state transition probability distribution $P(z^{k+1}|z^k)$ is selected as the proposal distribution $Q(z^{k+1}) = P(z^{k+1}|z^k)$ [154]. The weights of particles over iterations can be calculated with (5.19). Because the weights get equal after resampling, (5.19) simplifies to (5.20) [154].

Because this approach does not consider the latest observations for the calculation of weights, it can reduce the performance in the case of prior probability distribution does not match with the likelihood and make it sensitive to outliers [154, 155]. Considering the smoothness of the change of vessel curvature and vessel radius, these down sides may not be big problems for the localisation of retinal vessels with particle filters [155]. The major advantage of SIR filter is to provide a quick calculation of importance weights, which may be an important factor for clinical applications.

$$W(\left\{\mathbf{z}^{k+1}\right\}^n) \propto W(\left\{\mathbf{z}^{k}\right\}^n) P(y^{k+1}|\left\{\mathbf{z}^{k+1}\right\}^n) \tag{5.19}$$

$$\propto P(y^{k+1}|\left\{\mathbf{z}^{k+1}\right\}^n) \tag{5.20}$$

How to implement SIR sampling in Particle filter when the proposal distribution is the state transition probability distribution $P(z^{k+1}|z^k)$ is given with Algorithm 1. Also, Figure 5.2 visually demonstrates the working mechanism of a SIR particle filter. It can be deduced from this figure that the solitary usage of Importance Sampling can cause particles to get concentrated on most likely regions of state vector over iterations by leading to particle depletion because only a couple of particles become responsible for the representation of probability distributions. As seen from this figure, the introduction of resampling after importance sampling can prevent particle filtering from particle depletion. Even though some particles share the same state vectors after resampling, the variety of state vectors can be increased via the *Motion Model* (state transition function) again for the next tracking step.

5.4 Proposed Tracking System For Quantitative Analysis of Retinal Vasculature

In this research, $2D$ retinal vasculature is simplified to be modelled as a network of single vessels connected at landmark locations, which describes connection locations between vessels regardless of their types such as bifurcating or crossing points.

Tracking an individual vessel is a rather simple task, when compared with the tracking of a vessel tree fraction. The former one starts with the tracking of a vessel given an initial location and continues until this vessel disappears or gets connected with another one in a landmark location. On the other hand, the latter one also includes the estimation of existence of other vessels to this landmark location and collecting information about their initial locations.

Algorithm 1 Implementation of SIR sampling for Particle filter when the proposal distribution is state transition probability distribution $P(z^{k+1}|z^k)$ and resampling is realised after each importance sampling step.

1: **function** $\bar{\mathbf{z}}^{k+1}$=SIR SAMPLING($P(z^0)$, $P(z^{k+1}|z^k)$,N,y)
2: $\left\{\mathbf{z}^0\right\}_1^N \backsim P(z^0)$ ▷ Draw N samples from distribution $P(z^0)$.
3: $\left\{W^0\right\}^n \leftarrow \frac{1}{N}$ ▷ Assign a weight to each particle n, where $n = 1 : N$
4: **for** k = 0 to K **do** ▷ K is the maximum iteration number for tracking
5: $\left\{\dot{z}^{k+1}\right\}_1^N \backsim P\left(z^{k+1}|\left\{z^k\right\}^n\right)$ ▷ Draw N samples from the proposal distribution
6: $\left\{W^{k+1}\right\}^n \leftarrow P\left(y^{k+1}|\left\{\dot{z}^{k+1}\right\}^n\right)$ ▷ $P\left(y^{k+1}|\left\{\dot{z}^{k+1}\right\}^n\right)$ shows likelihood calculated from an observation model (see (5.27), (5.28), (5.29) and (5.30) for the proposed observation models).
7: Compute $\left\{\hat{W}^{k+1}\right\}^n$ with (5.16)
8: $\left\{\mathbf{z}^{k+1}\right\}_1^N \backsim \left\{\dot{z}^{k+1}\right\}_1^N$ ▷ Resample N particles with regard to $\left\{\hat{W}^{k+1}\right\}_1^N$
9: $\left\{W^{k+1}\right\}^n \leftarrow \frac{1}{N}$
10: Estimate $\bar{\mathbf{z}}^{k+1}$ with (5.21) ▷ $\bar{\mathbf{z}}^{k+1}$ represents the expectation of posterior probability distribution, approximated with the particle set $\left\{\mathbf{z}^{k+1}\right\}_1^N$

$$\bar{\mathbf{z}}^{k+1} = \sum_{n=1}^{N} \left\{\hat{W}^{k+1}\right\}^n \left\{\mathbf{z}^{k+1}\right\}^n = \frac{1}{N} \sum_{n=1}^{N} \left\{\mathbf{z}^{k+1}\right\}^n \tag{5.21}$$

11: **end for**
12: **end function**

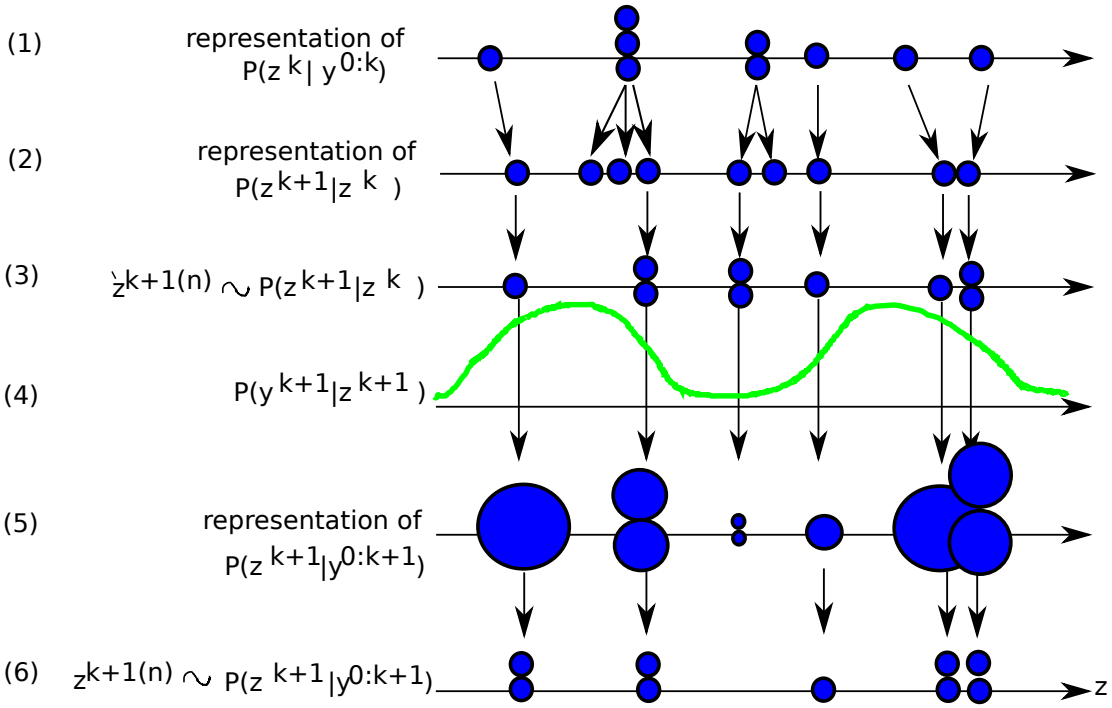
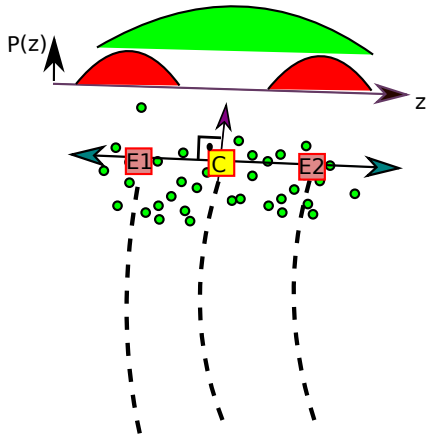
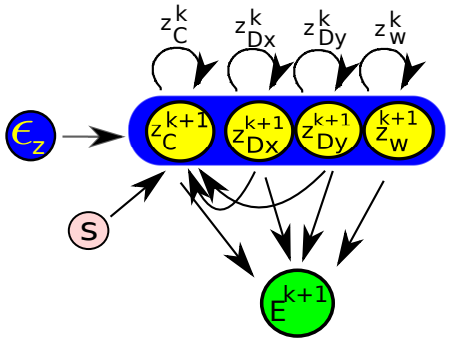


Figure 5.2: The sequential estimation of the posterior probability distribution of state vectors by using a Particle Filter, which samples from SIR and uses State Transition Probability Distribution as the proposal distribution. In the diagram, the weights of particles are visualised with their sizes. The particles with the same state vector (identical particles) are located top of another (adapted from [8]).

In order to deal with this complicated situation of the tracking of vessel tree fractions, vasculature tracking methods were usually devised with an additional branch detection technique, which checks the situation of vessel branching (exist or not) [99] or they investigate the probability of a vessel fitness in various cases such as normal, bifurcating or crossing [74, 104]. In this study, a multiple mode (working mode) Particle Filtering is proposed to tackle this issue. A multiple mode tracking can switch from one mode to another one when it faces a different situation during the tracking so it allows more flexible and simplified modelling. Some examples of multiple mode tracking in object tracking can be found in [156–158]. In this research, two modes of tracking are used, which are denoted *mode* 1 and *mode* 2. The first mode is regular vessel tracking, while the second mode contains functions for detecting landmark locations and for collecting information about the initialisation of future tracking of the last detected vessels. After giving initial conditions at the beginning of the tracking, the proposed tracking system works without user intervention. In the first mode, the tracker traces a single vessel. If the tracker has a strong clue about that the next tracking point is very close to a landmark location, the mode of tracking is switched to *mode 2*. In this mode, a possible landmark location is detected, if it exists. Then, all vessels connected to this location are identified. Possible initial states necessary for the tracking of these vessels are estimated and these initialisations are stored in the memory of the tracker. After completing these tasks, the mode of tracking is switched back to *mode 1*. This tracking continues until two conditions are



(a) *Particles represent the posterior probability distribution of the state of the tracking system given the motion model and observation model, which is a combination of vessel edge probability and vesselness probability maps.*



(b) *The proposed vessel tracking system is defined with the concurrent estimation of vessel centerline C^{k+1} and those of edge locations E_1^{k+1} and E_2^{k+1} depending on the estimates of states of the system $\mathbf{z}^{k+1} = \left[z_{Dx}^{k+1}, z_{Dy}^{k+1}, z_w^{k+1}\right]$; (yellow discs represent stochastic variables, the green disc a deterministic variable and blue disc noise and pink disc constant.)*

Figure 5.3: The proposed vessel tracking system; an example of (a) observation model and (b) motion model

satisfied: (i) there is no future tracking in the memory and (ii) the vessel on the current tracking disappears. In this work, the decison to continue the tracking is determined by checking the centreline likelihood of the estimated centreline location. If the likelihood is below a threshold, tracking stops, which is similar to [155].

Apart from the multi-modal structure of the proposed tracking system, another important contribution to the literature is the usage of outputs of a deep network as observations obtained from multiple measurement systems for retinal vessel tracking. These measurement systems are namely a vessel detector, a vessel centreline detector, a vessel boundary detector and a landmark location detector. How to combine these measurements from these systems during tracking will be explained in the following sections.

5.4.1 *Mode 1*: Tracking Vessel Segments

In the literature, the detection of vessel segments based on probabilistic tracking methods has been usually performed either by tracking only vessel boundaries [74, 104] or by estimating centrelines and width together [99]. In this study, I conform to the latter approach because edge information can be accessible through centreline and widths. Also, as will be explained later, this approach may facilitate a better modelling of vessel cross-section.

In this section, initially, general models for motion and observation models will be explained in order to give a general framework for this mode. Then, four models, which can be used for *mode 1*, will be introduced in Section 5.4.1. Figure 5.3a demonstrates this default mode (*mode 1*). As seen in this figure, the variations on vessel centreline locations, vessel widths and vessel orientations/directions are reflected to hypotheses (particles) during tracking. Also, the estimations of vessel centrelines and boundaries are made in a direction perpendicular to the direction of the tracker.

Motion Model

A motion model (it is also called prior, which was explained in Section 5.2.1) manages the internal dynamics of a tracking system; in other words, how to update system states. A motion model for the tracking system is demonstrated in Figure 5.3b.

As understood from this figure, the next estimates of centreline locations depend on next estimates of direction vector, those of vessel width and the previous estimates of centreline location. However, edge locations are deterministically calculated from next estimates of centreline location, those of widths and those of direction vector. The variables mentioned above constitute the state vector of the tracker, which can be written as $\mathbf{z}^k = \left[z_{Cx}^k, z_{Cy}^k, z_{Dx}^k, z_{Dy}^k, z_w^k\right]$, where $z_C^k = \left(z_{Cx}^k, z_{Cy}^k\right)$ represents an estimated centreline location. Similarly, $z_D^k = \left(z_{Dx}^k, z_{Dy}^k\right)$ denotes an estimated direction of the tracker and z_w^k an estimate of the vessel width. The motion model demonstrated in Figure 5.3b can be formulated with (5.22),(5.23) and (5.24).

$$z_D^{k+1} = z_D^k + \epsilon_{z_D} \tag{5.22}$$

$$z_w^{k+1} = z_w^k + \epsilon_{z_w} \tag{5.23}$$

$$z_C^{k+1} = z_C^k + z_D^{k+1} \cdot s + \epsilon_{z_C} \tag{5.24}$$

where ϵ_z represents a normal distributed noise variable (ϵ_{z_D} for direction vector, ϵ_{z_w} for width and ϵ_{z_C} for centreline location). s is a constant and denotes step size.

One of the aims of the tracker is to estimate vessel edges. Edge locations symmetrically located at each side of centreline locations can be calculated from (5.25) and (5.26).

$$E_1^{k+1} = C^{k+1} + \frac{1}{2} z_w^{k+1} \cdot (z_D^{k+1})_\perp \tag{5.25}$$

$$E_2^{k+1} = C^{k+1} - \frac{1}{2} z_w^{k+1} \cdot (z_D^{k+1})_\perp \tag{5.26}$$

where C^{k+1} shows an estimate of centreline location at iteration $k+1$. E_1^{k+1} and E_2^{k+1} are estimates of edge locations at iteration $k+1$. z_D^{k+1} is an estimate of the direction vector at iteration $k+1$ while $(z_D^{k+1})_\perp$ is a perpendicular direction vector to the estimated direction vector z_D^{k+1}. z_w^{k+1} is an estimate of the vessel width at iteration $k+1$.

The noise in the motion model is intuitively designed as a normal distributed random variable by relating it with the anatomy of vessels. As opposed to tracking a moving object or pedestrian in an image sequence [159], the direction of tracking along vessels does not change quickly but smoothly for a small step size. Similarly, the change on vessel width along a small step size of tracking is minute. A normal distributed variable can model the smooth variation effectively by giving larger chance to small changes but less chance to larger variations. This may be very effective way of estimating future state values, when the information received from measurements is very noisy.

Tracking starts from an initially known state vector $\mathbf{z}^0$, whose values can be assigned

manually or automatically. In this study, only, the user selects an initial location for a possible vessel centreline and enters an estimate of vessel width at this location. Then, an initial estimate of direction vector can be calculated from eigenvector analysis [17]. Apart from similar approaches, eigenvectors are calculated from vesselness probability maps instead of from gray fundus images. This was observed to provide a more accurate direction for initial estimation because these maps are much less noisy than their corresponding fundus images.

Observation Model

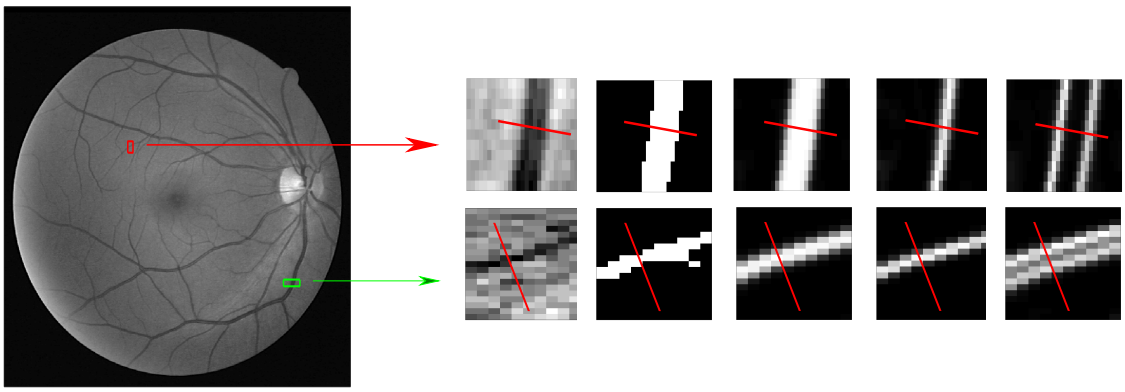
An observation model describes a noisy measurement system where the fitness of a hypothesis (presented with a particle) generated by a tracking system to actual measurements can be evaluated. A general formula for an observation model was given with (5.4).

A measurement system in retinal image analysis usually corresponds to vesselness measure, which can be obtained by fitting a model to vessel intensity cross-section profiles [74,97,99,104]. Tracking methods relying on this modelling usually accept that vessel boundaries are symmetric and background has a constant intensity [74,97,104]. However, mere information coming from vesselness may be not enough for tracking to evaluate the fitness of a hypothesis when a vessel cross-section profile does not resemble a Gaussian function; which may result in early stopping [74,97]. For example, uneven illumination or pathology may reduce the visibility of one of the vessel boundaries in fundus images [74] or both edges of a vessel may disappear in crossing regions [97]. On the other hand, using additional sources of information may empower the tracker to accurately estimate vessel properties [74], when any of them is not trustable at a specific time.

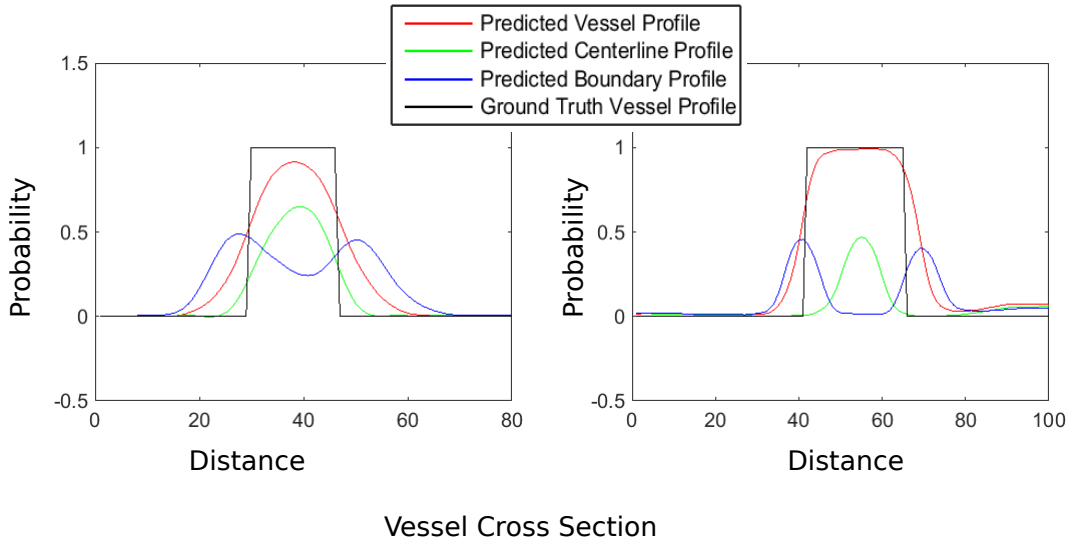
Apart from previous studies [74, 97, 99, 104], in this study, vesselness measure is not generated by model fitting but taken from probability maps generated by a deep network trained for segmentation (a multi-labelling TDBN which was introduced in Section 4.7 in Chapter 4). Moreover, in addition to vesselness, vessel edge and centreline measures are considered in the observation model aside from other studies. These measures are taken from probability profiles between estimated edge locations on related probability maps generated by the multi-labelling TDBN. Readers can refer to Figure 5.4 for examples of vesselness, centreline and edge profiles for thicker and thinner vessels. Vessel cross-section profiles shown in Figure 5.4b are obtained along the red lines located over vesselness, vessel centreline and boundary probability maps in Figure 5.4a.

When Figure 5.4b is examined closely, it is obvious that the maximum values of vesselness probabilities of these cross-sections are very close to 1 despite the contrast difference between these vessel segments in the original fundus image. Also, peaks at centreline profiles and those at boundary profiles almost overlap with vessel center and boundary locations in ground truth data consecutively. Despite the thickness difference between these vessels, the profiles of both vessels have very similar characteristics.

Due to using multiple sources of information for vessel description in this thesis, these



(a) *On the left side, an image from DRIVE dataset is given. On the right side, a raw image patch and manually labelled image mask are followed by generated vessel interior pixel, centerline and boundary labels for large and fine blood vessels (top images belong to a fine vessel and bottom images to a large vessel).*



(b) *The left vessel profiles belong to the thinner vessel and the right one to the larger one.*

Figure 5.4: The difference between the responses of the proposed network to thinner and thicker vessels.

sources can be combined in various configurations for likelihood calculation. Even though different sources are combined, main aims of the configurations stay the same. These aims are the maximisation of both the probability of the tracker to accurately estimate vessel edges and the probability of a tracker being inside vessel. Although edge locations can be estimated from only edge probability profiles by maximising edge probabilities, there is a high possibility that the tracker may change its way from inside to outside vessel if there is no information available to the tracker to discriminate vessel probabilities from background probabilities.

The maximisation of these two probabilities can be obtained by combining different characteristics of probability profiles. For example, edge probability profiles present two peaks at edge locations while very low probabilities at centreline locations. However, the complements of edge probability profiles also have high probabilities at centreline locations and can be used to check if the tracker inside vessel or not. Another example is that centreline locations can be obtained either from vesselness or centreline probability profiles because both have the highest probabilities around vessel centres. Augmenting these characteristics, four models for *mode 1* will be introduced in the following section. These models are generated by pairing different observation and motion models.

The Proposed Models for *Mode 1*

In this section, four observation models and an additional motion model will be introduced. These models are somehow related to each other in terms of more sophisticated ones being built on more primitive models and the models will be explained in this order.

***Model 1*:** This model assumes vessel boundaries around vessel centre to be symmetric, similar to previous methods [74, 97, 99, 104] so, it uses a general type of motion model described with (5.22),(5.23) and (5.24). The observation model uses vesselness and edge probability profiles for the calculation of likelihood. It is expected that while the maximisation of vesselness probability forces the tracker to stay inside vessel, the maximisation of edge probabilities pushes the tracker to probe maximum probabilities at edge locations.

Considering the symmetric boundary assumption, the maximum probability along a vesselness profile is mostly expected to be at the centreline location $z = C = \frac{E_1+E_2}{2}$. However, this profile may not have the maximum in vessel centre due to noise, particularly central light reflex, inherited from images. In order to mitigate the effect of noise in likelihood calculation, sampling from multiple locations in vesselness profile can be performed. As an example, mid locations at either side of estimates of vessel centre, $z = \frac{E_1+3E_2}{4}$ and $z = \frac{3E_1+E_2}{4}$, can be considered. Also, the complements of edge probabilities at these locations can strengthen the information about the tracker to be inside vessel.

Similarly, an edge probability profile is expected to have the maxima at both edge locations $z = E_1$ and $z = E_2$ while any probabilities close to centreline location, $z = \frac{E_1+3E_2}{4}$ and $z = \frac{3E_1+E_2}{4}$, or at this location $z = C = \frac{E_1+E_2}{2}$ are expected to be much lower values, virtually zero.

The likelihood for this model can be calculated from 5.27 by multiplying sampled probabilities mentioned above.

$$\begin{aligned} P^1(y^{k+1}|z^{k+1}) =& P_e(z = \grave{E}_1^{k+1}) P_e(z = \grave{E}_2^{k+1}) P_v\left(z = \frac{(3\grave{E}_1^{k+1} + \grave{E}_2^{k+1})}{4}\right) \cdot \\ & P_v\left(z = \frac{(\grave{E}_1^{k+1} + \grave{E}_2^{k+1})}{2}\right) P_v\left(z = \frac{(\grave{E}_1^{k+1} + 3\grave{E}_2^{k+1})}{4}\right) \\ & \left(1 - P_e\left(z = \frac{(3\grave{E}_1^{k+1} + \grave{E}_2^{k+1})}{4}\right)\right) \\ & \left(1 - P_e\left(z = \frac{(\grave{E}_1^{k+1} + \grave{E}_2^{k+1})}{2}\right)\right) \\ & \left(1 - P_e\left(z = \frac{(\grave{E}_1^{k+1} + 3\grave{E}_2^{k+1})}{4}\right)\right) \end{aligned} \tag{5.27}$$

where $P^1(y^{k+1}|z^{k+1})$ shows the likelihood for *Model 1*. $\grave{E}_1^{k+1}$ and $\grave{E}_2^{k+1}$ correspond to the first and the second edge locations estimated from the state transition probability distribution with the particle filter. P_e and P_v are profile functions respectively obtained from edge, vesselness probability maps.

