Medical image segmentation based on deep feature learning and multistage classification

Wang, Xiaohong

2020

Wang, X. (2020). Medical image segmentation based on deep feature learning and multistage classification. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/137579>

<https://doi.org/10.32657/10356/137579>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 22:34:34 SGT

**MEDICAL IMAGE SEGMENTATION BASED
ON DEEP FEATURE LEARNING AND
MULTISTAGE CLASSIFICATION**

WANG XIAOHONG

School of Electrical & Electronic Engineering

2020

MEDICAL IMAGE SEGMENTATION BASED ON DEEP FEATURE LEARNING AND MULTISTAGE CLASSIFICATION

WANG XIAOHONG

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

2020

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

15-Aug-2019

.....

Date

wang xiao hong

.....

Wang Xiaohong

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

15-Aug-2019

.....

Date

.....

Prof. Jiang Xudong

Authorship Attribution Statement

This thesis contains material from 3 paper(s) published in the following peer-reviewed journal(s) / from papers accepted at conferences in which I am listed as an author.

Chapter 3 is published as [X. Wang, X. Jiang, and J. Ren. Blood vessel segmentation from fundus image by a cascade classification framework. *Pattern Recognition* **88**, 331-341 \(2019\). DOI: 10.1016/j.patcog.2018.11.030.](#)

The contributions of the co-authors are as follows:

- Prof. Jiang Xudong supervised me on the overall research direction and edited the manuscript drafts.
- I developed the algorithm, conducted the experiments and prepared the manuscript drafts.
- Dr. Ren Jianfen provided valuable comments on the manuscript.

Chapter 4 is published as [X. Wang, X. Jiang. Retinal vessel segmentation by a divide-and-conquer funnel-structured classification framework. *Signal Processing* **165**, 104-114 \(2019\). DOI: 10.1016/j.sigpro.2019.06.018.](#)

The contributions of the co-authors are as follows:

- Prof. Jiang Xudong supervised me on the overall research direction and edited the manuscript drafts.
- I developed the algorithm, conducted the experiments and prepared the manuscript drafts.

Chapter 5 is published as [X. Wang, X. Jiang, H. Ding, J. Liu, Bi-directional dermoscopic feature learning network with multi-scale consistent decision fusion for skin lesion segmentation, *IEEE Transaction on Image Processing* **29**, 3039 - 3051 \(2019\). DOI:10.1109/TIP.2019.2955297.](#)

The contributions of the co-authors are as follows:

- Prof. Jiang Xudong supervised me on the overall research direction and edited the manuscript drafts.

-
- I developed the algorithm, conducted the experiments and prepared the manuscript drafts.
 - Mr. Ding Henghui and Dr. Liu Jun provided valuable comments on the manuscript.

15-Aug-2019

.....

Date

wang xiao hong

.....

Wang Xiaohong

Acknowledgements

I would like to express sincere gratitude to my supervisor, Professor Jiang Xudong, who provided me with invaluable guidance, kind encouragement, and tremendous support during my Ph.D. study. Professor Jiang has helped me a lot in how to do academic research, including problem analyzing, professional thinking, and manuscript writing. His rigorous and enthusiasm for research inspire me to build deep interest in image processing and pattern recognition. I really appreciate his insightful supervision and dedication during my four-year research.

I would also like to thank the colleagues and staffs in the Media Technology Laboratory, Dr. Huang Ling, Dr. Kong Qinglei, Dr. Liu Ankang, Mr. Wang Suchen, Mr. Guo Heng, Ms. Hu Mengjiao, Ms. Qiu Wei, Ms. Zhang Xinyi, Mr. Liu Hexin, Mr. Liu Chang, Mr. Xiao Xiao, etc, for their accompany and help in my daily life. Thankfulness is also given to Dr. Liu Jun, Mr. Ding Henghui, Mr. Mei Jianhan, and Dr. Ren Jianfeng in the Rapid-Rich Object Search Laboratory for their technical support and constructive suggestion. Besides, I would show the special thanks to my friends Dr. Ge Qianqian and Dr. Zhang Lili in China, for their understanding and warm-hearted care that remind me to stay true to myself.

At last, I would like to thank my parents, my nephew and all other family members for their love and support at each stage of my life. I really appreciate their unconditional trusting and encouragement that give me the strength to pursue my life bravely.

Abstract

Automatic segmentation of target organs in medical image plays a crucial role in the computer-aided diagnosis of human diseases, like retinal vessel from fundus image, and melanoma from dermoscopic image. Manual segmentation of those tissues is a time-consuming and labor-intensive task that is not feasible for clinicians to annotate a large amount of medical images by hand. Thus, a reliable medical image segmentation system with lower cost and less human interaction than the observation-based techniques is attractive in the field of medical image analysis. This thesis investigates medical image segmentation architectures through deep feature learning and multistage classification. Specifically, we proposed two novel multistage classification schemes for retinal vessel segmentation. For more complex and large dermoscopic image data analysis issue, we proposed a novel deep feature learning scheme for skin lesion segmentation. Details of the completed works are summarized as follows:

In the first scheme, a novel and robust retinal vessel segmentation framework is proposed, which envelops a set of computationally efficient Mahalanobis distance classifiers to form a highly nonlinear decision. Different from other nonlinear classifiers that need a predefined nonlinear kernel or an iterative training, the proposed cascade classification framework is trained by a one-pass feedforward process. Thus, the degree of nonlinearity of the proposed classifier is not predefined, but determined by the complexity of the data structure. Experimental evaluations on three publicly available datasets show that the proposed cascade classification framework achieves high vessel segmentation accuracy consistently on all three diverse datasets.

In the second scheme, a hierarchical architecture for retinal vessel segmentation based on a divide-and-conquer strategy is designed. Current works for retinal vessel segmentation typically train a global discriminative model for retinal vessel classification that is still not sufficient to fit the complex pattern of vessel structure. In fact, the large geometrical structure difference among retinal vessels with different scales and positions greatly limits the precision of the decision boundary of the global discriminative model. To overcome this problem, an efficient dividing algorithm, named multiplex vessel partition (MVP), is proposed to divide the retinal vessel into well constrained subsets where vessel samples with the same geometrical property are assigned together. Then, a set of homogeneous classifiers are trained in parallel to form discriminative decision for each subset. Moreover, a funnel-structured vessel segmentation (FsVS) framework is proposed to link the classification results from each disjoint subset. It reduces the probability of poor partition at the dividing phase and further enhances the discriminative capability of the decision model. Both quantitative and qualitative experimental comparisons on three publicly datasets demonstrate the flexibility and efficiency of the proposed work on retinal vessel segmentation.

In the last scheme, a bi-directional feature dermoscopic learning framework with multiscale consistent decision fusion is proposed for skin lesion segmentation. Previously published skin lesion segmentation works enhance lesion detection performance by using deep learning based methods like fully convolutional network (FCN). Nevertheless, relationship between skin lesions and their informative context, as well as the consistency of the decision from multiple classification layers, have not yet been well explored by these previous studies. Different from the naive way of FCN learning an abstract feature representation of image, this thesis proposes a bi-directional dermoscopic feature learning (biDFL) framework that produces rich dermoscopic feature maps by controlling information propagation from two complementary directions at high level parsing layer. With the integration of bi-directional feature information passing, the proposed biDFL module gives better insight to the network about the complex structure of the skin lesion.

Furthermore, this thesis proposes a multiscale consistent decision fusion (mCDF) that is capable of selectively focusing on the informative decisions generated from multiple classification layers. By analysis of the consistency of the decision at each position, mCDF automatically adjusts the reliability of decisions and thus allows a more insightful skin lesion delineation. With the embedding of consistency analysis to the decisions from each classification layer, the proposed mCDF assists the network to learn better about which scales of features are more desirable for each pixel. The comprehensive experimental results show the effectiveness of the proposed method on skin lesion segmentation, achieving state-of-the-art performance consistently on two publicly available dermoscopic image datasets.

Contents

Acknowledgements	ix
Abstract	xi
List of Figures	xix
List of Tables	xxiii
1 Introduction	1
1.1 Background	1
1.1.1 Fundus Image	2
1.1.2 Dermoscopic Image	3
1.2 Motivations and Objectives	5
1.3 Research Contributions	7
1.4 Thesis Outline	9
2 Literature Review	11
2.1 Unsupervised Segmentation	11
2.2 Supervised Segmentation	13
2.2.1 Feature Extraction	13
2.2.2 Classifier Learning	14
2.3 Deep Learning Segmentation	16
2.3.1 Fully Convolutional Network	16
2.3.2 Application on Medical Image Segmentation	18
2.4 Summary	21
3 Cascade Classification Framework for Fundus Image Segmentation	23
3.1 Introduction	23
3.2 Proposed Framework	25
3.2.1 Retinal Image Preprocessing	25
3.2.2 Color Channel Fusion with Dimensionality Reduction	27
3.2.3 Cascade Classification Network	30
3.2.4 Postprocessing	36

3.3	Experimental Evaluations	36
3.3.1	Materials	36
3.3.2	Parameter Settings	38
3.3.3	Ablation Evaluation	39
3.3.4	Comparison with Other Published Methods	43
3.3.4.1	Vessel Segmentation Performance	43
3.3.4.2	Pathological Retinal Image Analysis	47
3.4	Summary	47
4	Divided-and-Conquer Funnel-Structured Classification Framework for Fundus Image Segmentation	51
4.1	Introduction	51
4.2	Proposed Framework	53
4.2.1	Multiplex Vessel Partition	55
4.2.2	Funnel-structured Vessel Segmentation	58
4.3	Experimental Evaluations	61
4.3.1	Comparison with Single SVM Classifier	62
4.3.2	Comparison with Other Published Methods	63
4.3.2.1	Vessel Segmentation Performance	63
4.3.2.2	Peripapillary Vessel Analysis	70
4.3.2.3	Thin and Low-contrast Vessel Analysis	70
4.3.2.4	Pathological Retinal Image Analysis	71
4.4	Summary	72
5	Deep Feature Learning for Dermoscopic Image Segmentation	75
5.1	Introduction	75
5.2	Proposed Framework	79
5.2.1	Context Feature Map Generation with Multiple Dilation Rates	80
5.2.2	Bi-directional Dermoscopic Feature Learning	82
5.2.3	Multi-scale Consistent Decision Fusion	84
5.3	Experimental Evaluations	86
5.3.1	Materials	86
5.3.2	Evaluation Criteria	87
5.3.3	Implementation Details	88
5.3.4	Ablation Study	88
5.3.4.1	Evaluation of Individual Component	88
5.3.4.2	Evaluation of Parameter of Consistent Decision Fusion	92
5.3.5	Comparison with Other Published Methods	94
5.4	Summary	96
6	Conclusion and Future Work	97
6.1	Conclusion	97
6.2	Future Work	98

CONTENTS

Author's Publications	101
Bibliography	103

List of Figures

1.1	Examples of normal and diabetic retinopathy images [1].	2
1.2	Anatomic structure of skin and some examples of dermoscopic images. (a) Anatomic structure of skin. (b) Dermoscopic images with skin lesions.	4
2.1	FCN network for semantic segmentation [2].	17
2.2	Some examples for skin lesion segmentation based on fully convolutional network. (a) Fully convolutional residual networks [3]. (b) Step-wise integration of the segmented results from FCNs [4]. (c) Multistage fully convolutional network [5].	19
3.1	Example of background normalization and truncation filtering. (a) Three sample square surrounding bands (orange, green, purple) of pixel (x, y) (red) for $W = 3$ and $\delta = 1$. (b) Original sub-image. (c) Sub-image after background normalization. (d) Sub-image after truncation filtering.	27
3.2	Flowchart of feature extraction, color channel fusion and dimensionality reduction.	29
3.3	An illustration of a binary classification by the threshold T_o and a three-state classification by two thresholds T_v and T_b	32
3.4	Constructing and training the cascade classification network. $\mathfrak{D}_i^v$ and $\mathfrak{D}_i^b$ are classified vessel and background sets of training samples in the i^{th} stage. $\mathfrak{D}_i^x$ is the uncertain set to train the classifier of the $(i + 1)^{th}$ stage.	34
3.5	Prediction procedure of the proposed cascade classification network.	35
3.6	Classification error rates against the number of stages of the proposed cascade classification network on the training, validation and testing data sets of CHASE_DB1 dataset.	35
3.7	An example of the postprocessing procedure. (a) A subregion of retinal image with lesions. (b) Result of cascade classification network. (c) The skeleton of (b) (red pixels indicate the branch points). (d) The detected lesions. (e) Result after removing lesions. (f) Erosion result of (e). (g) Convex hull of the dilated significant connected component of (f). (h) The detected lesions. (i) Final result after lesion removal.	37

3.8	Examples of segmentation results. The first column to the third column: a sample image and its segmentation results on DRIVE, STARE and CHASE_DB1 datasets. The first row to the last row: original image, ground truth, and segmentation results of SSMDC, PCA+SSMDC, and the proposed method. Green region: missing retinal vessels; Pink region: spurious retinal vessels.	41
3.9	Comparisons with the state-of-the art methods under some specific challenging and difficult conditions of retinal image. The first column to the last column: subregions of retinal images, ground truth, segmented results by Soares <i>et al.</i> [6], Marin <i>et al.</i> [7], Nguyen <i>et al.</i> [8], Orlando <i>et al.</i> [9], and the proposed method, respectively. . .	45
4.1	Intensity profile of cross section of three representative vessel fragments. The first column: fundus image with three vessel fragments (vessel width: Fragment 1 < Fragment 2 < Fragment 3). The second column: intensity profile of cross section of fragments marked in fundus image.	52
4.2	The proposed divide-and-conquer funnel-structured architecture for retinal vessel segmentation.	54
4.3	Examples of vessel pixels dividing results based on different τ . The top to the bottom row of the first column: fundus image fragments with scales from small to large. The second to the last column: vessel pixels dividing results at $\tau=1, 0.75, 0.5$, respectively. Red color marks pixels assigned to the edge cluster and blue color marks pixels assigned to the center cluster.	58
4.4	Threshold searching process. Shaded areas in green and yellow color are the searching range $[T_{v0,C_k}-\delta, T_{v0,C_k}+\delta]$ and $[T_{b0,C_k}-\delta, T_{b0,C_k}+\delta]$ respectively. Pixels in the region Ω_χ remain to be further processed at the subsequent level.	60
4.5	Examples of segmentation results. The first row to the third row: images and their segmentation results on DRIVE, STARE and CHASE_DB1 datasets. The first column to the last column: for the best segmentation case, original image, ground truth, segmentation result; for the worst segmentation case, original image, ground truth, segmentation result.	62
4.6	Local ACC of different methods with the different ratio of the number of pixels of background to that of vessel. (a) DRIVE dataset. (b) STARE dataset. (c) CHASE_DB1 dataset.	64
4.7	Comparisons with the state-of-the art methods under some specific challenging and difficult conditions of retinal image. (a) Parallel vessels. (b) Vessels with low contrast. (c) Crossing vessels. (d) Lesions. The first row to the last row: subregions of retinal images, ground truths, segmented results by [6], [8], [10], [11], [12], [9], and the proposed method, respectively.	68

5.1	Examples of challenges for skin lesion delineation. Lesion position is marked as red color.	77
5.2	The structure of the proposed framework. Bi-directional feature learning framework investigates the complex correlation between skin lesions and their informative context to generate a discriminative feature representation. Multiscale consistent decision fusion technique achieves more reliable and precise skin lesion prediction by analyzing the consistency of decisions in a local area. Green solid arrow: feature learning from local to zooming-out region. Red dot arrow: feature learning from global to zooming-in region.	79
5.3	Workflow of the bi-directional dermoscopic feature learning. (a) Receptive fields of dilated convolution with three different rates. (b) Feature passing from local to zooming-out region. (c) Feature passing from global to zooming-in region.	82
5.4	Qualitative segmented result comparison on ISBI 2016 dataset. First row to third row: segmentation results by baseline (green line), baseline +biDFL (blue line) and baseline+biDFL+mCDF (yellow line) respectively, where ground truths are annotated as red line. Last row: detailed segmentation results within sub-region (marked by white arrow in the corresponding column).	91
5.5	Qualitative segmented result comparison on ISBI 2017 dataset. First row to third row: segmentation results by baseline (green line), baseline +biDFL (blue line) and baseline+biDFL+mCDF (yellow line) respectively, where ground truths are annotated as red line. Last row: detailed segmentation results within sub-region (marked by white arrow in the corresponding column).	92
5.6	Distribution comparison in terms of Jaccard index. (a) ISBI 2016 dataset. (b) ISBI 2017 dataset.	95

List of Tables

3.1	Performance comparison with two baselines (%)	42
3.2	Performance comparison with other methods in terms of ACC (%) .	44
3.3	Performance comparison with other methods in terms of F1-score (%) and MCC (%)	46
3.4	Performance comparison on the abnormal retinal images in the STARE dataset (%)	47
4.1	Segmentation performance on DRIVE, STARE and CHASE_DB1 datasets(%)	63
4.2	Performance comparison with other methods in terms of ACC (%) and ACC_l (%)	66
4.3	Performance comparison with other methods in terms of F1-SCORE (%) and MCC (%)	67
4.4	Peripapillary vessel segmentation performance on DRIVE, STARE and CHASE_DB1 datasets(%)	69
4.5	Performance comparison on thin vessel and low-contrast vessel in STARE dataset (%)	71
4.6	Performance comparison on pathological retinal image in STARE dataset (%)	72
5.1	Segmentation performance on ISBI 2016 and ISBI 2017 datasets (%)	89
5.2	Segmentation performance on melanoma and non-melanoma (%) . .	89
5.3	Component investigation of bi-directional dermoscopic feature learning (%)	90
5.4	Performance comparison of different σ^2 values (%)	92
5.5	Performance comparison with other methods on ISBI 2016 dataset (%)	93
5.6	Performance comparison with other methods on ISBI 2017 dataset (%)	93
5.7	Performance comparison with other methods on ISBI 2016 dataset for melanoma and non-melanoma cases (%)	94
5.8	Performance comparison with other methods on ISBI 2017 dataset for melanoma and non-melanoma cases (%)	94

Chapter 1

Introduction

1.1 Background

Medical imaging is a process that noninvasively creates a visual representation of the organ inside the human body, like ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI). Quantitative analysis of medical data from those imaging techniques provides insight into the organ anatomy or pathology, thus essential for the clinicians in disease diagnosis and treatment planning. Computer-aided diagnosis (CAD) system is an invaluable tool to assist clinicians in the interpretation of medical images and the detection of the human organ abnormality, such as architectural distortion in mammogram [13], liver nodule in CT [14] and brain tumor in MRI [15]. Segmentation of region of interest (ROI) in medical images, i.e. partition a medical image into multiple segments and delineate the outline of a target organ, is a fundamental step in CAD system. Actually, medical image segmentation provides an effective way to quantitatively analyze the anatomical structures in clinical applications. For instance, the segmentation result of the tumor in MRI is greatly useful to support the measurement of brain atrophy and allows for quantitative analysis of Alzheimer disease [16]. The left ventricle segmentation plays an important role in the evaluation of the heart functions by quantitative indicators like ejection fraction, stroke volume, and myocardium mass

FIGURE 1.1: Examples of normal and diabetic retinopathy images [1].

[17]. In this work, we investigate automatic segmentation architectures for two separate modalities of medical images, i.e. fundus image and dermoscopic image.

1.1.1 Fundus Image

The first modality of medical image is from fundus imaging which is the process whereby the reflected light is used to obtain a 2D representation of the 3D retinal semitransparent tissues projected on the imaging plane [18, 19]. Fig. 1.1 show some examples of fundus images which capture the structure of retinal blood vessel tree. Retinal blood supply comes from a branch of the ophthalmic artery, and then splits into two secondary branches (i.e. the superior and inferior branches) just before entering the eye [20, 21]. Those branches are further divided progressively, which results in the bifurcation of the large vessel into capillary.

Appearance of retinal vessel is a crucial diagnostic indicator for some ophthalmic pathologies, such as arteriosclerosis, vein occlusions, cardiovascular disease, and diabetic retinopathy [22, 23]. Those diseases often reflect on changes of branching angle, as well as retinal vessel caliber tortuosity [8, 24]. For example, diabetic retinopathy, the leading cause of blindness among working-age population in most developed countries, results in small outpunching in capillary vessel (also known as

microaneurysms) at the first stage and cotton wool spots, hemorrhages and macular edema as the disease progresses [25, 26]. The right image of Fig. 1.1 shows an example of fundus image with diabetic retinopathy. Choroidal neovascularization (CNV), also called age-related macular degeneration, is another visually threatening pattern which characterized by the ingrowth of a choroidal vascular structure into the macula with the increase of vascular permeability [18].

Early detection of retinal vessel provides opportunity to diagnose those diseases and prepare optimal treatment by minimizing the risk of visual loss. Moreover, segmentation and analysis of retinal vessel are also beneficial to identify the location of optic disk or fovea [27, 28], evaluate the retinopathy of prematurity [29, 30], and quantify retinal arteriovenous nicking [31]. Some reported algorithms, like the extraction of reference points for image registration [32, 33], the measurement of tortuosity or width of retinal vessel [34, 35], and the designation of computer-assisted retinal laser photocoagulation system [36], also depend on accurate retinal vessel detection. In the field of biometric identification, the geometric structures of retinal vessel are considered as a sort of valuable information for high security applications [37].