Figure 5.5 demonstrates the calculation of likelihood for two hypothesises (a strong and a weak one) for a vessel cross-section. Estimated edges are marked with stars. The estimates of edge locations in Figure 5.5a overlap with actual ones but not in Figure 5.5b. Figure 5.5c and Figure 5.5d show sampled probabilities according to (5.27), for the strong hypothesis and the weak one respectively. As seen from Figure 5.5d, lower probabilities obtained for the weak hypothesis scales down the over all likelihood; though, some probabilities are reasonably closer to those in the strong hypothesis. Likelihoods were calculated 0.070 for the strong hypothesis and 0.00045 for the weak hypothesis. In other words, the former likelihood was 1555 times the latter one, which means that the contribution of the strong hypothesis to the final decision (the expectation of posterior probability distribution) became much bigger than that of weak hypothesis.

***Model 2*:** *Model 1* accepts the symmetry of vessel boundaries; however, this assumption may prevent the tracker from estimating better edges when the maximum in a vesselness probability profile is not located in vessel centre due to imaging artefacts, pathologies or noise. In order to deal with this issue, a motion model is introduced to consider asymmetry of boundaries by using two state variables, instead of one, for width estimation. With this motion model, both width variables make estimates for a distance between estimated vessel centre and one of edges. Each variable is independent of the other and responsible for only the same side of vessel over a course of tracking. With this modification, the state vector changes to $\mathbf{z}^k = \left[z_{Cx}^k, z_{Cy}^k, z_{Dx}^k, z_{Dy}^k, z_{wr}^k, z_{wl}^k\right]$, where z_{wr}^k denotes the right side of vessel

and z_{wl}^k left side of vessel. Although the asymmetry of vessel boundaries was considered in some recent studies using model fitting for vessel width estimation [17, 50], there was no example, as far as I am aware, implemented for probabilistic tracking.

The likelihood for this model is calculated in the same way as given for *Model 1* from (5.27) with the contribution of another probability, which is the centreline probability at estimated centreline location. With this additional probability, it is expected that estimated centre locations can be moved closer to the maxima at centreline probability profiles.

$$P^2(y^{k+1}|z^{k+1}) = P^1(y^{k+1}|z^{k+1})P_v(z = \grave{C}) \tag{5.28}$$

where $P^2(y^{k+1}|z^{k+1})$ denotes the likelihood for *Model 2*. $P^1(y^{k+1}|z^{k+1})$ corresponds to likelihood calculated with (5.27). $P_v(z = \grave{C})$ shows vesselness probability at estimated centreline location $\grave{C}$.

***Model 3*:** Similar to *Model 2*, *Model 3* also assumes the asymmetry of vessel boundaries so uses the same motion model. On the other hand, *Model 3* replaces vesselness probability profiles in likelihood calculation in (5.27) with centreline probability profiles. The latter profiles usually show lower standard deviations than those of vessel probability profiles as shown in Figure 5.4 and the spread of centreline probability profile, almost, does not depend on vessel width. Also, it was observed that centreline probability profiles did not posses noise, especially central light reflex, as much as vessel probability profiles did. These characteristics of centreline probability profiles can make them better options than vesselness probability profiles in the localisation of centreline. Therefore, P_v in (5.27) is replaced with P_c, where P_c denotes the centreline location probability profiles.

$$P^3(y^{k+1}|z^{k+1}) = P^{1*}(y^{k+1}|z^{k+1})P_c(z = \grave{C}) \tag{5.29}$$

where $P^3(y^{k+1}|z^{k+1})$ denotes the likelihood for *Model 3*. $P^{1*}(y^{k+1}|z^{k+1})$ is the modified version of the likelihood calculated with (5.27), where vesselness probability profiles P_v are replaced with centreline location probability profiles P_c. $P_c(z = \grave{C})$ shows centreline probability at estimated centreline location $\grave{C}$.

***Model 4*:** This model shares the same motion model with *Model 2* and *Model 3*. Its observation model is also similar to *Model 3* but *Model 4* includes extra edge information coming from complements of centreline probability profiles. In addition, this model incorporates the similarity of estimates of direction vectors to the eigenvectors (one with lower eigenvalue [75]) at estimated centreline locations. These eigenvectors are calculated from vesselness probability maps because they are far less noisier than fundus images.

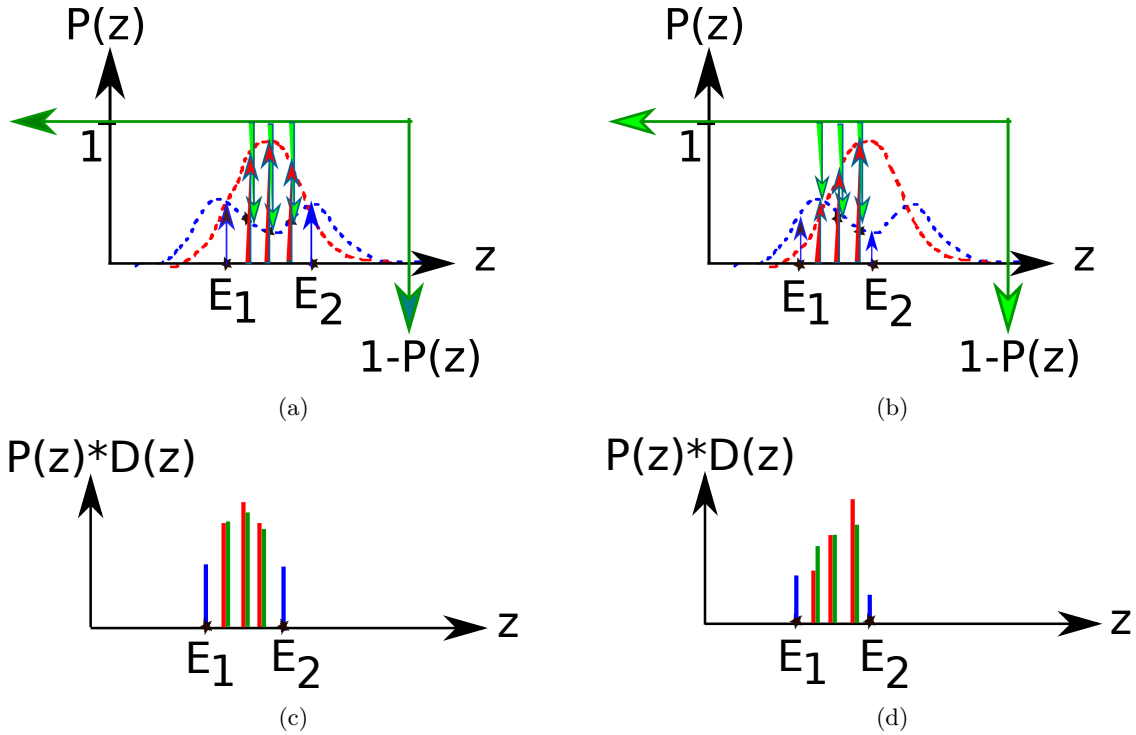


Figure 5.5: The calculation of likelihood for *Model 1* for (a),(c) a strong hypothesis (b),(d) a weak hypothesis, where E_1 and E_2 denote estimated edge locations. In (a) and (b), vesselness probability profile is shown with red dashed line and samples from this profile with red arrows. Similarly, edge profile with blue dashed line and blue arrows and samples from the complement of edge profile with green arrows. (c) and (d) shows samples from these distributions.

$$P^4(y^{k+1}|z^{k+1}) = P^{1*}(y^{k+1}|z^{k+1})P_c(z=\grave{C})(1-P_c(z=\grave{E}_1^{k+1})) \\ (1-P_c(z=\grave{E}_2^{k+1}))P_s \tag{5.30}$$

where $P^4(y^{k+1}|z^{k+1})$ denotes the likelihood for *Model 4*. $P_c(z=\grave{E}_1^{k+1})$ and $P_c(z=\grave{E}_2^{k+1})$ shows centreline probability at estimated edge locations $\grave{E}_1^{k+1}$ and $\grave{E}_2^{k+1}$ respectively. $P_s = |z_D^{k+1} \cdot \overrightarrow{E}(z=\grave{C})|$ denotes the similarity of estimates of vessel direction z_D^{k+1} to eigenvector $\overrightarrow{E}(z=\grave{C})$ at estimates of centreline locations.

Final Estimation Of Centreline and Boundary Locations

After evaluating the likelihood of each particle according to the observations, one needs to return the final decision about vessel centrelines and edge locations or vessel widths as a result of a tracking step. For this aim, one can calculate the expectation of the posterior probability distribution and use it for the final decision [146]. It was explained that the expectation of the distribution in discrete space can be calculated from the weighted average of estimates of particles [148]. Therefore, the expectation for centreline and edge locations can be calculated from (5.31) and (5.32), similar to (5.21).

$$\bar{C}^{k+1} = \frac{1}{N}\sum_{n}^{N}\left\{C^{k+1}\right\}^{n} \tag{5.31}$$

$$\bar{E}_{1,2}^{k+1} = \frac{1}{N}\sum_{n}^{N}\left\{E_{1,2}^{k+1}\right\}^{n} \tag{5.32}$$

5.4.2 *Mode 2*: Detecting Landmark Locations and Collecting Information for the Initialisation of Future Tracking of Connected Vessels

Particle filters theoretically can deal with multi-mode probability distributions emerged due to the tracking of multiple branches. However, the concurrent tracking of multiple branches may cause the deprivation of particles representing the less dominant mode in this multi-mode distribution because resampling usually gives more chance to the samples representing the dominant mode during tracking steps [155]. As a result of that, many researchers have explicitly detected new branches by clustering particles with clustering algorithms such as K-means clustering, a spectral clustering algorithm or mean-shift clustering [105, 155, 160–163]. However, the efficiency of branch detection with this clustering algorithms may be limited with how well particles represent the vasculature search space [155].

In this study, I introduced a method to identify daughter vessels, which provides the separation of vessel tracking from the detection of landmark locations (bifurcation/crossing locations). This method contains two steps of estimation. The first step detects a landmark location given a landmark location probability map, which is generated by the multi-labelling TDBN. The second step identifies possible vessel segments connected to this estimated landmark location with the mean-shift clustering algorithm.

1- Landmark Location Detection

There have been several methods specifically proposed for the detection of bifurcation/crossing locations in retinal vasculature [41, 48, 107–110]. Some of them relied on the connectivity of pixels in vasculature centreline. Apart from these methods, Fang *et al.* [109] used segmentation probability maps generated by a deep network in the place of vasculature skeleton images. Azzopardi and Petkov [41,108] trained a filter (COSFIRE) to detect only bifurcation locations on fundus photos.

In this work, we propose to detect landmark locations from landmark probability maps generated by the multi-labelling TDBN. Figure 5.6a demonstrates an example probability map. As seen from this figure, the network responds very strongly to landmark locations but also moderately responds to the main vessel centrelines as by-product. However, the responses of the network to landmark location or to vasculature can be differentiable. For example, centreline responses seem steeper hills, while responses to landmark locations

look like a sum of Gaussian-like functions. Because of these reasons, the detection of landmark location is related to the change of landmark location probabilities in vessel centreline. To be more precise, if the landmark location probability exceeds a threshold T_L at the centreline of a vessel during tracking, the mode of the tracking is switched to *mode* 2.

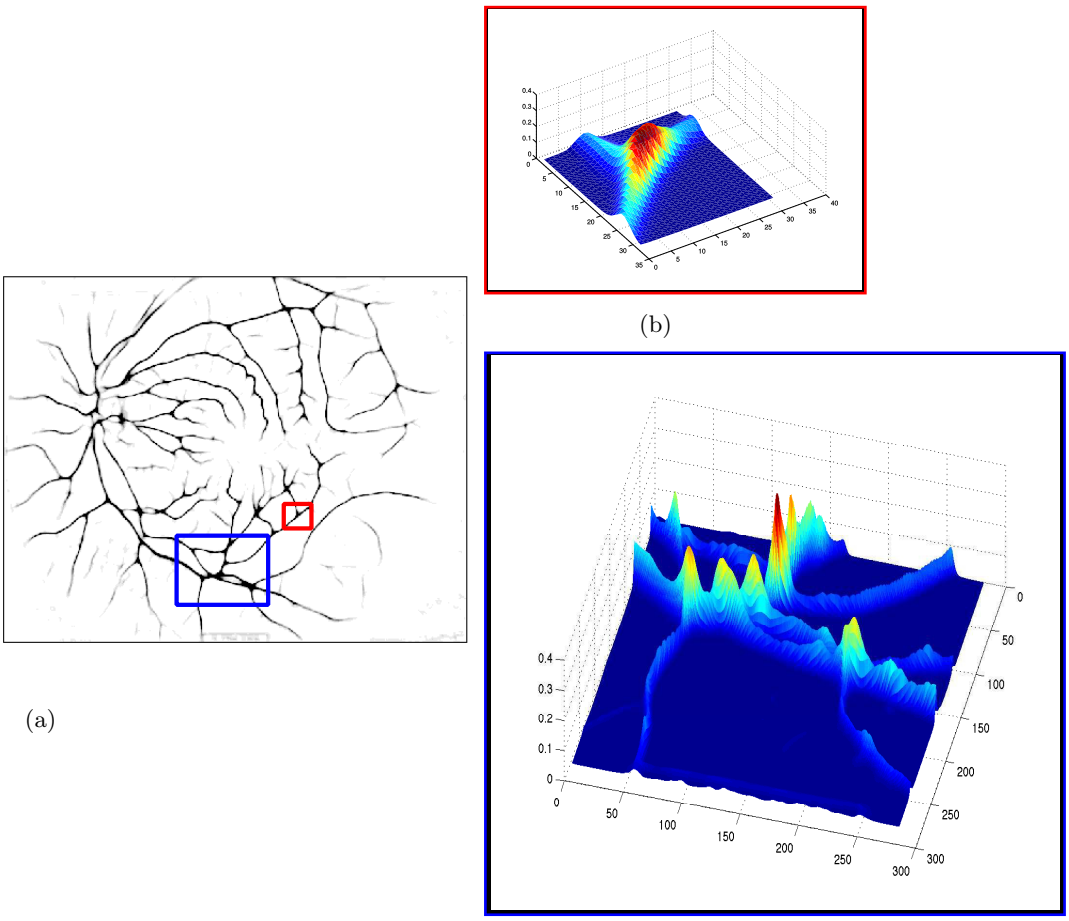
Although landmark location probability distributions belonging to a fundus image have distinctive characteristics for landmark locations, it is not easy to find a global threshold to determine these locations, because the heights of peaks at these locations vary as illustrated in Figure 5.6c. As a result, mode 2 is mostly about finding the peak location of a possible landmark location probability distribution. Figure 5.6b shows a single example of a landmark location probability distribution for a closer inspection.

The proposed landmark detection method is simply a peak detection algorithm using Particle Filters, which is inspired from gradient ascent optimisation method. This method uses landmark location probability maps as observations. The landmark location probabilities at estimates of centreline locations are directly used as weights of particles. This means that the estimates of centreline locations of particles are drawn to the peak locations of this probability distribution during tracking. After reaching a peak location, particles will not want to move beyond this point because likelihoods get smaller beyond the peak point; however some displacement can be observed due to the motion model and because returning of the particles to already traced path is prohibited during the tracking. Eventually, particles become stuck in a small region with tiny step sizes. When step sizes get reduced beyond a certain value, this tracking can be stopped. The final decision of landmark location can be made by evaluating the landmark location probabilities of estimated centrelines during this tracking. The centreline location with the highest landmark location probability is selected as *landmark location*. Figure 6.3a shows estimated centreline locations during the tracking when landmark location probability map used for the observation model. Although this method may not always return a very good estimate of the exact peak location of the related probability distribution, because of using step-wise estimation, the estimated location is usually closer to this peak location. A complete algorithm for the estimation of landmark locations is given in Algorithm 5 in Appendix.

After estimating a landmark location, one can estimate the location of branches passing from this location and connectivity type of the landmark (bifurcation or crossing). The discrimination of crossing locations from bifurcation locations is rather a difficult task due to the existence of various patterns [48, 110]. However, after locating branching/crossing points, one can estimate the type of connection from the number of vessels passing from the location.

2- Daughter Vessel Identification

The *daughter vessel identification* stage is about detecting the locations of modes in a multimodal probability distribution of vessel centreline locations around the estimated



(a)

(b)

(c)

Figure 5.6: (a) A bifurcation map belongs to the first image in VDIS dataset. (The contrast of images is inverted for better visualisation.) (b)-(c) $3D$ visualisation of some selected regions from (a), where bifurcating/crossing locations appear peaks. (Please, find matching colours of frames in (a) and (b)-(c) for region identification.)

landmark location. It is expected to have multiple modes in the centreline probability distribution because of the expectation of multiple vessels connected to the landmark location.

Lesage *et al.* used mean shift clustering for bifurcation detection and the detection of daughter vessels. However, they reported missing some small vessels connected to main aorta with larger angles because their state transition probability distribution did not cover this type of large changes [155]. In order to tackle this type of problems emerging due to unknown direction of a possible daughter vessel according to the main vessel being on the current tracking, the best practice to estimate possible locations of daughter vessels is to generate many particles with all possible tracking directions.

In this study, 7 different direction vectors are used. Among them, only one is the same as the direction of previous tracking before switched to *mode 2*. The other 6 directions are generated by consecutively rotating the initial direction vector as much as $\frac{\pi}{4}$ both clockwise and counter clockwise until a new direction vector aligns with the opposite direction of the initial direction. Generated direction vectors are given as follows:

$$D_{x,y}, rot\left(D_{x,y}, \frac{-\pi}{4}\right), rot\left(D_{x,y}, -\frac{2\pi}{4}\right), rot\left(D_{x,y}, \frac{-3\pi}{4}\right), rot\left(D_{x,y}, \frac{\pi}{4}\right)$$
$$, rot\left(D_{x,y}, \frac{2\pi}{4}\right), rot\left(D_{x,y}, \frac{3\pi}{4}\right)$$

where $D_{x,y}$ is the initial direction obtained in the previous tracking and $rot(D_{x,y}, \theta)$ means that rotate the direction vector $D_{x,y}$ as much as the angle θ.

A multimodal prior probability distribution is generated by using each direction vector with the other parameters of the previous tracking such as weighted mean of the step size $\bar{z}_s^k$ and weighted mean of the vessel width $\bar{z}_w^k$ as if they are mean values of a multivariate normal distribution $N\left(\left[D_x^f, D_y^f, \bar{z}_s^k, \bar{z}_w^k\right], \left[\sigma_{Dx}, \sigma_{Dy}, \sigma_s, \sigma_w\right]\right)$, where D_x^f and D_y^f represent the f^{th} direction vector and satisfy $f = 1, 2, 3, \cdots, 7$. In the end, one can have 7 different normal distributions, represented with particles, around the landmark location. These particles from 7 different normal distributions are put into the same set and the final large particle set is moulded according to the motion model so this set becomes the prior probability distribution of possible daughter vessels. A posterior probability distribution can be calculated by using the centreline location probability map as likelihood. With the calculation of this posterior probability distribution, particles with better estimates of centreline location dominate the particle set, which can appear with the increased particle number at possible centreline locations. Figure 6.5a illustrates the posterior probability distribution of centreline locations of possible daughter vessels. These possible branch locations and particles located in these locations can be estimated by using *Mean Shift Clustering* method,

Mean Shift Clustering Algorithm [164]: This algorithm treats a set of data points as if they are sampled from a probability distribution. According to this description, the

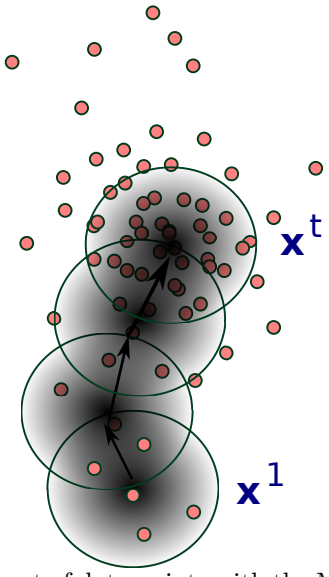


Figure 5.7: Detecting the mode of a set of data points with the Mean Shift algorithm. x^1 represents the first estimate of a mode location.

locations with larger number of data points are viewed as the modes of this probability distribution [165]. A mode in this distribution can be detected by finding a location where the change of the estimates of a mode location in two consecutive iterations is so small or zero during the application of a gradient ascent process on locally estimated density. The data points in the periphery of a mode location is viewed as the members of the cluster associated with this mode location.

A function $f(x)$ described with n data points can be approximated with the sum of a kernel $K(\cdot)$ with a bandwidth h in data locations as given in (5.33) [165].

$$\grave{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\big(\frac{x - x_i}{h}\big) \tag{5.33}$$

The gradient of this kernel density approximation to a function $f(x)$ can be written in (5.34).

$$\Delta\grave{f}(x) = \frac{c}{nh}\Bigg(\sum_{i=1}^{n} g\big(\frac{x - x_i}{h}\big)\Bigg)\Bigg(\frac{\sum_{i=1}^{n} x_i g\big(\frac{||x-x_i||^2}{h}\big)}{\sum_{i=1}^{n} g\big(\frac{||x-x_i||^2}{h}\big)} - x\Bigg) \tag{5.34}$$

where c shows a coefficient, which emerges due to the derivation of the kernel $K(\cdot)$. The derivation of the kernel function is shown with $g(\cdot)$.

In (5.34), the second term shows the direction towards the larger density and it is called the *mean shift vector*. This vector moves the estimates of a mode location to a stable estimate. Figure 5.7 demonstrates how the Mean Shift Clustering Algorithm detects the mode of a set of data points by iterating estimates of a mode location until reaching a stable location.

In this study, a *Flat Kernel* with a bandwidth h of 3 is selected to determine modes of the multi-modal centreline probability distribution around a landmark location. This

kernel function can be described with (5.35) [166].

$$K(x) = \begin{cases} 1 & if \quad x \leq h \\ 0 & if \quad x > h \end{cases} \tag{5.35}$$

If the number of modes is more than an expected number, the bandwidth h can be increased iteratively.

Figure 6.5b in Chapter 6 demonstrates particle sets representing modes of a multimodal probability distribution showing possible daughter vessels. Each mode in this probability distribution represents the centreline location of a possible daughter vessel. The particles representing a mode are put into a separate particle set and used for the initialisation of tracking of the next daughter vessel directly by starting from the centreline location calculated from the weighted mean of centreline locations obtained from particles in this set.

5.5 Conclusion

Quantitative vasculature analyses such as vessel width estimation has a lot of significance in clinical applications in terms of diagnosing eye-related diseases or using them as biomarkers to predict the risk factor or incidence of a disease. In this thesis, a flexible tracking method, particle filter, is combined in a tracking system with a powerful segmentation tool, the multi-labelling TDBN.

The way tracking is implemented is an example of a simulation-based probabilistic approach. Although there are other probabilistic tracking methods in the retinal vessel analysis literature, the proposed method is distinguished by being able to evaluate a large number of hypotheses and using the expectation of the simulated probability distribution with these hypotheses (due to using particle filters) in the decision of vessel characteristics, in contrast to maximising posteriori probability over some selected choices [74, 104]. Another difference of the proposed tracking system from other methods is how various vessel configurations (single, bifurcating or crossing) are modelled.

Regarding probabilistic tracking, some previous studies investigated the connectivity of vessels under various configurations such as 'single, bifurcating or crossing' vessels, along with the estimation of vessel parameters [74, 104]. Therefore, the likelihood calculation of possible pairs of edge locations across a vessel profile was calculated for three possible scenarios and the highest likelihood among them regardless of configuration was selected as the best fitting ones. This operation may increase the complexity of likelihood calculation, which may need to handle multiple factors. On the other hand, the proposed tracking system reduces the number of configurations to be evaluated to one ('single' vessels) by accepting vasculature consisted of single vessels connected at bifurcation/crossing locations. *Mode* 1 (corresponding to 'Single' configuration in other approaches [74, 104]) is utilised as default for tracking. The system switches to *mode* 2 (associated with the

identification of daughter vessels) if some conditions are satisfied during tracking in *mode* 1. This reduces computation memory for the estimation of regular vessel parameters by not engaging it with finding the right configuration. Also, the multiple mode structure of the system makes the tracking system flexible in terms of adopting different approaches for the detection of branches or tracking regular vessels. For example, bifurcating/ crossing locations of vessels are detected by using landmark location probability maps and the mean shift clustering algorithm, but regular vessel tracking is performed by using centreline/vesselness/edge probability maps and particle filters. Moreover, this approach may prevent tracking from 'particle depletion', which usually occurs when the tracker follows multiple branches at the same time and a branch, represented with the weak mode of a multimode probability distribution, does not receive enough number of particles to represent it [155].