1.1.2 Dermoscopic Image

The second modality of medical image under analysis in this work is dermoscopic image. Dermoscopy imaging technique, widely used in the clinical examination of melanoma, is a noninvasive imaging tool that provides a detailed visualization of the pigmented skin lesion appearance [38]. More specifically, dermoscopic image is captured by the digital imaging system like stereo microscope or dermatoscope to magnify the surface structure of skin lesion [39]. Dermoscopic images provide detailed lesion feature information, like streaks, dots, pigment networks and blue-white areas, which are not clearly visible to the naked human eye. Fig. 1.2 (a) (image from American cancer society¹) shows the anatomic structure of the

¹<https://www.cancer.org/cancer/melanoma-skin-cancer/about/what-is-melanoma.html>

FIGURE 1.2: Anatomic structure of skin and some examples of dermoscopic images. (a) Anatomic structure of skin. (b) Dermoscopic images with skin lesions.

skin. The top skin layer has three main kinds of cells, i.e. squamous cell, basal cell and melanocyte. When melanocytes grow in clusters alongside normal cells, melanocytic nevi or pigmented skin lesions will appear on the surface of the skin [40]. Fig. 1.2 (b) shows some examples of dermoscopic images with skin lesions.

Melanoma is one of the most lethal types of skin cancers through all over the world [41]. The mortality rate of melanoma also relates to its high potential to metastasize to other human organs [42]. In the advanced stages of melanoma (with signs of metastases), it is incurable by the clinical treatment like radiation therapy, chemotherapy, and immunotherapy [40].

During skin lesion screening procedures, clinicians usually apply certain criteria to decide whether a candidate lesion belongs to melanoma or non-melanoma. A common criterion for melanoma recognition is ABCD rule [40]:

- Asymmetry (A): assess contour, colors and structures.
- Border (B): abrupt cut-off of pigment pattern at the periphery.
- Color (C): presence of 6 defined colors, i.e. white, red, light-brown, dark-brown, blue-gray and black.

- Differential structures (D): presence of 5 differential structures, i.e. pigment network, structureless or homogeneous areas, streaks, dots and globules.

Another common criterion for melanoma recognition is 7-point checklist rule [40]: (1) atypical pigment network, (2) blue-whitish veil, (3) atypical vascular pattern, (4) irregular streaks, (5) irregular dots/globules, (6) irregular blotches, (7) regression structures. It is important to note that distinguishing melanoma from other type of skin lesion is not always easy. One reason is that different dermatologists may disagree with their diagnosis of skin lesion, since the assessment of skin lesion highly depends on the visual acuity and the experience of the expert.

Automatic segmentation of skin lesion in dermoscopic images separates the lesion area from the background (i.e. normal healthy skin), which allows for the desired lesion feature extraction, like asymmetry, border irregularity, colors and diameter [43]. This then further enhances the diagnosis of melanoma disease that largely relies on the performance of the lesion segmentation stage in some clinical scenarios. In addition, segmentation of the skin lesion in CAD system assists dermatologists to reduce the inconsistency of melanoma diagnosis.

1.2 Motivations and Objectives

In order to achieve reliable semantic interpretation of medical image in CAD system, segmentation of target organs should be as accurate as possible. However, manual segmentation of medical image is a difficult task since it needs human experts to have the necessary prior knowledge on the anatomical feature of the target organ. Another critical aspect of manual segmentation is the time consumption of the experts on the annotation of the large medical data.

In the last decades, many approaches with lower cost and less human interaction than the observation-based techniques, have been developed for automatic medical image segmentation. Although they have displayed the impressive success in

medical image processing, challenges still exist in fundus image and dermoscopic image segmentation. More specifically, the challenge for the task of fundus image segmentation derives from the variations of vessel caliber, the interference of noises from uniform illumination and inter-camera variability, the existence of other structures (e.g. irregular shaped lesions, optic disc and pathologies), and the low contrast between retinal vessel and background. For retinal vessel segmentation, we design two multistage classification architectures to segment retinal vessel from background, i.e. a cascade classification framework, and a divided-and-conquer funnel-structured classification framework. Both frameworks achieve more accurate retinal vessel segmentation on three different datasets than the existing state-of-the-art approaches.

Recently, deep learning based image segmentation has attracted a lot of research attention. Its strength has been shown in many medical image segmentation tasks. However, the lack of publicly available retinal vessel images limits the learning ability of the deep learning networks on retinal vessel segmentation. For example, three publicly available datasets DRIVE, STARE and CHASE_DB1 only provide 40, 20 and 28 retinal images. Hence, before training the deep neural network for retinal vessel segmentation, data augmentation by cropping the retinal image into different patches is needed. One limitation of the patch-based deep learning methods is that it produces a lot of redundant information in the available data. In addition, the optimal patch size is also a difficult problem because the large patch size provides more contextual information but leads to low localization ability of the network. We have tried to apply a fully convolutional network for retinal vessel segmentation, which produces the results without visible improvement. To investigate the ability of the deep learning methods on medical image segmentation, we take dermoscopic image segmentation as our second task. There are two publicly available datasets ISBI 2016 and ISBI 2017, which provide more than 2000 dermoscopic images. Those dermoscopic images have various skin lesions i.e. different color, shape and texture characteristics. The large dermoscopic datasets are beneficial to the investigation of the application of deep learning method on medical

image segmentation.

1.3 Research Contributions

The research in this thesis focus on the development of segmentation system for two modalities of medical images. Main contributions of this thesis are stated as follows:

- A novel and robust cascade classification framework is proposed for the retinal vessel delineation. This classification model envelops a set of computationally efficient Mahalanobis distance classifiers to form a highly nonlinear decision. In contrast to other nonlinear classifiers that need a predefined nonlinear kernel or need iterative training, the proposed cascade classification framework is trained through a one-pass feed forward process, which generates a definite training and testing result that can be exactly reproduced. In addition, the degree of nonlinearity of the proposed cascade classification system is not predefined but determined by the complexity of the data structure. Thus, it is an adaptive solution for medical image segmentation that can be flexibly extended to other pixel-wise segmentation tasks. Experimental evaluations on three publicly available datasets demonstrate the effectiveness of the proposed cascade classification architecture on retinal vessel segmentation. Qualitative and quantitative comparisons with other retinal vessel segmentation approaches show the superiority of the proposed framework on the detection of the challenging retinal structures.
- To improve the segmentation performance of retinal vessels with different scales and positions, a hierarchical architecture based on the divide-and-conquer strategy is proposed. Previous works for retinal vessel segmentation depend on the training a global discriminative model that treat all type of retinal vessel as equal regardless of the geometrical structure difference

among retinal vessel with different scales and positions. The proposed architecture overcomes this problem by exploiting a multiplex vessel partition (MVP) algorithm that separates retinal vessel pixels into well constrained subsets. It enables retinal vessels that have the same geometrical property are grouped to the same subsets, which benefits the classifier to form a discriminative decision for different retinal vessel pattern. Moreover, we propose a funnel-structured vessel segmentation (FsVS) framework to fuse the classification results from each disjoint subset. It not only adjusts the deviation of the retinal vessel partition but further enhances the discriminative capability of the decision model. Performance evaluations on three publicly datasets validate the effectiveness of the proposed segmentation framework for retinal vessels with different scales and positions.

- A bi-directional dermoscopic feature learning framework with multiscale consistent decision fusion is proposed, which achieves the accurate skin lesion segmentation from dermoscopic images. Current published works utilize the deep learning based models for skin lesion segmentation because of their powerful strength on discriminative feature learning for skin lesion structure. Nevertheless, skin lesion segmentation is still a difficult task that needs the investigation of the relationship between skin lesions and their informative context, as well as the consistency of the decision from multiple classification layers. To address this issue, we propose a bi-directional dermoscopic feature learning (biDFL) framework to generate a substantially rich and discriminative feature representation by integrating skin lesions with their informative context. By manipulating the feature propagation through two complementary directions among high-level layers, the proposed biDFL makes the network to construct a rich feature representation of the skin lesion. Furthermore, a multiscale consistent decision fusion (mCDF) is proposed to selectively focus on the informative decisions generated at different classification layers. By the consistency analysis of the decision in a local region, the proposed mCDF adaptively adjusts the reliability of decision and thus

produces more precise skin lesion prediction. Experimental evaluations on two publicly dermoscopic image datasets show the validity of the proposed network on skin lesion segmentation.

1.4 Thesis Outline

The organization of the rest of the thesis is shown as follows:

In Chapter 2, we briefly review the related works on medical image segmentation, in terms of the previous works using unsupervised and supervised methods, as well as deep learning on fundus image and dermoscopic image segmentation.

In Chapter 3, a novel and robust cascade segmentation framework is proposed to detect retinal vessel from fundus image. We also investigate the enhancement of fundus image quality for preprocessing, and the fusion with asymmetric principal component analysis (APCA) dimensionality reduction for feature extraction.

In Chapter 4, a multiplex vessel partition (MVP) strategy is presented to divide retinal vessels with the same geometrical property into the well constrained subsets. A funnel-structured vessel segmentation framework is further proposed for the enhancement of the retinal vessel delineation accuracy.

In Chapter 5, a bi-directional dermoscopic feature learning framework is proposed for the feature learning of the complex structure of skin lesion. Moreover, a multiscale consistent decision fusion is also proposed for the improvement of the reliability of the decision on skin lesion segmentation.

In Chapter 6, the conclusions and future research works are identified.

Chapter 2

Literature Review

In this chapter, we first briefly review the medical image segmentation based on unsupervised methods. Then we introduce the existing supervised methods for medical image segmentation. At last, we review the application of deep learning on medical image segmentation.

2.1 Unsupervised Segmentation

The unsupervised segmentation methods on medical image provide an automatic way to subdivide a medical image into different components like target organ or background based on their inherent patterns and characteristics. Popular unsupervised methods for retinal vessel and skin lesion segmentation are presented as follows:

- Tracking: Tracking methods mainly start with the selection of seed points (automatically or manually), from which a model of retinal vessel is recursively expanded. In [44], Tolias *et al.* located the vessel seed pixels around the region of the optic nerve, and then tracked vessel pixels by a clustering method. Erilk *et al.* tracked vessel edges through a biologically inspired multi-orientation analysis [45].

- **Thresholding:** Segmentation using thresholding enjoys a crucial position in applications of medical image analysis, because of its simplicity of implementation and validity of segmentation. Celebi *et al.* [46] combined four thresholding methods for skin lesion delineation. Yüksel and Borlu [46] utilized automatic threshold determination with the type-2 fuzzy logic technique for skin boundary detection. Adaptive thresholding [10, 47] were applied to segment retinal vessel tree from fundus image.
- **Morphological operation:** The principal of morphological operation is to simplify the medical image data by preserving specific organ characteristics (such as boundary and skeleton) and removing other unrelated elements. Some classical morphological approaches for retinal vessel segmentation were described in [48, 49], where several basic morphological operators with different structure elements were introduced. In addition, the morphological component analysis (MCA) was adopted to separate retinal vessel and lesion in [50]. In [51], morphological operators were used to smooth the skin lesion segmentation.
- **Deformable model:** Segmentation methods based on deformable model detect target organ in medical image by minimizing the energy function defined within an image domain. Zhou *et al.* [52] applied a mean shift based gradient vector flow (GVF) algorithm for skin lesion segmentation. Ma and Tavares [53] simulated the process of the skin lesion segmentation with a geometric deformable model. “Ribbon of Twins” active contour model [54], region-scalable fitting based active model [55], infinite perimeter active contour model [56], were presented to detect blood vessel. In [57], the energy function formulated for retinal vessel segmentation was based on the geodesic active contour model where the edge indicator was replaced with the defined anisotropy approximation.

Other unsupervised methods, like multifractal analysis [58], multiresolution hermite model [59], trainable COSFIRE filter [11], and fuzzy c-means clustering [60,

61], were also applied for medical image segmentation.

2.2 Supervised Segmentation

Due to the critical property of medical image processing, segmentation of human organs should be accurate enough to support the further step of computer-aided disease diagnosis. Hence, many supervised segmentation algorithms are developed to mimic the strength of experts with prior knowledge about the organ characteristic. Supervised segmentation methods exploit the extracted features in the labeled training data to train a classification model which learns a set of rules to distinguish target organ from background. In this section, we review the feature extraction and classifier learning on medical image segmentation.

2.2.1 Feature Extraction

Feature vector contains quantifiable measures for classifier to recognize each candidate pixel as target organ or background. During the last decades, many feature descriptors have been developed to extract robust and discriminative information of the target organ in medical images. Feature descriptors for retinal vessel representation are illustrated as follows:

- Frangi filter: It is known that the geometrical structure of retinal vessels can be regarded as tubular, a powerful feature for delineating vessels within 2D image. Frangi filter [62], is a process that detects this tubular structure based on eigenvalue analysis of the Hessian matrix. More specifically, the eigenvalues of the Hessian matrix indicate the principal curvatures of the intensity surface at each pixel, and the sign of eigenvalues determines whether a pixel belongs to local maximum or minimum.
- Gray-level-based feature: Gray-level variation in a local region is also helpful for retinal vessel representation. For instance, five statistic measurements

based on gray-level information were formulated in [63], including intensity mean, standard deviation, differences between the pixel and the mean, maximum and minimum of the local area in fundus image.

- Matched filter: The kernel of the matched filter is usually applied to model the profile of the retinal vessel structure. For feature extraction with matched filter [64], fundus image was convolved with a set of filters at different orientations. The maximum convolution response among all angles is extracted as the feature representation.
- Difference of Gaussian (DoG): This feature descriptor interprets the relation of a pixel with respect to its surroundings. DoG kernels with one baseline scale and five different smoothing scales were designed for fundus image analysis [65].

Other feature descriptors like 2-D Gabor wavelet transformation [66], color histogram information [51], and gray-level co-occurrence matrix (GLCM) [67] were also applied for object feature representation in the tasks of retinal vessel and skin lesion segmentation.

2.2.2 Classifier Learning

Given a set of training samples (well represented by the feature vector), each marked as belonging to target organ or background, a classification model can learn the specific rule to assign a new test sample as one of two classes. Many classifiers have been developed for medical image segmentation:

- Support vector machine (SVM): SVM is one of the most popular classifiers in the segmentation of the medical image. It achieves class separation by building a hyperplane in a high-dimensional space that has the largest distance to the nearest training data of any class [68]. In [69], a 74-dimension texture feature vector was extracted and then fed into the support vector

machine for skin lesion segmentation. You *et al.* [70] classified retinal vessel feature vector derived from the steerable complex wavelet by SVM classifier. In [71], a SVM classifier was applied to learn the decision model for retinal vessel segmentation. Published works [9, 72–74], also exploited SVM for the delineation of retinal vessel tree.

- K-nearest neighbor (KNN): KNN is a non-parametric method used for classification, where an object is classified by a plurality vote of its neighbors, i.e. the object is assigned to the class most common among its k nearest neighbors measured by a distance function [75]. Niemeijer *et al.* [76] extracted feature vectors based on Gaussian filters, and used KNN classifier to determine the probability of a candidate pixel as retinal vessel. Staal *et al.* [77] computed the feature vector encoding the information of the line elements, and implemented retinal vessel classification by a KNN classifier based on sequential forward feature selection.
- Neural network (NN): NN is a type of learning algorithms based on the structure of human's brain on vision, which is constructed by lots of neural units connected with adjustable weights. In the training phase, ground truth is used to adjust these weights so that the whole network can generate proper filters for object segmentation at the output layer of the network. Many literatures apply neural network to classify the hand-crafted features for medical organ. For instance, Marin *et al.* [7] designed a 7-D feature vector encoding the gray-level information for retinal vessel classification by a neural network. Vega *et al.* [78] reported a lattice neural network with dendritic processing to classify feature vector extracted as [7]. Wang *et al.* [51] deployed 20 descriptive features for skin lesion segmentation with a neural network classifier.

Other supervised machine learning method using the extracted feature information were also proposed for medical image segmentation. For example, features in [6] were composed of the pixel intensities and 2-D Gabor wavelet transformation responses, followed by classification with a Gaussian mixture model (GMM)

classifier. Lupascu *et al.* [79] applied a feature-based AdaBoost classifier for vessel segmentation. Fraz *et al.* [66] constructed a 9-D feature vector for vessel pixel classification by an ensemble classifier with the boosted and bagged decision trees. In [80], a set of eight features based on pixel neighborhood and different-order gradient images were extracted, along with classification by GMM classifier. Barkana *et al.* [81] fused the classification results obtained from fuzzy logic, ANN and SVM by employing majority voting.

2.3 Deep Learning Segmentation

Deep learning based methods have demonstrated their superior performance in the area of image recognition and computer vision, such as semantic segmentation [82, 83], and image classification [84]. Segmentation task is different from category classification task since it needs to predict the class for each pixel of the input image and thus recover the context and location of the objects in the input image. Recently, deep learning based methods have also been successfully applied to medical image segmentation following their success in other semantic segmentation tasks, such as deep learning for multi-modal brain tumor segmentation [85, 86], gland segmentation [87, 88], liver tumor detection [89, 90], and bone segmentation [91, 92].

2.3.1 Fully Convolutional Network

Fully convolutional network (FCN) is one of the most commonly baseline model applied in the medical image segmentation [2, 93]. FCN contains a series of locally connected layers like pooling, unsampling, convolution and deconvolution. The construction of FCN system makes it work regardless of the input image size and the fixed number of units at each stage. In order to complete the segmentation task, two crucial paths of FCN are needed [2]:

FIGURE 2.1: FCN network for semantic segmentation [2].

- Downsampling path: capture semantic information and interpret the context.
- Upsampling path: recover spatial information and localize the object.

Fig. 2.1 shows the fully convolutional network in [2]. Feature map of each layer is a three-dimensional array of size $\tilde{h} \times \tilde{w} \times \tilde{c}$. Here, $\tilde{h}$ and $\tilde{w}$ are spatial dimensions, and $\tilde{c}$ is channel dimension. Suppose $x_{i,j}$ for the data vector at location (i, j) in a particular layer, the output $o_{i,j}$ is computed by [2]

$$o_{i,j} = f_{k,s}(\{x_{si+\delta_i, sj+\delta_j}\}_{0 \leq \delta_i, \delta_j \leq k}) \quad (2.1)$$

where k represents the kernel size. s is the stride or subsampling factor. $f_{k,s}$ denotes the layer type, like a matrix multiplication for convolution, a spatial max for pooling, or an activation function. Given v as the output through the deconvolution network, the segmentation loss of FCN architecture can be formulated as [94]

$$L = - \sum_{p \in S} \sum_{h=1}^c y_h^p \log(v_h^p), \quad (2.2)$$

where y_h^p is the binary label indicator for the pixel located at position p , and p ranges over all the spatial position S . $h \in 1, 2, \dots, c$ and c is the number of class. The architecture of FCN eradicates the need of hand-crafted feature extraction and

leads to an end-to-end training system. In addition, it also automatically learns discriminative features from raw image and simultaneously achieves the objective of the required segmentation task.

2.3.2 Application on Medical Image Segmentation

The great success of FCN on semantic segmentation is primarily attributed to the capability of FCN that learns the feature representation for a high-level semantic information. Nevertheless, a series of consecutive pooling and convolution striding operations in FCN reduce the feature map resolution [2]. In order to alleviate the effect of this issue on the medical image segmentation, a variety of more advanced FCN-based methods have been developed. Fig. 2.2 shows some examples of those advanced FCN-based architectures for medical image segmentation.

In the literature [3] (Fig. 2.2 (a)), the authors proposed a deep fully convolutional residual network (FCRN) for medical image segmentation. To exploit the object-level features and the low-level spatial information in multiple layers of the network, they integrated multiscale contextual information at different layers in FCRN system. This designation enhanced the network capability for the delineation of object boundary.

In the literature [4] (Fig. 2.2 (b)), the authors trained the ResNet [84] based FCN model by a deep class-specific learning. In contrast to the traditional FCN based methods that apply a single-class information, this method quantified the subtle visual property of different types of objects and integrated complementary class-specific information, thus enhancing skin lesion segmentation performance for both melanoma and non-melanoma segmentation.

In the literature [5] (Fig. 2.2 (c)), Bi *et al* applied a multistage FCN (mFCN) approach to improve the accuracy and consistency of medical image segmentation. Multistage FCN has ability of learning from the training data as well as the

of using the channels from Hue-Saturation-Value color space were also discussed. Al-masni *et al.* [98] proposed a full resolution convolutional networks (FrCn) that eliminated the subsampling layers in the networks and made the convolutional layers to learn the full spatial features of medical image data. Inspired by RefineNet [99], Li *et al.* [100] devised a dense deconvolutional network (DDN) for skin lesion segmentation based on residual learning.

For the segmentation task with insufficient medical image samples, like retinal vessel segmentation from fundus image, deep learning based methods mostly focus on learning feature from the separated patches. For example, DRIVE or STARE dataset only contains 20 fundus images for retinal vessel segmentation, which provides insufficient image samples to train a deep learning model. To interpret the detailed characteristic of retinal vessel tree, many methods have divided each image into a large amount of patches and then train the deep classification model. In [101], the authors proposed a deep neural network to indicate the relationship between retinal image and vessel map. Liskowski *et al.* [102] applied a convolutional neural network with structured prediction to classify each pixel to be retinal vessel or not. Yan *et al.* [103] trained the deep learning network for retinal vessel segmentation by jointly adopting both the segment-level and the pixel-level losses. Oliveira *et al.* [104] delineated the retinal vessel tree by the combination of the stationary wavelet transform with a multiscale fully convolutional neural network. Hu *et al.* [105] utilized a multiscale convolutional neural network with an improved cross-entropy loss for retinal vessel detection. Yan *et al.* [106] separated the vessel segmentation task into three sub-tasks, and each of them was trained by a deep learning model with a unique pixel-wise loss. Yu *et al.* [107] hierarchically divided retinal vessel tree based on the segmentation result from deep neural network. Feng *et al.* [108] segmented retinal vessel with a cross-connected convolutional neural network (CcNet).