The most striking difference between the proposed method and other probabilistic methods is how a vessel cross-section is approximated. In previous work [74, 104], a vessel intensity cross-section profile was assumed to be Gaussian and background intensity to be constant. However, these assumptions may oversimplify these profiles, leading to the failure of tracking when one of vessel boundaries has a considerably different intensity level due to non-even illumination or pathologies. On the other hand, some recent studies adopted model fitting approach for vessel width estimation, suggested to model vessel cross-section profiles with non-Gaussian and asymmetric shapes [17, 50], which seems to give better estimation of vessel parameters. The way vessel profiles are mimicked in this thesis is in the same line with the latter studies. Vessel mid-line is accepted to be around the centre of vessel boundaries but not strictly at the center, which allows the tracker to cope with shifts of peak locations in vessel profile.

To date, the modelling of vessel intensity cross-section has been an important step to determine vessel boundaries and to calculate vessel widths [17, 34, 50, 74, 92, 94–97, 99, 104, 167, 167]. In this study, vessel cross-sections are defined with three probability profiles: edge, centreline and vesselness. Also, the maximization of boundary and centreline probabilities is taken into account in the decision of vessel boundaries without conforming to any well-defined vessel cross-section model. These probability profiles are simply extracted from probability maps generated by the multi-labelling TDBN, which does not require the estimation of probability profiles according to local intensities [104]. Apart from previous studies using a dynamic search window, whose size varied depending on estimated width and the change on vessel orientation [74, 104], a fixed size of search line is used in the calculation of likelihood, where the scales of probability profiles are adapted to the length of the line. This facilitates the use of the same locations in the search line as hypothetical edge locations, which are selected to be both ends of the line. This simplifies sampling from the probability profiles by allowing faster computation. With these aforementioned advantages, the proposed tracking system can trace vessels consistently, even, in challenging images such as those affected from central light reflex or pathologies; even though, probability profiles do not have very sharp peaks or do not conform to any well-defined

vessel cross-section model.

Chapter 6

Tracking Performance Evaluation

6.1 Introduction

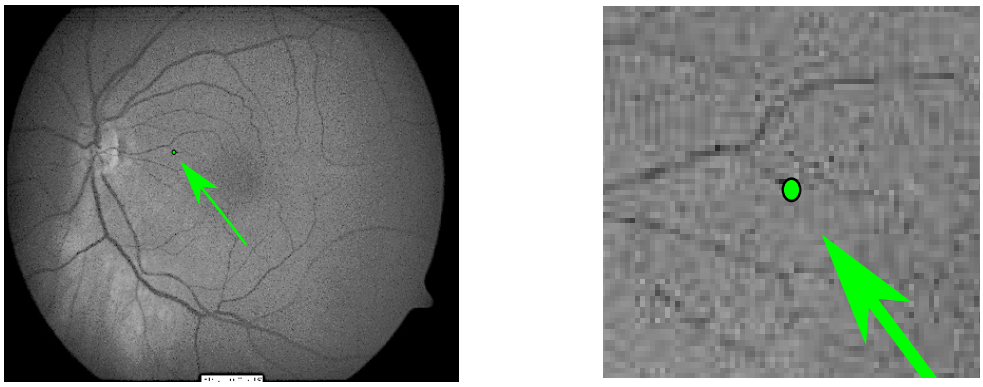
In the previous chapter, a tracking system was presented for retinal vessels from a theoretical point of view. This chapter will concentrate on applications of this tracking system to vessel width estimation and to the generation of a vasculature network. Initially, the behaviour of the complete system will be examined in a vessel bifurcation region, which will allow readers to observe how the system is working when it is in different modes. Then, the focus will be shifted to the first mode of the system, regular vessel tracking, which will be run for the width estimation of vessel segments. As mentioned before, the change on vessel widths is one of the most important biomarkers in the diagnosis of retinal and other diseases [11, 24, 26, 27, 30, 31]. This section will compare the performance of the tracking system on this task, when using different observation models. This experiment will be performed on the REVIEW dataset. Although generating a complete vasculature network in a fundus image is beyond the aims of this thesis, the tracker (the tracking system) will be utilised for tracing a vessel tree fraction in order to demonstrate the ability of the tracking system on landmark location estimation and daughter vessel detection.

6.2 Demonstration of Vessel Tree Tracking

This section aims to visualise how the tracking is performed over a small part of vessel tree fraction, where the tracking of a regular vessel, detecting landmark locations and identifying daughter vessels at this location will be visually demonstrated.

6.2.1 Experimental Settings

For this tracking, 400 particles were employed. The initial estimates of both a centreline location and a vessel width were made with the user. Figure 6.1 shows this estimated location, which is over a capillary. The initial direction for the tracking was estimated from the eigenvector of Hessian matrix at this location.



(a) (b)

Figure 6.1: (a)A fundus image from VDIS dataset: Green arrow shows the starting location of the tracking, which is shown with a green dot (b) Zoomed version of (a)

This direction is demonstrated with an arrow in Figure 6.1. The standard deviations for direction vector, step size and vessel width were experimentally selected and were 0.1, 0.3 and 0.5 respectively. Both the threshold for being a landmark location T_L and the one to stop tracking according to centreline probability T_c were 0.3. These values were also determined empirically. As observation model, *Model 1* was used.

The image used in this experiment belongs to VDIS subset of the REVIEW dataset. Because the REVIEW dataset does not have ground truth vessel maps, the network trained with the DRIVE dataset, which was explained in Section 4.7 in Chapter 4, was used for the segmentation of VDIS dataset. Due to remarkable difference between the resolution of the DRIVE and that of the VDIS, the resolution of VDIS dataset was reduced to half of its original resolution before feeding the dataset into the network. The tracking of vessels in VDIS dataset was started after returning its reduced resolution back to the original one. To be clear, tracking was realised at the original resolution of the VDIS dataset; though, the the segmentation was realised on its downsampled images.

6.2.2 *Mode 1*: Regular Vessel Segment Tracking

Because the starting point is located on a regular vessel segment, the tracking starts with *mode 1* (*Regular Vessel Tracking Mode*). After moving the particles with the initial distribution according to the motion model, the posterior probability distributions are calculated given observations. Figure 6.2a shows estimations of centreline and edge locations with the particles at iteration $k = 1$. Because the memory of the tracking has not yet been built, the estimate at this first step relies mostly on the initial information. The likelihoods of particles with the best predictions and those with the worst predictions can be examined by engaging with the vesselness and edge profile functions in Figure 6.2b. Because of initialising the tracking with good estimates of the centreline and width, both centreline and edge profiles corresponding to the predictions of particles are centred in the windows in many cases. The amount of the shift from the center increases towards the particles with lower likelihoods.

The estimates in the second step of tracking $k = 2$ are given in Figure 6.2c. The likelihoods of some particles are illustrated in Figure 6.2d. After each tracking step, the probability of a landmark location is checked. If it is larger than the threshold T_L, the mode of tracking changes from *mode 1*, *vessel tracking*, to *mode 2*, *landmark location detection*. Regarding current tracking, this threshold was exceeded after $k = 2$ so the mode was changed to *mode 2*.

6.2.3 *Mode 2*: Landmark Location Detection

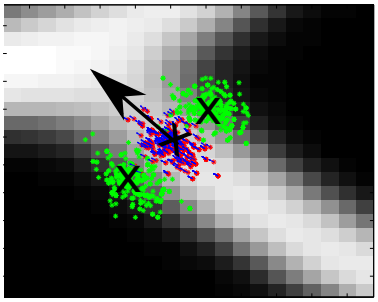
This mode has two functions. The first one is to estimate a possible location of a landmark, which can be bifurcation or crossing. The second one is to identify daughter vessels connected to this landmark.

A landmark location is estimated by maximising the posterior probability of being landmark location through particles. For this process, the same particles used in the last *regular vessel detection* are used for searching the peak of landmark location probability map in a region whose landmark location probability is higher than the threshold T_L. This search is made along vessel centrelines because of the shape of landmark location probability functions (see Figure 5.6 in Chapter 5); though, the likelihood of a hypothesis/particle relates to the landmark location probability at these locations. Figure 6.3a visualises four steps of estimates for a landmark location during this search. As seen from Figure 6.3b, the location with the highest landmark location probability (the yellow dot) is selected as the landmark location in this region.

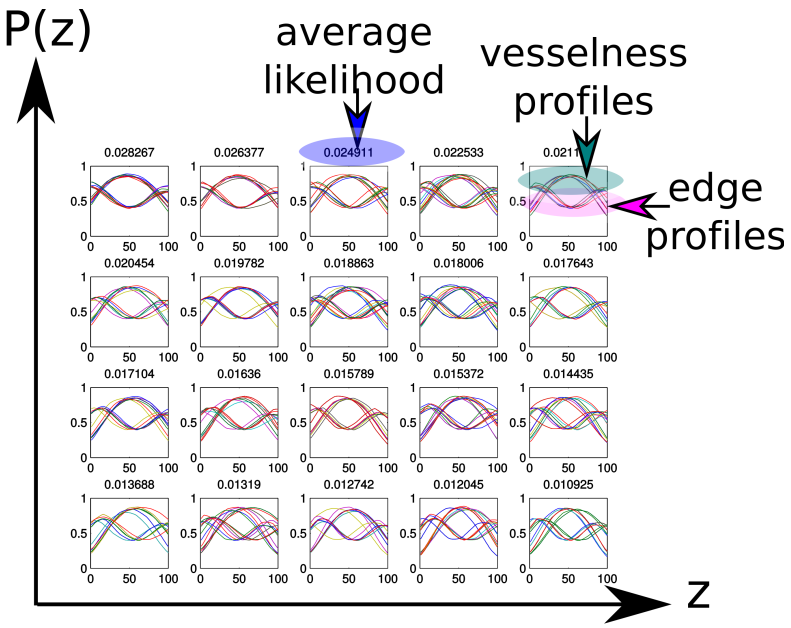
6.2.4 Detecting Daughter Vessels

The second function in *mode 2*, the *landmark location detection* mode, is the identification of daughter vessels connected to the estimated landmark location. In order to predict which daughter vessels are connected to the estimated landmark location, a probabilistic approach is asserted. In this approach, new sets of particles are initialised. The initial centreline locations for all sets of particles are assigned as the estimated landmark location, whereas particles are allowed to maintain their states for other parameters such as step size, vessel width estimated in the last step of *mode 1*, *regular vessel tracking* mode. On the other hand, the direction assigned to each set was selected from one of 7 directions shown in Figure 6.4. After passing the particles from all sets through the motion model, the estimates of centreline locations and those of directions are demonstrated in Figure 6.4. The variety in direction vector is assumed to increase odds for the prediction of daughter vessels elongated in any direction.

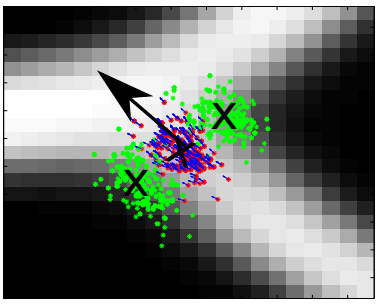
Figure 6.5a shows the updated states of particles representing the posterior probability distribution given likelihood, which is centreline probability. The observation model for this stage was made to be related to centreline probability map because this probability gives stronger suggestions for the existence of a vessel regardless of its width. As understood from this figure, there are two stronger suggestions for possible daughter vessels,



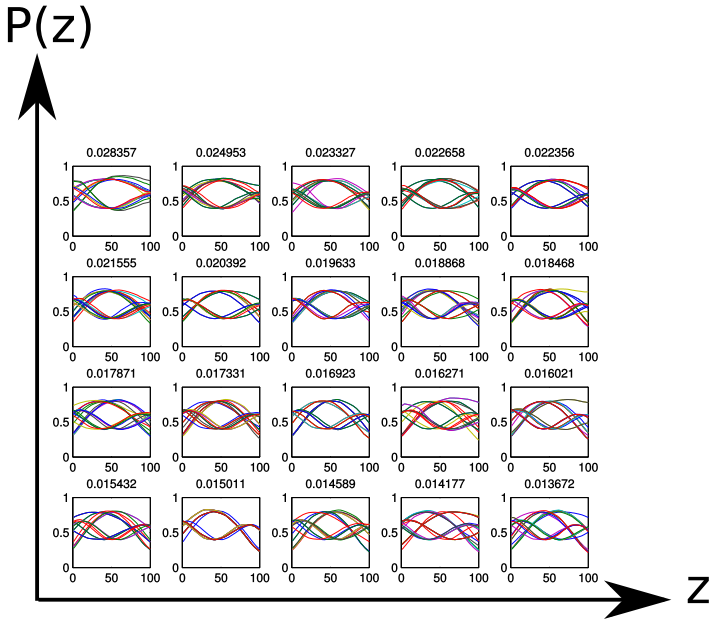
(a) $k = 1$



(b)

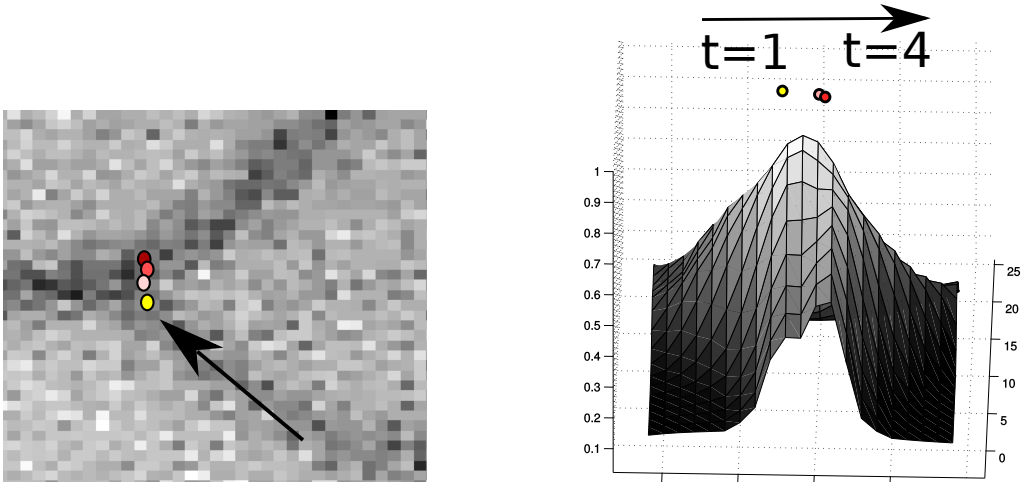


(c) k=2



(d)

Figure 6.2: Visualisation for the tracking of a regular vessel segment: (a)-(c) Estimates of each particle regarding centreline and edge locations are respectively shown with red and green dots. The expectations of centreline and edge posterior probability distributions are illustrated with a black cross. The expectation for the direction of tracking is illustrated with a big black arrow and the direction of each particle is demonstrated with a blue small arrow connected to its centreline location. (b)-(d) Each subfigure shows both vesselness (curves appear as peaks) and edge (curves appear as valleys) profiles for 10 particles. The first and second row shows the first 100 particles with highest likelihood while the third and the last row belong to the last 100 particles in the same order. The number on the top of each subfigure shows the average of the likelihoods of 10 particles in this subfigure.



(a) (b)

Figure 6.3: An estimation for bifurcation location with the Landmark Detection Function: The estimated landmark location is visualised with a yellow disc and other possible candidates with red discs (a) on the fundus image (b) on the landmark location probability map from $3D$ view. The estimation time for the yellow disc was $t = 1$ and for others $t = 2 : 4$ (the increase in the time is illustrated with the darker tones of red). The landmark location probabilities at these locations are respectively 0.799,0.796, 0.768 and 0.742 from $t = 1$ to $t = 4$.

which appear darker blue due to increased number of direction vectors in these regions. The mean shift clustering also identifies two particle clusters for possible daughter vessels as demonstrated in Figure 6.5b. In order to use any particle set suggesting a daughter vessel directly for the initialisation of the tracking of this vessel, two steps of tracking in *mode 1* are used to validate the result of Mean Shift Cluster and to improve the estimates of particles.

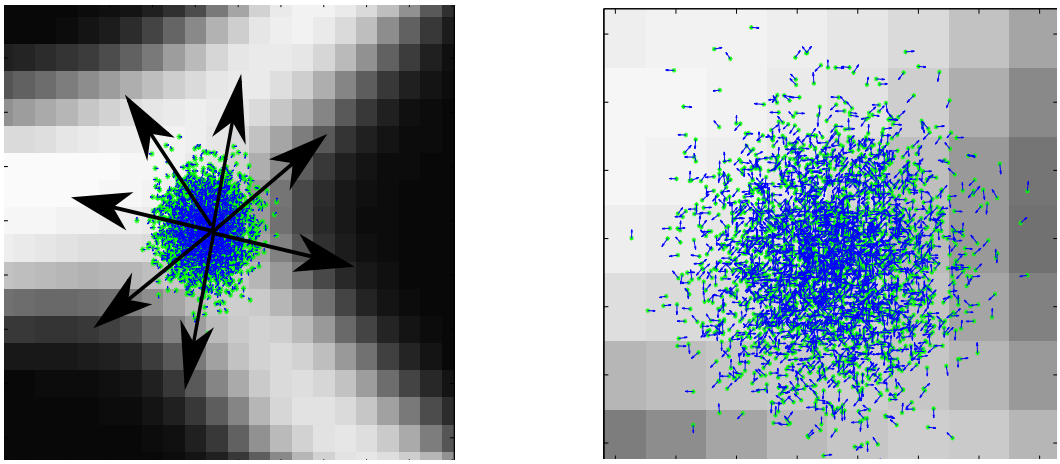
Figure 6.6a and Figure 6.6c show the tracking of daughter vessels in this validation stage. As seen from Figure 6.6b, the vesselness probability profiles are not so smooth, which shows the increased noise level in the vesselness probabilities in this vessel part.

The tracking of a vessel continues until the centreline probability at an estimated centreline location reduces beyond a threshold, T_c. In rare cases, the direction of tracking can change from vessel surface to background image due to noise or a step size that is larger than vessel radius. Regular checking of the probabilities at the centreline locations can stop the tracking in these cases before the tracking starts to behave unexpectedly.

6.3 Quantitative Vasculature Analysis

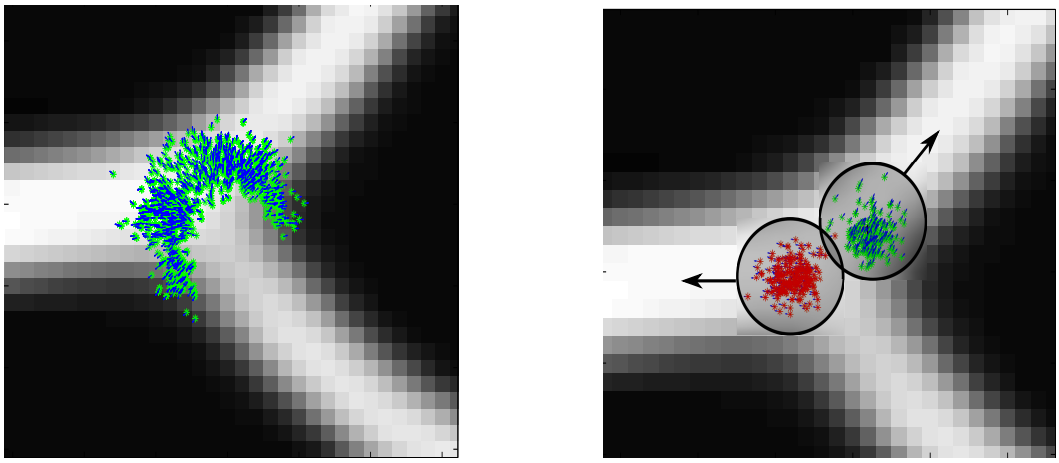
6.3.1 Introduction

In this section, the proposed tracking system will be utilised for the quantification of retinal vasculature. Firstly, the reliability of this system on width estimation will be evaluated on the REVIEW dataset and will be compared with that of previous studies. Then, the performance of the system on the detection of landmarks and daughter branches will be



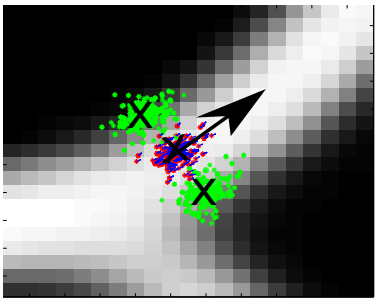
(a) (b)

Figure 6.4: (a)Prior Probability Distribution of daughter vessel existence: (b) Zoomed version of (a); where blue arrows show the direction vector for each particle and green dots centreline locations of possible daughter vessels.

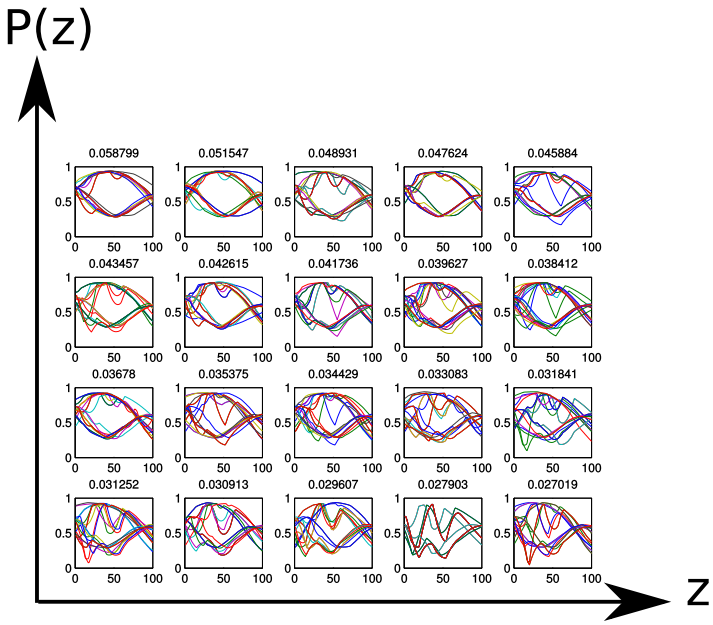


(a) (b)

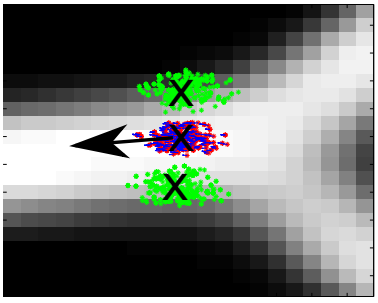
Figure 6.5: (a) Posterior probability distribution of being a part of the centreline of a daughter vessel: which is the result of one step of Particle Filtering ($P(z^{k+1}|y^{k+1})$). Each particle is represented with its direction vector and centreline location. (b) Estimated particle sets corresponding to possible daughter vessels by using *Mean Shift Clustering Function*, where each particle set is shown with a black circle and a direction arrow.



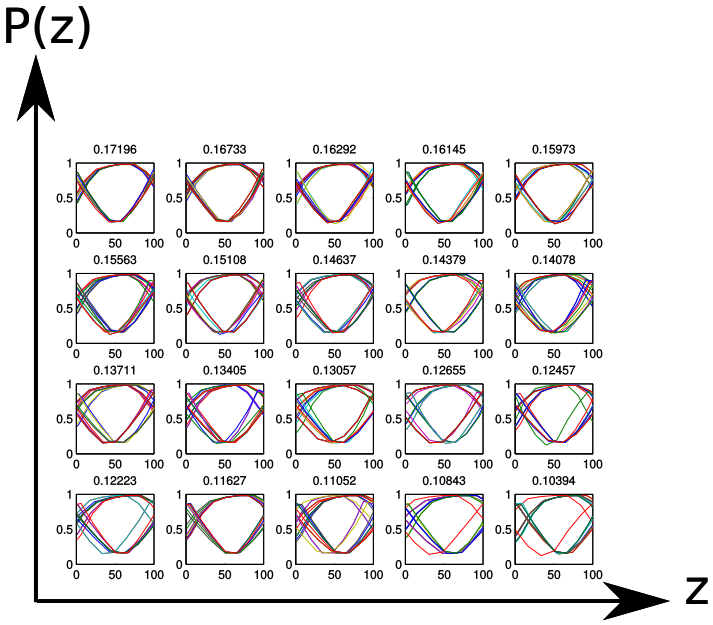
(a) *Tracking of the right daughter vessel at* $k = 1$.



(b)



(c) *Tracking of the left daughter vessel at* $k = 1$.



(d)

Figure 6.6: Starting the tracking of estimated daughter vessels (Refer back to Figure 6.2).

assessed on the DRIVE dataset.

6.3.2 Experimental Settings

Regarding the tracking, the parameters, which were given in Section 6.2.1, were used for the tracking of the DRIVE and the REVIEW datasets, unless stated otherwise.

For the DRIVE dataset

In this experiment, the tracking was started from a manually selected seed location and the tracked path, estimates of landmark locations during this tracking and estimates of centreline and boundary locations were examined in detail as follows. The image used for the tracking was the 10^{th} test image of the DRIVE dataset. The probability maps of vesselness, vessel centreline, vessel boundary and landmark locations were generated by using the proposed multi-labelling network. The probability maps were generated in the same resolution as the original images.

The initial estimate of centreline location was ($x = 462,\ y = 316$) and that of vessel width was 7 pixels. For this tracking, the step size was fixed to 2 pixels. The initial estimate of the direction vector was the eigenvector of the Hessian matrix at this estimate of centreline location.

For the REVIEW dataset

Because the REVIEW dataset has ground truth data for vessel boundary locations and one can calculate vessel centreline locations and vessel widths from these boundary locations, the initial estimates of both centreline and boundary locations and that of vessel width were obtained from the ground truth data. Also, the initial estimates of tracking direction were defined with vectors directed from the first centreline locations to the second ones, similar to Zhang *et al.*'s application [74].

Because the REVIEW dataset does not have any ground truth vessel maps, the probability maps of vesselness, vessel centreline and edge locations and landmark locations were generated by using the network trained on the DRIVE dataset for multi-labelling. Figure 6.7 shows an example of probability maps generated for the REVIEW dataset. However, it is worth mentioning that because the resolution of the DRIVE dataset, which is 768 by 584, is much lower than the sub-datasets of the REVIEW dataset (VDIS dataset has a resolution of 1360 by 1024 pixels, HRIS of 3584 by 2438 pixels, CLRIS of 2160 by 1440 pixels and KPIS of 3300 by 2600 pixels), the probability maps were synthesised in lower resolutions than the original resolutions of images in the REVIEW dataset after sub-sampling them by a factor of 2 for VDIS, 3 for CLRIS, 4 for HRIS. Later, the probability maps were up-sampled to the original resolutions of these datasets.

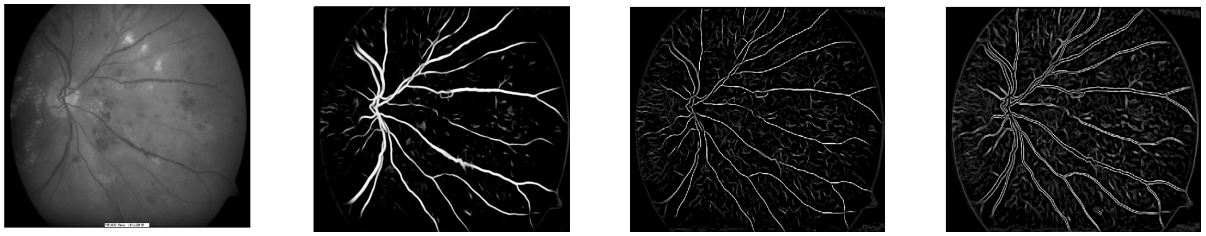


Figure 6.7: The segmentation of blood vessels in VDIS dataset (8^{th} image): The images from left to right are respectively a fundus image and the probability maps of its vessel interior pixel, centerline and boundary segmentation

6.3.3 Evaluation Criteria

For Vessel Width Estimation

Precision and *accuracy* are the most commonly used measures to evaluate the estimates of vessel widths [47, 50, 88, 168]. These measures can be calculated from (6.1) and (6.2) respectively [168].

$$Precision = std(w_r - w_e) \tag{6.1}$$

$$Accuracy = mean(|w_r - w_e|) \tag{6.2}$$

where w_r denotes reference width while w_e shows an estimate of the vessel width.

Depending on the way vessel widths estimated, mean of vessel widths over a vessel segment may vary among methods [6]. Because not actual vessel width but the relative vessel width over a fundus image is considered to be important [34], consistency in vessel width estimates is more important than the accurate estimates of actual vessel widths. The consistency in vessel width estimates is indicated with *precision* so it has more importance than *accuracy*. On the other hand, the bias of an estimator calculated by *accuracy* can be linearly removed either by subtracting a constant from it or dividing/multplying it with a scalar [6]. Because of these reasons, even though *accuracy* is reported in following experiments, it is to inform readers but will not be used for the performance evaluation.

In addition to *precision* and *accuracy*, the percentage of traced vessels was also reported in previous studies [17, 34, 47, 88, 97, 168]. This measure shows the ability of a method to successfully return vessel widths for given centreline locations in ground truth data [88] and can be used as an indication of early stopping.

For Vessel Tree Extraction and Landmark Location Estimation

For related experiments, only visual demonstration will be provided for the performance evaluation.

6.3.4 Vessel Width Estimating in the REVIEW dataset

In this section, the prediction reliability of four models on vessel width estimation, introduced in the previous chapter, will be assessed on the REVIEW dataset. In order to evaluate the performance of these models independent of that of tracking, two sets of experiments will be presented. The first one assumes boundary locations (which will be used to extract probability profiles) for the estimation of vessel width are known; while the other accepts that only starting information for the tracking is known. After evaluating the performance of the models in these experiments, this performance will be compared with that of previous studies.

The First Experiment

This experiment assesses the reliability of the four models on the estimation of vessel widths, while tracking is performed by using ground truth centreline locations (the mid points between boundary locations) and the directions between boundary locations. Therefore, the difference between vessel width estimations can be accepted to be only due to the usage of different models.

Table 6.1 tabulates the performance of each model on subsets of the REVIEW dataset (CLRIS, HRIS, VDIS, KPIS) by considering a comparison over vessel profiles (not vessel segments). As seen in this table, all four models successfully estimate vessel widths for each subset. Among them, *Model* 1 produces higher precision than that of the other three models. This model assumes the symmetry of vessel boundaries according to centreline and uses vesselness probability profile for the evaluation of the tracker to be inside vessel or outside it. The second model considers asymmetry of vessel boundaries so the model uses a state variable to estimate vessel width for either side of centreline. When compared with the performance of the first model, the second model provides improvement in precision for the image sets. The third model replaces vesselness probability profiles in the first and the second model with centreline probability profiles. This change in the likelihood calculation leads to further improvement of consistency in CLRIS, HRIS and VDIS image sets. Particularly, a remarkable decrease in precision for CLRIS is present. This may be due to centreline probability profiles to mostly maintain the smooth curve appearance despite of the existence of central light reflex or other pathologies; on the other hand, vesselness probability profiles may be affected from this noise components to some extent. Finally, the forth model further improves the consistency of CLRIS. Adding edge strengthening information (increased number of sampling from edge probability profiles in the calculation of likelihood) to likelihood appears to decrease precision for this set; while it does not lead to any improvement of precision for HRIS and VDIS, where the latter one has blurry edges. In spite of decreases in precision values for CLRIS, HRIS and VDIS after replacing vesselness probability profiles with centreline probability profiles, the precision calculated for KPIS increases. This may be due to the usage of 'kick points' for locating vessel edges in this dataset, which is different from other sets. The best model detect-

Table 6.1: The performance of the proposed method on vessel width estimation when it follows centreline location given in ground truth data

	CLRIS			HRIS			VDIS			KPIS		
Models	Acc.	Prec.	%	Acc.	Prec.	%	Acc.	Prec.	%	Acc.	Prec.	%
Model 1	2.1695	1.1388	100	0.7979	0.5639	100	0.9197	1.1446	100	0.6258	0.3939	100
Model 2	1.6376	1.2003	100	0.7694	0.4464	100	0.8715	1.0920	100	0.6742	**0.3080**	100
Model 3	0.6516	0.8326	100	0.3296	**0.4130**	100	1.0095	**0.9347**	100	1.6894	0.3510	100
Model 4	0.7395	**0.8244**	100	0.4027	0.4182	100	1.5176	0.9595	100	1.9808	0.4427	100

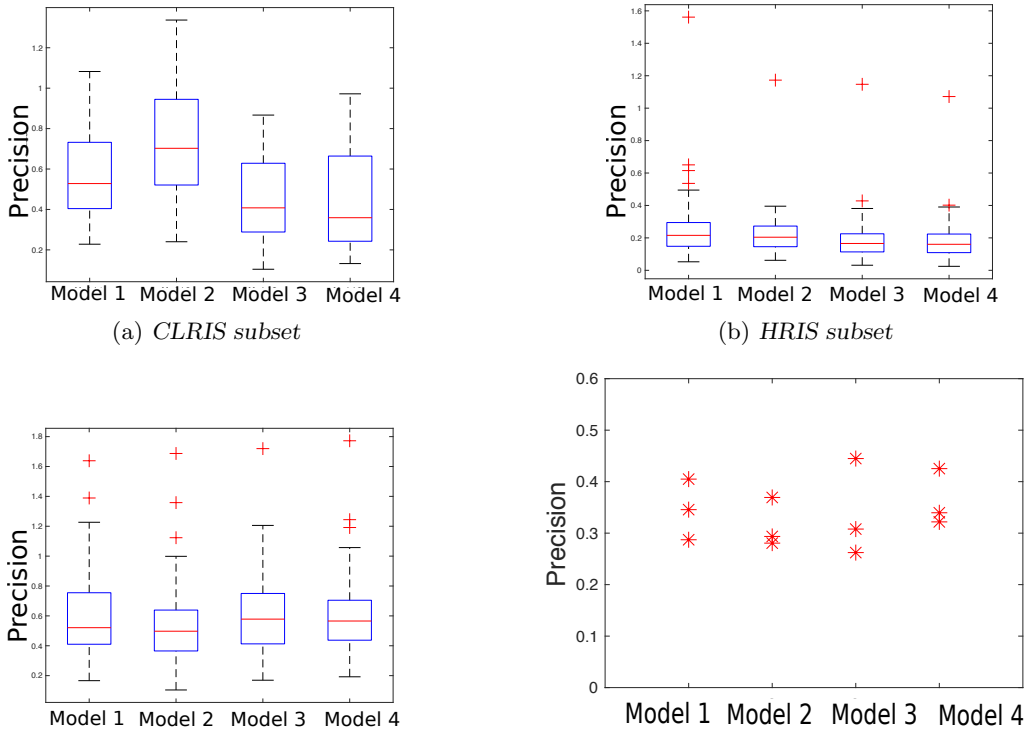
ing edge locations closer to 'kick points' seems to be the second model, which considers vesselness probability profiles.

Although the general approach to assess the performances of methods on vessel width estimation is to make the comparison for each vessel profile given in the ground truth, this approach does not show the success of methods on individual vessel segments. Especially, a disagreement between predicted widths and reference widths may be specific to individual vessel segments but not related to general consistency of the method on width estimation. In order to evaluate the consistency of the method on vessel segments, average precision for each vessel segment was calculated. Figure 6.8 compares average precision for the four models. As seen in this figure, there is almost strong disagreement between reference and predicted widths for a couple of vessel segments for HRIS and VDIS datasets. Examples of these segments are given in Figure 6.9. As seen in this figure, the disagreement may be due to (i) different definition of vessel edges between reference data [88] and the models as demonstrated in Figure 6.9a or (ii) immediate changes in vessel width as a result of pathologies (e.g. vessel beading in Figure 6.9b and 6.9c). It should be noted that there was still a dependency between previous states of particle filtering and its current states and this dependency should be taken into account in the performance evaluation of the proposed method; though, centreline locations were taken from the ground truth. The adaptation of the tracking system to quick changes in vessel width may be enhanced by making estimations for locations between given centrelines.

The Second Experiment

In contrast to the first experiment, here, only initial values of centreline location, vessel width and vessel direction were taken from the ground truth data. Then, the tracker automatically estimated centreline and boundary locations with a step size of 1 pixel. Because reference centrelines were not used to infer vessel widths, the same number of widths were sampled from both estimated and reference profiles along a vessel segment in order to find common locations between estimations and reference data for performance evaluation. For this reason, bi-cubic splines were used to interpolate vessel widths between either predictions or reference in the same number (which was 100).

In this experiment, the four models in the first experiments were used again. However, *Model* 3 and *Model* 4 were modified to incorporate direction information in the observation model, where previously the estimation of direction vectors depended on only the motion



(a) *CLRIS subset* (b) *HRIS subset*

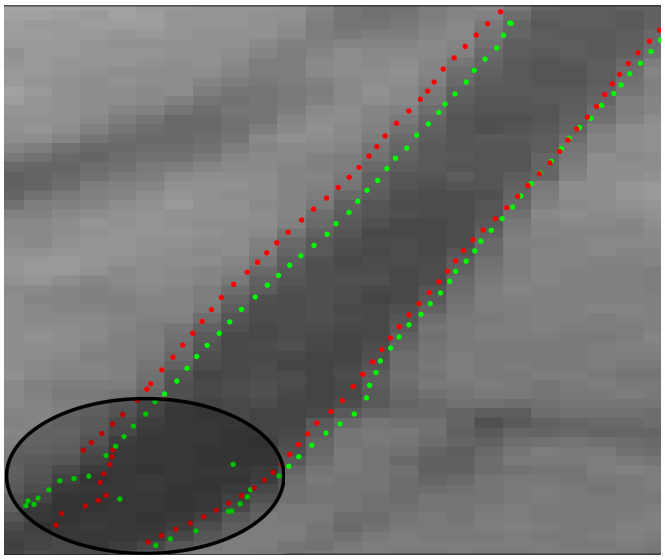
(c) *VDIS subset* (d) *KPIS subset*

Figure 6.8: The precison distribution for individual vessel segments: vessel widths are estimated given centreline locations in the ground truth data (Whiskers represent $+/-2.7\sigma$ in (a)-(c). '*' indicates the result of a individual image in (d)).

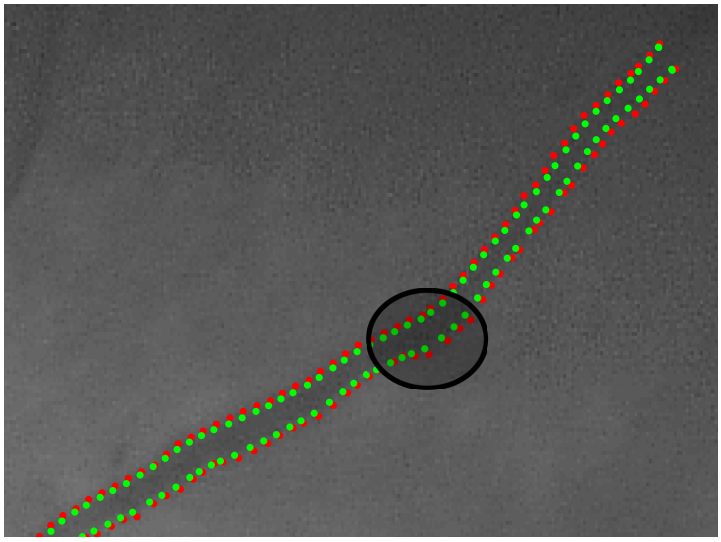
models. This modification may be necessary to make consistent estimation of widths on estimates of vessel cross-sections, whose alignment depends on estimates of vessel direction. Unless estimates of vessel directions are made accurately, a range of vessel widths may be estimated from various estimates of cross-sections passing from the same centreline. In order to observe if there is any performance change occurring in *Model* 3 when the direction vector is observed, the performance of *Model* 3 without making this change was also assessed under the name of *Model* 3*.

Table 6.2 compares the performance of the models. Similar to Table 6.1, the consistency of width estimates improves parallel to the increasing complexity of the models such as incorporating boundary asymmetry in the motion model, direction vector in the observation model and by emphasising vessel boundaries in the likelihood with extra boundary probabilities. The lowest precision values belong to *Model* 4 for CLRIS and VDIS, *Model* 3 for HRIS and *Model* 1 for KPIS. Interestingly, the consistency of width estimation for HRIS and VDIS, when their widths are inferred by centrelines estimated during the tracking, is found slightly better than those given in Table 6.1.

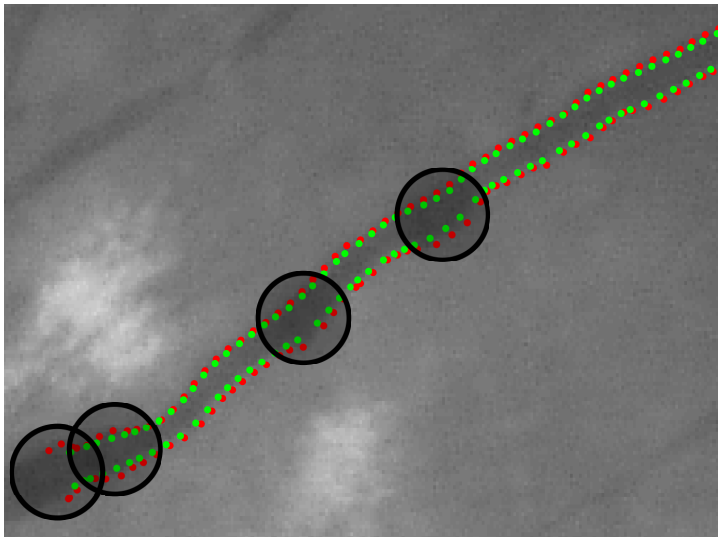
The results in Table 6.2 also align with box plots in Figure 6.10. Among the four models, *Model* 4 gives lower median and less scattered precision distribution for all image sets. Figure 6.11 and 6.12 illustrate examples of tracking results for vessel segment, which produce the largest and the lowest precision in each set of data respectively. In order to compare



(a) *HRIS*$(1-26)$ *estimated with Model* 4



(b) *VDIS*$(8-10)$ *estimated with Model* 4 *(upper part)*



(c) *VDIS*$(8-10)$ *estimated with Model* 4 *(lower part)*

Figure 6.9: Vessel segments with the worst average precision from HRIS and VDIS. Estimated locations are shown with green dots and reference locations with red dots. The black ellipse in (a) shows a mispredicted vessel boundary in the reference data. The black circles in (b) and (c) encircle vessel beading.

the variation of estimation depending on the model, the most common segments with the largest and the lowest precision among the three models were selected for demonstration.

For the example segment from CLRIS in Figure 6.11a, the assumption of asymmetric vessel boundaries appears to facilitate vessel boundaries to be followed better by the second and the third model when compared with those traced by the first model. On the other hand, centreline locations estimated by the first model align better with the ground truth. The three models could not catch the narrowing of vessel width in Figure 6.11d, 6.11e and 6.11f.

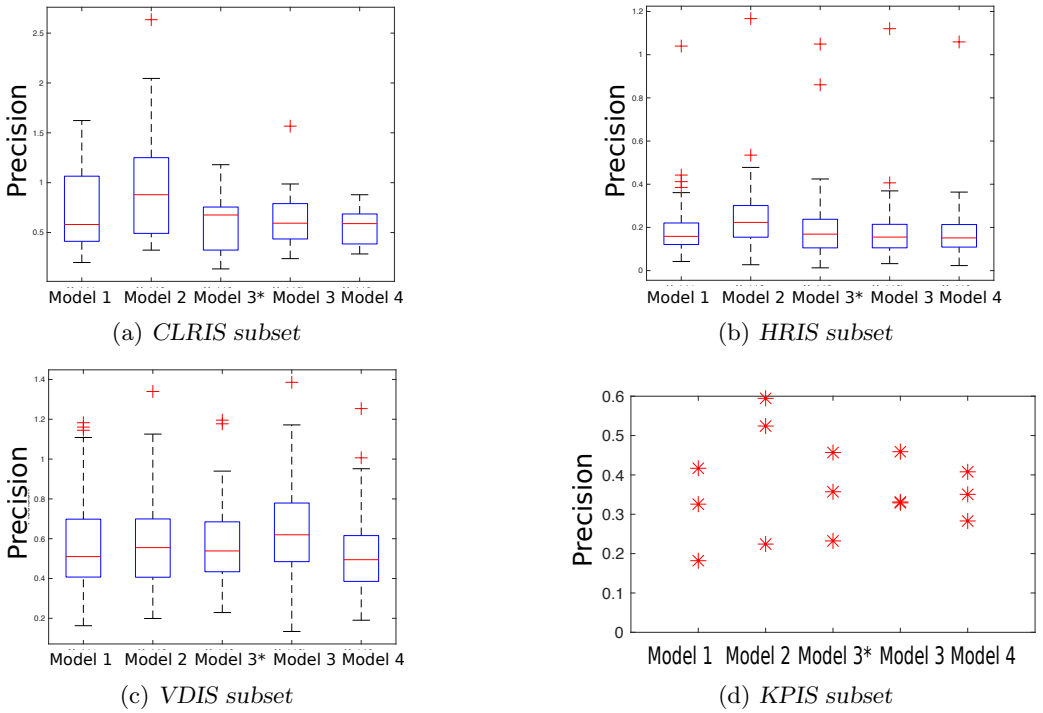
Figure 6.12 illustrates tracking results with the lowest precision. As expected, the models estimate vessel boundaries consistently for healthy and high resolution images as demonstrated in Figure 6.12d-6.12f and 6.12j-6.12l. Moreover, the three models achieve to predict vessel boundaries accurately and consistently despite the presence of a strong central light reflex in Figure 6.12a-6.12c. Likewise, a vessel affected from moderate noise and central light reflex in Figure 6.12g-6.12i is traced successfully with three models.

The central light reflex may cause trouble for some methods [34,88]. Figure 6.13a-6.13b demonstrate an example of tracking in the presence of central light reflex. As seen from the figures, central light reflex does not affect edge probability profiles so vessel widths can still be estimated accurately and consistently. Another issue related to vessel analysis is that close vessels were found to be difficult for tracking or width estimation [76, 88]. However, as demonstrated in Figure 6.13c, the proposed method successfully traces a vessel without being distracted by the close one. Also, Figure 6.13d shows this vessel for close examination. Although one of vessel boundaries have very low probability and very diffused, the tracker manages to completely trace this vessel. Another problem encountered in width estimation/tracking is the negative effect of vessel junctions [76,97], where regular profile of vessel cross-section changes/disappears and boundaries become indistinguishable. However, the proposed method can successfully deal with junctions as demonstrated in Figure 6.13k. Although edge probabilities become very low in junction regions, the vessel is accurately traced due to information preserved in the prior probability distribution.

The noisy and low contrast appearance of vessels may be challenges for human to locate vessel boundaries accurately. Figure 6.13e and 6.13l show examples of this situation, where reference profiles do not reflect the changing curliness of vessels as much as the tracker do. Also, Figure 6.13f ,6.13g and 6.13h show the limitation of human ability in the detection of vessel boundaries, where locations for one of boundaries are not located appropriately. Moreover, in these examples, reference vessel boundaries are not strictly perpendicular to vessel direction, which was reported in [6]. This may affect the performance of the method on width estimation. In Figure 6.13i and 6.13j, vessel boundaries are located more consistently by the tracking system/the tracker than reference data.

Table 6.2: The performance of the proposed method on vessel width estimation when it is used as an automatic tracker.

	CLRIS			HRIS			VDIS			KPIS		
Models	Acc.	Prec.	%	Acc.	Prec.	%	Acc.	Prec.	%	Acc.	Prec.	%
Model 1	2.4754	1.5156	100	0.8128	0.4592	100	0.9671	1.2037	100	0.5631	**0.3566**	100
Model 2	1.9546	1.5414	100	0.7049	0.4794	100	0.9060	1.1220	100	0.5313	0.4889	100
Model 3*	0.8703	1.1120	100	0.3217	0.4031	100	1.5034	0.9483	100	1.4731	0.4403	100
Model 3	1.2989	1.2177	100	0.3110	**0.3815**	100	2.3142	1.0545	100	2.0614	0.4343	100
Model 4	1.0349	**1.0308**	100	0.4034	0.4082	100	1.6273	**0.9044**	100	1.9623	0.4142	100



(a) *CLRIS subset* (b) *HRIS subset*

(c) *VDIS subset* (d) *KPIS subset*

Figure 6.10: The precison distribution for individual vessel segments: vessel widths are autonomously estimated by the tracking system. (Whiskers represent $+/-2.7\sigma$ in (a)-(c). '*' indicates the result of a individual image in (d)).