2.4 Summary

In this chapter, we briefly review the works related to medical image segmentation. Unsupervised segmentation methods are implemented easily but have the limitation on the segmentation accuracy of the target organ. The performance of supervised methods is better than those unsupervised methods. The advantage of supervised methods over unsupervised methods is mainly from the knowledge of organ segmentation gained from the manual segmentation by experts. Deep learning based methods have displayed their powerful strengths on medical image segmentation. Though there exist some limitations of those deep learning based techniques on medical image processing, like the scarcity of medical data, deep learning based methods with more sophisticated architectures are still popular investigation trends in medical image segmentation.

Chapter 3

Cascade Classification Framework for Fundus Image Segmentation

In our first work, we focus on retinal vessel segmentation from fundus images. A new cascade classification framework that envelops a set of computationally efficient Mahalanobis distance classifiers, is introduced for this task. In addition, we also present the enhancement of fundus image quality for preprocessing, the fusion with asymmetric principal component analysis (APCA) dimensionality reduction for feature extraction in this chapter.

3.1 Introduction

In order to facilitate semantic interpretation of fundus images, one of the most important technique is the retinal vessel segmentation [50, 109]. As stated in Section 2.4, automatic segmentation system with different learning algorithms have been developed for accurate retinal vessel detection. High performance of retinal vessel segmentation most relies on the generation of the decision boundary through supervised learning at one stage classification, such as SVM [9, 70] and neural network [7, 78].

Though those approaches have displayed strength on fundus image segmentation, accurate retinal vessel detection remains a very difficult problem due to many factors. First, the non-perfect imaging environment causes poor quality of retinal images. To tackle this problem, we propose to enhance the image by utilizing some prior knowledge of retinal image such as the small percentage of vessel pixels in the image and that a vessel can be a curve as narrow as just one pixel width. Second, among various features developed in the previous works, it is difficult to judge which features are better than the others due to the complex vessel shape. Our strategy in this chapter is utilizing as many features as possible in all 3 color channels and applying an effective dimensionality reduction technique to reduce the number of features before classification.

Moreover, both vessel and background pixels are distributed much more complex than Gaussian so that the accurate segmentation of them is a highly nonlinear classification problem. Although the SVM classifier can handle non-Gaussian data, its training time is extremely long if a large number of samples are used in training. Thus, the previous approaches only use a very small portion of data in SVM training, which limits the classification performance and its certainty. Although some nonlinear classifiers can form a very complex decision boundary, such as neural networks, it is difficult to solve under-fitting (trapped in local minimum) and over fitting problems. They also need predefined network size and iterative training, whose classification results are significantly different from different network size and initialization. In this chapter, we propose a novel cascade classification network to solve these problems. The network size is not predefined but determined by the complexity of the training data. It is trained by a one-pass feed forward process, which produces a definite training and testing result that can be exactly reproduced. This greatly reduces the uncertainty of its usage in the both training and testing and hence it is user-friendly. Last, pathological region is a challenging issue for accurate retinal vessel classification. This chapter also suggests a pathological region removal method based on the vessel geometric properties for postprocessing.

3.2 Proposed Framework

We design a novel cascade classification framework for retinal vessel segmentation where the network is constructed and trained stage by stage according to the complexity of the training data. The core of this framework is to construct and train a powerful cascade classification network which classifies a feature vector fused from all 3 color channels with dimensionality reduction. The framework includes: (a) image preprocessing; (b) color channel fusion and feature dimensionality reduction; (c) constructing and training cascade classification network; (d) postprocessing for pathological region removal.

3.2.1 Retinal Image Preprocessing

Retinal images often have poor quality resulted from non-ideal imaging environment during image acquisition process [73]. This may lead to a considerable bias to retinal vessel segmentation. Two prominent causes leading to poor image quality are intensity fluctuation and random noise. Thus, we propose two steps of image enhancement targeting at alleviating these two problems: 1) background normalization; 2) detail preserving noise removal.

A. Background Normalization

We propose a background normalization to alleviate the image intensity fluctuation. With the prior knowledge of retinal image that vessel pixels take only a small portion of the image (around 12%), a median filter will capture the value of background if the filter size is large enough. Therefore, the background of retinal image can be normalized by the median filter. Let $I(x, y)$ be the original image, the background-normalized image $\bar{I}(x, y)$ is computed by

$$\bar{I}(x, y) = \frac{I(x, y)}{\text{median } I(x - n_1, y - n_2)}, -\omega \leq n_1, n_2 \leq \omega, \quad (3.1)$$

where ω is the window size of median filter. The median filter removes the retinal vessels, thus generates an intensity map that is close to retinal image background. By dividing the original retinal image with this background map, the image intensity fluctuation is considerably reduced.

B. Details Preserving Noise Removal

Fundus images often contain some isolated spots resulted from artificial noises during medical imaging process. However, there are very thin vessels in the retinal image (as thin as just one pixel width) that should be preserved from filtering. Thus, we apply an image detail-preserving filter, truncation filter, which attenuates isolated noise inside an enclosed surrounding band but preserves image details that stretch to this band. For each pixel (x, y) , we define B_k as the band between the inner window WI_k that covers the pixel (x, y) with the size of $W \times W$ and the outer window WO_k that centers as WI_k with the size of $(W + \delta) \times (W + \delta)$, $\delta \geq 1$. Here, $k = 1, 2, \dots, N$ ($N = W \times W$), denote N enclosed square surrounding bands of a pixel (x, y) . Let u_k and v_k be the maximal and minimal intensity values of a surrounding band B_k that encloses a pixel (x, y) , respectively, i.e.

$$u_k = \max_{x,y}(\bar{I}(x, y) | (x, y) \in B_k), \quad (3.2)$$

$$v_k = \min_{x,y}(\bar{I}(x, y) | (x, y) \in B_k). \quad (3.3)$$

The output of the truncation filter of the pixel is expressed as [110]:

$$\hat{I}(x, y) = \begin{cases} \mu_{min}, & \text{if } \bar{I}(x, y) > \mu_{min}, \\ \nu_{max}, & \text{if } \bar{I}(x, y) < \nu_{max}, \\ \bar{I}(x, y), & \text{otherwise,} \end{cases} \quad (3.4)$$

where $\mu_{min} = \min_{x,y}(\mu_k | k = 1, 2, \dots, N)$ and $\nu_{max} = \max_{x,y}(\nu_k | k = 1, 2, \dots, N)$. Fig. 3.1 shows an example of background normalization and truncation filtering. A pixel (x, y) and three of its nine square surrounding bands for $W = 3$ and $\delta = 1$ are displayed as Fig. 3.1 (a). Background normalization reduces the image intensity fluctuation as shown by the dark background region (white arrow) in Fig. 3.1 (b).

FIGURE 3.1: Example of background normalization and truncation filtering. (a) Three sample square surrounding bands (orange, green, purple) of pixel (x, y) (red) for $W = 3$ and $\delta = 1$. (b) Original sub-image. (c) Sub-image after background normalization. (d) Sub-image after truncation filtering.

Truncation filter attenuates noise spot (in yellow circles in Fig. 3.1 (b-c)) enclosed by any one of its surrounding bands and preserves vessel curve of even one pixel width as long as it stretches to its surrounding bands [110].

3.2.2 Color Channel Fusion with Dimensionality Reduction

Feature vector contains quantifiable measures for classifier to recognize each pixel as retinal vessel or background. In this work, we evaluate our proposed method with features that have been widely used in the literature of retinal image analysis, including Gabor wavelet transformation (total 8 features) [66], matched filter (total 32 features) [111], gray-level-based feature (total 35 features) [7], Frangi filter (total 20 features) [62] and Difference of Gaussian (total 5 features) [65].

Although some of the above described 100 features may be more effective than others, none of them is totally useless for vessel segmentation. Most previous approaches work on the green channel as red and blue channels show very low contrast and deficiency for vessel segmentation. We utilize all the 100 features on

all 3 color channels and applies a dimensionality reduction technique to select more reliable features for classification.

Suppose the 100 features extracted in red, green and blue channels are normalized and represented by vectors $\mathbf{f}_r$, $\mathbf{f}_g$ and $\mathbf{f}_b$, respectively. Here, each feature has been normalized to have zero mean and unit variance over the field of view (FOV) of its image as done in [6]. Then features of the 3 color channels are weighted and concatenated into a 300-dimensional feature vector $\mathbf{x}$ as

$$\mathbf{x} = [w_r \mathbf{f}_r^T \quad w_g \mathbf{f}_g^T \quad w_b \mathbf{f}_b^T]^T. \quad (3.5)$$

This work chooses $w_r = 1$, $w_g = 2$ and $w_b = 0.5$ to make the green channel higher priority and blue channel lower priority in the following feature selection.

This 300-dimensional feature vector $\mathbf{x}$ collects rich information for retinal vessel segmentation. However, it inevitably embeds some redundant information that results in some unreliable dimensions, which adversely affects the robust classification of retinal vessel. Principal component analysis (PCA) is an effective tool to remove the unreliable dimensions for classification [112, 113]. However, the data of the retinal image pixels are highly asymmetric, since the number of vessel samples is much smaller than that of background samples, e.g. only 9%-14% of pixels of image are retinal vessel [114]. This makes PCA inefficient in removal of the unreliable dimensions because PCA is an unsupervised method that works on the whole data without awareness of the data class membership [115]. To remove the unreliable dimensions of feature vector effectively, we introduce the asymmetric principal component analysis (APCA) [116].

Let $\mathfrak{D}_v$ and $\mathfrak{D}_b$ be two training data sets that contain N_v and N_b samples of vessel and background, respectively. The covariance matrices of $\mathfrak{D}_v$ and $\mathfrak{D}_b$ are computed by:

$$\Sigma_v = \frac{1}{N_v} \sum_{\mathbf{x} \in \mathfrak{D}_v} (\mathbf{x} - \boldsymbol{\mu}_v)(\mathbf{x} - \boldsymbol{\mu}_v)^T, \quad (3.6)$$

FIGURE 3.2: Flowchart of feature extraction, color channel fusion and dimensionality reduction.

and

$$\Sigma_b = \frac{1}{N_b} \sum_{\mathbf{x} \in \mathcal{D}_b} (\mathbf{x} - \boldsymbol{\mu}_b)(\mathbf{x} - \boldsymbol{\mu}_b)^T, \quad (3.7)$$

where $\boldsymbol{\mu}_v$ and $\boldsymbol{\mu}_b$ denote the class-conditional mean of samples of data sets $\mathcal{D}_v$ and $\mathcal{D}_b$, respectively. The covariance mixture of the asymmetric PCA is computed by

$$\Sigma_\epsilon = \tau \Sigma_v + (1 - \tau) \Sigma_b + \Sigma_s, \quad (3.8)$$

where Σ_s is the covariance matrix of the two class means. τ is a weighting factor which is determined by the reliability of Σ_v and Σ_b . If $\tau = N_v/(N_v + N_b)$, Σ_ϵ will be the same as the total scatter matrix used by PCA, which puts unduly high weightage on the background class. To remove the unreliable dimension effectively, we take $\tau = 0.5$, much higher value than $N_v/(N_v + N_b)$ (from 0.09 to 0.14) used in PCA.

Eigen-decomposition of the covariance mixture Σ_ϵ is performed

$$\Sigma_\epsilon = \Phi \Lambda \Phi^T, \quad (3.9)$$

where Φ and Λ contain the eigenvectors and eigenvalues of Σ_ϵ , respectively. By selecting m eigenvectors Φ_m corresponding to the m largest eigenvalues, we project a d -dimensional feature vector $\mathbf{x}$ to a m -dimensional feature vector $\mathbf{z}$:

$$\mathbf{z} = \Phi_m^T \mathbf{x}, m < d. \quad (3.10)$$

By dimensionality reduction with the above APCA, 300 features from red, green and blue channel are fused into a more reliable and compact 100-dimensional vector. Fig. 3.2 shows the flowchart of our proposed feature extraction approach, where features extracted from three different color channels are weighted and fused together by asymmetric PCA.

3.2.3 Cascade Classification Network

Retinal vessel segmentation is performed by classifying a pixel represented by feature vector $\mathbf{z}$ to either vessel Ω_v or background Ω_b . The minimum probability of misclassification is achieved by assigning $\mathbf{z}$ to one class Ω_c that has the maximum posterior probability $P(\Omega_c|\mathbf{z})$, where c indexes v or b . Since the posterior probability $P(\Omega_c|\mathbf{z}) = P(\Omega_c)p(\mathbf{z}|\Omega_c)p^{-1}(\mathbf{z})$ and $p(\mathbf{z})$ is not a function of Ω_c , the discriminant function can be defined as

$$g_c(\mathbf{z}) = \ln p(\mathbf{z}|\Omega_c) + \ln P(\Omega_c). \quad (3.11)$$

Under the consideration that $p(\mathbf{z}|\Omega_c)$ is a multivariate Gaussian distribution, maximizing the discriminant function (3.11) is equivalent to minimize the Mahalanobis distance [115]

$$d_c(\mathbf{z}) = (\mathbf{z} - \bar{\mathbf{z}}_c) \Sigma_c^{-1} (\mathbf{z} - \bar{\mathbf{z}}_c)^T, \quad (3.12)$$

with a proper threshold. Here $\bar{\mathbf{z}}_c$ and Σ_c are respectively the class mean and covariance matrix. After applying eigen-decomposition, Eq. (3.12) is simplified as [115]

$$d_c(\mathbf{z}) = \sum_{k=1}^m \frac{(\mathbf{h}_k - \bar{\mathbf{h}}_k)^2}{\lambda_k}, \quad (3.13)$$

where $\mathbf{h}_k$ and $\bar{\mathbf{h}}_k$ are respectively the projections of $\mathbf{z}$ and $\bar{\mathbf{z}}_c$ on the eigenvector corresponding to the eigenvalue λ_k of Σ_c . We omit the class index in $\bar{\mathbf{h}}_k$ and λ_k for symbolic simplicity.

Since eigenvalue λ_k is an estimate of the class population variance based on the available training data, the decision rule from Eq. (3.13) will overfit the training samples if eigenvalue λ_k deviates from the population variance, which leads to a poor generalization or prediction on the testing data. To compensate the effect of the deviation of the eigenvalues, a regularized eigenspectrum $\hat{\lambda}_k$ is given by [117]

$$\hat{\lambda}_k = \begin{cases} \lambda_k, & k < r, \\ \frac{\alpha}{k+\beta}, & r \leq k \leq m, \end{cases} \quad (3.14)$$

where

$$\alpha = \frac{\lambda_1 \lambda_r (r-1)}{\lambda_1 - \lambda_r}, \quad \beta = \frac{r \lambda_r - \lambda_1}{\lambda_1 - \lambda_r}.$$

Eq. (3.13) with the regularized eigenspectrum $\hat{\lambda}_k$ is used for classification. If we define

$$\Delta d(\mathbf{z}) = d_v(\mathbf{z}) - d_b(\mathbf{z}), \quad (3.15)$$

the classifier makes decision with a threshold T_o as

$$\mathbf{z} \in \begin{cases} \Omega_v, & \text{if } \Delta d(\mathbf{z}) < T_o, \\ \Omega_b, & \text{if } \Delta d(\mathbf{z}) \geq T_o. \end{cases} \quad (3.16)$$

This is a nonlinear classifier, which forms a quadratic decision boundary to minimize the classification error if the feature vector is class-conditional Gaussian distributed. However, the real distribution of the feature vector is much more complex

FIGURE 3.3: An illustration of a binary classification by the threshold T_o and a three-state classification by two thresholds T_v and T_b .

than Gaussian, which makes the classifier not optimal. To solve this problem, we propose to find two thresholds T_v and T_b to classify the data into 3 classes, vessel Ω_v , background Ω_b and an uncertain class Ω_x as

$$\mathbf{z} \in \begin{cases} \Omega_v, & \text{if } \Delta d(\mathbf{z}) < T_v, \\ \Omega_b, & \text{if } \Delta d(\mathbf{z}) > T_b, \\ \Omega_x, & \text{if } T_v \leq \Delta d(\mathbf{z}) \leq T_b. \end{cases} \quad (3.17)$$

As illustrated in Fig. 3.3, by allowing an uncertain region, we can have much higher classification accuracy in the two regions $\Delta d < T_v$ and $\Delta d > T_b$ if T_v and T_b are far away from T_o . Feature vectors falling in the uncertain region, $T_v \leq \Delta d \leq T_b$, are further classified by a new classifier trained by the training samples falling in the same region. This process continues to construct a cascade classification network until the final binary classification generate a satisfactory result.

In the training procedure, classifier of each stage learns from the training samples falling in the uncertain region of the previous stage and partitions its training

samples $sp(\mathbf{z})$ into 3 data sets $\mathfrak{D}_v$, $\mathfrak{D}_b$ and $\mathfrak{D}_x$, i.e.

$$sp(\mathbf{z}) \in \begin{cases} \mathfrak{D}_v, & \text{if } \Delta d(\mathbf{z}) < T_v, \\ \mathfrak{D}_b, & \text{if } \Delta d(\mathbf{z}) > T_b, \\ \mathfrak{D}_x, & \text{if } T_v \leq \Delta d(\mathbf{z}) \leq T_b. \end{cases} \quad (3.18)$$

This training procedure constructs a cascade classification network as shown in Fig. 3.4. The classifier of the first stage learns from the whole training data and that of the last stage classifies its training samples into two classes.

For each cascade classification stage i except for the last stage, we adjust the two thresholds $T_{v,i}$ and $T_{b,i}$ so that the classification error rates in $\mathfrak{D}_i^v$ and $\mathfrak{D}_i^b$ are lower than two predefined values e_b and e_v . Clearly, smaller values of e_b and e_v generate deeper classification network. The network size also depends on the complexity of the training data. For easily classified data, training converges to a few stages while a very deep network is required for difficultly classified data. Training samples are only accessed to compute the covariance matrices and their Mahalanobis distances to the two class means to determine the thresholds. This training process is a non-iterative one-pass feed forward process, thus the large number of training samples won't result in the significant increase of time consumption and memory usage in constructing and training our proposed classification framework. The cascade classification network produces a definite training and testing result that can be exactly reproduced. This greatly reduces the uncertainty of its usage in the both training and testing and hence it is user-friendly.

For prediction, starting from the first stage, a pixel represented by its feature vector $\mathbf{z}$ is classified into one of the three classes (Ω_v , Ω_b or Ω_x). It will be passed to the next stage for further classification only if it is assigned to the class Ω_x by the current stage. Once the pixel is assigned to either class Ω_v or Ω_b at any stage, the classification process is completed. Fig. 3.5 shows this prediction procedure of the proposed cascade classification network. In fact, a large portion of pixels are recognized rapidly in the first few stages while a very small number of pixels that are difficult to be distinguished go through the last stage. Fig. 3.6 shows the

FIGURE 3.4: Constructing and training the cascade classification network. $\mathfrak{D}_i^v$ and $\mathfrak{D}_i^b$ are classified vessel and background sets of training samples in the i^{th} stage. $\mathfrak{D}_i^x$ is the uncertain set to train the classifier of the $(i+1)^{th}$ stage.

classification error rates against the number of stages of the proposed cascade classification network on the training, validation and testing data sets of CHASE_DB1 dataset. It shows that the proposed cascade classification network greatly enhances the accuracy from the single stage of classifier.

The classifier segments a test image into a vessel map I_o where vessel pixels have value one and background zero. For the problem of retinal vessel segmentation, the best overall accuracy is achieved at the operation point where the error rate of vessel pixels e_{vo} is much higher (around 12 times higher) than that of background because the retinal image has much more background pixels than vessel. This results in large number of vessel pixels missing in the classification result. We can increase the classification thresholds to detect more vessel pixels at a price of higher segmentation error of background. Let $T_{x,i}$ denote the thresholds in stage i , $T_{v,i}$, $T_{b,i}$ or $T_{o,M}$ in the last stage. Each of them increases by

$$T_{x,i}^\gamma = T_{x,i} + \gamma(T_{b,i} - T_{v,i}), \quad (3.19)$$

where γ is a parameter that is determined so that the network produces a certain smaller error rate of vessel $e_{v\gamma}$ on validation data, $e_{v\gamma} < e_{vo}$. With a slightly increased threshold, $\gamma = \gamma_1$, the classification network will segment some more vessel pixels in the vessel map I_{γ_1} such as very thin and low contrast vessels. We can extract the skeleton of this vessel map $S[I_{\gamma_1}]$ to remove the spurious vessel

FIGURE 3.5: Prediction procedure of the proposed cascade classification network.

pixels while keep the thin and low contrast vessels that are missing in I_o . Thick vessels appear being smooth ridges while the detected vessels in I_o could be rough as the vessel is detected pixel by pixel separately. With a further increased threshold, $\gamma = \gamma_2$, $\gamma_2 > \gamma_1$, the classification network will segment much more vessel pixels in the vessel map I_{γ_2} to increase the vessel width. The erosion of it, $E[I_{\gamma_2}]$, smooths the vessel edges and removes small and isolated spurious vessels. The final vessel map I_{out} is then the union of them

$$I_{out} = I_o \cup S[I_{\gamma_1}] \cup E[I_{\gamma_2}]. \quad (3.20)$$

FIGURE 3.6: Classification error rates against the number of stages of the proposed cascade classification network on the training, validation and testing data sets of CHASE_DB1 dataset.