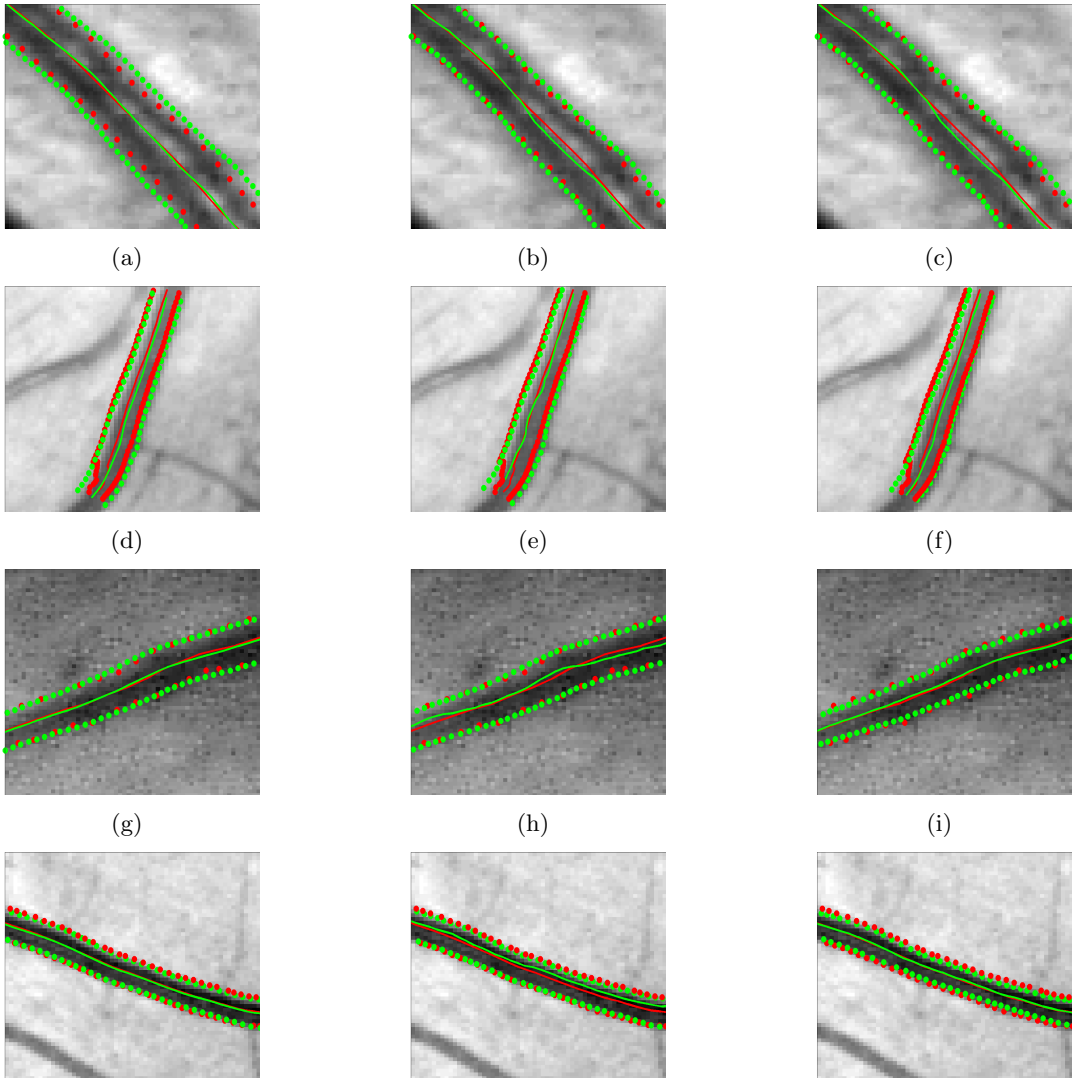


Figure 6.11: Performance comparison of the three models of vessel centreline and boundary locations (vessel segments with the worst precision). The top row belongs to the CLRIS dataset, the second one to the HRIS dataset, the third one to the VDIS dataset and the bottom one to the KPIS dataset; while the first column to *Model* 1, the second column to *Model* 2 and the third column to *Model* 3. Estimated locations are marked with green and reference locations with red.

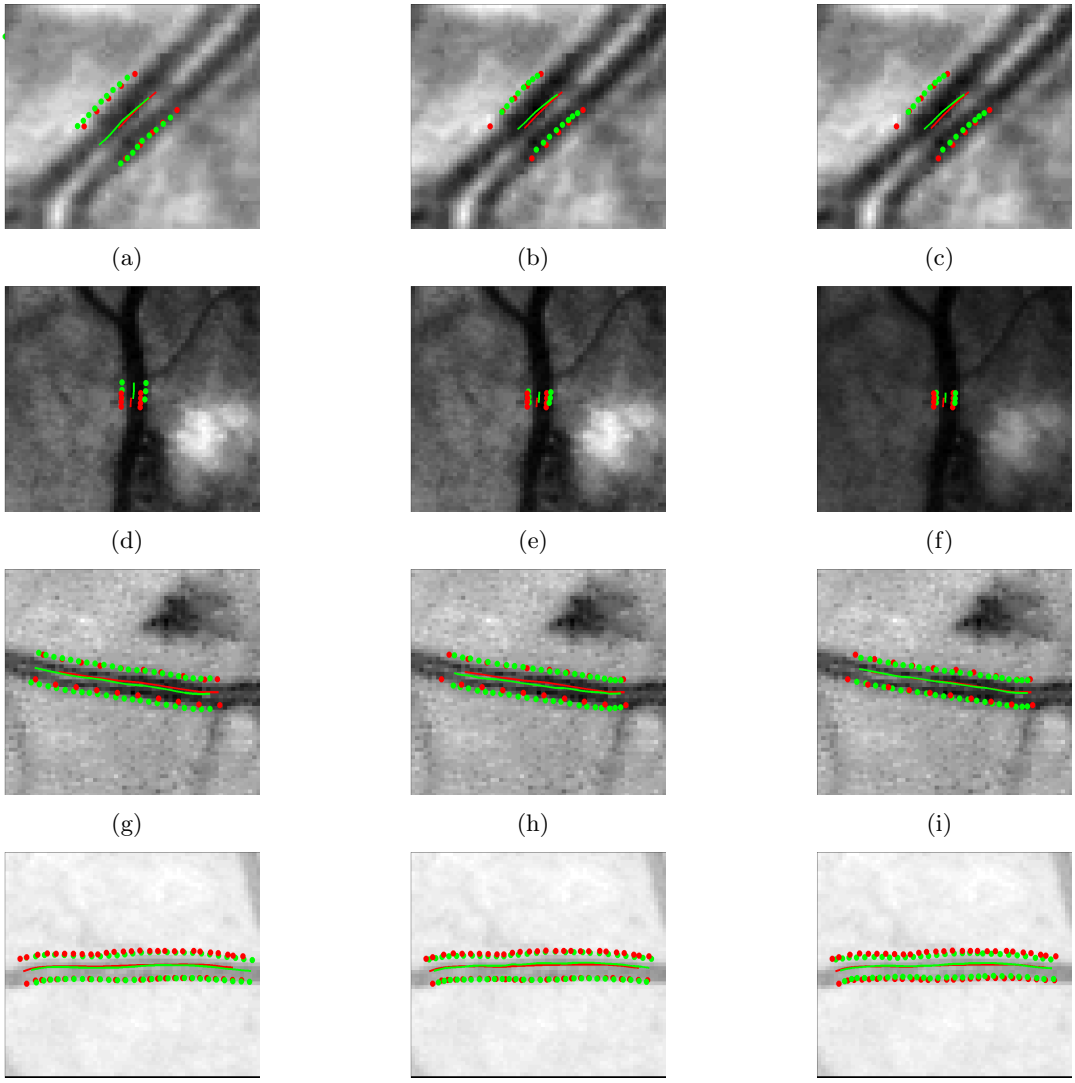


Figure 6.12: Performance comparison of the three models of vessel centreline and boundary locations (vessel segments with the best precision). The top row belongs to the CLRIS dataset, the second one to the HRIS dataset, the third one to the VDIS dataset and the bottom one to the KPIS dataset; while the first column to *Model* 1, the second column to *Model* 2 and the third column to *Model* 3. Estimated locations are marked with green and reference locations with red.

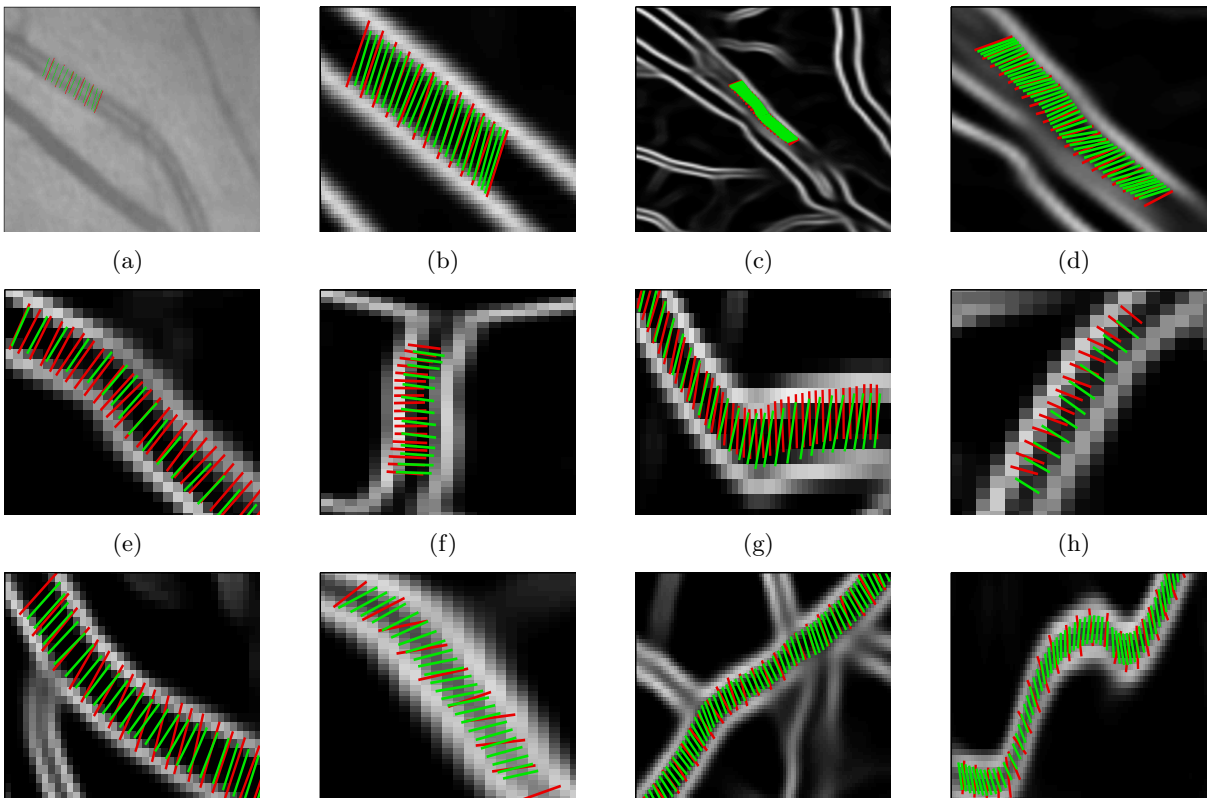


Figure 6.13: Some examples of predicted boundaries by *Model* 4 and reference boundaries (a)-(d) from CLRIS (e)-(h)from HRIS, (i) from KPIS and (j)-(l) from VDIS (green lines show predicted profiles and red lines reference profiles).

Performance Comparison With Previous Studies

In this section, I will compare the performance of the proposed tracking system when applied independently of centreline reference data (refer to Section 6.3.4), on the estimation of vessel widths in the REVIEW dataset with that of previous work. Table 6.3 tabulates the performance of methods in terms of *accuracy*, *precision* and *percentage of vessel completeness/the success rate* (which will be shown with % in the table).

Among the methods whose performances are compared in Table 6.3, recent tracking methods [74,104] generally have better consistency on width estimation than that of other methods given in this table, apart from that of the Araújoa *et al.* [50] under the category of supervised methods. Although a disadvantage of tracking methods over other methods regarding the dependency of estimated vessel parameters to those obtained in a previous step, which may lead to early stopping [74], the success rates of tracking methods on the estimation of widths over given profiles seems to be comparable to those of other methods.

Among tracking methods, Zhang *et al.* and Yin *et al.* used a probabilistic approch in the decision of vessel widths [74, 104], similar to the proposed method. When compared with the performance of these methods, *Model* 4 in CLRIS and *Model* 1 in KPIS present more consisteny in width estimation with a precision of 1.03 for CLRIS and 0.36 for KPIS. Also, the proposed models returned vessel widths for the entire set of profiles in these image sets, while Zhang *et al.* was able to return 98.3% of vessel profiles [74] in CLRIS. *Model* 3 has the second lowest precision in HRIS with 0.38 by following Zhang *et al.* with a precision of 0.3 [74]. It should be noted that width estimation in HRIS was made on the down-sampled images by the factor of 4 not on the original resolution as supposed to be [6]. The tracking on the original resolution may improve the consistency of estimated vessel widths for the proposed models. For VDIS, the proposed models could not reach the same precision level demonstrated by Yin *et al.* [104], with a precision of 0.56, and Zhang *et al.* [74], with a precision of 0.59; however, the precision of 0.90 generated by *Model* 4 is lower than that reported in many other studies in the table. Also, the proposed system achieves to completely trace all vessel segments in this image set while Zhang *et al.* [74] could not estimate vessel widths for %5.8 of profiles. A reason behind poor performance of the proposed method may be because of slow adaptation of the proposed system to quick changes in vessel width occurring due to vessel beading.

Despite the success of tracking methods on width estimation, the best performance regarding precision was reported by Araújoa *et al.* [50] in the four image sets, to the best of my knowledge. Their performance, even, outperformed the precision returned from some observers. On the other hand, the performance of other supervised methods reported by Aliahmad and Kumar [168] and Lupaşcu *et al.* [17] was inferior than that of tracking methods in the used datasets apart from KPIS, where Lupaşcu *et al.* [17] has the second best precision. Contary to tracking methods, supervised methods have a disadvantage in that their performance mostly relies on the characteristics of datasets used for training. As argued before by Lupaşcu *et al.* [17] and Araújoa *et al.*, an unbalanced

training set in terms of using smaller numbers of samples for some values of vessel widths may reduce the reliability of the trainable estimators. For example, the proportion of profiles for large vessels in VDIS set is considerably lower than that of small vessels, which increased the precision for vessels with widths of $13 - 19$ pixels to almost twice of that for smaller widths, as reported by Lupaşcu *et al.* [17]. Also, a large range of vessel widths in the retinal images may exacerbate the performance of supervised methods because this variety in vessel width may reduce the similarity of samples in training and test sets in the case of having a small or unbalanced dataset [17].

6.3.5 The Tracking of A Vessel Tree Fraction

This section demonstrates the performance of the proposed tracking method on the tracking of a vessel tree fraction. The performance of the method is visually examined in three ways: (i) continuously tracking vessel segments regardless of their thickness, (ii) reasonable estimation of landmark locations, (iii) identification of daughter vessels.

For this experiment, *Model* 1 was used for the tracking of vasculature in the DRIVE dataset. The tracking was started from a seed and continued until vesselness probability measured in vessel centreline was smaller than a threshold. During the tracking, the step size was fixed to 2 pixels. Figure 6.14a demonstrates the starting location for tracking and the traced path by the tracker. The details of tracking (centreline and edge locations) are given in Figure 6.14c-6.14d. As seen in these figures, vessels from a range of vessel thickness are traced smoothly. Even, almost invisible capillaries to the human eye are traced correctly.

The landmark locations detected during this tracking are illustrated in Figure 6.14b. According to this figure, the landmark locations are mostly estimated in reasonable locations, even for capillaries. Despite the detection of landmark locations correctly, some branches are not traced as seen in Figure 6.14c- 6.14d. This may be due to the failure in the identification of daughter vessels with *Mean Shift Clustering* (see the 5^{th} circle in Figure 6.14d) or due to wrong estimates of direction vector at these locations (see the 2^{nd}-3^{th} circles in Figure 6.14c and 4^{th} circle in Figure 6.14d), which eventually leads the tracking to outside vessel and causes these vessel branches to be not included in the tracking. Also, the estimates of landmark locations to be inaccurate locations (see the 1^{st} circle in Figure 6.14c) may lead to exclude possible daughter vessels from the tracking.

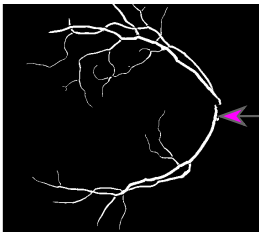
6.4 Conclusion

In this chapter, the working mechanism of the tracking system was illustrated for each mode of tracking by giving snapshots from the states of tracking system at various locations, which covered the tracking of regular vessels, the detection of landmark locations and the identification and tracking of daughter vessels. Then, the performance of the proposed vessel tracking system was evaluated on vessel width estimation, which may be

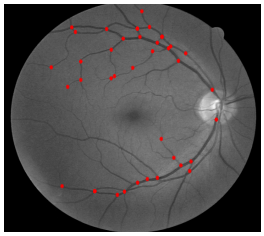
Table 6.3: The comparision of the performance of the proposed method with that of previous studies regarding the accuracy of vessel width estimates in subsets of the REVIEW dataset .

			CLRIS			HRIS			VDIS			KPIS		
	Method	Year	Acc.	Prec.	%	Acc.	Prec.	%	Acc.	Prec.	%	Acc.	Prec.	%
	Observer 1		0.61	0.567		0.23	0.288		0.35	0.543		0.34	0.417	
	Observer 2		0.11	0.698		0	0.256		0.06	0.621		0.11	0.317	
	Observer 3		0.72	0.566		0.23	0.285		0.3	0.669		0.23	0.326	
	Model 1	2017	2.47	1.52	**100**	0.81	0.46	**100**	0.97	1.20	**100**	0.56	**0.36**	**100**
	Model 2	2017	1.95	1.54	**100**	0.71	0.48	**100**	0.90	1.12	**100**	0.53	0.49	**100**
Tracking M.	Model 3	2017	1.30	1.22	**100**	0.31	0.38	**100**	2.31	1.05	**100**	2.06	0.43	**100**
	Model 4	2017	1.03	**1.03**	**100**	0.40	0.40	**100**	1.62	0.90	**100**	1.96	0.41	**100**
	Zhang *et al.* [74]	2014	0.37	1.13	98.3	0.08	**0.30**	**100**	1.37	0.59	94.2	0.74	0.37	**100**
	Yin *et al.* * [104]	2012	0.77	1.41	-	0.01	0.39	-	1.41	**0.56**	-	0.69	0.43	-
	Zhou *et al.* ** [97]	1994	7.5	4.14	98.6	0.54	0.90	99.6	3.07	2.11	99.9	2.57	0.4	**100**
	Araújoa *et al* [50]	2017	0.01	**0.56**	-	0.00	**0.22**	-	0.00	**0.69**	-	0.00	**0.30**	-
	Aliahmad and Kumar [168]	2016	0.33	1.56	98	0.24	0.65	99.4	0.45	1.14	97.8	0.72	0.45	**100**
Supervised M.	Lupaşcu *et al.* [17]	2013	0.00	1.154	**100**	0.00	0.44	**100**	0.02	1.07	**100**	0.02	0.32	**100**
	Xu *et al.* [47]	2011	0.08	1.78	94.3	0.21	0.567	**100**	0.53	1.43	96	1.14	0.67	99.4
	Al-Diri *et al.* [88]	2009	1.9	1.47	93	0.28	0.42	99.7	0.05	0.77	99.6	0.96	0.33	**100**
Unsupervised M.	Lowell *et al.* ** [34]	2004	6.8	6.02	26.7	0.17	0.70	98.9	2.26	1.33	77.2	1.65	0.34	**100**

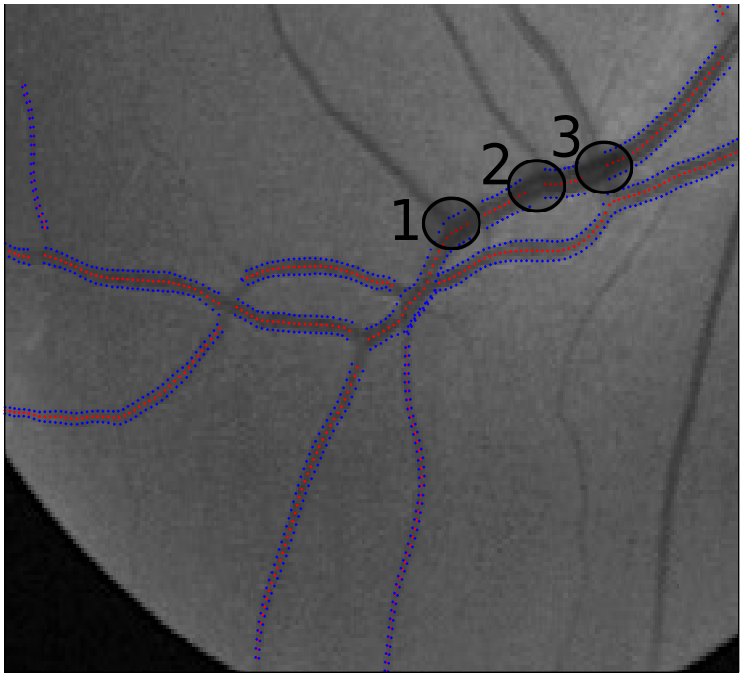
* These results were taken from [74]. ** These results were taken from [168].



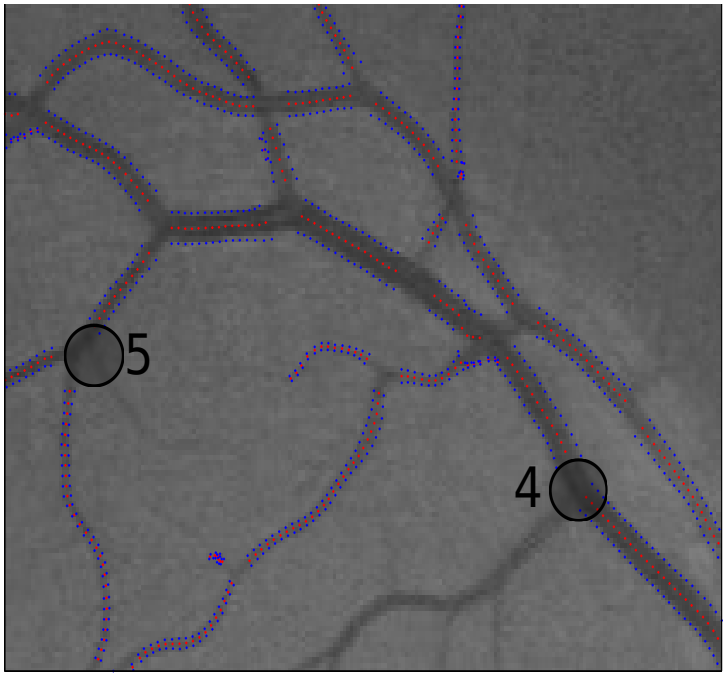
(a)



(b)



(c)



(d)

Figure 6.14: The tracking of a part of vasculature in 10^{th} test image in the DRIVE dataset: (a) The traced path (the starting point for tracking is shown with the pink arrow.) (b) Estimated landmark locations during the tracking (c)-(d) Estimates of vessel centreline and edge locations, where estimates of boundary locations are marked with blue and those of centreline locations are with red. (c) shows the bottom left corner of the image and (d) shows its top right corner (Some missed branches during tracking are illustrated with circles).

considered as an important task regarding its common usage in medical research as a biomarker. These experiments were followed by another one, where the tracking system was utilised to estimate regular vessel parameters, landmark locations and daughter vessel parameters of a vasculature network fraction by starting from a point near the optic disc. This final section will summarise findings of these experiments and the contributions of this research to the literature.

6.4.1 The Performance Evaluation

When compared with previous studies, the consistency of estimated widths by the proposed tracking system was found better or comparable to similar studies. For KPIS and HRIS datasets, the consistency was very close to that of observers and was the lowest among that reported by tracking methods. In challenging datasets containing the central light reflex and pathologies, namely CLRIS and VDIS, the precision of widths of the proposed method was better than that of many previous methods. Whereas methods sharing similar precision with the proposed method failed to return width estimation for some profiles, the proposed method was able to estimate widths for all vessel profiles given in the dataset [74, 88, 168].

Among surveyed studies, the best performance was presented by a supervised method [50]. However, there were some missing details about how they divided data for training and for evaluation. For example, how many samples were used for training or performance evaluation were not given. Because the REVIEW dataset is unbalanced regarding vessel widths, the selection of samples for training/evaluation seems an important task for a supervised method. According to a previous study using a supervised method for vessel width estimation [17], the discrepancy between the range of vessel widths in training and that in evaluation was shown to reduce the reliability of a trainable estimator on estimated widths. Also, an unbalanced training set, where the samples from a specific width range may dominate the training data, resulted in skewed estimation of widths. Given the large range of vessel widths in retina (from 2 pixels to 19 pixels for VDIS set [17]), and scarcity of large vessels in an image [17], a supervised method may require a large dataset to predict various vessel widths without adding extra bias for vessels with specific widths.

Despite having a trainable part (multi-labelling TDBN) in the proposed tracking system, this part does not require re-training to adapt to an unseen dataset. After being trained once, it can be used for the generation of segmentation probability maps for different datasets. Although the training of this network takes a while (the training time of the enhanced TDBN for STARE dataset was 5 hours), the segmentation time of an image is very short (see Table 4.11 in Chapter 4). As shown in the experiments, tracking vasculature on a dataset whose resolution and characteristics were different from that the network was trained with, did not degrade the performance of the tracking system on width estimation. To be more precise, the width estimation was completed successfully on such datasets which suffer from central light reflex and severe pathological lesions even though training

dataset (the DRIVE) does not contain these challenges at all or to a small extent. Similarly, the resolution difference between training dataset and that used for tracking did not make the tracking system less competitive among other methods.