3.2.4 Postprocessing

Pathological region with irregular shape is a challenging issue for accurate retinal vessel segmentation [74, 114]. In this work, we propose a pathological region removal method based on the vessel geometric properties. Firstly, we group the detected vessel pixels into connected components and work on each component. Vessels are smooth curves and so are their skeletons. Thus, the ratio r_s of the number of the branch points to that of the total pixels of the vessel skeleton is small. A pathological component is detected if it is higher than a threshold, $r_s > T_s$, as shown in Fig. 3.7 (c). Moreover, we can formulate the positions of pixels in each connected component as an ellipse. If a vessel stretches out, the ratio r_λ between the smaller and the larger eigenvalues of the covariance matrix of pixel locations is small. Even if otherwise the vessel forms a round path or cross shape, the ratio r_c of the number of pixels on the vessel to that in its convex hull is small. Thus, a pathological component is detected if both of them are larger than some thresholds, $r_\lambda > T_\lambda$ & $r_c > T_c$, as shown in Fig. 3.7 (g). For the pathological regions connected to vessel, we first break the connection between the pathological regions and vessel tree by morphological erosion operation, and then apply dilation with the same structure element as that used in erosion, followed by the above process on each significant connected component. Fig. 3.7 shows an example of our postprocessing procedure.

3.3 Experimental Evaluations

3.3.1 Materials

The proposed framework is tested on three publicly available datasets that contain digital color retinal images with 8 bit per color channel:

- The DRIVE dataset [77] comprises 20 training and 20 testing images of size 565×584 pixels, captured by Canon CR5 nonmydriatic 3 charge-coupled-device (CCD)

FIGURE 3.7: An example of the postprocessing procedure. (a) A subregion of retinal image with lesions. (b) Result of cascade classification network. (c) The skeleton of (b) (red pixels indicate the branch points). (d) The detected lesions. (e) Result after removing lesions. (f) Erosion result of (e). (g) Convex hull of the dilated significant connected component of (f). (h) The detected lesions. (i) Final result after lesion removal.

cameras at 45° field of view. Vessel pixels were manually segmented by three observers trained by an experienced ophthalmologist. The testing set was segmented twice, resulting in 577 649 pixels marked as vessel in set A, against 556 532 in set B. In this chapter, set A is used as the ground truth for the performance evaluation.

- The STARE dataset [118] contains 20 images of size 700×605 pixels, captured by a TopCon TRV-50 fundus camera at 35° field of view. Here, ten images with pathologies are included in STARE dataset. Two observers manually segmented all the retinal images in the STARE dataset. The first observer marked 615 726 pixels

as vessel, against 879 695 for the second observer. Results of the first observer are used as the ground truth.

- The CHASE_DB1 dataset [66] consists of 28 images of size 960×999 pixels, captured from 14 school children by a hand-held Nidek NM-200-D fundus camera at 30° field of view. Segmentation results of the first of the two observers are deployed as the ground truth.

For the DRIVE dataset, the training and testing images were predefined in the dataset. For the STARE dataset, we apply the leave-one-out strategy: every image is tested by the classification model trained by the other 19 images [6, 9, 77, 80, 101, 102, 114]. For the CHASE_DB1 dataset, training is performed on the first 20 images and the remaining 8 images are used for testing [101]. For optimization of some parameters, a validation set is separated from the training set by randomly taking 10% training samples.

3.3.2 Parameter Settings

For image preprocessing, the value of ω of the median filter should not be smaller than the width of the thickest vessel. We take $\omega = 15$ for DRIVE and STARE datasets and $\omega = 25$ for CHASE_DB1 dataset as image resolution of CHASE_DB1 dataset is almost double as that of DRIVE and STARE datasets. The surrounding bound of the truncation filter consists of edges of square windows of size 7×7 and 9×9 for DRIVE and STARE datasets, and 13×13 and 15×15 for CHASE_DB1 dataset. For postprocessing, parameters are set empirically based on analysis of the geometrical structure (e.g. vessel width and tortuosity) of retinal vessel on 3 datasets. Parameters for DRIVE and CHASE DB1 datasets are fixed to be the same, i.e. $T_s = 0.09$, $T_\lambda = 0.3$ and $T_c = 0.65$. For STARE dataset, the thresholds T_λ and T_c are reduced to 0.275 and 0.475 because the vessel ground truth from the first observer is only 10%, much less than that of the second observer of 15% [77].

For each cascade classification stage i except for the last stage, the two thresholds $T_{v,i}$ and $T_{b,i}$ are automatically determined by the two predefined classification error rates on training data e_b and e_v . Larger values of e_b and e_v result in less classification stages but larger classification error on training data. Smaller values of e_b and e_v lead to higher classification accuracy on training data but more classification stages. The optimal values of e_b and e_v are dataset dependent. For a fair comparison with other methods, they are fixed at $e_v = 0.45\%$ and $e_b = 0.05\%$ in all experiments of this work. We choose e_v 9 times as large as e_b because the number of background pixels is 6 to 9 times higher than that of vessel and because at the operation point that generates the highest total accuracy, the final error rate of vessel is 10 to 14 times higher than that of background in most published papers. We construct and train the maximal 125 stages of classifiers and take the number of stages that produces the lowest misclassification rate on a validation set. The binary decision threshold $T_{o,M}$ of the last stage classifier is set to have the lowest misclassification rate on the validation set. The classification regularization parameter r starting from the first stage of 95 decreases to 20 of the 90th stage and is kept at 20 for all stages higher than 90. Fig. 3.6 shows the classification error rates against the number of the stages M of the classification network. It shows that the error rates decrease rapidly first and then tend to be stable with the increasing classification stages. As addressed earlier, at the thresholds for the lowest overall misclassification rate, the vessel missing detection rate is much higher than the background error rate. We set the parameter γ_1 and γ_2 to increase the thresholds so that we detect 6.5% and 10% more correct vessel pixels on validation set, respectively in vessel maps I_{γ_1} and I_{γ_2} than that in I_o .

3.3.3 Ablation Evaluation

Most previous approaches are evaluated by segmentation accuracy (ACC), sensitivity (SEN) and specificity (SPE). ACC is the ratio of the number of correctly segmented vessel and background pixels to the total number of pixels within the field of view (FOV), while SEN and SPE are the segmentation accuracies of vessel

and background, respectively. By changing the decision thresholds, we can always get very high Sen at a price of very low SPE or vice versa. Thus, ACC is the most critical performance indicator among the three while SEN and SPE are mainly for information only. Some recent works also evaluate their approaches using F1-score (F1) and Matthews correlation coefficient (MCC) [9, 11, 78, 114, 119]. F1 and MCC are effective metrics for evaluating the classification performance on the unbalanced dataset where positive and negative classes are highly asymmetric [9]. They are given by

$$F1 = \frac{2(TP/(TP + FP)) \times (TP/(TP + FN))}{TP/(TP + FP) + TP/(TP + FN)}, \quad (3.21)$$

and

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}, \quad (3.22)$$

where TP and TN denote the number of pixels correctly classified as vessel and background, respectively; FN and FP represent the number of pixels incorrectly classified as the background and vessel, respectively.

The proposed framework works on features suggested and utilized in several previous works. For high-dimensional feature vector, PCA is widely used for dimensionality reduction. Each stage of the proposed cascade classification network employs the essential Mahalanobis distance classifier. To test how much performance gain the proposed framework achieves, we take a single stage Mahalanobis distance classifier on the 300-dimensional feature vector used in this work and on the 100-dimensional feature vector reduced by the standard PCA as two baselines because the proposed framework is built on these two approaches. These ablation testing results on DRIVE, STARE and CHASE_DB1 datasets are shown in Table 3.1. It shows that PCA improves the performance, though marginally, consistently over all the 5 performance indicators and over all the 3 datasets. Comparing with the performance gain achieved by the standard PCA, Table 3.1 demonstrates that the proposed framework significantly enhances the retinal vessel segmentation performance consistently over all the 5 performance indicators and over all the 3 datasets.

FIGURE 3.8: Examples of segmentation results. The first column to the third column: a sample image and its segmentation results on DRIVE, STARE and CHASE_DB1 datasets. The first row to the last row: original image, ground truth, and segmentation results of SSMDC, PCA+SSMDC, and the proposed method. Green region: missing retinal vessels; Pink region: spurious retinal vessels.

TABLE 3.1: Performance comparison with two baselines (%)

dataset	Method	SEN	SPE	F1	MCC	ACC
DRIVE	SSMDC	59.89	97.50	67.47	64.30	92.71
	PCA+SSMDC	63.82	97.68	70.82	67.80	93.36
	Proposed method	76.48	98.17	80.93	78.51	95.41
STARE	SSMDC	59.73	96.95	63.78	60.50	93.07
	PCA+SSMDC	59.74	97.59	65.46	62.68	93.66
	Proposed method	75.23	98.85	81.25	79.58	96.40
CHASE_DB1	SSMDC	52.67	96.25	55.07	51.15	92.22
	PCA+SSMDC	55.28	96.83	58.89	55.35	92.98
	Proposed method	77.30	97.92	78.09	75.91	96.03

SSMDC: Single stage Mahalanobis distance classifier.

Fig. 3.8 shows some examples of segmentation results on the 3 datasets, which provides visual evidence of where the proposed framework enhances the performance from the two baselines. For example, the proposed method detects more retinal vessels correctly, especially retinal vessels located at the boundary. It better preserves integral geometrical structure of retinal vessel trees, where some thin vessels lost by SSMDC and PCA+SSMDC are detected in the vessel trees by the proposed method. Image in the second column contains some lesions that are hard to be distinguished by SSMDC and PCA+SSMDC, which shows that the proposed approach is also more robust to lesions.

In addition, this section investigates the time and space complexities of the proposed framework. The complexity of the proposed cascade vessel classification system is dominated by the computation of the discriminant function at each cascade stage. Given N_T training samples with feature dimensionality m , the time and space complexities of the proposed method are $O(MN_Tm^2)$ and $O(N_Tm + m^2)$ respectively. Since M is the number of the cascade stages that is much smaller than

N_T , the time complexity of the proposed method can be rewritten as $O(N_T m^2)$. The proposed method has advantages in dealing with the large-scale dataset, without the power increase of time and space complexities.

3.3.4 Comparison with Other Published Methods

3.3.4.1 Vessel Segmentation Performance

We compare our proposed framework with other state-of-the-art algorithms on the 3 datasets. Table 3.2 shows the comparison in terms of the segmentation accuracy, where the segmentation results of other methods are reported by their respective publications. Since the CHASE_DB1 dataset was built in 2012, no result on this dataset is available before this year. The evaluation performance is computed on pixels inside the FOV. For DRIVE dataset, like most previous works, we use the FOV provided by the dataset, while the FOV masks for STARE and CHASE_DB1 are computed by the same program as that used in [6, 64, 66, 101]. Note that different FOVs may produce significantly different results, which are not comparable. Therefore, Table 3.2 lists the available results that use the same FOV, including those whose employed FOVs are not clearly mentioned in their papers. Table 3.2 shows that the proposed framework delivers the highest segmentation accuracy, ACC, among all algorithms consistently for all the 3 datasets. Deep learning techniques have shown outstanding performance in semantic segmentation [120]. Compared to recent powerful deep learning works on vessel segmentation [101, 102], we also marginally outperform their segmentation accuracy.

As the retinal vessel segmentation is an unbalanced classification problem, besides the segmentation accuracy, some recent approaches are also evaluated in terms of F1 and MCC. Table 3.3 records the F1 and MCC scores of the proposed method and those of other algorithms that published them. Results of Hoover *et al.* [118], Soares *et al.* [6], Marin *et al.* [7], Nguyen *et al.* [8] are computed by the segmented images from their corresponding papers. Comparative performance in Table 3.3

TABLE 3.2: Performance comparison with other methods in terms of ACC (%)

Year	Methods	DRIVE	STARE	CHASE_DB1
Unsupervised				
2000	Hoover <i>et al.</i> [118]	N.A	92.67	N.A
2006	Mendonca and Campilho [121]	94.52	94.40	N.A
2010	lam <i>et al.</i> [122]	94.72	95.67	N.A
2013	Wang <i>et al.</i> [123]	94.61	95.21	N.A
2013	Nguyen <i>et al.</i> [8]	94.07	93.24	N.A
2014	Zhao <i>et al.</i> [55]	94.77	95.09	N.A
2015	Yin <i>et al.</i> [21]	94.33	93.25	N.A
2015	Azzopardi <i>et al.</i> [11]	94.42	94.97	N.A
2016	Zhang <i>et al.</i> [23]	94.76	95.54	94.52
2017	Rezaee <i>et al.</i> [124]	94.63	95.21	N.A
Supervised				
2004	Staal <i>et al.</i> [77]	95.16	N.A	N.A
2006	Soares <i>et al.</i> [6]	94.66	94.80	N.A
2011	You <i>et al.</i> [70]	94.34	94.97	N.A
2011	Marin <i>et al.</i> [7]	94.52	95.26	N.A
2012	Fraz <i>et al.</i> [66]	94.80	95.34	94.69
2015	Roychowdhury <i>et al.</i> [80]	95.20	95.10	95.30
2015	Vega <i>et al.</i> [78]	94.12	94.83	N.A
2016	Li <i>et al.</i> [101]	95.27	96.28	95.81
2016	Liskowski and Krawiec [102]	95.15	N.A	N.A
2017	Barkana <i>et al.</i> [81]	95.02	95.53	N.A
2017	Zhang <i>et al.</i> [114]	94.66	95.47	95.02
—	Proposed method	95.41	96.40	96.03

FIGURE 3.9: Comparisons with the state-of-the-art methods under some specific challenging and difficult conditions of retinal image. The first column to the last column: subregions of retinal images, ground truth, segmented results by Soares *et al.* [6], Marin *et al.* [7], Nguyen *et al.* [8], Orlando *et al.* [9], and the proposed method, respectively.

shows that the proposed framework outperforms those retinal vessel segmentation methods in terms of the both F1 and MCC scores consistently over the 3 datasets.

Besides the quantitative evaluation by the statistical measurements on the whole

TABLE 3.3: Performance comparison with other methods in terms of F1-score (%) and MCC (%)

Year	Method	DRIVE		STARE		CHASE_DB1	
		F1	MCC	F1	MCC	F1	MCC
2000	Hoover <i>et al.</i> [118]	N.A	N.A	65.42	61.34	N.A	N.A
2006	Soares <i>et al.</i> [6]	77.59	74.81	74.06	71.23	N.A	N.A
2011	Marin <i>et al.</i> [7]	76.16	73.53	N.A	N.A	N.A	N.A
2013	Nguyen <i>et al.</i> [8]	76.06	72.71	70.99	67.74	N.A	N.A
2015	Azzopardi <i>et al.</i> [11]	N.A	74.75	N.A	73.35	N.A	68.02
2015	Vega <i>et al.</i> [78]	68.84	66.17	60.82	59.27	N.A	N.A
2017	Zhang <i>et al.</i> [114]	79.53	76.73	78.15	76.08	75.81	73.24
2017	Gu <i>et al.</i> [119]	78.86	75.89	79.53	77.24	72.02	69.28
2017	Orlando <i>et al.</i> [9]	78.57	75.56	76.44	74.17	73.32	70.46
—	Proposed method	80.93	78.51	81.25	79.58	78.09	75.91

dataset, ACC, F1 and MCC, it is interesting and informative to examine qualitatively the segmentation results under some specific challenging and difficult conditions of retinal image. We show four specific challenging and difficult conditions: (1) low contrast vessels, especially thin vessels that are easily missed during segmentation process; (2) bifurcate vessels that often cause disconnected segmentation results and reduce the integrity of vessel tree; (3) parallel vessels that tend to be detected as one wide vessel if they are close to each other; (4) lesions. Fig. 3.9 shows sample segmentation results under these 4 conditions by our proposed method and methods by Soares *et al.* [6], Marin *et al.* [7], Nguyen *et al.* [8] and Orlando *et al.* [9]. Compared to the segmentation results of these 4 methods under the 4 challenging image conditions, the proposed method is better to produce more accurate low contrast vessel, preserve the structure bifurcate vessel, distinguish two closely parallel vessels, and remove pathological regions. Results in Fig. 3.9 visually show that the proposed method is better to maintain the integrity of vessel tree under

TABLE 3.4: Performance comparison on the abnormal retinal images in the STARE dataset (%)

Year	Methods	ACC	F1	MCC
2000	Hoover <i>et al.</i> [118]	92.12	61.01	56.91
2006	Soares <i>et al.</i> [6]	94.18	68.36	65.15
2013	Nguyen <i>et al.</i> [8]	91.93	64.16	61.05
2015	Azzopardi <i>et al.</i> [11]	94.76	72.17	69.29
2017	Orlando <i>et al.</i> [9]	93.93	68.35	65.05
—	Proposed method	96.24	77.90	76.22

these challenging and difficult conditions of retinal image.

3.3.4.2 Pathological Retinal Image Analysis

We quantitatively evaluate the segmentation performance on the ten pathological images given in the STARE dataset. For fair comparison, we compute the ACC, F1 and MCC in the same FOV produced by [6] for all approaches in Table 3.4. Image pixel segmentation results of other approaches are provided by their authors. Performance comparison in Table 3.4 shows that our method significantly improves vessel segmentation result with the performance gain about 2% (ACC), 6% (F1), 7% (MCC). It shows that the proposed framework is more robust to deal with the pathological retinal images.

3.4 Summary

In this chapter, we propose a novel and robust cascade classification framework for retinal vessel segmentation from fundus image. The proposed framework deploys several techniques to the retinal vessel segmentation: background normalization and detail-preserving filter for preprocessing, weighted color channel fusion with

APCA dimensionality reduction for feature extraction, and skeleton and convex hull analysis for postprocessing. The backbone of our framework is the newly proposed cascade classification network that possesses the capability to form a highly nonlinear decision boundary by enveloping a set of computationally efficient Mahalanobis distance classifiers. Different from other nonlinear classifiers, the proposed cascade classification network is constructed and trained by a one-pass feed forward process, which produces a definite training and testing result that can be exactly reproduced. This greatly reduces the uncertainty of its usage in the both training and testing and hence it is user-friendly.

Experimental results demonstrate that the proposed cascade classification framework outperforms the state-of-the-art approaches in terms of all the four performance indicators: segmentation accuracy, segmentation accuracy from the cross dataset training, F1 score and Matthews correlation coefficient, consistently over the three public datasets, DRIVE, STARE and CHASE_DB1. This framework also marginally outperforms some recent powerful deep learning methods in terms of segmentation accuracy. The proposed method performs well for pathological images with different kinds of lesions, making it useful for computer-aided screening for patients with eye diseases. Comparison between the segmented images with the ground truth shows that the proposed method is capable of maintaining the integrity of vessel trees and geometrical structure of retinal vessels.

The demonstrated performance, effectiveness and robustness in the experiments allow the proposed framework to be an advanced tool in the automated retinal image analysis system for real clinical application. Though the proposed method needs some parameters to control the retinal vessel segmentation process, when facing a new dataset, we can tune those parameters on the basis of the image resolution as well as the property of the targeted objects. While techniques employed in our preprocessing and post-processing steps are domain specific, the newly proposed cascade classification network is a novel general approach that can be also applied to solve many other biomedical image recognition problems. Especially, the degree of nonlinearity of the proposed cascade classifier and the network size

is not predefined. They are determined by the complexity of the data structure. This makes the proposed cascade classification network adaptive to the complexity of the data structure and hence flexibly being applied to solve the both easy and difficult classification problems.

Chapter 4

Divided-and-Conquer Funnel-Structured Classification Framework for Fundus Image Segmentation

In the previous chapter, a cascade framework that envelops Mahalanobis distance classifiers is introduced for retinal vessel segmentation. Nevertheless, it does not well investigate the property of the retinal vessels with different scales and positions. In this chapter, we design a divided-and-conquer funnel-structured classification framework to enhance the detection performance of retinal vessel structure.

4.1 Introduction

Since accurate segmentation of retinal vessel plays an important role in the computer-aided diagnosis of retinal diseases, this section still investigates the probability of the performance enhancement of retinal vessel delineation. Chapter 3 introduces a cascade classification framework for retinal vessel segmentation, which shows the

FIGURE 4.1: Intensity profile of cross section of three representative vessel fragments. The first column: fundus image with three vessel fragments (vessel width: Fragment 1 < Fragment 2 < Fragment 3). The second column: intensity profile of cross section of fragments marked in fundus image.

capability of multistage learning on retinal vessel detection. Generally, supervised methods usually perform better than unsupervised ones on retinal vessel segmentation. Nevertheless, the decision boundary between vessel and non-vessel is sometimes hard to be determined without considering the property of retinal vessels with different scales and positions. Fig. 4.1 shows intensity profile of cross section of three representative vessel samples. Intensity profile varies with the vessel scale, e.g. vessels with large scale tend to be darker than those with small scale. In addition, since vessels almost never have ideal step edges [111], intensity of vessels near the center differ from those located at edge. Previous methods treat all vessel pixels equally and train a global discriminative model for retinal vessel detection. The large geometrical structure difference among retinal vessel with different scales and positions greatly limits the precision of the decision boundary of the global discriminative model.

Motivated by the extreme difficulty to handle the large variation of vessel pixels

by learning a single global discriminative model, we develop a novel divide-and-conquer funnel-structured classification framework for retinal vessel segmentation. Main contributions of our work include: first, we propose a multiplex vessel partition method to divide the retinal vessel pixels into well constrained subsets where vessel pixels with similar geometrical property (scale and position) are assigned together. This multiplex vessel partition method decomposes the complex vessel pixel pattern into a number of homogeneous patterns, making each of them easier to be classified from the non-vessel pixels. Thus, a set of homogeneous classifiers are trained in parallel to form discriminative model for each subset. Training a retinal vessel model in each subset leads to two benefits: one is that it reduces the computation cost like time and storage and the other is that it makes the vessel pixels more separable from the non-vessel pixels as vessel samples in the same subset have smaller variation. Second, we propose a funnel-structured vessel segmentation framework to reclassify the uncertain samples caused by imperfect data partition in the dividing phase. This further enhances the discriminative capability of the decision model. Our approach achieves high performance for retinal vessel segmentation consistently on all three diverse datasets, better than the state-of-the-art methods.

4.2 Proposed Framework

In this section, we describe the details of our proposed hierarchical architecture for retinal vessel segmentation based on divide-and-conquer strategy. The proposed divide-and-conquer funnel-structured architecture is shown in Fig. 4.2. First, we propose a multiplex vessel partition (MVP) method to divide retinal image pixels into 6 subsets on the basis of their geometric property, i.e. three scales (small, medium, and large) and two positions (edge and center) for each scale. For each of the 6 subsets, a classifier is trained to classify the data into vessel pixels, non-vessel pixels and uncertain pixels. Second, we feed the classified uncertain samples from each subset to the proposed funnel-structured vessel segmentation (FsVS)

FIGURE 4.2: The proposed divide-and-conquer funnel-structured architecture for retinal vessel segmentation.

framework, which concatenates the classification decision at three levels. The 1st level of FsVS utilizes the information of 6 subsets separately; the 2nd level of FsVS merges the uncertain samples of six subsets to three subsets, i.e. combine the pixels with the same scale but different position into one subset; the 3rd level of FsVS groups all uncertain samples to be a global subset. With this funnel-structured architecture, we reduce the classification error caused by the imperfect partition at the dividing phase and enhance the discriminative capability of the decision model.

4.2.1 Multiplex Vessel Partition

To divide retinal vessel into a compact subset, we propose a multiplex vessel partition method based on Gabor wavelet transformation. Gabor filter has been widely applied to retinal image analysis [6, 55, 66, 73] thanks to its biological relevance and superior mathematic property. A 2-D Gabor wavelet kernel is the product of an elliptical Gaussian envelope and a complex plane wave, expressed as [66]

$$\varphi_{\lambda,\sigma,\gamma,\theta}(x, y) = \exp\left(-\frac{x'^2 + \gamma y'^2}{2\sigma^2}\right) \exp\left(i2\pi\frac{x'}{\lambda}\right), \quad (4.1)$$

where $x' = x\cos\theta + y\sin\theta$ and $y' = -x\sin\theta + y\cos\theta$. θ denotes the orientation of Gabor wavelet. λ is the wavelength of the sinusoidal factor. σ and γ symbolize the scale of the Gaussian envelope and the ellipticity of the Gabor filter respectively.

For a given pixel at (x, y) in an image $I \in \mathbb{R}^2$ ($\mathbb{R}^2$ denotes the real plane), the Gabor wavelet transformation of $I(x, y)$ is obtained by convolving $I(x, y)$ with Gabor kernel $\varphi_{\lambda,\sigma,\gamma,\theta}$, given by

$$J_{\lambda,\sigma,\gamma,\theta}(x, y) = I(x, y) * \varphi_{\lambda,\sigma,\gamma,\theta}(x, y). \quad (4.2)$$

The magnitude of $J_{\lambda,\sigma,\gamma,\theta}(x, y)$ is computed as

$$A_{\lambda,\sigma,\gamma,\theta}(x, y) = \sqrt{\text{Re}(J_{\lambda,\sigma,\gamma,\theta}(x, y))^2 + \text{Im}(J_{\lambda,\sigma,\gamma,\theta}(x, y))^2}, \quad (4.3)$$

where $\text{Re}(J_{\lambda,\sigma,\gamma,\theta}(x, y))$ and $\text{Im}(J_{\lambda,\sigma,\gamma,\theta}(x, y))$ represent the real and imaginary parts of $J_{\lambda,\sigma,\gamma,\theta}(x, y)$ respectively. Generally, the real part of a Gabor wavelet kernel is used as a smooth filter while its imaginary part is regarded as edge detector. As the Gabor wavelet transformation is applied in a finite local window centered at (x, y) , we normalize both the real and the imaginary parts of Gabor filter coefficients to be zero mean and the unit sum of the absolute values, respectively.

Our proposed multiplex vessel partition (MVP) method partitions vessel into different subsets based on two crucial properties of Gabor wavelet transformation. First, Gabor filter can be tuned to capture the vessel width, since the Gabor wavelet transformation with the scale that best matches for vessel yields the maximum magnitude value $A_{\lambda,\sigma,\gamma,\theta}(x, y)$. Second, Gabor wavelet transformation produces higher absolute value of its real part than its imaginary part, i.e. $|\text{Re}(J_{\lambda,\sigma,\gamma,\theta}(x, y))| > |\text{Im}(J_{\lambda,\sigma,\gamma,\theta}(x, y))|$ for pixels located near the vessel center. Similarly, for pixels near the vessel edge we have $|\text{Im}(J_{\lambda,\sigma,\gamma,\theta}(x, y))| > |\text{Re}(J_{\lambda,\sigma,\gamma,\theta}(x, y))|$.

Algorithm 1 shows the details of our proposed multiplex vessel partition. For each vessel pixel $I(x, y)$, we are interested in the maximum magnitude of the Gabor wavelet transformation over different orientations at a given scale. In this chapter, a Gabor wavelet kernel is rotated within twelve orientations at step of $\pi/12$, i.e. $\theta = [0, \pi/12, \dots, 11\pi/12]$. The maximum magnitude of the Gabor wavelet transform over all orientations for each scale σ_h is marked as $A'_{\lambda,\sigma_h,\gamma,\theta_p}(x, y)$. By capturing the maximum value of $A'_{\lambda,\sigma_h,\gamma,\theta_p}(x, y)$ over all scales $\sigma_h, h \in 1, 2, \dots, m$, we assign the pixel $I(x, y)$ into the subset with the scale that best matches the vessel width. In this chapter, scale $\sigma_h (h \in 1, 2, \dots, m)$ is selected from the set of $[2, 4, 6]$ that span the possible widths of vessels throughout the retinal images. Parameters γ , and λ are set to be the same as those used in [55] for a good delineation of the vessel geometrical structure. To check a retinal vessel $I(x, y)$ near to the vessel center or edge, we propose a position decision rule: if the absolute value of the real part of the Gabor wavelet transformation $|\text{Re}(J_{\lambda,\sigma,\gamma,\theta}(x, y))|$ is larger than that of its imaginary part $|\text{Im}(J_{\lambda,\sigma,\gamma,\theta}(x, y))|$, the pixel $I(x, y)$ is assigned to a center cluster; otherwise it is assigned to an edge cluster.

To make our proposed method produce a balanced number of pixels between the vessel center cluster and the edge cluster for noisy and low contrast retinal image, a weight τ is introduced to control the partitioned proportion of center and edge clusters. Fig. 4.3 shows the partition results based on different τ . Clearly, a smaller value of τ leads to more pixels assigned to edge cluster. From this observation, we set τ to be 0.5 to have a balanced partition. Moreover, Fig. 4.3 visibly exemplifies

Algorithm 1 Multiplex Vessel Partition

Require: Sample $I(x, y)$, Gabor wavelet kernel $\varphi_{\lambda, \sigma, \gamma, \theta}$;
Ensure: Subsets C_{qg} and C_{qe} ;

- 1: **for** $h = 1$ to m **do**
- 2: **for** $w = 1$ to n **do**
- 3: $J_{\lambda, \sigma_h, \gamma, \theta_w}(x, y) = I(x, y) * \varphi_{\lambda, \sigma_h, \gamma, \theta_w}(x, y)$;
- 4: **end for**
- 5: Decide: $\theta_p = \underset{\theta_w, w \in 1, 2, \dots, n}{\operatorname{argmax}} A_{\lambda, \sigma_h, \gamma, \theta_w}$;
- 6: $A'_{\lambda, \sigma_h, \gamma, \theta_p}(x, y) = |J_{\lambda, \sigma_h, \gamma, \theta_p}(x, y)|$;
- 7: **end for**
- 8: Decide: $\sigma_q = \underset{\sigma_h, h \in 1, 2, \dots, m}{\operatorname{argmax}} A'_{\lambda, \sigma_h, \gamma, \theta_p}(x, y)$;
- 9: **if** $\tau \times |\operatorname{Re}(J_{\lambda, \sigma_q, \gamma, \theta_p}(x, y))| > |\operatorname{Im}(J_{\lambda, \sigma_q, \gamma, \theta_p}(x, y))|$ **then**
- 10: $I(x, y)$ is assigned to cluster C_{qg} , g indicates a center cluster;
- 11: **else**
- 12: $I(x, y)$ is assigned to cluster C_{qe} , e indicates an edge cluster;
- 13: **end if**

that the proposed method can effectively divide vessel pixels into clusters based on their geometrical property. Even for the small scale retinal vessel (fragment in the first row of Fig. 4.3), the proposed method also performs well to divide pixels into the center and edge clusters.

The proposed multiplex vessel partition (MVP) method is targeted at dividing the complex vessel patterns of large variation into a number of homogeneous groups of patterns. Non-vessel pixels are also portioned according the same rules. It is an unsupervised dividing strategy that is conducted on training and test samples independently, without learning from data. By dividing the samples with different geometrical property into a compact subset efficiently and robustly, we decompose the complex vessel pattern into a number of homogeneous patterns. This greatly reduces the variation of the vessel pattern in each group. Therefore, the difficult vessel segmentation problem can be conquered by decomposing it into a number of easier sub-problems and solve (conquer) them one by one.

FIGURE 4.3: Examples of vessel pixels dividing results based on different τ . The top to the bottom row of the first column: fundus image fragments with scales from small to large. The second to the last column: vessel pixels dividing results at $\tau=1, 0.75, 0.5$, respectively. Red color marks pixels assigned to the edge cluster and blue color marks pixels assigned to the center cluster.

4.2.2 Funnel-structured Vessel Segmentation

Based on the proposed MVP scheme, the entire training sample is grouped into a number of subsets C_{qg} and C_{qe} (Algorithm 1). Let C_k index the subset C_{qg} or C_{qe} . k is the index of qg or qe . Our next step is to conquer each subset C_k separately by training a discriminative model locally. With each subset C_k , we train a non-linear SVM classifier because of its well-known good generalization and popular application in binary vessel classification problem [64, 70, 81]. To circumvent the unavoidable partition error by the MVP scheme, we propose a funnel-structured vessel segmentation (FsVS) framework. It classifies the data in each subset C_k into three classes (vessel pixels, non-vessel pixels and uncertain pixels) and concatenates the uncertain pixels for further classification. The first level of FsVS utilizes information of each subset separately. The second level of FsVS merges the uncertain samples of six subsets to three subsets for further classification by combining the vessel with the same scale but different positions into the one subset. The last level of FsVS groups all uncertain samples to be

a global subset to finally classify it into two classes. With this funnel-structured architecture, we reduce the classification error caused by the imperfect partition at the dividing phase and enhance the discriminative capability of the whole decision model.

Let $S_{\mathbf{x}_j, C_k}$, $j \in [1, 2, \dots, N_{C_k}]$, represent the training sample in subset C_k , where $\mathbf{x}_j$ is the feature vector of the training sample. Assigning $S_{\mathbf{x}_j, C_k}$ to be vessel Ω_v , background Ω_b or uncertain class Ω_χ , is decided by

$$S_{\mathbf{x}_j, C_k} \in \begin{cases} \Omega_v, & \text{if } P(S_{\mathbf{x}_j, C_k}) > T_{v, C_k}, \\ \Omega_b, & \text{if } P(S_{\mathbf{x}_j, C_k}) < T_{b, C_k}, \\ \Omega_\chi, & \text{if } T_{b, C_k} \leq P(S_{\mathbf{x}_j, C_k}) \leq T_{v, C_k}, \end{cases} \quad (4.4)$$

where $P(S_{\mathbf{x}_j, C_k})$ is the output of SVM classification that denotes the probability of the pixel belonging to the vessel. Thresholds T_{v, C_k} and T_{b, C_k} control the decision of $S_{\mathbf{x}_j, C_k}$ belonging to the vessel or background. Samples falling in the class Ω_χ remain to be further processed at the higher level of funnel-structured vessel segmentation framework. In this chapter, we design a threshold searching method to select appropriate thresholds T_{v, C_k} and T_{b, C_k} . First, the probability distribution of vessel and background of training data in the subset C_k are sorted as ascending order respectively, resulting in two sequences Q_{v, C_k} and Q_{b, C_k} . Then, we define two initial thresholds T_{v0, C_k} and T_{b0, C_k} whose values yield about 60% and 30% training samples removal (being classified into either vessel Ω_v or background Ω_b) in the first two levels of FsVS, with remaining 10% training pixels for training decision model in the last level of FsVS.

To fine-tune both T_{v0, C_k} and T_{b0, C_k} , a searching scheme is formulated as:

$$\text{Decide : } T_{v, C_k} = \underset{T'_v}{\operatorname{argmin}} \frac{\Gamma[Q_{v, C_k} > T'_v]}{\Gamma[Q_{v, C_k} > T'_v] + \Gamma[Q_{b, C_k} > T'_v]}, \quad (4.5)$$

and

$$\text{Decide : } T_{b, C_k} = \underset{T'_b}{\operatorname{argmin}} \frac{\Gamma[Q_{b, C_k} < T'_b]}{\Gamma[Q_{v, C_k} < T'_b] + \Gamma[Q_{b, C_k} < T'_b]}, \quad (4.6)$$

FIGURE 4.4: Threshold searching process. Shaded areas in green and yellow color are the searching range $[T_{v0,C_k} - \delta, T_{v0,C_k} + \delta]$ and $[T_{b0,C_k} - \delta, T_{b0,C_k} + \delta]$ respectively. Pixels in the region Ω_χ remain to be further processed at the subsequent level.

where $T'_v \in [T_{v0,C_k} - \delta, T_{v0,C_k} + \delta]$ and $T'_b \in [T_{b0,C_k} - \delta, T_{b0,C_k} + \delta]$. $\Gamma[\cdot]$ is the function of counting the length of sequence. δ is the searching range around T_{v0,C_k} and T_{b0,C_k} . Fig. 4.4 shows the proposed process of the threshold searching. By searching two thresholds in a small range centered at T_{v0,C_k} and T_{b0,C_k} , we can select two thresholds T_{v,C_k} and T_{b,C_k} that produce the low classification errors. According to the searching rule of Eqs. (4.5-4.6), we achieve the smallest classification error for vessel and background within the range of 2δ centered at T_{v0,C_k} and T_{b0,C_k} respectively. Parameter δ is limited to a small value that makes threshold searching among 200 samples around T_{v0,C_k} and T_{b0,C_k} compared to approximately 0.1~1 million vessel and background training samples. The threshold at the last level of the FsVS is set to be the one that produces the maximum accuracy on the training set. Our proposed funnel-structured vessel segmentation framework greatly reduces the classification error caused by the imperfect partitions at the dividing phase. Moreover, the proposed hierarchal classification structure enhances the discriminative capability of the decision model comparing to a single SVM model.

In the test procedure, a testing sample is assigned to one of the six subsets based on the proposed multiplex vessel partition scheme in the same way as that of the

training samples. It will be classified into either vessel pixel or background pixel in one and only one of the 20 outputs of the proposed funnel-structured architecture as shown in Fig. 4.2.

4.3 Experimental Evaluations

To evaluate the performance of our proposed vessel segmentation framework, extensive experiments are conducted on three publicly available retinal vessel datasets: DRIVE, STARE and CHASE_DB1. Preprocessing, feature fusion with dimensionality reduction, postprocessing in Chapter 3 are also utilized for the improvement of fundus image quality and the representation and delineation of retinal vessel tree.

Performance of the proposed method is quantitatively analyzed by comparing the segmented result with the ground truth provided by each dataset. Four different measurements for vessel segmentation performance evaluation are employed in this chapter, including the segmentation accuracy (ACC), local segmentation accuracy (ACC_l), F1-score (F1) and Matthews correlation coefficient (MCC). The definition of ACC, F1 and MCC can see Section 3.3. Since background pixels that are easily detected occupy more than 80% pixels of FOV, different methods only have a small discrimination in terms of the global accuracy ACC [8]. ACC_l evaluates the segmentation accuracy of the local area which consists of vessel and its neighbourhood within a certain distance from the center of vessel [8]. We generate this local area by the morphological dilation operation with a disc structuring element. The radius of structuring element is set to be the value that forms the local area where the number of vessel and background pixels are balanced.

FIGURE 4.5: Examples of segmentation results. The first row to the third row: images and their segmentation results on DRIVE, STARE and CHASE_DB1 datasets. The first column to the last column: for the best segmentation case, original image, ground truth, segmentation result; for the worst segmentation case, original image, ground truth, segmentation result.

4.3.1 Comparison with Single SVM Classifier

The segmentation performance of the proposed framework and a single SVM classifier with the same preprocessing, feature representation and postprocessing on the three public datasets are shown in Table 4.1. In addition, the segmentation performance of the second human observer is also computed in Table 4.1 for reference. It shows that the segmentation performance of the proposed method are better than the single SVM classifier consistently over the 4 different evaluation criteria and consistently over the 3 different data sets. The comparison of the proposed framework to the single SVM classifier shown in Table 4.1 demonstrates the consistent performance gain brought by the proposed framework in this chapter as they use the same methods of preprocessing, feature representation, classifier and postprocessing. Note that other techniques of the preprocessing, feature representation, classifier and postprocessing can also be applied in our proposed framework.

Fig. 4.5 shows the best and worst segmentation cases in DRIVE, STARE and CHASE_DB1 datasets respectively. The ACC, ACC_l , F1 and MCC of the best case for DRIVE dataset are 96.70%, 87.69%, 86.54%, and 84.68%, while those measures of the worst case are 94.45%, 81.74%, 79.54% and 76.62%, respectively. For the STARE dataset, the values of the best case are 97.65%, 83.75%, 79.41%

and 78.22%, and the values of the worst case are 94.26%, 79.74%, 76.14% and 74.26%, respectively. For the CHASE_DB1 dataset, the best case measures are 97.03%, 82.73%, 80.52% and 78.92%, and the worst case measures are 95.30%, 79.30%, 77.50% and 74.88%, respectively. Fig. 4.5 visualizes that the proposed method produces a high matching confidence between the segmentation result and the ground truth, integrating complex structures to the vessel tree with small false negative error.

TABLE 4.1: Segmentation performance on DRIVE, STARE and CHASE_DB1 datasets(%)

dataset	Method	ACC	ACC _l	F1	MCC
DRIVE	2nd human observer	94.73	82.10	78.91	75.91
	Singel SVM	95.16	82.41	78.78	76.77
	Proposed method	95.47	83.72	81.44	78.95
STARE	2nd human observer	93.49	81.10	74.18	72.08
	Singel SVM	96.10	84.67	79.51	77.74
	Proposed method	96.46	85.79	81.92	80.13
CHASE_DB1	2nd human observer	96.13	81.58	79.73	77.67
	Singel SVM	95.69	78.66	75.46	73.19
	Proposed method	96.17	80.70	78.63	76.55

4.3.2 Comparison with Other Published Methods

4.3.2.1 Vessel Segmentation Performance

Table 4.2 shows the performance of different vessel segmentation methods in terms of ACC and ACC_l on DRIVE, STARE and CHASE_DB1 datasets. CHASE_DB1 dataset was built in 2012, thus no result before this year is available. All methods used for comparison report their segmentation accuracy ACC. We compute the local accuracy ACC_l of the published methods based on the segmented images provided by their respective publications, like [6], [7], [8], [9], [10], [11], [12], [118].

FIGURE 4.6: Local ACC of different methods with the different ratio of the number of pixels of background to that of vessel. (a) DRIVE dataset. (b) STARE dataset. (c) CHASE_DB1 dataset.

Table 4.2 shows that the proposed framework achieves better segmentation performance than those state-of-the-art methods in terms of the both ACC and ACC_l consistently over all the 3 datasets. Though the improvement of segmentation accuracy ACC between the proposed and other methods is marginal, the local

segmentation accuracy ACC_l is visibly higher, about 2%-3% enhancement. High value of ACC_l is important for vessel segmentation algorithms because it indicates their discriminative ability of vessel from its boundary. High ACC_l produced by the proposed method exemplifies the effectiveness of it in retinal vessel segmentation. Recently, many powerful deep learning based approaches perform well on the separation of vessel and background from fundus images. Our proposed method produces better segmentation performance than most of the compared deep learning based approaches. Though our segmentation accuracy is slightly lower than that of Oliveira *et al.* [104] (i.e. ACC only 0.5% less than it on all 3 datasets), we achieve more efficient vessel segmentation architecture without the need of training large number of parameters, extracting patches from each image and predicting the retinal map throughout every single patch. Comparison results between deep learning based methods and the proposed technique demonstrate the high potential of our designed none-deep learning model in dealing with vessel segmentation problem.

Fig. 4.6 shows the local accuracy of different methods with the different ratio of the number of pixels of background to that of vessel, obtained by increasing radius of structuring element from 1 to 15 on DRIVE, STARE and CHASE_DB1 datasets respectively. Local accuracy at the small ratio of background to vessel indicates the ability of the segmentation method to distinguish pixels around vessel edges. The errors for local region are often resulted from the merging of parallel vessel, the missing of low contrast vessel, and the coarse delineation of vessel edge. With the increasing ratio of background to vessel, more background pixels far away from vessel are counted in the computation of the accuracy and the accuracy curve tends to be stable. Accuracy improvement of our segmentation method with respect to other segmentation methods at different ratio of background to vessel is clearly shown in Fig. 4.6. Consistently better performance yielded by the proposed segmentation framework demonstrates its capability to generate accurate decision boundary between vessel and background.