In previous studies, two dominant approaches (the first two approaches in Figure 6.15) have been used for tracking regardless of the characteristics of the tracking system. As seen in Figure 6.15, the first approach inputs a grey scale image to a tracking system/a tracker, while the second approach sends a binary vessel map to the tracker. On the other hand, the third approach, which was proposed in this thesis, initially generates probability maps of related vessel parts for a given image then inputs the resulting probability maps to the tracker. Among these approaches, although the second approach appears to simplify tracking by giving the tracker a clue of vessel locations, this approach inherits errors emerged from segmentation such as missing vessels or false positive vessels, which can not be avoided during tracking due to losing related information as a result of thresholding. On the other hand, the third approach maintains most information from gray fundus images in a probabilistic way, which is somehow similar to the first approach regarding the richness of information for tracking. The most distinctive difference between the first and the third approach is the difficulty and quality of representing vessels in observation models required by the tracker. The latter approach simplifies the calculation of likelihood significantly by directly using probabilities from the maps in the calculation. Also, because these probability maps define several parts of a vessel concurrently, this approach may provide better probability distribution of vesselness than that generated from fundus images with model fitting to fundus images.

The success of the third approach is brought out with the usage of a probabilistic tracking system based on particle filtering. Particle filters were not previously used for width estimation on retinal images despite being used for tracking [105] (on the other hand, there are some studies predicting widths for $3D$ vasculature with particle filters [155]). This probabilistic approach allows the tracking system to use uncertainties in the estimation of vessel boundaries/width. As shown with experiments, the uncertainty obtained from vessel boundary/centreline/vesselness segmentation probability maps was beneficial to deal with challenges faced in retinal images such as a range of resolution of images or the presence of pathology and central light reflex without needing to adjust the parameters of tracking accordingly. Using probabilities provided by vessel interior, centreline and boundary vessel probability maps, instead of generating them by fitting a model to vessel cross-section intensities for likelihood calculation (used in the first approach) may provide less noisy information for tracking and the proposed approach can be used for other probabilistic tracking methods too. Also, low probabilities of small vessels, which may be missed due to segmentation (as may happen in the second approach), can still be used for tracking. This may open doors to analyse tiny and low contrast vessels in fundus photos.

Although the proposed method was observed to deal with challenges mentioned above, it has some limitations because it is mostly designed for images in the publicly available datasets. Firstly, the method is aimed to work on reasonably well captured images, which

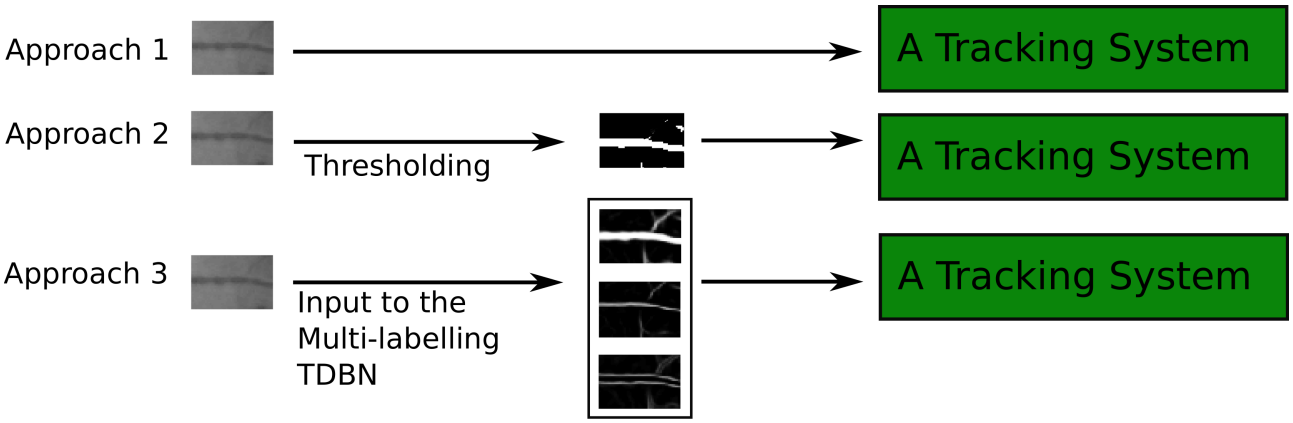


Figure 6.15: Approaches to vessel tracking regarding their observation spaces: the observation space is the image space in approach 1, the binary image in approach 2 and a set of probability maps in approach 3.

may exclude some images collected in screening programs, classified as ungradable. In particular, the images are assumed to be free of any blockage such as eye-lashes and, well-focused. Any pathologies in the images are assumed to be in a range that they should not intervene with the information related to centreline and edge locations in the probability maps to a large extent. Also, the design of the method has been affected by the ethnicities and age groups represented in the publicly available datasets, which were collected in Europe and from certain age groups. Therefore, the method considers the characteristics of fundus images related to these populations.

On the other hand, because the working mechanism of the method mostly rely on probability distributions and Bayesian calculation, I expect that the method can also work on large datasets after evaluating the responses of the method to potential problems with further analysis. To be more precise, the design of the method does not directly depend on image intensity distributions but the probability maps of some vessel parts such as vessel interior, centreline and edges, which suppress noise, pathologies and the central light reflex to a large extent. For the large datasets, which are captured by the same characteristics of the camera (e.g. resolution), the design parameters of the methods (e.g. the standard deviations of prior probability distributions) can be only adjusted in the beginning of the analysis. The probability maps of vessel interior, centreline and edges can be obtained with knowledge transfer, as performed in this study. However, in this stage, it is not possible to extrapolate the performance of the method to unseen datasets, especially which include poorly captured images, those with severe pathologies, and those whose some parts are blocked by a structure such as eye-lashes. The performance of the proposed method mostly relies on the accuracy and smoothness of probability maps generated by the multi-labelling TDBN. If the network fails to differentiate vasculature from other pathologies or noise and produces higher probabilities for them in probability maps, the performance of the particle filtering may significantly drop.

6.4.2 Vessel Cross-Section Modelling

The most informative part of a vessel in terms of estimating widths has been the intensity profiles of vessel cross-sections. Either methods fitted $1D$ or $2D$ models to describe the profiles or used the behaviour of edges [34, 72, 92, 93, 96–98]. These models were initially more simple, usually Gaussian models (twin Gaussian functions for modelling central light reflex). Recently, these models evolved to become more sophisticated to better describe vessel properties from intensity profiles [17,50]. These new models accepted the asymmetry of intensity profiles around vessel boundaries (which means one edge may have larger intensities than the other) [50] or the asymmetry around the centres of profiles (which means peak locations of intensity profiles may not be at the centres of the profiles) [17]. However, probabilistic methods did not catch up with this trend to date [74, 104] perhaps because it may be hard to model non-Gaussian probability distribution from intensity profiles.

On the other hand, the proposed method in the thesis takes probability distributions directly from probability maps generated by the network so it does not struggle with the modelling of vessel cross-sections. The proposed method calculates the likelihood of a hypothesis from three types of probability distribution over vessel cross-section intensity profiles: centreline, edge and vesselness probability distributions. These distributions are smooth, and specific to each intensity profile. Using multiple probability profiles may relate to an increase in the robustness of the method when compared with probabilistic methods using only vesselness profiles (cross-section intensity profiles), which was also shown by Zhang *et al.* [74]. As reported before, the proposed method successfully estimated all vessel widths in the REVIEW dataset while the method proposed by Zhang *et al.* [74] failed to estimate vessel widths for some vessel profiles (see Table 6.3).

Although probability profiles obtained from the probability maps are informative enough for the localisation of vessel boundaries, readers should be aware that the probability profiles are not uniform over a vessel map. For example, edge locations in edge probability profiles have the maximum values but a pair of edge locations may not have the same certainty value at both edge locations. Similarly, peaks in centreline probability profiles may not have the same height. For an observation model aiming to maximise the likelihood including centreline and edge probabilities, these factors may reduce the accuracy of estimations. In order to reduce the variation on likelihood due to changing probabilities at peak locations and to make the likelihood more related to the accuracies of estimated locations, the asymmetry of probability profiles can be considered. For now, 'the asymmetry of vessel interior/centreline probability profiles around vessel centreline' is implemented in particle filters, where two variables in state vectors are responsible for width estimation. Each variable predicts the distance between an estimate of centreline location and an estimate of boundary of interest independent of the other one. In the experiments conducted, an improvement in consistency regarding width estimations were observed. Particularly, the consideration of the asymmetry in tracking was found to be useful for images having

central light reflex or pathologies (CLRIS and VDIS), which may resemble more to images encountered in ophthalmology clinics. I anticipate that considering 'the asymmetry of vessel edge probability profiles on vessel boundaries' in tracking may also improve the performance of particle filtering in width estimation. However, this is left for future work.

A disadvantage of using particle filters in width estimation is that the estimates of vessel directions depend on previous estimates and the noise model, and may not align well with actual vessel directions. Although the method may still generate high likelihoods for hypotheses where hypothesised directions do not match with the actual vessel direction but correct estimates for boundary locations are made, the consistency of width estimation may not be provided due to measuring widths along shorter or longer cross-sections. However, in order to estimate widths consistently, the same angle between actual vessel directions and estimates of these directions (if a cross-section is defined as the shortest distance between vessel edges [88] for tracking, the angle tolerance drops to zero.) must be maintained. In order to improve the accuracy of estimated directions, the similarity of estimated directions to eigenvectors of Hessian matrices (corresponding to the mimimum eigenvalue) is taken into account in the likelihood calculation. Although this consideration may not appear as an improvement in consistency of estimated widths, regarding the results of the experiments, this is mostly due to reference boundary locations not being perpendicular to vessel directions [6] and not always being successful at locating vessel boundaries.

An advantage of using particle filters in vessel width estimation is that even if the visibility of vessel boundaries in images lessens due to pathologies such as beading (encountered in DR) or being in a bifurcation region, the tracker can still estimate vessel widths in a reasonable level due to using vessel widths estimated in the previous step. On the other hand, this may sometimes slow down the adaptation of the tracker to quick changes in vessel curvature unless it progresses with very small step size. Also, the slow adaptation to changes may lead to a slight reduction in the consistency of estimated widths at the beginning of the tracking, where estimates mostly depend on initial values given by a user because this initial information may be different from the vessel description which the tracking system is modelled for. For example, vessel boundaries given by the user may not be located close to peak locations in edge probability profiles. This may cause the tracker to take a while to find the correct course for itself. This situation also may occur when the tracker starts to trace a new branch whose properties may not be initially correctly estimated. If there are not sufficient number of hypothesis/particles (may depend on noise level in motion model) to put the tracker into an acceptable track, tracking may stop. This problem was demonstrated for the experiment in Section 6.3.5, where some daughter vessels (shown with the 2^{nd}-3^{th} circles in Figure 6.14c and 4^{th} circle in Figure 6.14d) were missed.

Chapter 7

Discussion and Conclusion

7.1 Introduction

Monitoring the retina is easy and less costly when compared to other part of internal anatomy. Changes on retinal vasculature morphology, particularly those on vessel diameter, have been associated with the incidence or progression of eye-related diseases, such as diabetic retinopathy, and systematic/vasculature diseases, for example diabetes, hypertension and stroke, in many studies [11, 21, 24–26, 28]. The analysis of retinal vasculature in fundus photos may be realised in two ways: the presence of vasculature (segmentation) and the quantitative analysis of vasculature (e.g. the estimation of vessel diameters, branching angles or tortuosity).

Although many studies have been devoted to the segmentation of fundus images [33, 39, 40] they were mostly validated on low resolution and healthy fundus images. However, the segmentation of pathological images or high resolution images is still an open area for research [33]. In this thesis, the segmentation of both low and high resolution images with the same method is targeted. For this aim, two applications of an enhanced TDBN are presented on the segmentation of low resolution images and high resolution images, where the network is named super-resolution TDBN. Furthermore, the quantitative analysis of vasculature is performed to obtain more quantitative information about retinal vasculature morphology. Another application of the enhanced TDBN, which is called multi-labelling TDBN, generates probability maps for various parts of vasculature: vessel interior pixels, centreline, boundary and landmark location patterns on vessel centreline probability maps. These maps become observation models for a probabilistic tracking system. This tracking system is used for vessel width estimation and exploring the connectivity of vasculature network. The former analysis is validated on datasets from a range of image resolution and in the presence of any pathologies or artefacts.

In this chapter, firstly, the contribution of this thesis to the literature will be summarised, which will be followed by the limitation of the proposed methods. Finally, possible directions for future work will be given.

7.2 Contributions

The contributions of this thesis to the retinal vasculature analysis literature will be examined in two categories: vasculature segmentation and quantitative vasculature analysis.

7.2.1 Vasculature Segmentation

In vessel segmentation, recently, deep learning approaches have presented more successful results than other approaches [33]. This leads me to use a deep network for vasculature segmentation. Although the general tendency in vasculature segmentation is the classification of each pixel in a fundus image as a vessel pixel and a non-vessel pixel, the consideration of spatial connectivity of label pixels in the output of the network has been observed to increase segmentation performance [36]. In this thesis, the spatial connectivity of label pixels is considered by generating vessel masks for given input image patches as a result of complying with the cross-modality learning approach [43]. In order to improve spatial connectivity of label pixels further, the training of the TDBN [1] is improved with a denoising, which was found to result in more flat probability responses to vasculature and to image background. Although this effect may not be considered as an improvement regarding segmentation, where probabilities over a threshold would be labelled with the same class, the same effect can be valuable for the robustness of the proposed tracking method.

The enhanced TDBN is found to be versatile in terms of applicability to different tasks by slightly modifying its architecture and without significantly changing the way of training the network. When it is used as it is, the network performed comparable or better on the segmentation of low resolution images, when compared with similar methods. Later, the TDBN is converted to a segmentation tool for high resolution images, where deep learning methods have been rarely applied, by combining super-resolution with segmentation. Although the main expectation from the design of this network (super-resolution TDBN) is to reduce segmentation time for high resolution images, its performance is also found to be better than that of a similar cross-modality learning based segmentation network [43] on the same dataset (CHASE_DB1) and better than or comparable to that of previous studies. Finally, multi-labelling is realised with the TDBN to label vessel interior, centreline, boundary pixels and landmark location patterns on vessel centreline. The output of the network is directly used to guide a probabilistic tracker, without thresholding it. This strategy is found to be beneficial in the identification of thin vessels and dealing with pathologies/noises.

7.2.2 Quantitative Vasculature Analysis

In this thesis, a tracking system is proposed to trace retinal vasculature in fundus images. This system has two modes (i) to track vessels and (ii) to identify bifurcation/crossing locations and to suggest initial estimates of tracking states for new branches. The first

mode uses particle filters for vessel segment tracking and the second mode contains a new method for landmark location detection and another one for the identification of new branches, which is based on a clustering algorithm, Mean Shift Clustering.

Both modes fetch information from probability maps generated by the multi-labelling TDBN. The first mode uses vesselness, centreline and edge probability maps and the second one landmark location and centreline probability maps. To the best of my knowledge, this is the first time when probability maps for segmentation are used as observation models for a probabilistic tracking systems in retinal image analysis. The usage of probability maps, instead of binary vesselness maps, were observed to easily identify thin vessels which might disappear in binary vessel maps due to the their vesselness probabilities being lower than a threshold.

Although particle filters have ability to trace multiple branches and some studies used them for this purpose [141, 160], particle filters in this study are used for tracking single branches. On the contrary to studies tracing multiple branches simultaneously with particle filters, particle depletion (which may be observed in the tracking of thin branches originating from larger ones and may be result in early stopping for thin branches) was not observed during tracking.

Although particle filters were used for tracking retinal vessels in [68, 105], their performance on the estimations of retinal vessel characteristics has not been reported, to the best of my knowledge. Although particle filters are powerful estimators in terms of modelling non-Gaussian and non-linear systems, they may not produce good estimates of vessel properties such as vessel width and orientation unless they are devised with good observation models. In this thesis, four observation models are presented to consider probabilities related to different vessel parts/properties. These models are common in aiming for the simultaneous maximisation of vessel centre and boundary probabilities for the accurate estimation of vessel centrelines and widths. A superiority of this method to previous studies is that proposed observation models have an access to a good representation of vessel boundary probability profiles; on the contrary to previous methods which obtained boundary information from image intensity gradients or from model fitting to vessel cross-section intensity distributions [47, 74, 88, 104].

Apart from other probabilistic tracking methods using Gaussian vessel intensity cross-section models obtained by model fitting to fundus images [74, 104], the likelihood of a hypothesis explaining vessel parameters is calculated from the profiles obtained from probability maps generated by a segmentation network, the multi-labelling TDBN. Aside from fundus images, these probability maps filter out noise and pathology in fundus images to some extent, which may make these maps better candidates for likelihood calculation. Also, probabilities sampled from these maps can be directly used in likelihood calculation without requiring any further processing such as model fitting. The results of the experiments on images with central light reflex and pathologies also suggest that the tracker guided by these probability maps may have more accurate and consistent estimation of vessel boundaries so widths than that of human observers and that of other methods (see

Table 6.3 and Figure 6.13 in Chapter 6).

Moreover, the assumption of vessel cross-section to be a Gaussian function is replaced with that of vessel boundaries to be asymmetric around vessel centre in order to obtain better estimation of vessel widths. This assumption is found to improve the detection of vessel boundaries when supported with the usage of centreline probability maps in the place of vesselness probability maps, and the similarity of estimated directions by the tracker, to directions calculated from eigenvectors. This improvement is observed, particularly, in images with central light reflex. On the other hand, the slight reduction in the performance on other image sets is found to be related to the limits of human beings in the identification of vessel boundaries in noisy images. This also may reinforce the superiority of the proposed method over supervised methods, which may not be aware of human mistakes in ground truth data and may learn how to estimate widths from these mistakes.

7.3 Limitations

This section will summarise the limitations of the proposed methods under two groups: vasculature segmentation and quantitative vasculature analysis.

7.3.1 Vasculature Segmentation

The limitations of the proposed method on vasculature segmentation appear as those due to data and those related to the architecture of the networks. These limitations will be explained as follows.

Although deep learning methods have been shown to be successful in image segmentation, their performance has been mostly related to availability of large amount of labelled data. Regarding TDBN and its derivatives introduced here, the segmentation method has the same limitation. When encountered with a new dataset, retraining of any networks requires labelled data for training. If labelled data is not available and it is necessary to use a network trained with another dataset for the segmentation of the new dataset, the performance of the network may not reach the same level if the new image set have different resolution or pathologies or noise types than the dataset the networks trained with. This emphasises the necessity for large labelled datasets for segmentation with deep learning methods, which should reflect the variety of images encountered in ophthalmology clinics regarding image resolution, pathologies and different noise levels.

Similarly, lacking of labelled data, binary vessel maps, was encountered in the generation of probability maps for the REVIEW dataset. Although this dataset has vessel boundary locations for some vessel segments, it does not contain binary vessel maps. In this thesis, the network trained with the DRIVE dataset was used for the segmentation of the REVIEW dataset without retraining. Although the proposed tracking system can tolerate vague information in probability maps, occurred as a result of training and test

datasets to be from different image sets, to some extent, using probability maps generated with a network trained with the REVIEW may further improve the performance of the tracker on vessel width estimation, which can also allow a performance evaluation of the method on the REVIEW.

Another limitation about data is related to limited capacity of human being to accurately determine vessel pixels in fundus images. As reported for publicly available datasets used in this thesis, some observers may ignore some thin vessels or become more vigilant for vessel boundary pixels for vessel segmentation. Using only one observer's labels as ground truth in performance comparison may not be an objective way to do that, which may favour some methods closer to the observer's bias. In addition to bias, ground truth data may include misinformation, which may not appear explicitly without a rigorous examination. For instance, the proposed tracking system was found to be the least successful in VDIS dataset among other datasets. When possible reasons for this result were investigated, it became obvious that discrepancy between some ground truth data and corresponding estimates of boundary locations was due to misidentified boundary pixels by the observers; not due to inconsistent estimates by the proposed method.

Apart from data, the suggested approach for segmentation has its own limitations. Although the TDBNs have large number of units in the output layer may facilitate to maintain spatial interaction between label pixels, this naturally large output layer may limit the applicability of this network to high resolution images in the case of training a multi-labelling network. Even though the best option for segmentation of a high resolution image is to use a super-resolution TDBN, when the network is used for multi-labelling, already very large output layer of the super-resolution TDBN may be required to be quadrupled for the generation of vessel interior, centreline, edge and landmark location probability maps. It may be hard or not possible to train such a network.

In addition, the cross-modality learning approach may not be appropriate for the segmentation of either vessel centreline, boundary or landmark locations, if this is the only purpose to train the network. The network may respond weakly to these locations or may generate very noisy output because the relation between fundus images and these vessel parts is more complicated than that between fundus images and vessel interior. In this case, a different loss function may be a better choice for training rather than L_2 loss such as weighted cross-entropy loss.

7.3.2 Quantitative Vasculature Analysis

The limitations related to quantitative vasculature analysis are mostly related to the imperfections of probability maps generated by the multi-labelling TDBN and the natural limits of tracking methods.

Although edge locations have larger probabilities than other regions in edge probability profiles, the values at the local maxima for a pair of edge locations may not be the same. This may attract the tracker (the tracking system) towards edges with higher edge

probabilities. Although vessel width may be estimated consistently due to tracking on relatively similar probabilities for each edge over consecutive iterations, in other words, in the same ratio of the distance between vessel boundaries, the tracker may not predict actual vessel boundaries. Also, using upscaled probability maps for tracking may slightly reduce the performance of the tracker by preventing it from predicting accurate sub-pixel locations because the maxima in the probability maps may be close to but not at actual locations due to generating these maps on lower resolutions.

Although tracking vessel segments for given initial centreline locations and widths was explored in detail in this thesis, the same attention was not able to be be paid to the tracking of vessel tree fractions and it is left for future work. In particular, the switch from *mode 1* to *mode 2* during tracking was crude because of using a binary decision depending on a fixed threshold level. Although the threshold was selected after observing the landmark location probability maps of the DRIVE dataset, this manual selection inserted subjectivity to the tracking system and the performance of the tracking system on landmark detection may be affected from this threshold to some extend. In other words, if this threshold is set higher than any peak of landmark probability distribution, a landmark location may be not recognised, which in turn may cause missing branches at this landmark location. Similarly, selecting a low threshold may lead to false detection of landmark locations; though, false estimates may be identifiable after completing tracking then can be removed. A more sophisticated switching between these modes may enhance the performance of the network on the tracking of vessel tree fractions.

The performance of tracking systems on the current iteration may depend on their performance on previous iterations. Although a tracker may recover from a misestimated vessel description occurring due to pathology or noise, the recovery may take a few iterations. As being a tracking method, the proposed method also has the same weakness. This situation may be observed not only in the presence of pathology and noise but also after initialising the tracking of a vessel. This is mostly due to mismatch between initial and actual vessel parameters. Due to relying on initial parameters in the beginning of tracking, the proposed tracking system may not find the right course to track if only one of initial edge locations does not overlap with the locations of high probabilities on edge probability profiles. This is because of the assumption of the maximisation of both probabilities of edge locations and the tracker being inside vessel.

Wrong initialisation of vessel parameters may happen due to human factor in the beginning of tracking or due to the tracking system to not correctly estimate daughter vessel parameters over the course of tracking. For instance, human observers may not correctly locate vessel boundaries when images are noisy and vessel contrast is very low, which may lead wrong initialisation of vessel parameters. Another example is related to large changes on vessel parameters, which is beyond the parameter search region of the tracking system. New branches may change their directions very quickly after leaving landmark locations so the initial estimate of direction may not be valid for the next tracking step. Moreover, the parameters of a previously traced main vessel may not be a good indicator of daughter

vessel parameters, which are suggested for initialising daughter vessel parameters. If the difference between a previously estimated width of a main vessel close to a landmark location and a currently estimated width of a daughter branch emerging from this landmark location is not affordable with the noise model, the noise model may fail to model the parameters of the daughter vessel correctly.

7.4 Future Work

Although the performance of the proposed methods (the networks and the tracking system) were assessed over various datasets (DRIVE, STARE, CHASE_DB1, HRF and REVIEW), these datasets have limited presentation ability of clinical data. Many images encountered in ophthalmology clinics may be in far worse conditions regarding noise, artefacts and the severity/prevalence of pathologies than those selected for the performance evaluation of segmentation methods. Without evaluating the limits of these methods over severe conditions such as severe diabetes, high noise levels or other artefacts, the real impact of these methods to the literature may not be identified efficiently.