TABLE 4.2: Performance comparison with other methods in terms of ACC (%) and ACC_l (%)

Year	Method	DRIVE		STARE		CHASE_DB1	
		ACC	ACC_l	ACC	ACC_l	ACC	ACC_l
2016	Li <i>et al.</i> [101]*	95.27	N.A	96.28	N.A	95.81	N.A
2016	Fu <i>et al.</i> [125]*	94.70	N.A	95.45	N.A	N.A	N.A
2016	Fu <i>et al.</i> [12]*	95.23	81.44	95.85	82.62	94.89	75.70
2018	Yan <i>et al.</i> [103]*	95.42	N.A	96.12	N.A	96.10	N.A
2018	Oliveira <i>et al.</i> [104]*	95.76	N.A	96.94	N.A	96.53	N.A
2018	Hu <i>et al.</i> [105]*	95.33	N.A	96.32	N.A	N.A	N.A
2019	Yan <i>et al.</i> [106]*	95.38	N.A	96.38	N.A	96.07	N.A
2019	Yu <i>et al.</i> [107]*	95.24	N.A	96.13	N.A	N.A	N.A
2019	Feng <i>et al.</i> [108]*	95.28	N.A	96.33	N.A	N.A	N.A
2000	Hoover <i>et al.</i> [118]	N.A	N.A	92.67	77.54	N.A	N.A
2004	Staal <i>et al.</i> [77]	95.16	N.A	N.A	N.A	N.A	N.A
2006	Soares <i>et al.</i> [6]	94.66	80.97	94.80	81.58	N.A	N.A
2007	Perez <i>et al.</i> [22]	93.16	76.70	91.96	76.07	N.A	N.A
2011	You <i>et al.</i> [70]	94.34	N.A	94.97	N.A	N.A	N.A
2011	Marin <i>et al.</i> [7]	94.48	80.50	95.26	N.A	N.A	N.A
2012	Fraz <i>et al.</i> [66]	94.80	N.A	95.34	N.A	94.69	N.A
2013	Nguyen <i>et al.</i> [8]	94.07	81.73	93.24	81.37	N.A	N.A
2015	Roychowdhury <i>et al.</i> [10]	94.90	77.17	95.60	79.49	94.67	N.A
2015	Azzopardi <i>et al.</i> [11]	94.42	80.20	94.97	82.16	93.87	74.53
2017	Orlando <i>et al.</i> [9]	94.54	81.80	94.89	83.08	94.59	78.02
2017	Zhang <i>et al.</i> [114]	94.66	N.A	95.47	N.A	95.02	N.A
—	Proposed method	95.47	83.72	96.46	85.79	96.17	80.70

*Result produced by deep learning based method.

Table 4.3 shows the comparison of F1 and MCC between the proposed method and other state-of-the-art methods on DRIVE, STARE and CHASE_DB1 datasets. Values of F1 and MCC for [118], [6], [7], [8], [10], [11] and [12], are computed by the segmented images from their corresponding publications. Our method reports a statistically noticeable improvement of F1 and MCC consistently over the 3 datasets, displaying advantage of the proposed framework for vessel classification where the samples in two classes (vessel and background) are unbalanced.

TABLE 4.3: Performance comparison with other methods in terms of F1-SCORE (%) and MCC (%)

Year	Method	DRIVE		STARE		CHASE_DB1	
		F1	MCC	F1	MCC	F1	MCC
2000	Hoover <i>et al.</i> [118]	N.A	N.A	65.42	61.34	N.A	N.A
2006	Soares <i>et al.</i> [6]	77.59	74.81	74.06	71.23	N.A	N.A
2011	Marin <i>et al.</i> [7]	76.16	73.53	N.A	N.A	N.A	N.A
2013	Nguyen <i>et al.</i> [8]	76.06	72.71	70.99	67.74	N.A	N.A
2015	Roychowdhury <i>et al.</i> [10]	72.85	69.45	72.97	70.36	N.A	N.A
2015	Azzopardi <i>et al.</i> [11]	77.64	74.75	76.18	73.35	70.90	68.02
2015	Vega <i>et al.</i> [78]	68.84	66.17	60.82	59.27	N.A	N.A
2016	Fu <i>et al.</i> [12]*	78.75	76.05	78.41	76.17	73.72	70.90
2017	Zhang <i>et al.</i> [114]	79.53	76.73	78.15	76.08	75.81	73.24
2017	Gu <i>et al.</i> [119]	78.86	75.89	79.53	77.24	72.02	69.28
2017	Orlando <i>et al.</i> [9]	78.57	75.56	76.44	74.17	73.32	70.46
—	Proposed method	81.44	78.95	81.92	80.13	78.63	76.55

*Result produced by deep learning based method.

The proposed method works well on vessel segmentation of fundus images, including some complex structures that are difficult to be correctly detected by many

FIGURE 4.7: Comparisons with the state-of-the-art methods under some specific challenging and difficult conditions of retinal image. (a) Parallel vessels. (b) Vessels with low contrast. (c) Crossing vessels. (d) Lesions. The first row to the last row: subregions of retinal images, ground truths, segmented results by [6], [8], [10], [11], [12], [9], and the proposed method, respectively.

state-of-the-art methods. Four specific challenging and difficult conditions of retinal images are shown in the first row of Fig. 4.7, i.e. closely parallel vessels, low contrast vessels, crossing vessels, and lesions. In Fig. 4.7, we compare the segmentation results under those four conditions of our proposed method with methods by [6], [8], [10], [11], [12] and [9]. Better segmentation performance of the proposed vessel segmentation framework is visible in Fig. 4.7, e.g. distinguish closely parallel vessels, detect vessels with low contrast, and preserve the structure between

crossing vessels and remove pathological regions. One reason is that we design a hierarchical architecture for retinal vessel segmentation, thus vessels at branch and boundary that differ from the majority of vessels are able to be correctly detected, avoiding the merging of closely parallel vessels, or the coarse delineation of branch and boundary vessels, which are challenging problems in many compared methods. The aforementioned representative examples shown in Fig. 4.7 also provide a visual evidence that our proposed method is capable of maintaining the integrity of vessel tree under challenging and difficult conditions of retinal image.

TABLE 4.4: Peripapillary vessel segmentation performance on DRIVE, STARE and CHASE_DB1 datasets(%)

dataset	Method	ACC	ACC _l	F1	MCC
DRIVE	Soares <i>et al.</i> [6]	87.91	79.54	81.21	72.69
	Marin <i>et al.</i> [7]	88.94	80.65	81.76	75.30
	Nguyen <i>et al.</i> [8]	88.05	82.55	82.53	73.45
	Roychowdhury <i>et al.</i> [10]	89.76	82.13	83.15	77.25
	Azzopardi <i>et al.</i> [11]	90.19	83.39	85.55	78.14
	Fu <i>et al.</i> [12]*	90.24	83.28	86.42	79.04
	Orlando <i>et al.</i> [9]	89.30	82.73	85.17	77.08
	Proposed method	91.80	86.66	87.88	81.70
STARE	Hoover <i>et al.</i> [118]	83.40	72.21	65.78	54.86
	Soares <i>et al.</i> [6]	83.24	76.71	67.54	56.34
	Nguyen <i>et al.</i> [8]	85.15	79.14	72.27	62.50
	Azzopardi <i>et al.</i> [11]	88.18	78.70	73.72	66.83
	Fu <i>et al.</i> [12]*	88.69	79.52	76.94	69.46
	Orlando <i>et al.</i> [9]	84.74	80.61	71.46	61.40
	Proposed method	89.71	81.31	77.71	84.25
CHASE_DB1	Azzopardi <i>et al.</i> [11]	89.53	79.90	81.65	74.35
	Fu <i>et al.</i> [12]*	86.66	79.90	81.65	74.35
	Orlando <i>et al.</i> [9]	88.69	79.62	80.79	72.83
	Proposed method	91.33	83.36	85.10	79.13

*Result produced by deep learning based method.

4.3.2.2 Peripapillary Vessel Analysis

Vessel in and around the optic disk are referred to be the peripapillary vessel, whose structure changes can lead to many retinal abnormalities like glaucoma, central retinal vein occlusion. Existing methods often perform less well due to the effect of packed nerve heads [119]. We carry out the proposed framework on those peripapillary vessel and compare our performance with other state-of-the-art methods. A mask for peripapillary region is delineated manually around the optic disk in each fundus image on DRIVE, STARE and CHASE_DB1 datasets respectively. For STARE dataset, we only perform peripapillary vessel analysis on the first 19 images because the last image from the STARE dataset does not contain the peripapillary region [10]. Table 4.4 shows the comparison result of peripapillary vessel detection. Values of the state-of-art methods are computed from the images provided by the corresponding works. Compared to those well-known vessel segmentation algorithms, our proposed framework exhibits the best performance on the detection of peripapillary vessel consistently on three datasets.

4.3.2.3 Thin and Low-contrast Vessel Analysis

We investigate the segmentation effectiveness of the proposed method on two challenging vessels, i.e. thin vessel and low-contrast vessel. Vessels with width less than 3 pixels are delineated as thin vessel. Background regions around thin vessel within 2 pixels from the vessel edges are taken to assess the proposed method on thin vessel segmentation. Table 4.5 shows the performance comparison results on thin vessel segmentation, where the proposed method outperforms other methods consistently on three measurements, i.e. achieving performance gains of 2.2% for ACC, 4.2% for F1 and 6.2% for MCC from the best performances of the compared existing approaches. To evaluate the segmentation performance on low-contrast retinal vessel, we search a set of subregions containing low contrast vessel throughout all retinal images, and report their segmentation performance in Table 4.5.

The results show that our proposed method produces the best segmentation performance for low-contrast vessel consistently on three measurements. From the detailed quantitative experimental analyses in Table 4.5, we exemplify the effectiveness of the proposed method on the detection of the challenging thin vessel and low-contrast vessel.

TABLE 4.5: Performance comparison on thin vessel and low-contrast vessel in STARE dataset (%)

Method	Thin vessel			Low-contrast vessel		
	ACC	F1	MCC	ACC	F1	MCC
Hoover <i>et al.</i> [118]	72.64	51.50	32.83	94.13	53.73	50.66
Soares <i>et al.</i> [6]	81.22	60.05	48.40	96.22	61.95	61.72
Nguyen <i>et al.</i> [8]	76.01	60.72	45.07	96.03	68.67	66.63
Azzopardi <i>et al.</i> [11]	76.71	58.79	42.99	96.10	65.96	64.00
Fu <i>et al.</i> [12]*	79.30	55.19	42.53	96.26	61.33	61.69
Orlando <i>et al.</i> [9]	80.64	59.43	47.14	96.49	64.96	64.78
Proposed method	83.38	64.89	54.57	97.09	72.29	71.73

*Result produced by deep learning based method.

4.3.2.4 Pathological Retinal Image Analysis

To evaluate the performance of the proposed method on pathological retinal image, we conduct experiment on ten abnormal images in STARE dataset. Those abnormal images have lesions such as hemorrhages and exudates that are difficult for segmentation system to correctly detect retinal vessel. Table 4.6 shows our comparison results with other methods, where we achieve the significantly improvement on pathological retinal image segmentation. Especially for MCC, the proposed method yields an improvement of 3.6% that indicates the proposed method can effectively distinguish vessel from complex background. The high segmentation performance on pathological retinal image also shows the capacity of the proposed method on clinical diagnosis of retinal vessel disease.

TABLE 4.6: Performance comparison on pathological retinal image in STARE dataset (%)

Method	ACC	ACC _l	F1	MCC
Hoover <i>et al.</i> [118]	92.12	77.06	61.01	56.91
Soares <i>et al.</i> [6]	94.18	79.64	68.36	65.15
Nguyen <i>et al.</i> [8]	91.93	79.92	64.16	61.05
Azzopardi <i>et al.</i> [11]	94.76	80.89	72.17	69.29
Orlando <i>et al.</i> [9]	93.93	80.59	68.35	65.05
Fu <i>et al.</i> [12]*	95.70	80.52	75.38	73.21
Proposed method	96.29	83.68	78.71	76.85

*Result produced by deep learning based method.

4.4 Summary

In this chapter, we have proposed a new divide-and-conquer funnel-structured classification framework for retinal vessel segmentation. By dividing the retinal vessel pixels into well constrained subsets where samples with similar geometrical property are grouped together, the sample variation in each subset is largely reduced. This decomposes a complex classification problem into a number of relatively simpler ones. Training a set of homogeneous classifiers in parallel not only reduces the computational cost like time and storage, but also simplifies the classification complexity because samples at the same subset have smaller variation. With the proposed funnel-structured vessel segmentation framework, we alleviate the problem caused by the imperfect data partition in the dividing phase. This further enhances the discriminative capability of the proposed whole decision model. Experiments on three diverse datasets show the superiority of the proposed framework for vessel segmentation, achieving the best performance consistently in terms of four different evaluation metrics on all three datasets. Even for complex vessel structures, the proposed framework also displays very good detection ability. High

accuracy for peripapillary blood vessel extraction makes the proposed framework a suitable tool for computer-aided screening of early disease detection like glaucoma, and vein occlusion. The demonstrated flexibility, robustness and effectiveness suggests a good potential of the proposed framework in real clinic application of the automated retinal image analysis.

Chapter 5

Deep Feature Learning for Dermoscopic Image Segmentation

Chapter 3 and Chapter 4 introduce the retinal vessel segmentation based on multistage classification. In this chapter, we focus on a more challenging task, i.e. automatic segmentation of skin lesion from dermoscopic images, and propose a bi-directional deep feature learning framework with multiscale consistent decision fusion for lesion delineation.

5.1 Introduction

Medical image analysis has attracted a lot of research attention in the last decades. With the development of digital data acquisition, a large amount of medical image data from different imaging modalities are produced. Nevertheless, it is impracticable for experts to quantitatively analyze those large datasets. Recently, deep learning have demonstrated their superior performance on medical image analysis, following their successful application in the area of image recognition and computer vision [126, 127]. Dermoscopic image is one of the medical image modalities that aids the detection of melanoma [38, 128]. According to the investigation of

skin cancer [41, 129], melanoma has become the most lethal type of skin cancer. Automatic segmentation of skin lesion assists clinicians for the analysis of melanoma [130, 131]. Deep learning based methods, like fully convolutional neural networks (FCNs) [4, 5, 95], have displayed promising segmentation performance for dermoscopic image. These approaches show the powerful feature representation ability that encodes both low-level appearance information and high-level semantic information. Nevertheless, the spatial resolution of feature maps are reduced by a series of pooling and down-sampling operations, thus yielding the insufficient interpretation of skin lesion. For example, the FCNs fail to extract valid feature information for some small skin lesions since they are so inconspicuous compared to background. In addition, FCNs also limit the localization of skin lesion boundary because of the loss of the detailed feature representation. Some advanced convolutional neural networks have been developed for dealing with the reduction of feature resolution, like atrous convolutional neural network that resolves this issue with multiple parallel dilated convolutional layers [132, 133]. Nevertheless, it is still challenging for atrous convolutional neural network to learn a desirable feature representation for skin lesion, since there exist non-negligible heterogeneous characteristic of different skin lesions. Fig. 5.1 shows some examples of dermoscopic images with different appearance, e.g. the shape and color of skin lesion. Noise like hairs and air bubbles in dermoscopic images also limit the detection of the skin lesions.

As reviewed in Chapter 1, when melanocytes grow in clusters alongside normal cells, pigmented lesions or melanocytic nevi appear on the surface of the skin. Then, skin lesions progressively metastasize to other human organs and result in high mortality rate of melanoma for patients. Usually, the spatial configuration of the skin lesion reflects the velocity and severity of the lesion growths. The correlation between candidate pixels and their surrounding contextual regions benefits the network to interpret the feature of skin lesion. To better capture the complex correlation between skin lesions and their informative neighborhood over dermoscopic images, this chapter presents a bi-directional dermoscopic feature learning (biDFL)

FIGURE 5.1: Examples of challenges for skin lesion delineation. Lesion position is marked as red color.

framework for skin lesion feature representation. It combines the object lesion pixels with their informative context and thus leads to a substantially meaningful description of lesion structure. Main advantage of the proposed bi-directional dermoscopic feature learning framework over the classical FCNs [2] is that it produces more discriminative dermoscopic feature maps by controlling information propagation from two complementary directions at high level parsing layer. Thus, the proposed bi-directional dermoscopic feature learning framework makes the network learn better about the complex structure of the skin lesion.

Score map generated by neural network represents the probability of each pixel belonging to the skin lesion or background. Aggregation of multiscale score information from skip layers have shown their strengths in semantic image segmentation [2, 94]. Nevertheless, it does not indicate pixel scores at all skip layers are reliable and informative for skin lesion segmentation. For example, skin lesion have different sizes and thus the selection from multiscale feature representation is necessary for the description of lesion. In addition, it is more reliable for pixels far

away from the lesion edge to be classified by using larger scale of features, while for pixels near the lesion boundary, the decision with smaller scale of features is a reliable result for lesion segmentation. Thus, it is beneficial to selectively focus on the reliable decision and try to reduce the effect of unreliable one, since the latter contributes little for skin lesion segmentation, and even leads to inconsistent segmentation performance. In order to achieve more consistent and trusted skin lesion segmentation, we propose a multiscale consistent decision fusion scheme that selectively aggregates score maps by controlling the reliability of multiscale feature representation among multiple classification layers. Our multi-scale consistent decision fusion embeds the consistency information around local decision context to adjust the confidence of decision and thus allows more reliable and precise skin lesion delineation. If the decision for a candidate pixel is consistent to its local context decision, then the network gives high confidence for this decision; otherwise the network reduces the confidence of this decision.

Main contributions of our proposed work are outlined as follows:

- We propose a bi-directional feature propagation architecture for skin lesion segmentation. It generates a rich and discriminative feature representation by encoding the information from different context around the skin lesions through two directions.
- We propose a multiscale consistent decision fusion strategy that controls the summation of the score maps at all individual block of the network selectively, which further ensures the effective and accurate skin lesion structure delineation and complex boundary localization.
- We achieve a state-of-the-art skin lesion segmentation performance consistently for melanoma and non-melanoma dermoscopic images on two publicly available datasets.

FIGURE 5.2: The structure of the proposed framework. Bi-directional feature learning framework investigates the complex correlation between skin lesions and their informative context to generate a discriminative feature representation. Multiscale consistent decision fusion technique achieves more reliable and precise skin lesion prediction by analyzing the consistency of decisions in a local area. Green solid arrow: feature learning from local to zooming-out region. Red dot arrow: feature learning from global to zooming-in region.

5.2 Proposed Framework

The structure of the proposed network for skin lesion segmentation is shown in Fig. 5.2, where we apply a FCN-like architecture with ResNet50 (pre-trained on ImageNet [134]) as the baseline network. As illustrated in Fig. 5.2, feature information are propagated at two complementary directions. The first direction depicts the feature information passing from local to zooming-out region (marked as green solid arrow in Fig. 5.2), while the second direction denotes the feature information passing from global to zooming-in region (marked as red dot arrow in Fig. 5.2). This feature information propagation simulates the human visual perceptual process that dominates the field of the anatomical, physiological and

behavioural studies of vision system [135]. For example, cerebral cortex usually interprets visual stimulus from the extraction of local feature in the lower visual pathways, and then integrates the local feature into more global information in the higher visual pathways [136]. In addition, cerebral cortex also encodes visual stimulus feature from global to local [137, 138]. By the integration of dermoscopic feature information from two directions, local and contextual features are combined to generate a rich description of skin lesion. Moreover, the proposed multiscale consistent decision fusion (shown in the yellow box in Fig. 5.2) controls the fusion of the reliable score maps at multiple layers selectively, which further ensures the effective and accurate tiny skin lesion structure delineation and complex boundary localization.

5.2.1 Context Feature Map Generation with Multiple Dilation Rates

Fully convolutional neural networks have achieved promising results in semantic image segmentation by extracting semantic information at high level layer of network. Nevertheless, a series of consecutive pooling and convolution striding operations reduce the feature map resolution, which limits their performance on segmentation task that needs accurate pixel-wise prediction on the whole image. In order to improve the feature representation on skin lesion, it is important to generate more rich and meaningful feature maps at high-level layer. As discussed in Chapter 1, the skin lesion progressively invade their nearby tissues that indicates there exists correlation between different spatial configuration of skin lesion. Therefore, we propose to explore the integration of information from each part of the lesion structure which assists network to learn better about the characteristic of the skin lesion.

Suppose that F_0 are the feature maps generated at the top layer of the network, the multiscale context feature representation is produced by a series of dilated

convolutions with ascending dilation rates γ_j :

$$F_j = \mathcal{F}_{\gamma_j}(F_0, \Theta_j), j \in [1, 2, \dots, J], \quad (5.1)$$

where $\mathcal{F}_{\gamma_j}$ denotes a 3×3 convolution with the dilation rate γ_j . Θ_j are the respective convolution parameters. J is the number of dilation rates. With different dilation rate γ , we can effectively control the field of view of multiple convolution filters to incorporate semantic information and generate the feature maps with complementary contextual information. We investigate the relationship between dilation rate γ and different levels of spatial contextual information:

- *Local*: Dilated convolution with small rate concentrates on the detection of feature information around the local region. Thus, a detailed feature representation is generated for the description of skin lesion. Though small dilation rates preserve locality of features, they are sensitive to artificial signals during dermoscopic image acquisition, such as noises and air bubbles. In addition, small dilation rates have ability to detect more accurate boundary information while also have probability of the prediction of many isolated fake lesion tissues.
- *Regional*: When dilation rates increase progressively, the receptive field of dilated convolution filters are enlarged to extract more larger spatial contextual information of skin lesions. In contrast to small dilation rates, medium dilation rates allow the network to harness larger scale of contextual information more efficiently and thus generate prediction more robust to the challenging artificial noise, irregular lesion tissues, and some inhomogeneous dermoscopic images. Nevertheless, for the size of extremely tiny or large skin lesions, dilated convolution filters with medium dilation rates are not suitable to capture their characteristic precisely.
- *Global*: Dilated convolution filters with large dilation rates are fit for the feature extraction of lesion with large scale. Though they have ability to

FIGURE 5.3: Workflow of the bi-directional dermoscopic feature learning. (a) Receptive fields of dilated convolution with three different rates. (b) Feature passing from local to zooming-out region. (c) Feature passing from global to zooming-in region.

see more global skin lesion pattern, their power of capturing detailed local information are decreased that may leads to the coarse prediction for skin lesion.