Similarly, the parameters of many segmentation methods depend on image resolution. In the case of deep learning being used as a segmentation tool, segmentation methods may require not only re-training of the networks but also the adaptation of the size of receptive field according to the resolution. In the case of cross-modality learning, the latter also means the modification of output layer size. Regarding the variety of image cameras used for screening programs or in clinics, the immediate use of a segmentation method may be limited. A follow up study may focus on this issue by investigating the range of resolution, which a segmentation method can tolerate without much reduction in its performance, or may search for ways to make the network compatible with new resolution.

On the other hand, the tracking system was found to be more flexible in dealing with resolution difference. Similarly, the limits of this system would be further examined over a large dataset consists of various qualities of images. Aside from segmentation, the performance of the tracker on vessel width estimation was found to be associated with the reliability of ground truth data. Because the detection of vessel boundaries is not easy for a human being and is usually related to how an observer perceives vessel boundaries, ground truth data collected from human observers may not be completely reliable for performance evaluation. Therefore, simulated data where actual vessel boundaries, imaging conditions and noise properties are known, may be a better choice for performance evaluation.

In this thesis, the multi-labelling TDBN was used for the generation of probability maps required for tracking. It may be interesting to examine the compatibility of the tracking system to probability maps generated by other methods such as random forests, which may accelerate the adaptability of the system to different image sets. Another direction for future work may be the integration of online processing of image patches for probability map generation in the tracking system. Therefore, probability maps required for the calculation of likelihood may be calculated for only related pixel groups. Such an

addition to the proposed tracking system may reduce the segmentation time spent by the networks for whole fundus images, where pixels representing vasculature are much lower than non-vessel pixels.

A potential problem for the proposed tracking system may arise due to different maxima values for a pair of edge locations. The tracker may be attracted by stronger boundaries and the other edge may be traced inside vessel but not at actual edge locations, which means only a proportion of actual widths may be returned. In addition to suggestions given above, observation models may be modified to consider asymmetry of vessel probabilities at boundaries in order to tackle this problem.

In this thesis, the tracker was mostly evaluated on width estimation of vessel segments, whose initial and final boundary locations were known. It may be useful to observe the behaviour of the tracker on tracing vessel trees in fundus images by starting from seed locations around the optic discs for a large dataset of fundus images. This analysis may be followed by the performance analysis of the tracker on estimation of landmark locations. Also, crossing or bifurcation locations may need to be classified accurately in order to provide vessel connectivity maps. Ultimately, the tracking system may be used for estimation of other vasculature descriptors such as branching angles, tortuosity or length to diameter ratio.

Bibliography

[1] I. Fasel and J. Berry, "Deep belief networks for real-time extraction of tongue contours from ultrasound during speech," in *Proc. International Conference on Pattern Recognition (ICPR)*. IEEE, 2010, pp. 1493–1496.

[2] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, 2004.

[3] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Transactions on Medical imaging*, vol. 19, no. 3, pp. 203–210, 2000.

[4] J. Odstrcilik, R. Kolar, A. Budai, J. Hornegger, J. Jan, J. Gazarek, T. Kubena, P. Cernosek, O. Svoboda, and E. Angelopoulou, "Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database," *IET Image Processing*, vol. 7, no. 4, pp. 373–383, 2013.

[5] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. Barman, "An ensemble classification-based approach applied to retinal blood vessel segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 9, pp. 2538–2548, 2012.

[6] B. Al-Diri, A. Hunter, D. Steel, M. Habib, T. Hudaib, and S. Berry, "Review-a reference data set for retinal vessel profiles," in *Proc. the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2008, pp. 2262–2265.

[7] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, pp. 3371–3408, 2010.

[8] A. Doucet, N. de Freitas, and N. Gordon, "An introduction to Sequential Monte Carlo methods," in *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds. Springer New York, 2001, pp. 3–14.

[9] T. MacGillivray, E. Trucco, J. Cameron, B. Dhillon, J. Houston, and E. Van Beek, "Retinal imaging as a source of biomarkers for diagnosis, characterization and prognosis of chronic illness or long-term conditions," *The British Journal of Radiology*, vol. 87, no. 1040, p. 20130832, 2014.

[10] H. Kolb. (2005 May 1 [Updated 2012 Jan 31]) Simple anatomy of the retina. [Online]. Available: http://europepmc.org/books/NBK11533;jsessionid=B79738A925EB1BC08E45C6B72F19DE8E

[11] E. M. Boerman, J. E. Everhart, and S. S. Segal, "Small vessels–big problems: Novel insights into microvascular mechanisms of diseases: Advanced age decreases local calcium signaling in endothelium of mouse mesenteric arteries in vivo," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 310, no. 9, p. H1091, 2016.

[12] P. H. Scanlon, "The English national screening programme for diabetic retinopathy 2003–2016," *Acta Diabetologica*, vol. 54, no. 6, pp. 515–525, 2017.

[13] M. A. Pérez, B. B. Bruce, N. J. Newman, and V. Biousse, "The use of retinal photography in non-ophthalmic settings and its potential for neurology," *The Neurologist*, vol. 18, no. 6, p. 350, 2012.

[14] R. Besenczi, J. Tóth, and A. Hajdu, "A review on automatic analysis techniques for color fundus photographs," *Computational and Structural Biotechnology Journal*, vol. 14, pp. 371–384, 2016.

[15] N. Panwar, P. Huang, J. Lee, P. A. Keane, T. S. Chuan, A. Richhariya, S. Teoh, T. H. Lim, and R. Agrawal, "Fundus photography in the 21st century–a review of recent technological advances and their implications for worldwide healthcare," *Telemedicine and e-Health*, vol. 22, no. 3, pp. 198–208, 2016.

[16] E. Trucco, A. Ruggeri, T. Karnowski, L. Giancardo, E. Chaum, J. P. Hubschman, B. Al-Diri, C. Y. Cheung, D. Wong, M. Abramoff *et al.*, "Validating retinal fundus image analysis algorithms: Issues and a proposal validating retinal fundus image analysis algorithms," *Investigative Ophthalmology & Visual Science*, vol. 54, no. 5, pp. 3546–3559, 2013.

[17] C. A. Lupaşcu, D. Tegolo, and E. Trucco, "Accurate estimation of retinal vessel width using bagged decision trees and an extended multiresolution hermite model," *Medical Image Analysis*, vol. 17, no. 8, pp. 1164–1180, 2013.

[18] Causes of blindness and visual impairment. WHO. Access time: 25 May 2017. [Online]. Available: http://www.who.int/blindness/causes/en/

[19] M. D. Abràmoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Reviews in Biomedical Engineering*, vol. 3, pp. 169–208, 2010.

[20] G. Liew, M. Michaelides, and C. Bunce, "A comparison of the causes of blindness certifications in England and Wales in working age adults (16–64 years), 1999–2000 with 2009–2010," *British Medical Journal (BMJ) Open*, vol. 4, no. 2, 2014.

[21] C. Wilkinson, F. L. Ferris, R. E. Klein, P. P. Lee, C. D. Agardh, M. Davis, D. Dills, A. Kampik, R. Pararajasegaram, and J. T. Verdaguer, "Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales," *Ophthalmology*, vol. 110, no. 9, pp. 1677–1682, 2003.

[22] R. Agrawal, N. Joachim, L.-J. Li, J. Lee, A. Agarwal, D. A. Sim, P. A. Keane, G. Liew, and C. E. Pavesio, "Assessment of retinal vascular calibres as a biomarker of disease activity in birdshot chorioretinopathy," *Acta Ophthalmologica*, vol. 95, no. 2, 2017.

[23] E. W. Steyerberg, M. J. Pencina, H. F. Lingsma, M. W. Kattan, A. J. Vickers, and B. Van Calster, "Assessing the incremental value of diagnostic and prognostic markers: a review and illustration," *European Journal of Clinical Investigation*, vol. 42, no. 2, pp. 216–228, 2012.

[24] M. K. Ikram, C. Y. Cheung, M. Lorenzi, R. Klein, T. L. Jones, and T. Y. Wong, "Retinal vascular caliber as a biomarker for diabetes microvascular complications," *Diabetes Care*, vol. 36, no. 3, pp. 750–759, 2013.

[25] V. Daien, I. Carriere, R. Kawasaki, J.-P. Cristol, M. Villain, P. Fesler, K. Ritchie, and C. Delcourt, "Retinal vascular caliber is associated with cardiovascular biomarkers of oxidative stress and inflammation: the pola study," *PLOS ONE*, vol. 8, no. 7, p. e71089, 2013.

[26] J. Ding, K. L. Wai, and K. McGeechan, "Retinal vascular caliber and the development of hypertension: a meta-analysis of individual participant data," *Journal of Hypertension*, vol. 32, no. 2, p. 207, 2014.

[27] B. Gopinath, J. Chiha, A. J. Plant, A. Thiagalingam, G. Burlutsky, P. Kovoor, G. Liew, and P. Mitchell, "Associations between retinal microvascular structure and the severity and extent of coronary artery disease," *Atherosclerosis*, vol. 236, no. 1, pp. 25–30, 2014.

[28] A. McGowan, G. Silvestri, E. Moore, V. Silvestri, C. C. Patterson, A. P. Maxwell, and G. J. McKay, "Evaluation of the retinal vasculature in hypertension and chronic kidney disease in an elderly population of irish nuns," *PLOS ONE*, vol. 10, no. 9, p. e0136434, 2015.

[29] F. Berisha, G. T. Feke, C. L. Trempe, J. W. McMeel, and C. L. Schepens, "Retinal abnormalities in early Alzheimer's disease," *Investigative Ophthalmology & Visual Science*, vol. 48, no. 5, pp. 2285–2289, 2007.

[30] C. Y.-l. Cheung, Y. T. Ong, M. K. Ikram, S. Y. Ong, X. Li, S. Hilal, J.-A. S. Catindig, N. Venketasubramanian, P. Yap, and D. Seow, "Microvascular network alterations in the retina of patients with Alzheimer's disease," *Alzheimer's & Dementia*, vol. 10, no. 2, pp. 135–142, 2014.

[31] H.-Q. Wu, H. Wu, L.-L. Shi, L.-Y. Yu, L.-Y. Wang, Y.-L. Chen, J.-S. Geng, J. Shi, K. Jiang, and J.-C. Dong, "The association between retinal vasculature changes and stroke: a literature review and meta-analysis," *International Journal of Ophthalmology*, vol. 10, no. 1, p. 109, 2017.

[32] J. Kaur and D. Mittal, "A generalized method for the detection of vascular structure in pathological retinal images," *Biocybernetics and Biomedical Engineering*, vol. 37, no. 1, pp. 184–200, 2017.

[33] C. L. Srinidhi, P. Aparna, and J. Rajan, "Recent advancements in retinal vessel segmentation," *Journal of Medical Systems*, vol. 41, no. 4, p. 70, 2017.

[34] J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, and R. L. Kennedy, "Measurement of retinal vessel widths from fundus images based on 2-D modeling," *IEEE Transactions on Medical Imaging*, vol. 23, no. 10, pp. 1196–1204, 2004.

[35] C. Y. Cheung, M. K. Ikram, R. Klein, and T. Y. Wong, "The clinical implications of recent studies on the structure and function of the retinal microvasculature in diabetes," *Diabetologia*, vol. 58, no. 5, pp. 871–885, 2015.

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

[37] A. Tufail, V. V. Kapetanakis, S. Salas-Vega, C. Egan, C. Rudisill, C. G. Owen, A. Lee, V. Louw, J. Anderson, and G. Liew, "An observational study to assess if automated diabetic retinopathy image assessment software can replace one or more steps of manual imaging grading and to determine their cost-effectiveness," *Health Technology Assessment*, vol. 20, no. 92, pp. 1–72, 2016.

[38] M. Sofka and C. V. Stewart, "Retinal vessel centerline extraction using multiscale matched filters, confidence and edge measures," *IEEE Transactions on Medical Imaging*, vol. 25, no. 12, pp. 1531–1546, 2006.

[39] C. Kirbas and F. K. Quek, "Vessel extraction techniques and algorithms: a survey," in *Proc. Bioinformatics and Bioengineering (BIBE)*. IEEE, 2003, pp. 238–245.

[40] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "Blood vessel segmentation methodologies in retinal images–a survey," *Computer Methods and Programs in Biomedicine*, vol. 108, no. 1, pp. 407–433, 2012.

[41] G. Azzopardi and N. Petkov, "Detection of retinal vascular bifurcations by trainable V4-like filters," in *Computer Analysis of Images and Patterns*. Springer, 2011, pp. 451–459.

[42] D. Gadriye, G. Khandale, and R. Nawkhare, "System for diagnosis of diabetic retinopathy using neural network," *International Journal of Recent Technology and Engineering*, vol. 3, pp. 81–86, 2014.

[43] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, and T. Wang, "A cross-modality learning approach for vessel segmentation in retinal images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 1, pp. 109–118, 2015.

[44] S. Wang, Y. Yin, G. Cao, B. Wei, Y. Zheng, and G. Yang, "Hierarchical retinal blood vessel segmentation based on feature and ensemble learning," *Neurocomputing*, vol. 149, pp. 708–717, 2015.

[45] E. Cheng, L. Du, Y. Wu, Y. J. Zhu, V. Megalooikonomou, and H. Ling, "Discriminative vessel segmentation in retinal images by fusing context-aware hybrid features," *Machine Vision and Applications*, vol. 25, no. 7, pp. 1779–1792, 2014.

[46] A. Hunter, J. Lowell, R. Ryder, A. Basu, and D. Steel, "Tram-line filtering for retinal vessel segmentation." Prague, Czech Republic: Proc. European Medical and Biological Engineering Conference (EMBEC), 20-24 November 2005.

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

[48] A. Bhuiyan, B. Nath, J. Chua, and K. Ramamohanarao, "Automatic detection of vascular bifurcations and crossovers from color retinal fundus images," in *Proc. Signal-Image Technologies and Internet-Based System (SITIS)*. IEEE, 2007, pp. 711–718.

[49] Q. Li, J. You, and D. Zhang, "Vessel segmentation and width estimation in retinal images using multiscale production of matched filter responses," *Expert Systems with Applications*, vol. 39, no. 9, pp. 7600–7610, 2012.

[50] T. Araújoa, A. M. Mendonçaa, and A. Campilhoa, "Estimation of retinal vessel caliber using model fitting and random forests," in *Proc. SPIE Medical Imaging*, 2017, pp. 101 341K–101 341K.

[51] U. T. Nguyen, A. Bhuiyan, L. A. Park, and K. Ramamohanarao, "An effective retinal blood vessel segmentation method using multi-scale line detection," *Pattern Recognition*, vol. 46, no. 3, pp. 703–715, 2013.

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

[53] D. Marín, A. Aquino, M. E. Gegúndez-Arias, and J. M. Bravo, "A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features," *IEEE Transactions on Medical Imaging*, vol. 30, no. 1, pp. 146–158, 2011.

[54] A. Lahiri, A. G. Roy, D. Sheet, and P. K. Biswas, "Deep neural ensemble for retinal vessel segmentation in fundus images towards achieving label-free angiography," in *Proc. the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2016, pp. 1340–1343.

[55] D. Maji, A. Santara, S. Ghosh, D. Sheet, and P. Mitra, "Deep neural network and random forest hybrid architecture for learning to detect retinal vessels in fundus images," in *Proc. the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 3029–3032.

[56] D. Maji, A. Santara, P. Mitra, and D. Sheet, "Ensemble of deep convolutional neural networks for learning to detect retinal vessels in fundus images," *CoRR*, vol. abs/1603.04833, 2016, preprint. [Online]. Available: http://arxiv.org/abs/1603.04833

[57] H. Fu, Y. Xu, S. Lin, D. W. K. Wong, and J. Liu, "DeepVessel: Retinal vessel segmentation via deep learning and conditional random field," in *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2016, pp. 132–139.

[58] J. H. Tan, U. R. Acharya, S. V. Bhandary, K. C. Chua, and S. Sivaprasad, "Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network," *Journal of Computational Science*, vol. 20, pp. 70–79, 2017.

[59] B. D. Barkana, I. Saricicek, and B. Yildirim, "Performance analysis of descriptive statistical features in retinal vessel segmentation via fuzzy logic, ANN, SVM, and classifier fusion," *Knowledge-Based Systems*, vol. 118, pp. 165–176, 2017.

[60] P. Arbeláez and L. Van Gool, "Deep retinal image understanding," in *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 9901. Springer, 2016, p. 140.

[61] J. I. Orlando and M. Blaschko, "Learning fully-connected CRFs for blood vessel segmentation in retinal images," in *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2014, pp. 634–641.

[62] W. S. Oliveira, J. V. Teixeira, T. I. Ren, G. D. Cavalcanti, and J. Sijbers, "Unsupervised retinal vessel segmentation using combined filters," *PLOS ONE*, vol. 11, no. 2, p. e0149943, 2016.

[63] J. Zhang, B. Dashtbozorg, E. Bekkers, J. P. Pluim, R. Duits, and B. M. ter Haar Romeny, "Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores," *IEEE Transactions on Medical Imaging*, vol. 35, no. 12, pp. 2631–2644, 2016.

[64] Y. Bengio, "Learning deep architectures for AI," *Foundations and Trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[65] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014, preprint. [Online]. Available: http://arxiv.org/abs/1409.1556

[66] Q. Li, L. Xie, Q. Zhang, S. Qi, P. Liang, H. Zhang, and T. Wang, "A supervised method using convolutional neural networks for retinal vessel delineation," in *Proc. Congress on Image and Signal Processing (CISP)*, Oct 2015, pp. 418–422.

[67] Y. Ganin and V. Lempitsky, "Nˆ 4-fields: Neural network nearest neighbor fields for image transforms," in *Proc. Asian Conference on Computer Vision*. Springer, 2014, pp. 536–551.

[68] A. Wu, Z. Xu, M. Gao, M. Buty, and D. J. Mollura, "Deep vessel tracking: A generalized probabilistic approach via deep learning," in *Proc. International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2016, pp. 1363–1367.

[69] D. Crisan and K. Li, "Generalised particle filters with Gaussian mixtures," *Stochastic Processes and Their Applications*, vol. 125, no. 7, pp. 2643–2673, 2015.

[70] C. Sutton and A. McCallum, "An introduction to conditional random fields," *Foundations and Trends® in Machine Learning*, vol. 4, no. 4, pp. 267–373, 2012.

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

[72] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters," *IEEE Transactions on Medical imaging*, vol. 8, no. 3, pp. 263–269, 1989.

[73] S. Konishi, A. L. Yuille, J. M. Coughlan, and S. C. Zhu, "Statistical edge detection: Learning and evaluating edge cues," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 1, pp. 57–74, 2003.

[74] J. Zhang, H. Li, Q. Nie, and L. Cheng, "A retinal vessel boundary tracking method based on Bayesian theory and multi-scale line detection," *Computerized Medical Imaging and Graphics*, vol. 38, no. 6, pp. 517–525, 2014.

[75] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 1998, pp. 130–137.

[76] E. Bekkers, R. Duits, T. Berendschot, and B. ter Haar Romeny, "A multi-orientation analysis approach to retinal vessel tracking," *Journal of Mathematical Imaging and Vision*, vol. 49, no. 3, pp. 583–610, 2014.

[77] D. E. Goldberg and J. H. Holland, "Genetic algorithms and machine learning," *Machine Learning*, vol. 3, no. 2, pp. 95–99, 1988.

[78] A. Budai, R. Bock, A. Maier, J. Hornegger, and G. Michelson, "Robust vessel segmentation in fundus images," *International Journal of Biomedical Imaging*, vol. 2013, 2013.

[79] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi, "Iterative vessel segmentation of fundus images," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 7, pp. 1738–1749, 2015.

[80] L. C. Neto, G. L. Ramalho, J. F. R. Neto, R. M. Veras, and F. N. Medeiros, "An unsupervised coarse-to-fine algorithm for blood vessel segmentation in fundus images," *Expert Systems with Applications*, vol. 78, pp. 182–192, 2017.

[81] R. Annunziata, A. Garzelli, L. Ballerini, A. Mecocci, and E. Trucco, "Leveraging multiscale Hessian-based enhancement with a novel exudate inpainting technique for retinal vessel segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 4, pp. 1129–1138, 2016.

[82] E. Imani, M. Javidi, and H.-R. Pourreza, "Improvement of retinal blood vessel detection using morphological component analysis," *Computer Methods and Programs in Biomedicine*, vol. 118, no. 3, pp. 263–279, 2015.

[83] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 6, pp. 679–698, 1986.

[84] J. Kittler and J. Illingworth, "Minimum error thresholding," *Pattern Recognition*, vol. 19, no. 1, pp. 41–47, 1986.

[85] M. E. Martinez-Perez, A. D. Hughes, S. A. Thom, A. A. Bharath, and K. H. Parker, "Segmentation of blood vessels from red-free and fluorescein retinal images," *Medical Image Analysis*, vol. 11, no. 1, pp. 47–61, 2007.

[86] Y. Q. Zhao, X. H. Wang, X. F. Wang, and F. Y. Shih, "Retinal vessels segmentation based on level set and region growing," *Pattern Recognition*, vol. 47, no. 7, pp. 2437–2446, 2014.

[87] C. Li, C.-Y. Kao, J. C. Gore, and Z. Ding, "Minimization of region-scalable fitting energy for image segmentation," *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1940–1949, 2008.

[88] B. Al-Diri, A. Hunter, and D. Steel, "An active contour model for segmenting and measuring retinal vessels," *IEEE Transactions on Medical imaging*, vol. 28, no. 9, pp. 1488–1497, 2009.

[89] M. Barchiesi, S. H. Kang, T. M. Le, M. Morini, and M. Ponsiglione, "A variational model for infinite perimeter segmentations based on Lipschitz level set functions: denoising while keeping finely oscillatory boundaries," *Multiscale Modeling & Simulation*, vol. 8, no. 5, pp. 1715–1741, 2010.

[90] Y. Zhao, L. Rada, K. Chen, S. P. Harding, and Y. Zheng, "Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images," *IEEE Transactions on Medical Imaging*, vol. 34, no. 9, pp. 1797–1807, 2015.

[91] G. Läthén, J. Jonasson, and M. Borga, "Blood vessel segmentation using multi-scale quadrature filtering," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 762–767, 2010.

[92] N. Chapman, N. Witt, X. Gao, A. Bharath, A. Stanton, S. Thom, and A. Hughes, "Computer algorithms for the automated measurement of retinal arteriolar diameters," *British Journal of Ophthalmology*, vol. 85, no. 1, pp. 74–79, 2001.

[93] O. Brinchmann-Hansen and H. Heier, "Theoretical relations between light streak characteristics and optical properties of retinal vessels," *Acta Ophthalmologica*, vol. 64, no. S179, pp. 33–37, 1986.

[94] O. Brinchmann-Hansen and O. Engvold, "Microphotometry of the blood column and the light streak on retinal vessels in fundus photographs," *Acta Ophthalmologica*, vol. 64, no. S179, pp. 9–19, 1986.

[95] S. Rassam, V. Patel, O. Brinchmann-Hansen, O. Engvold, and E. M. Kohner, "Accurate vessel width measurement from fundus photographs: a new concept." *British Journal of Ophthalmology*, vol. 78, no. 1, pp. 24–29, 1994.

[96] P. Gregson, Z. Shen, R. Scott, and V. Kozousek, "Automated grading of venous beading," *Computers and Biomedical Research*, vol. 28, no. 4, pp. 291–304, 1995.

[97] L. Zhou, M. S. Rzeszotarski, L. J. Singerman, and J. M. Chokreff, "The detection and quantification of retinopathy using digital angiograms," *IEEE Transactions on Medical Imaging*, vol. 13, no. 4, pp. 619–626, 1994.

[98] X. Gao, A. Bharath, A. Stanton, A. Hughes, N. Chapman, and S. Thom, "A method of vessel tracking for vessel diameter measurement on retinal images," in *Proc. the International Conference on Image Processing (ICIP)*, vol. 2. IEEE, 2001, pp. 881–884.

[99] O. Chutatape, L. Zheng, and S. M. Krishnan, "Retinal blood vessel detection and tracking by matched Gaussian and Kalman filters," in *Proc. the IEEE Engineering in Medicine and Biology Society (EMBC)*, vol. 6. IEEE, 1998, pp. 3144–3149.

[100] B. Al-Diri, A. Hunter, D. Steel, and M. Habib, "Joining retinal vessel segments," in *Proc. BioInformatics and BioEngineering (BIBE)*. IEEE, 2008, pp. 1–6.

[101] L. Wang, A. Bhalerao, and R. Wilson, "Analysis of retinal vasculature using a multiresolution Hermite model," *IEEE Transactions on Medical Imaging*, vol. 26, no. 2, pp. 137–152, 2007.

[102] B. Aliahmad, D. K. Kumar, and R. Jain, "Automatic analysis of retinal vascular parameters for detection of diabetes in Indian patients with no retinopathy sign," *International Scholarly Research Notices*, vol. 2016, 2016.

[103] A. Can, H. L. Tanenbaum, C.-L. Tsai, C. V. Stewart, H. Shen, B. Roysam, J. N. Turner, and K. H. Fritzsche, "Automated model-based segmentation, tracing, and analysis of retinal vasculature from digital fundus images," in *Angiography and Plaque Imaging: Advanced Segmentation Techniques*, 2003, pp. 225–297.

[104] Y. Yin, M. Adel, and S. Bourennane, "Retinal vessel segmentation using a probabilistic tracking method," *Pattern Recognition*, vol. 45, no. 4, pp. 1235–1244, 2012.

[105] B. Nayebifar and H. A. Moghaddam, "A novel method for retinal vessel tracking using particle filters," *Computers in Biology and Medicine*, vol. 43, no. 5, pp. 541–548, 2013.

[106] C. Sammut and G. I. Webb, *Encyclopedia of machine learning*. Springer Science & Business Media, 2011.

[107] D. Calvo, M. Ortega, M. G. Penedo, and J. Rouco, "Automatic detection and characterisation of retinal vessel tree bifurcations and crossovers in eye fundus images," *Computer Methods and Programs in Biomedicine*, vol. 103, no. 1, pp. 28–38, 2011.

[108] G. Azzopardi and N. Petkov, "Automatic detection of vascular bifurcations in segmented retinal images using trainable COSFIRE filters," *Pattern Recognition Letters*, vol. 34, no. 8, pp. 922–933, 2013.

[109] T. Fang, R. Su, L. Xie, Q. Gu, Q. Li, P. Liang, and T. Wang, "Retinal vessel landmark detection using deep learning and Hessian matrix," in *Proc. Congress on Image and Signal Processing (CISP)*. IEEE, 2015, pp. 387–392.

[110] S. Morales, V. Naranjo, J. Angulo, A. Legaz-Aparicio, and R. Verdú-Monedero, "Retinal network characterization through fundus image processing: signicant point identication on vessel centerline," *Signal Processing: Image Communication*, 2017.

[111] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[112] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[113] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, 2016.

[114] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.

[115] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.

[116] N. Srivastava and R. Salakhutdinov, "Learning representations for multimodal data with deep belief nets," in *Proc. International Conference on Machine Learning Workshop*, 2012.

[117] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *Proc. International Conference on Machine Learning (ICML)*, 2011, pp. 689–696.

[118] Y. Wang, H. Yao, and S. Zhao, "Auto-encoder based dimensionality reduction," *Neurocomputing*, vol. 184, pp. 232–242, 2016.

[119] Y.-l. Boureau, Y. L. Cun *et al.*, "Sparse feature learning for deep belief networks," in *Advances in neural information processing systems*, 2008, pp. 1185–1192.

[120] P. Smolensky, "Information processing in dynamical systems: Foundations of harmony theory," DTIC Document, Tech. Rep., 1986.

[121] Y. Freund and D. Haussler, "Unsupervised learning of distributions on binary vectors using two layer networks," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 1992, pp. 912–919.

[122] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Computation*, vol. 14, no. 8, pp. 1771–1800, 2002.

[123] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[124] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. International Conference on Machine Learning (ICML)*, 2010, pp. 807–814.

[125] G. Hinton, "A practical guide to training restricted Boltzmann machines," *Momentum*, vol. 9, no. 1, p. 926, 2010.

[126] K. Cho, A. Ilin, and T. Raiko, "Improved learning of Gaussian-Bernoulli restricted Boltzmann machines," in *Proc. International Conference on Artificial Neural Networks*. Springer, 2011, pp. 10–17.

[127] G. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[128] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21–36, 2003.

[129] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.

[130] A. Salazar-Gonzalez, D. Kaba, Y. Li, and X. Liu, "Segmentation of the blood vessels and optic disk in retinal images," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 6, pp. 1874–1886, 2014.

[131] M. M. Fraz, A. Basit, and S. Barman, "Application of morphological bit planes in retinal blood vessel extraction," *Journal of Digital Imaging*, vol. 26, no. 2, pp. 274–286, 2013.

[132] M. M. Fraz, S. Barman, P. Remagnino, A. Hoppe, A. Basit, B. Uyyanonvara, A. R. Rudnicka, and C. G. Owen, "An approach to localize the retinal blood vessels using bit planes and centerline detection," *Computer Methods and Programs in Biomedicine*, vol. 108, no. 2, pp. 600–616, 2012.

[133] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2016.

[134] A. Rueda, N. Malpica, and E. Romero, "Single-image super-resolution of brain MR images using overcomplete dictionaries," *Medical Image Analysis*, vol. 17, no. 1, pp. 113–132, 2013.

[135] J. Gao, Y. Guo, and M. Yin, "Restricted Boltzmann machine approach to couple dictionary training for image super-resolution," in *Proc. the International Conference on Image Processing (ICIP)*. IEEE, 2013, pp. 499–503.

[136] H. Yu, S. Barriga, C. Agurto, G. Zamora, W. Bauman, and P. Soliz, "Fast vessel segmentation in retinal images using multiscale enhancement and second-order local entropy," in *Proc. SPIE Medical Imaging*, 2012, pp. 83 151B–83 151B.

[137] Z. M. Rodriguez, P. Kenny, and L. Gaynor, "Improved characterisation of aortic tortuosity," *Medical Engineering & Physics*, vol. 33, no. 6, pp. 712–719, 2011.

[138] A. Fathi and A. R. Naghsh-Nilchi, "Automatic wavelet-based retinal blood vessels segmentation and vessel diameter estimation," *Biomedical Signal Processing and Control*, vol. 8, no. 1, pp. 71–80, 2013.

[139] R. Welikala, M. Fraz, P. Foster, P. Whincup, A. R. Rudnicka, C. G. Owen, D. Strachan, and S. A. Barman, "Automated retinal image quality assessment on the UK Biobank dataset for epidemiological studies," *Computers in Biology and Medicine*, vol. 71, pp. 67–76, 2016.

[140] B. Kayalibay, G. Jensen, and P. van der Smagt, "CNN-based segmentation of medical imaging data," *CoRR*, vol. abs/1701.03056, 2017, preprint. [Online]. Available: http://arxiv.org/abs/1701.03056

[141] D. Lesage, E. D. Angelini, I. Bloch, and G. Funka-Lea, "A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes," *Medical Image Analysis*, vol. 13, no. 6, pp. 819–845, 2009.

[142] M. Schaap, I. Smal, C. Metz, T. van Walsum, and W. Niessen, "Bayesian tracking of elongated structures in 3D images," in *Proc. Information Processing in Medical Imaging (IPMI)*. Springer, 2007, pp. 74–85.

[143] S. Worz, W. J. Godinez, K. Rohr, J. P. Pluim, and B. M. Dawant, "Segmentation of 3D tubular structures based on 3D intensity models and particle filter tracking," in *Proc. SPIE*, vol. 7259, 2009, p. 72591P.

[144] C. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer,New York, 2007.

[145] J. D. Hamilton, "State-space models," in *Handbook of Econometrics*. Elsevier, 1994, vol. 4, pp. 3039–3080.

[146] M. Johannes and N. Polson, "Particle filtering," in *Handbook of Financial Time Series*. Springer, 2009, pp. 1015–1029.

[147] F. Gustafsson, "Particle filter theory and practice with positioning applications," *IEEE Aerospace and Electronic Systems Magazine*, vol. 25, no. 7, pp. 53–82, 2010.

[148] Z. Chen, "Bayesian filtering: From Kalman filters to particle filters, and beyond," *Statistics*, vol. 182, no. 1, pp. 1–69, 2003.

[149] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[150] Y. Ho and R. Lee, "A Bayesian approach to problems in stochastic estimation and control," *IEEE Transactions on Automatic Control*, vol. 9, no. 4, pp. 333–339, 1964.

[151] C. M. Grinstead and J. L. Snell, *Introduction to probability.* American Mathematical Soc., 2012.

[152] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.

[153] N. J. Gordon, D. J. Salmond, and A. F. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," in *Proc. IEE Proceedings F - Radar and Signal Processing*, vol. 140, no. 2. IET, 1993, pp. 107–113.

[154] S. Maskell and N. Gordon, "A tutorial on particle filters for on-line nonlinear/non-Gaussian Bayesian tracking," in *Target tracking: Algorithms and Applications (Ref. No. 2001/174), IEE.* IET, 2001, pp. 2–1.

[155] D. Lesage, E. D. Angelini, G. Funka-Lea, and I. Bloch, "Adaptive particle filtering for coronary artery segmentation from 3D CT angiograms," *Computer Vision and Image Understanding*, vol. 151, pp. 29–46, 2016.

[156] T. Tung and T. Matsuyama, "Visual tracking using multimodal particle filter," *International Journal of Natural Computing Research (IJNCR)*, vol. 4, no. 3, pp. 69–84, 2014.

[157] H. L. de Micheaux, C. Ducottet, and P. Frey, "Online multi-model particle filter-based tracking to study bedload transport," in *Proc. the International Conference on Image Processing (ICIP).* IEEE, 2016, pp. 3489–3493.

[158] Y. Zhai, M. B. Yeary, S. Cheng, and N. Kehtarnavaz, "An object-tracking algorithm based on multiple-model particle filtering with state partitioning," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 5, pp. 1797–1809, 2009.

[159] H. Yang, L. Shao, F. Zheng, L. Wang, and Z. Song, "Recent advances and trends in visual tracking: A review," *Neurocomputing*, vol. 74, no. 18, pp. 3823–3831, 2011.

[160] C. Florin, N. Paragios, and J. Williams, "Particle filters, a quasi-Monte Carlo solution for segmentation of coronaries," in *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI).* Springer, 2005, pp. 246–253.

[161] C. Florin, N. Paragios, and J. Williams, "Globally optimal active contours, sequential Monte Carlo and on-line learning for vessel segmentation," in *Proc. European Conference on Computer Vision (ECCV)*. Springer, 2006, pp. 476–489.

[162] K. Allen, C. Yau, and J. Noble, "A recursive, stochastic vessel segmentation framework that robustly handles bifurcations," 2008.

[163] O. Friman, M. Hindennach, C. Kühnel, and H.-O. Peitgen, "Multiple hypothesis template tracking of small 3D vessel structures," *Medical Image Analysis*, vol. 14, no. 2, pp. 160–171, 2010.

[164] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Transactions on Information Theory*, vol. 21, no. 1, pp. 32–40, 1975.

[165] K. G. Derpanis, "Mean shift clustering," online, https://pdfs.semanticscholar.org/da1b/8039bd3ea75404369da61846e70326ea9708.pdf, 2005, Lecture Notes.

[166] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 790–799, 1995.

[167] X. W. Gao, A. Bharath, A. Stanton, A. Hughes, N. Chapman, and S. Thom, "Quantification and characterisation of arteries in retinal images," *Computer Methods and Programs in Biomedicine*, vol. 63, no. 2, pp. 133–146, 2000.

[168] B. Aliahmad and D. K. Kumar, "Adaptive Higuchi's dimension-based retinal vessel diameter measurement," in *Proc. the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2016, pp. 1308–1311.

Chapter 8

Appendix

Algorithm 2 Vessel Tree Fraction Tracking with Particle Filter

function VESSEL TREE FRACTION TRACKING FUNCTION(x)

$C^0 \leftarrow \left(C_x^0, C_y^0\right)$ ▷ Select an initial centreline location $\left(C_x^0, C_y^0\right)$.

$z_w^0 \leftarrow w(C^0)$ ▷ Assign an estimate of vessel width $w(C^0)$ at the location C^0.

$s \leftarrow s^0$ ▷ Assign a step size for tracking.

$z_D^0 \leftarrow \left(z_{Dx}^0, z_{Dy}^0\right)$; ▷ Assign the eigenvector with the minimum eigenvalue at C^0 as the initial direction vector

$\left[\sigma_{Dx}, \sigma_{Dy}, \sigma_w\right] \leftarrow \left[\sigma_{Dx}^0, \sigma_{Dy}^0, \sigma_w^0\right]$ ▷ Define standard deviations for direction vectors and vessel widths

$\left\{z^0\right\}_1^N \leftarrow N\left(\left[z_{Dx}^0, z_{Dy}^0, z_w^0\right], \left[\sigma_{Dx}, \sigma_{Dy}, \sigma_w\right]\right)$ ▷ Generate a Particle Set by drawing N samples from the normal distribution.

$\mathbf{S}_P \leftarrow \left\{\left(\left\{C^0\right\}_1^N, \left\{z^0\right\}_1^N\right)\right\}$ ▷ $\mathbf{S}_P$ is a seed set. An element of the $\mathbf{S}_P$ contains a set of particle state vectors $\left\{z^0\right\}_1^N$ and a set of their centreline locations $\left\{C^0\right\}_1^N$.

$y \leftarrow O$ ▷ O denotes observations.

do

$k \leftarrow 0$ ▷ k denotes a tracking step

$\left(\left\{C^0\right\}_1^N, \left\{z^0\right\}_1^N\right) \leftarrow \mathbf{S}_P\{1\}$

$\left\{C^k\right\}_1^N \leftarrow \left\{C^0\right\}_1^N$

$\left\{z^k\right\}_1^N \leftarrow \left\{z^0\right\}_1^N$

do

$\left[\left\{z^{k+1}\right\}_1^N, \left\{C^{k+1}\right\}_1^N, \left\{E_{1,2}^{k+1}\right\}_1^N\right] \leftarrow$ *Regular Vessel Segment Tracking Mode*$\left(\left\{z^k\right\}_1^N, \left\{C^k\right\}_1^N, \mathbf{y}\right)$

Calculate $\bar{C}^{k+1}$ from (5.31) ▷ $\bar{C}^{k+1}$ refers to the weighted mean of the estimated centreline locations.
Calculate $\bar{E}_{1,2}^{k+1}$ from (5.32) ▷ $\bar{E}_{1,2}^{k+1}$ refers to the weighted mean of the first and the second edge locations.
Return $P_L(z = \bar{C}^{k+1})$ ▷ $P_L(z = \bar{C}^{k+1})$ denotes the landmark probability at $\bar{C}^{k+1}$.
Return $P_c(z = \bar{C}^{k+1})$ ▷ $P_c(z = \bar{C}^{k+1})$ shows the centreline probability at $\bar{C}^{k+1}$.
if $P_L(z = \bar{C}^{k+1}) > T_L$ **then**
$\mathbf{S}_P \leftarrow$*Daughter Vessel Tracking Mode*$(\left\{z^{k+1}\right\}_1^N, \mathbf{S}_P, \mathbf{y})$
end if
Display $\bar{C}^{0:k+1}$, $\bar{E}_{1,2}^{0:k+1}$ and detected landmark locations.
$k \leftarrow k + 1$
while $P_c(z = \bar{C}^{k+1}) > T_c$
Remove $\mathbf{S}_P\{1\}$ from $\mathbf{S}_P$.
while $\mathbf{S}_P \neq \emptyset$
end function

Algorithm 3 Regular Vessel Segment Tracking Mode

function $\left[\left\{\mathbf{z}^{k+1}\right\}_1^N, \left\{C^{k+1}\right\}_1^N, \left\{E_{1,2}^{k+1}\right\}_1^N\right]$ =Regular Vessel Segment Tracking Mode$(\left\{\mathbf{z}^k\right\}_1^N, \left\{C^k\right\}_1^N, \mathbf{y})$
Estimate $\mathbf{z}^{k+1}$ from Algorithm 1.
Calculate $\left\{C^{k+1}\right\}^n$ for each particle n from (5.24)
Calculate $\left\{E_1^{k+1}\right\}^n$ and $\left\{E_2^{k+1}\right\}^n$ from (5.25) and (5.26) consecutively.
Return $\left\{\mathbf{z}^{k+1}\right\}_1^N$, $\left\{C^{k+1}\right\}_1^N$ and $\left\{E_{1,2}^{k+1}\right\}_1^N$.
end function

Algorithm 4 Daughter Vessel Tracking Mode

function $\mathbf{S}_P$ = Daughter Vessel Tracking Mode$(\left\{\mathbf{z}^k\right\}_1^N, \mathbf{S}_P, \mathbf{y})$
$\bar{C}^L \leftarrow$*Landmark Detection Function*$(\left\{\mathbf{z}^k\right\}_1^N, \mathbf{y})$
$\mathbf{S}_P \leftarrow$*Identify Connected Branches Function*$(\left\{\mathbf{z}^k\right\}_1^N, \bar{C}^L, \mathbf{S}_P)$
end function

Algorithm 5 Landmark Detection Function

function $\bar{C}^L$ =Landmark Detection Function($\left\{\mathbf{z}^k\right\}_1^N, \mathbf{y}$)
  $t = 0$
  $\left\{\mathbf{z}^t\right\}_1^N = \left\{\mathbf{z}^k\right\}_1^N$
  do
    Estimate $P(\mathbf{z}^{t+1}|\mathbf{y}^{t+1})$ from 5.21 and 8.1 ▷

$$P\left(y^{t+1}|\left\{z^{t+1}\right\}^n\right) = P_c\left(z = \left\{\grave{C}^{t+1}\right\}^n\right) \tag{8.1}$$

  where $n \in [1, \cdots, N]$ and $\left\{\grave{C}^{t+1}\right\}^n$ shows the calculated centreline for particle n according to (5.24) by using the estimates of direction with respect to state transition probability distribution $P(z^{t+1}|z^t)$.
    Calculate $\left\{C^{t+1}\right\}^n$ from (5.24)
    Calculate $\bar{C}^{t+1}$ from (5.31).
    $\mathbf{S}_L \leftarrow \left[\bar{C}^{t+1}, P_L(z = \bar{C}^{t+1})\right]$ ▷ Push the couple of possible landmark location $\bar{C}^{t+1}$ and its probability to be a landmark $P_L(z = \bar{C}^{t+1})$ to $\mathbf{S}_L$, which is a temporary set to collect these couples, for each time step $t + 1$.
    $t = t + 1$
  while $P_L(z = \bar{C}^{t+1}) > T_L$
  $P_{max} \leftarrow max\left(P_L(z = \bar{C}^1), \cdots, P_L(z = \bar{C}^{t+1})\right)$
  Find $\bar{C}^L$ satisfying that $\left(P_L(z = \bar{C}^L) == P_{max}\right)$
  Return $\bar{C}^L$
end function

Algorithm 6 Identify Connected Branches Function

1: **function** $\mathbf{S}_P$=IDENTIFY CONNECTED BRANCHES FUNCTION $(\left\{\mathbf{z}^k\right\}_1^N, \bar{C}^L, \mathbf{S}_P)$

2: (i) Calculate $\bar{z}_w^k$, $\bar{z}_{Dx}^k$ and $\bar{z}_{Dy}^k$ from (8.2), (8.3) and (8.4) respectively.

$$\bar{z}_w^k = \sum_n^N \left(\left\{\hat{W}^k\right\}^n \cdot \left\{z_w^k\right\}^n \right) \tag{8.2}$$

$$\bar{z}_{Dx}^k = \sum_n^N \left(\left\{\hat{W}^k\right\}^n \cdot \left\{z_{Dx}^k\right\}^n \right) \tag{8.3}$$

$$\bar{z}_{Dy}^k = \sum_n^N \left(\left\{\hat{W}^k\right\}^n \cdot \left\{z_{Dy}^k\right\}^n \right) \tag{8.4}$$

3: $\left\{ \left\{\mathbf{z}_0^k\right\}_1^N, \cdots, \left\{\mathbf{z}_6^k\right\}_1^N \right\} \backsim N(\left[z_{Dx}^0, z_{Dy}^0, z_w^0\right], \left[\sigma_{Dx}, \sigma_{Dy}, \sigma_w\right])$ ▷ Initialise 7 new particles sets by drawing N samples from a multivariate normal distribution for each set by considering two conditions as follows

1. $(z_D)_f \in \left\{ z_D^k, rot(z_D^k, \frac{\pi}{4}), rot(z_D^k, \frac{-\pi}{4}), rot(z_D^k, \frac{\pi}{2}), rot(z_D^k, \frac{-\pi}{2}), rot(z_D^k, \frac{3\pi}{4}), rot(z_D^k, \frac{-3\pi}{4}) \right\}$, ▷ where $rot(z_D, \theta)$ means that rotate the direction vector z_D as much as the angle θ.
2. $\left[z_{Dx}^0, z_{Dy}^0, z_w^0\right] = \left[(z_{Dx})_f, (z_{Dy})_f, \bar{z}_w^k\right]$,

4: $\left\{C_{0:6}^0\right\}_1^N \leftarrow \bar{C}^L$ ▷ Assign initial tracking location for each particle in each cluster

5: Estimate $P(z^{k+1}|y^{k+1})$ from (8.5).

$$P(y^{k+1}|z^{k+1}) = \left(1 - P_c(z = \grave{E}_1^{k+1})\right)\left(1 - P_c(z = \grave{E}_2^{k+1})\right) \tag{8.5}$$
$$P_c(z = \frac{(3\grave{E}_1^{k+1} + \grave{E}_2^{k+1})}{4}) P_c(z = \frac{(\grave{E}_1^{k+1} + \grave{E}_2^{k+1})}{2}) P_c(z = \frac{(\grave{E}_1^{k+1} + 3\grave{E}_2^{k+1})}{4})$$

6: Calculate $\left\{C_{0:6}^{k+1}\right\}_1^N$ with (5.24)

7: $\left\{\underbrace{z_m}\right\}_1^{\hat{N}} \leftarrow$ *Mean Shift Cluster Function*$(K,h,T_M,\left\{z_{0:6}^{k+1}\right\}_1^N,\left\{C_{0:6}^{k+1}\right\}_1^N)$

▷ Calculate the number of clusters in $\left\{C_{0:6}^{k+1}\right\}_1^N$ and the members (particles) of each cluster

8: **for** m=1:M **do** ▷ Assume the number of maintained cluster is M

9: **if** $\left\{C_m^{k+1}\right\}_1^N \neq \left\{C_m^{0:k}\right\}$ **then** ▷ $\left\{C_m^{0:k}\right\}$ shows centrelines already traced.

10: $\left\{\underbrace{z_m^{k+1}}\right\}_1^N \backsim \left\{\underbrace{z_m^{k+1}}\right\}_1^{\hat{N}}$ ▷ Draw N samples from m^{th} particle cluster $\left\{\underbrace{z_m^{k+1}}\right\}_1^{\hat{N}}$ with respect to current weights of particles

11: Ensure $\sum_{n=1}^N \left\{\hat{W}^{k+1}\right\}^n = 1$

12: Calculate $\bar{z}_D^{k+1}$ from (8.3) and (8.4) and $\bar{z}_w^{k+1}$ from (8.2)

13: $\left\{z_m^{k+1}\right\}_1^N \backsim N(\left[\bar{z}_{Dx}^{k+1}, \bar{z}_{Dy}^{k+1}, \bar{z}_w^{k+1}\right], \left[\sigma_{Dx}, \sigma_{Dy}, \sigma_w\right])$. ▷ Initialise a new particle set $\left\{z_m^{k+1}\right\}_1^N$

14: $\left\{\hat{W}_m^{k+1}\right\}_1^{\hat{N}} \leftarrow \frac{1}{N}$

15: Return $\left\{z_m^{k+3}\right\}_1^N$ from Algorithm 1.

16: $\mathbf{S}_P \leftarrow \left\{\bar{C}^L, \left\{\mathbf{z}_m^{k+3}\right\}_1^N\right\}$

17: **end if**

18: **end for**

19: Return $\mathbf{S}_P$

20: **end function**

Algorithm 7 Mean Shift Cluster Function

function $\left\{\underbrace{z_m}\right\}_1^{\hat{N}}$ = MEAN SHIFT CLUSTER FUNCTION(K,h,T_M,$\{z_{0:6}\}_1^N$,$\{C_{0:6}\}_1^N$)
  K is a radially symmetric kernel
  h is the bandwidth of the kernel K
  T_M is the threshold to stop mode location search
  $x^1 \leftarrow x$; $x \in \{C_{0:6}\}_1^N$ ▷ Select an initial estimate of a mode location
  while $||x^t - x^{t-1}|| > T_M$ **do**
    Calculate $d(x^{t+1})$ with (8.6). ▷ $d(x^{t+1})$ represents the mean shift vector.

$$d(x^{t+1}) = \frac{\sum_{i=1}^{n} x_i g(\frac{||x^t - x_i||^2}{h})}{\sum_{i=1}^{n} g(\frac{||x^t - x_i||^2}{h})} - x^t \tag{8.6}$$

  where x^t is the estimate of a mode location at time step t and h is a bandwidth and g is the derivative of the kernel function.
    Calculate $x^{t+1} \leftarrow x^t + d(x^{t+1})$
    $x^t \leftarrow x^{t+1}$
    $t \leftarrow t + 1$
  end while
  Return $\left\{\underbrace{z_m}\right\}_1^{\hat{N}} \subset N(x^t, \sigma_{x^t})$ ▷ Return the members of cluster $\left\{\underbrace{z_m}\right\}_1^{\hat{N}}$, whose mode location over centreline locations is x^t
end function