5.2.2 Bi-directional Dermoscopic Feature Learning

To exploit information of feature maps extracted with multiple receptive fields, the fusion of those feature maps are investigated in this section. A simple way for the combination of feature maps at multiscale rates is the concatenation of them directly. Though it achieves the integration of those feature maps, the increasing feature dimensionality of this manner will result in the degradation of the generalization capability of the feature maps for classification. Herein, we explore to propose a bi-directional dermoscopic feature learning module that perceives the information of lesion characteristics by controlling the propagation of the feature flow among the complex spatial configuration of the skin lesion. The workflow of the feature information passing along two directions is illustrated in Fig. 5.3, where

three feature maps obtained from dilated convolution with different rates are given in Fig. 5.3 (a). There are two feature propagation directions for the learning of the complex correlation between skin lesion and their informative context. One feature propagation direction from local to zooming-out region (Fig. 5.3 (b)), while the other from global to zooming-in region (Fig. 5.3 (c)). By the combination of both directional feature learning propagation, our proposed biDFL framework allows network to give better insight about skin lesion.

When the feature information propagates from local to zooming-out region, the feature maps F_j are refined by

$$F'_j = \mathcal{F}(\mathcal{C}(F'_{j-1}; F_j), \hat{\Theta}_j), j > 1, j \in [2, \dots, J], \quad (5.2)$$

where $\mathcal{C}$ denotes the process of feature concatenation process. $\mathcal{F}$ represents the 1×1 convolution, and $\hat{\Theta}_j$ are the respective parameters. Through the 1×1 convolution process, the channel dimension F'_j is kept the same as that of F_j , which integrates feature maps without the increase of the computation memory. By Eq. (5.2), the informative contexts $F_1, \dots, F_{j-1}, F_j$ are progressively fused into F'_j , making F'_j have more powerful representation for the complex structure of skin lesion. Especially for the case that feature maps F_j are generated from a large receptive field, the proposed feature information propagation have capability to alleviate the degradation of representation ability of those feature maps.

Another feature passing direction is from global to local region, which is opposite to the previous feature propagation direction. The feature maps F''_j in this direction are expressed as

$$F''_j = \mathcal{F}(\mathcal{C}(F''_{j+1}; F_j), \tilde{\Theta}_j), j < J, j \in [1, 2, \dots, J-1]. \quad (5.3)$$

As formulated in Eqs. (5.2) and (5.3), the classification network learns the correlation between skin lesions and their informative context through a series of modulators ($\mathcal{C}$ and $\mathcal{F}$). More specifically, the feature maps F'_j , receiving the messages through the forward feature learning process, fuse a set of zooming-out context feature information from F_l ($l < j$). In addition, the feature maps F''_j , obtaining messages through the backward feature learning process, integrate a series of zooming-in context feature information from F_g ($g > j$). By the combination of both forward and backward feature learning directions, the proposed biDFL generates more rich and discriminative feature representation for skin lesion.

Meanwhile, stacking the convolution layers together also enriches the interpretation of image at multiple receptive domains, since the characteristics of each lesion pixel can be more distinctively represented by propagating the information stream between those successive convolution layers. For example, classification for some skin lesions that have a similar appearance to non-lesion area needs the support of information from the appropriate receptive field. With the sequentially aggregation of feature maps from multiple receptive domains, more meaningful interpretation of skin lesion is achieved, which is important for challenging lesion segmentation.

5.2.3 Multi-scale Consistent Decision Fusion

Skin lesions in dermoscopic images usually have a large variation of size. We calculate the size of the skin lesion relative to that of the respective dermoscopic image in two publicly available datasets ISBI 2016 and ISBI 2017. The maximum value of the lesion size compared to the whole image size are 0.9954 and 0.9522 for ISBI 2016 and ISBI 2017 datasets, respectively. In contrast, the minimum value of the lesion size compared to the whole image size are small, i.e. 0.0027 and 0.0030 for ISBI 2016 and ISBI 2017 datasets, respectively. This indicates the existence of the relatively tiny and inconspicuous lesions compared to health skin tissues, which is hard for traditional networks like VGG net [139] and Resnet-50 [84] to capture the information of the aforementioned skin lesions. To improve the performance of skin

lesion, we exploit to utilize feature information from shallow layers because they extract more detailed information of the inconspicuous skin lesions and complex boundary structure.

Many methods for the integration of the score maps have been proposed for semantic image segmentation [2, 94, 140]. Nevertheless, they usually combine the score maps with non-selective summation, which decreases the precision of the skin lesion prediction since some unreliable scores are fused inappropriately. In this section, we propose a multiscale consistent decision fusion strategy which is an adaptive score map aggregation incorporating the selection of reliable scale features for each pixel. The proposed multiscale consistent decision fusion achieves the selective score maps aggregation by controlling the reliability of multiscale feature representation among multiple classification layers.

For a candidate pixel in dermoscopic image, network outputs K score maps S_p^k generated by K skip layers from different scales of features. Here, p denotes the spatial position and $k \in K$. Let α_p^k be the coefficient of multiscale consistent decision fusion. In this chapter, α_p^k is formulated as a Gaussian function of the standard deviation of the scores S_p^k over a spatial local region L_p centered at p , i.e.

$$\alpha_p^k = e^{\frac{-(\sigma_p^k)^2}{\sigma^2}}, \quad (5.4)$$

where

$$\sigma_p^k = \sqrt{\text{mean}_{L_p}(S_p^k - \text{mean}_{L_p}(S_p^k))^2}. \quad (5.5)$$

For each position p , multi-scale consistent coefficients α_p^k for K skip layers are computed. Parameter σ denotes the sensitivity of the consistency coefficient α_p^k to the variation of the score S_p^k in the local region. α_p^k indicates the reliability of decision in a local region centered at the k th skip layer. For example, a smaller coefficient α_p^k implies that the decision S_p^k at the k th skip layer for position p is less reliable. Hence, the effect of this S_p^k should be reduced during the score fusion process. A larger coefficient α_p^k means that the decision S_p^k at the k th skip layer for

position p is more reliable. Consequently, the effect of S_p^k with a larger coefficient $\alpha_p^{k,c}$ should be highlighted to ensure the desirable prediction.

With the manipulation of consistent coefficient α_p^k , score maps from multiple layers are selectively integrated by

$$S_p = \sum_{k=1}^K \alpha_p^k S_p^k. \quad (5.6)$$

The contribution of the decision S_p^k at each skip layer is adaptively adjusted by consistent coefficient α_p^k , without complicated heuristic learning process. Compared to the previous simple score summation, the proposed multiscale consistent decision fusion combines the score map at each skip layer selectively by embedding consistency analysis. In addition, network with our proposed multiscale consistent decision fusion also implicitly learns to inhibit irrelevant noises, which results in more accurate and smooth skin lesion delineation.

5.3 Experimental Evaluations

5.3.1 Materials

The proposed framework is evaluated on two publicly available datasets (ISBI 2016 [141] and ISBI 2017 [142]). They are two challenge datasets of “Skin Lesion Analysis toward Melanoma Detection” hosted by the International Symposium on Biomedical Imaging (ISBI) in 2016 and 2017 respectively.

- ISBI 2016, consists of 900 training images and 379 test images in JPEG format. The size of images in ISBI 2016 dataset changes from 540×722 to 2848×4288 . ISBI 2016 supports ground truths (encoded as single-channel 8-bit PNGs) for skin lesion segmentation. There are 727 non-melanoma and 173 melanoma in training set, as well as 304 non-melanoma and 75 melanoma in test set.

- ISBI 2017, the extension of ISBI 2016, includes 2000 training images and 600 test images in JPEG format. The size of images in ISBI 2016 dataset changes from 540×722 to 4499×6478 . Ground truths for skin lesion segmentation have also been provided. There are 374 melanoma, 254 seborrheic keratosis, and 1372 benign nevi in training set. Test set contains 117 melanoma, 90 seborrheic keratosis, and 393 benign nevi.

5.3.2 Evaluation Criteria

We evaluate the segmentation performance of the proposed method by comparing the skin lesion segmented result with the ground truth. Four different measurements are applied for skin lesion segmentation performance evaluation, including the segmentation accuracy (ACC), Jaccard index (JAC), Dice coefficient (DIC), and G-mean (GME). ACC is the ratio of correctly detected pixels to total pixels. JAC and DIC denote the similarity between the lesion segmented result and ground truth. GME estimates the imbalance between segmentation sensitivity and specificity by computing their geometric mean. Metrics of those criteria are defined as:

$$\begin{aligned}
 \text{SEN} &= \text{TP}/(\text{TP} + \text{FN}), \\
 \text{SPE} &= \text{TN}/(\text{TN} + \text{FP}), \\
 \text{GME} &= \sqrt{\text{SE} \times \text{SP}}, \\
 \text{JAC} &= \text{TP}/(\text{TP} + \text{FN} + \text{FP}), \\
 \text{DIC} &= 2\text{TP}/(2\text{TP} + \text{FN} + \text{FP}), \\
 \text{ACC} &= (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN}),
 \end{aligned} \tag{5.7}$$

where TP and TN represent the number of pixels correctly segmented as skin lesion and background pixels, respectively. FP and FN are the number of pixels incorrectly segmented as skin lesion and background pixels, respectively. JAC was taken as the most important criterion for segmentation comparison in ISBI 2016 and 2017 challenges where participants for competition were ranked based on it.

5.3.3 Implementation Details

In this chapter, we utilize a FCN-like architecture with ResNet50 as our fine-tune network. For the last two blocks of network, the resolution of feature maps are kept unchanged while the dilation rates of convolution layers at the two blocks are set to be 2 and 4 respectively. Thus the pre-trained weights for the network can be reused. We exploit the “poly” learning rate policy where the current learning rate is multiplied by $(1 - \frac{iter}{max_iter})^{power}$ [120]. The initial learning rate and power are set to be 10^{-3} and 0.9 respectively. On the top layer of the network, five sets of context feature maps with five different dilation rates (i.e. 3, 6, 12, 18, 24) are generated. Convolution transpose kernels [120] is applied for the up-sampling operation. Augmentation strategy, like horizontal and vertical flipping, and random scaling, are also conducted to produce more diverse training data.

5.3.4 Ablation Study

5.3.4.1 Evaluation of Individual Component

In order to assess the effect of each component of the proposed architecture on skin lesion segmentation, a step-by-step ablation experiment on dermoscopic images is designed for ISBI 2016 and ISBI 2017 datasets. Table 5.1 shows the detailed quantitative experimental results, from which we can observe that each proposed contribution collectively improves the baseline network on skin lesion segmentation. By encapsulating bi-directional feature learning in the baseline network, segmentation performance JAC for ISBI 2016 and ISBI 2017 datasets are enhanced by 2.7% and 3.1% respectively. By introducing the multi-scale consistent decision fusion, segmentation performance JAC on the both datasets are improved by 1.5%-1.6% respectively. Integration of bi-directional dermoscopic feature learning and multi-scale consistent decision fusion results in the JAC enhancement of 3.7%-5.7% on the both datasets. Moreover, Table 5.2 also illustrates the segmentation performance

TABLE 5.1: Segmentation performance on ISBI 2016 and ISBI 2017 datasets (%)

dataset	Method	Overall			
		JAC	DIC	ACC	GME
ISBI 2016	Baseline	84.47	90.97	95.59	93.94
	Baseline+biDFL	87.19	92.56	96.21	94.63
	Baseline+mCDF	86.09	91.88	95.91	94.82
	Baseline+biDFL+mCDF	88.12	93.33	96.75	95.19
ISBI 2017	Baseline	75.74	84.10	93.13	88.47
	Baseline+biDFL	78.87	86.51	94.01	89.73
	Baseline+mCDF	77.24	85.45	93.24	90.21
	Baseline+biDFL+mCDF	81.47	88.54	94.65	92.67

TABLE 5.2: Segmentation performance on melanoma and non-melanoma (%)

dataset	Method	Melanoma				Non-Melanoma			
		JAC	DIC	ACC	GME	JAC	DIC	ACC	GME
ISBI 2016	Baseline	85.03	91.56	94.57	93.73	84.33	90.83	95.84	93.99
	Baseline+biDFL	87.24	92.74	95.13	93.96	87.18	92.52	96.48	94.79
	Baseline+mCDF	87.03	92.86	95.10	94.45	85.86	91.64	96.11	94.90
	Baseline+biDFL+mCDF	88.38	93.62	95.95	94.53	88.06	93.26	96.95	95.35
ISBI 2017	Baseline	69.24	79.12	90.09	84.22	77.32	85.30	93.87	89.49
	Baseline+biDFL	72.91	82.03	91.01	86.05	80.32	87.59	94.74	90.61
	Baseline+mCDF	73.25	82.88	90.15	87.10	78.20	86.08	93.99	90.97
	Baseline+biDFL+mCDF	77.26	85.35	92.02	89.74	82.49	89.31	95.28	93.37

on melanoma and non-melanoma cases, where our proposed method produces better segmentation than the baseline model consistently in ISBI 2016 and ISBI 2017 datasets.

We also qualitatively analyze the validity of the proposed architecture on skin lesion

TABLE 5.3: Component investigation of bi-directional dermoscopic feature learning (%)

Method	ISBI 2016				ISBI 2017			
	JAC	DIC	ACC	GME	JAC	DIC	ACC	GME
Baseline	84.47	90.97	95.59	93.94	75.74	84.10	93.13	88.47
Baseline+DFL ₊	86.01	91.99	95.83	93.82	78.37	86.21	93.58	89.58
Baseline+DFL ₋	85.92	91.93	96.16	94.03	78.03	85.98	93.83	89.76
Baseline+biDFL	87.19	92.56	96.21	94.63	78.87	86.51	94.01	89.73

prediction in Figs. 5.4 and 5.5, which show experimental results on some challenging dermoscopic images, i.e. skin lesions with variable scale, low contrast, artifact, as well as irregular convex and concave boundary. Compared to the segmentation performance of baseline architecture, the proposed network produces more accurate delineation of the skin lesion. For instance, images at the last column of Fig. 5.4 shows skin lesion with the low contrast compared to background. It is difficult for baseline model to distinguish skin lesion from background. Our proposed network overcomes this limitation and generates desirable lesion segmentation results. For skin lesion with concave boundary (e.g. images at the third column of Fig. 5.4) and convex boundary (e.g. images at the second column of Fig. 5.5), the proposed network also produces accurate detection of the geometrical structure of the skin lesion. Qualitative analysis of the segmentation performance in Figs. 5.4 and 5.5 exemplify the effectiveness of the proposed network on skin lesion detection.

To assess the detailed contribution of the proposed biDFL strategy on skin lesion segmentation, we design some controllable experiments with variable conditions for bi-directional feature learning scheme. ISBI 2017 dataset is taken as the illustration example, since it includes more challenging lesion cases than ISBI 2016 dataset. Table 5.3 shows the performance of dermoscopic feature learning with message propagation in single forward and backward directions (denoted by DFL₊ and DFL₋, respectively). Compared to the baseline model, the proposed framework with feature passing along single direction (forward or backward) improves the JAC by 1.5%-2.6%, which validates the effectiveness of directional feature propagation

FIGURE 5.4: Qualitative segmented result comparison on ISBI 2016 dataset. First row to third row: segmentation results by baseline (green line), baseline + biDFL (blue line) and baseline + biDFL + mCDF (yellow line) respectively, where ground truths are annotated as red line. Last row: detailed segmentation results within sub-region (marked by white arrow in the corresponding column).

on skin lesion segmentation. Moreover, the combination of feature passing along two complementary directions further boosts the high-level parsing performance of the network, resulting in an additional increase of the JAC by 0.5%-1.2%. We also observe that feature passing along the forward direction performs slightly better than feature passing along the backward direction on lesion detection, i.e. improving the JAC by 0.1%-0.3%.

FIGURE 5.5: Qualitative segmented result comparison on ISBI 2017 dataset. First row to third row: segmentation results by baseline (green line), baseline + biDFL (blue line) and baseline + biDFL + mCDF (yellow line) respectively, where ground truths are annotated as red line. Last row: detailed segmentation results within sub-region (marked by white arrow in the corresponding column).

TABLE 5.4: Performance comparison of different σ^2 values (%)

Parameter	JAC	DIC	ACC	GME
$\sigma^2=1$	79.02	86.62	93.85	90.07
$\sigma^2=10$	81.47	88.54	94.65	92.67
$\sigma^2=100$	80.14	87.49	94.29	91.42

5.3.4.2 Evaluation of Parameter of Consistent Decision Fusion

As formulated in Eq. 5.5, the parameter σ controls the sensitivity of the consistency coefficient with respect to the variation of the decision in a local region. Table 5.4 shows the segmentation results based on parameter σ with different values. With

TABLE 5.5: Performance comparison with other methods on ISBI 2016 dataset (%)

Method	JAC	DIC	ACC
Team-EXB [141]	84.30	91.00	95.30
Team-CUMED [3]	82.90	89.70	94.90
Team-Rahman [141]	82.22	89.50	95.20
Team-SFU [141]	81.11	88.50	94.40
Team-TMU [141]	81.10	88.80	94.60
Yuan <i>et al.</i> [95]	84.70	91.2	95.50
Bi <i>et al.</i> [5]	84.64	91.18	95.51
Yuan <i>et al.</i> [97]	84.90	91.30	95.70
Bi <i>et al.</i> [4]	85.92	91.77	95.78
Proposed method	88.12	93.33	96.75

TABLE 5.6: Performance comparison with other methods on ISBI 2017 dataset (%)

Method	JAC	DIC	ACC
Team-Yuan [97]	76.50	84.90	93.40
Team-Berseth [142]	76.20	84.70	93.20
Team-popleyi [142]	76.00	84.40	93.40
Team-Ahn [142]	75.80	84.40	93.40
Team-RECOD [142]	75.40	83.90	93.10
Lin <i>et al.</i> [96]	65.00	79.00	N.A
Bi <i>et al.</i> [4]	77.73	85.66	94.08
Proposed method	81.47	88.54	94.65

the small parameter like $\sigma^2 = 1$, network is too sensitive to particular layer and thus only leads to a slight improvement of the segmentation performance. As the increase of parameter σ , network tends to produce more stable segmentation result, since a large σ is more robust against the artificial noise. In our experiment, the best lesion segmentation is obtained at around $\sigma^2 = 10$.

TABLE 5.7: Performance comparison with other methods on ISBI 2016 dataset for melanoma and non-melanoma cases (%)

Method	Melanoma		Non-Melanoma	
	JAC	DIC	JAC	DIC
Team-EXB [141]	82.94	90.11	84.64	91.18
Team-CUMED [3]	82.90	89.98	82.95	89.68
Team-Rahman [141]	82.65	89.93	82.04	89.44
Team-SFU [141]	81.88	89.44	80.88	88.32
Team-TMU [141]	82.31	89.68	80.73	88.58
Bi <i>et al.</i> [5]	85.84	92.03	84.34	90.97
Bi <i>et al.</i> [4]	85.62	91.72	85.80	91.78
Proposed method	88.38	93.62	88.06	93.26

TABLE 5.8: Performance comparison with other methods on ISBI 2017 dataset for melanoma and non-melanoma cases (%)

Method	Melanoma		Non-Melanoma	
	JAC	DIC	JAC	DIC
Team-Yuan [97]	71.20	81.04	77.78	85.81
Team-Berseth [142]	68.82	79.07	78.02	86.07
Team-popleyi [142]	69.28	79.62	77.60	85.50
Team-Ahn [142]	69.06	79.48	77.45	85.38
Team-RECOD [142]	68.78	79.08	77.00	85.11
Bi <i>et al.</i> [4]	72.18	81.65	79.07	86.63
Proposed method	77.26	85.35	82.49	89.31

5.3.5 Comparison with Other Published Methods

We also compare our method with the approaches from top 10 different teams in the competitions of challenge of ISBI 2016 and 2017 datasets and 7 other top published methods, as shown in Tables 5.5 and 5.6. Compared to those state-of-the-arts approaches, the proposed model achieves the best lesion segmentation results for ISBI 2016 and 2017 dataset consistently. We also show the segmentation performance of different methods on melanoma and non-melanoma cases in Tables

FIGURE 5.6: Distribution comparison in terms of Jaccard index. (a) ISBI 2016 dataset. (b) ISBI 2017 dataset.

5.7 and 5.8 separately, where the proposed method shows superiority on the lesion segmentation in both cases consistently.

The distribution comparison in terms of Jaccard index on two skin lesion datasets is shown in Fig. 5.6. For ISBI 2016 dataset, the proposed architecture produces 57.8% segmentation results with Jaccard index higher than 90%. It is a noticeable enhancement than other compared methods, e.g. 17.4% higher than the second best Team-EXB [141]. For more difficult ISBI 2017 dataset, the proposed architecture

generates 70.3% segmentation results with Jaccard index higher than 80%, leading to the increase of 12.6% compared to the second best Team-Yuan [97]. Analysis of distribution of Jaccard index indicates the utility and stability of our proposed method on skin lesion segmentation, where segmentation results with high Jaccard index occupy the majority of both ISBI 2016 and 2017 datasets, better than other methods used for comparison.

5.4 Summary

In this chapter, we propose a bi-directional dermoscopic feature learning framework with multiscale decision fusion for dermoscopic image segmentation. By the feature information propagation through two complementary directions among high-level layers, we generate a rich and discriminative feature representation for skin lesion. By the multiscale consistent decision fusion, we achieve more reliable and precise skin lesion delineation. Experimental results have demonstrated the effectiveness of the proposed method on skin lesion segmentation, which yields the state-of-the-art performance on two publicly available datasets consistently.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

Medical image segmentation plays a crucial rule in the computer-aided diagnosis of diseases. This dissertation focuses on the development of automatic segmentation systems for medical image data, i.e. multistage classification for fundus images, and deep feature learning for dermoscopic images. More detailed information for those automatic segmentation systems are summarized as follows.

First, we present a novel and robust cascade classification framework that generates a valid decision boundary for retinal vessel and background. This classification framework envelops a set of computationally efficient Mahalanobis distance classifiers for multistage classification. The proposed nonlinear decision model is trained through a one-pass feed forward process, which benefits the network to generate a definite training and testing result that can be exactly reproduced. Furthermore, this classification framework through multistage learning is flexibly extended to other pixel-wise segmentation tasks.

Second, a hierarchical classification architecture, named divide-and-conquer funnel-structured classification framework, is proposed to boost the segmentation performance of retinal vessels with different scales and positions. It overcomes the

limitation of the previous works for retinal vessel segmentation, i.e. they usually train a global discriminative model thus regardless of the geometrical structure difference between different retinal vessel. The proposed multiplex vessel partition algorithm separates retinal vessels with the same geometrical property into the well constrained subsets, which enhances the capability of the network on the detection of the retinal vessels with different scales and positions. In addition, a funnel-structured vessel segmentation framework is proposed to further improve the discriminative capability of the decision model by integrating classification results from each disjoint subset.

Last, a deep leaning based classification framework is proposed for dermoscopic image segmentation. We introduce a bi-directional feature dermoscopic learning for the rich and discriminative feature representation of the skin lesion. It aggregates feature information generated at multiple receptive fields along two complementary directions, which gives the network improved insight of the complex structure of the skin lesion. Moreover, a multiscale consistent decision fusion is presented to selectively adjust the confidence of the decision generated at multiple layers, which enhances the reliability of the network on skin lesion detection.

6.2 Future Work

Though promising results have been achieved for medical image segmentation in this thesis, there is still much effort that should be made for more accurate, efficient and user-friendly computer-aided diagnosis system. In the future, we aim to investigate the cooperation of the pixel-wise skin lesion segmentation and melanoma recognition.

Recent classification system integrates global average pooling layer for object recognition. Global average pooling takes the average of each feature map, thus robust to spatial translation of the input [143]. Nevertheless, global average pooling is insufficient for the area of the skin lesion recognition. The first reason is that taking

the averaging of each feature map extracts some redundant even misleading feature information from background for the representation of skin lesions. For example, there are many dermoscopic images that contain color disc and rulers for the measurement of the appearance of the skin lesion. Feature messages from those areas definitely reduce the discriminative ability of the network. The second reason is that global average pooling focuses on the whole lesion texture while ignores more discriminative feature information of the lesion recognition. For example, feature information from the periphery region of the skin lesion is important for melanoma recognition. Clinical investigations have found that the visual characteristics of the periphery region of the lesion provide crucial information for the diagnosis of melanoma in dermoscopic images, especially two pronounced structures, i.e. pigment networks and streaks [144].

Thus, for skin lesion segmentation, we investigate a diagnosis guidance feature fusion which exploits the information from specific skin lesion class, as melanoma and non-melanoma have different pathological feature representation. For melanoma recognition, we explore the incorporation of melanoma recognition and the signal from lesion segmentation framework. The first benefit of incorporating information related to the segmentation process into melanoma recognition is that we capture the reliable feature information from the candidate lesion area and thus reduce the interference of the misleading information from the background. The second benefit is that we achieve more discriminative lesion representation for melanoma recognition by taking the feature information from peripheral and center region of the skin lesion.

Author's Publications

Journal Papers

- [J1] **X. Wang**, X. Jiang, and J. Ren, “Blood vessel segmentation from fundus image by a cascade classification framework”, *Pattern Recognition*, vol. 88, pp. 331-341, 2019.
- [J2] **X. Wang**, X. Jiang, “Retinal vessel segmentation by a divide-and-conquer funnel-structured classification framework”, *Signal Processing*, vol. 165, pp. 104-114, 2019.
- [J3] **X. Wang**, X. Jiang, H. Ding, J. Liu, “Bi-directional dermoscopic feature learning network with multi-scale consistent decision fusion for skin lesion segmentation”, *IEEE Transaction on Image Processing*, vol. 29, pp. 3039-3051, 2019.

Conference Papers

- [C1] **X. Wang**, H. Ding and X. Jiang, “Dermoscopic image segmentation through the enhanced high-level parsing and class weighted loss”, in *Proceedings of the 26th IEEE International Conference on Image Processing (ICIP)*, 2019.
- [C2] **X. Wang**, and X. Jiang, “Post-processing for retinal vessel detection”, in *Proceedings of the 10th International Conference on Digital Image Processing (ICDIP)*, p.1080656, 2018.

- [C3] **X. Wang**, and X. Jiang, “Enhancing retinal vessel segmentation by color fusion”, in Proceedings of the 42nd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 891C895, 2017.

- [C4] **X. Wang**, and X. Jiang, “Nonlinear retinal image enhancement for vessel detection”, in Proceedings of the 9th International Conference on Digital Image Processing (ICDIP), p. 104202M, 2017.

Bibliography

- [1] P. Powar and C. Jadhav, “Comparative study of classifiers for diagnosis of microaneurysm,” *Int. J. Comput. Intell. Res.*, vol. 12, no. 2, pp. 139–148, 2016. [xix](#), [2](#)
- [2] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440. [xix](#), [16](#), [17](#), [18](#), [77](#), [85](#)
- [3] L. Yu, H. Chen, Q. Dou, J. Qin, and P. Heng, “Automated melanoma recognition in dermoscopy images via very deep residual networks,” *IEEE Trans. Med. Imaging*, vol. 36, no. 4, pp. 994–1004, 2017. [xix](#), [18](#), [19](#), [93](#), [94](#)
- [4] L. Bi, J. Kim, E. Ahn, A. Kumar, D. Feng, and M. Fulham, “Step-wise integration of deep class-specific learning for dermoscopic image segmentation,” *Pattern Recogn.*, vol. 85, pp. 78–89, 2019. [xix](#), [18](#), [19](#), [76](#), [93](#), [94](#)
- [5] L. Bi, J. Kim, E. Ahn, A. Kumar, M. Fulham, and D. Feng, “Dermoscopic image segmentation via multi-stage fully convolutional networks,” *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2065–2074, 2017. [xix](#), [18](#), [19](#), [76](#), [93](#), [94](#)
- [6] 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 Trans. Med. Imag.*, vol. 25, no. 9, pp. 1214–1222, 2006. [xx](#), [15](#), [28](#), [38](#), [43](#), [44](#), [45](#), [46](#), [47](#), [55](#), [63](#), [66](#), [67](#), [68](#), [69](#), [71](#), [72](#)

- [7] 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 Trans. Med. Imag.*, vol. 30, no. 1, pp. 146–158, 2011. [xx](#), [15](#), [23](#), [27](#), [43](#), [44](#), [45](#), [46](#), [63](#), [66](#), [67](#), [69](#)
- [8] U. T. Nguyen, A. Bhuiyan, L. A. Park, and K. Ramamohanarao, “An effective retinal blood vessel segmentation method using multi-scale line detection,” *Pattern Recognit.*, vol. 46, no. 3, pp. 703–715, 2013. [xx](#), [2](#), [43](#), [44](#), [45](#), [46](#), [47](#), [61](#), [63](#), [66](#), [67](#), [68](#), [69](#), [71](#), [72](#)
- [9] J. Orlando, E. Prokofyeva, and M. Blaschko, “A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images,” *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 16–27, 2017. [xx](#), [15](#), [23](#), [38](#), [40](#), [45](#), [46](#), [47](#), [63](#), [66](#), [67](#), [68](#), [69](#), [71](#), [72](#)
- [10] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi, “Iterative vessel segmentation of fundus images,” *IEEE Trans. Biomed. Eng.*, vol. 62, no. 7, pp. 1738–1749, 2015. [xx](#), [12](#), [63](#), [66](#), [67](#), [68](#), [69](#), [70](#)
- [11] G. Azzopardi, N. Strisciuglio, M. Vento, and N. Petkov, “Trainable cosfire filters for vessel delineation with application to retinal images,” *Med. Image Anal.*, vol. 19, no. 1, pp. 46–57, 2015. [xx](#), [12](#), [40](#), [44](#), [46](#), [47](#), [63](#), [66](#), [67](#), [68](#), [69](#), [71](#), [72](#)
- [12] H. Fu, Y. Xu, S. Lin, D. Wong, and J. Liu, “Deepvessel: retinal vessel segmentation via deep learning and conditional random field,” in *International Conference on Medical Image Computing and Computer-assisted Intervention*, 2016, pp. 132–139. [xx](#), [63](#), [66](#), [67](#), [68](#), [69](#), [71](#), [72](#)
- [13] J. Baker, E. Rosen, J. Lo, E. Gimenez, R. Walsh, and M. Soo, “Computer-aided detection (cad) in screening mammography: sensitivity of commercial cad systems for detecting architectural distortion,” *American Journal of Roentgenology*, vol. 181, no. 4, pp. 1083–1088, 2003. [1](#)

- [14] H. Kobatake, “Future cad in multi-dimensional medical images:project on multi-organ, multi-disease cad system,” *Comput. Med. Imag. Grap.*, vol. 31, no. 4-5, pp. 258–266, 2007. [1](#)
- [15] E. El-Dahshan, H. Mohsen, K. Revett, and A. Salem, “Computer-aided diagnosis of human brain tumor through mri: a survey and a new algorithm,” *Expert Syst. Appl.*, vol. 41, no. 11, pp. 5526–5545, 2014. [1](#)
- [16] H. Chen, Q. Dou, L. Yu, J. Qin, and P. Heng, “Voxresnet: deep voxelwise residual networks for brain segmentation from 3d mr images,” *NeuroImage*, vol. 170, pp. 446–455, 2018. [1](#)
- [17] B. Xiang, *Knowledge-based image segmentation using sparse shape priors and high-order MRFs*, Ph.D. thesis, Châtenay-Malabry, Ecole centrale de Paris, 2013. [2](#)
- [18] M. Abràmoff, M. Garvin, and M. Sonka, “Retinal imaging and image analysis,” *IEEE Rev. Biomed. Eng.*, vol. 3, pp. 169–208, 2010. [2](#), [3](#)
- [19] G. S. Ramlugun, V. K. Nagarajan, and C. Chakraborty, “Small retinal vessels extraction towards proliferative diabetic retinopathy screening,” *Expert Syst. Appl.*, vol. 39, no. 1, pp. 1141–1146, 2012. [2](#)
- [20] A. Christodoulidis, *Segmentation and characterization of small retinal vessels in fundus images using the tensor voting approach*, Ph.D. thesis, Ecole Polytechnique, Montreal (Canada), 2017. [2](#)
- [21] B. Yin, H. Li, B. Sheng, X. Hou, Y. Chen, W. Wu, P. Li, R. Shen, Y. Bao, and W. Jia, “Vessel extraction from non-fluorescein fundus images using orientation-aware detector,” *Med. Image Anal.*, vol. 26, no. 1, pp. 232–242, 2015. [2](#), [44](#)
- [22] 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,” *Med. Image Anal.*, vol. 11, no. 1, pp. 47–61, 2007. [2](#), [66](#)

- [23] J. Zhang, B. Dashtbozorg, E. Bekkers, J. Pluim, R. Duits, and B. ter Haar Romeny, “Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores,” *IEEE Trans. Med. Imag.*, vol. 35, no. 12, pp. 2631–2644, 2016. [2](#), [44](#)
- [24] G. Kovács and A. Hajdu, “A self-calibrating approach for the segmentation of retinal vessels by template matching and contour reconstruction,” *Med. Image Anal.*, vol. 29, pp. 24–46, 2016. [2](#)
- [25] M. Niemeijer, B. Van Ginneken, M. J. Cree, A. Mizutani, G. Quellec, C. I. Sánchez, B. Zhang, R. Hornero, M. Lamard, C. Muramatsu, X. Q. Wu, G. Cazuguel, J. You, A. Mayo, Q. Li, Y. Hatanaka, B. Cochener, C. Roux, F. Karray, M. García, H. Fujita, and M. D. Abrámoff, “Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs,” *IEEE Trans. Med. Imag.*, vol. 29, no. 1, pp. 185–195, 2010. [3](#)
- [26] A. Rocha, T. Carvalho, H. F. Jelinek, S. Goldenstein, and J. Wainer, “Points of interest and visual dictionaries for automatic retinal lesion detection,” *IEEE Trans. Biomed. Eng.*, vol. 59, no. 8, pp. 2244–2253, 2012. [3](#)
- [27] M. Niemeijer, M. D. Abrámoff, and G. B. Van, “Fast detection of the optic disc and fovea in color fundus photographs,” *Med. Image Anal.*, vol. 13, no. 6, pp. 859–870, 2009. [3](#)
- [28] A. Hoover and M. Goldbaum, “Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels,” *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 951–958, 2003. [3](#)
- [29] C. Heneghan, J. Flynn, M. O’Keefe, and M. Cahill, “Characterization of changes in blood vessel width and tortuosity in retinopathy of prematurity using image analysis,” *Med. Image Anal.*, vol. 6, no. 4, pp. 407–429, 2002. [3](#)
- [30] C. M. Wilson, K. D. Cocker, M. J. Moseley, C. Paterson, S. T. Clay, W. E. Schulenburg, M. D. Mills, A. L. Ells, K. H. Parker, G. E. Quinn, A. R. Fielder, and J. Ng, “Computerized analysis of retinal vessel width and tortuosity in

- premature infants,” *Invest. Ophthalmol. Vis. Sci.*, vol. 49, no. 8, pp. 3577–3585, 2008. 3
- [31] U. T. V. Nguyen, A. Bhuiyan, L. Park, R. Kawasaki, T. Y. Wong, J. J. Wang, P. Mitchell, K. Ramamohanarao, T. Y. Wong, and J. J. Wang, “An automated method for retinal arteriovenous nicking quantification from color fundus images,” *IEEE Trans. Biomed. Eng.*, vol. 60, no. 11, pp. 3194–3203, 2013. 3
- [32] G. K. Matsopoulos, P. Asvestas, N. Mouravliansky, and K. K. Delibasis, “Multimodal registration of retinal images using self organizing maps,” *IEEE Trans. Med. Imag.*, vol. 23, no. 12, pp. 1557–1563, 2004. 3
- [33] B. Fang and Y. Y. Tang, “Elastic registration for retinal images based on reconstructed vascular trees,” *IEEE Trans. Biomed. Eng.*, vol. 53, no. 6, pp. 1183–1187, 2006. 3
- [34] E. Grisan, M. Foracchia, and A. Ruggeri, “A novel method for the automatic grading of retinal vessel tortuosity,” *IEEE Trans. Med. Imag.*, vol. 27, no. 3, pp. 310–319, 2008. 3
- [35] 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 Trans. Med. Imag.*, vol. 23, no. 10, pp. 1196–1204, 2004. 3
- [36] A. M. Broehan, T. Rudolph, C. Amstutz, and J. H. Kowal, “Real-time multimodal retinal image registration for a computer-assisted laser photocoagulation system,” *IEEE Trans. Biomed. Eng.*, vol. 58, no. 10, pp. 2816–2824, 2011. 3
- [37] S. M. Lajevardi, A. Arakala, S. A. Davis, and K. J. Horadam, “Retina verification system based on biometric graph matching,” *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3625–3635, 2013. 3

- [38] P. Kharazmi, M. I. AlJasser, H. Lui, Z. J. Wang, and T. K. Lee, “Automated detection and segmentation of vascular structures of skin lesions seen in dermoscopy, with an application to basal cell carcinoma classification,” *IEEE J. Biomed. Health Inform.*, vol. 21, no. 6, pp. 1675–1684, 2017. [3](#), [75](#)
- [39] T. Satheesha, D. Satyanarayana, M. Prasad, and K. Dhruve, “Melanoma is skin deep: a 3d reconstruction technique for computerized dermoscopic skin lesion classification,” *IEEE J. Transl. Eng. He.*, vol. 5, pp. 1–17, 2017. [3](#)
- [40] K. Korotkov, J. Quintana, S. Puig, J. Malveyh, and R. Garcia, “A new total body scanning system for automatic change detection in multiple pigmented skin lesions,” *IEEE Trans. Med. Imag.*, vol. 34, no. 1, pp. 317–338, 2014. [4](#), [5](#)
- [41] L. Ma and R. C. Staunton, “Analysis of the contour structural irregularity of skin lesions using wavelet decomposition,” *Pattern Recogn.*, vol. 46, no. 1, pp. 98–106, 2013. [4](#), [76](#)
- [42] C. Barata, M. E. Celebi, and J. S. Marques, “Development of a clinically oriented system for melanoma diagnosis,” *Pattern Recogn.*, vol. 69, pp. 270–285, 2017. [4](#)
- [43] M. Bhuiyan, I. Azad, and M. Uddin, “Image processing for skin cancer features extraction,” *Int. J. Sci. Eng. Res.*, vol. 4, no. 2, pp. 1–6, 2013. [5](#)
- [44] Y. Tolia and S. M. Panas, “A fuzzy vessel tracking algorithm for retinal images based on fuzzy clustering,” *IEEE Trans. Med. Imag.*, vol. 17, no. 2, pp. 263–273, 1998. [11](#)
- [45] E. Bekkers, R. Duits, T. Berendschot, and B. ter Haar Romeny, “A multi-orientation analysis approach to retinal vessel tracking,” *J. Math. Imaging Vision*, vol. 49, no. 3, pp. 583–610, 2014. [11](#)
- [46] M. Yüksel and M. Borlu, “Accurate segmentation of dermoscopic images by image thresholding based on type-2 fuzzy logic,” *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 4, pp. 976–982, 2009. [12](#)

- [47] X. Jiang and D. Mojon, “Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 1, pp. 131–137, 2003. [12](#)
- [48] F. Zana and J. C. Klein, “Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation,” *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 1010–1019, 2001. [12](#)
- [49] L. Rodrigues and M. Marengoni, “Segmentation of optic disc and blood vessels in retinal images using wavelets, mathematical morphology and hessian-based multi-scale filtering,” *Biomed. Signal Process. Control*, vol. 36, pp. 39–49, 2017. [12](#)
- [50] E. Imani, M. Javidi, and H. R. Pourreza, “Improvement of retinal blood vessel detection using morphological component analysis,” *Comput. Methods Programs Biomed.*, vol. 118, no. 3, pp. 263–279, 2015. [12](#), [23](#)
- [51] H. Wang, R. Moss, X Chen, R Stanley, W Stoecker, M Celebi, J Malter, J. Grichnik, A. Marghoob, H. Rabinovitz, et al., “Modified watershed technique and post-processing for segmentation of skin lesions in dermoscopy images,” *Comput. Med. Imag. Grap.*, vol. 35, no. 2, pp. 116–120, 2011. [12](#), [14](#), [15](#)
- [52] H. Zhou, G. Schaefer, M. Celebi, F. Lin, and T. Liu, “Gradient vector flow with mean shift for skin lesion segmentation,” *Comput. Med. Imag. Grap.*, vol. 35, no. 2, pp. 121–127, 2011. [12](#)
- [53] Z. Ma and J. M. R. Tavares, “A novel approach to segment skin lesions in dermoscopic images based on a deformable model,” *IEEE J. Biomed. Health Inform.*, vol. 20, no. 2, pp. 615–623, 2016. [12](#)
- [54] B. Al-Diri, A. Hunter, and D. Steel, “An active contour model for segmenting and measuring retinal vessels,” *IEEE Trans. Med. Imag.*, vol. 28, no. 9, pp. 1488–1497, 2009. [12](#)

- [55] Y. Q. Zhao, X. H. Wang, X. F. Wang, and F. Y. Shih, “Retinal vessels segmentation based on level set and region growing,” *Pattern Recognit.*, vol. 47, no. 7, pp. 2437–2446, 2014. [12](#), [44](#), [55](#), [56](#)
- [56] Y. T. Zhao, Lavdie Rada, Ke Chen, Simon P Harding, and Yalin Zheng, “Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images,” *IEEE Trans. Med. Imag.*, vol. 34, no. 9, pp. 1797–1807, 2015. [12](#)
- [57] A. Gooya, H. Liao, K. Matsumiya, K. Masamune, Y. Masutani, and T. Dohi, “A variational method for geometric regularization of vascular segmentation in medical images,” *IEEE Trans. Image Process.*, vol. 17, no. 8, pp. 1295–1312, 2008. [12](#)
- [58] T. Stošić and B. D. Stošić, “Multifractal analysis of human retinal vessels,” *IEEE Trans. Med. Imag.*, vol. 25, no. 8, pp. 1101–1107, 2006. [12](#)
- [59] L. Wang, A. Bhalerao, and R. Wilson, “Analysis of retinal vasculature using a multiresolution hermite model,” *IEEE Trans. Med. Imag.*, vol. 26, no. 2, pp. 137–152, 2007. [12](#)
- [60] P. Schmid, “Segmentation of digitized dermatoscopic images by two-dimensional color clustering,” *IEEE Trans. Med. Imaging*, vol. 18, no. 2, pp. 164–171, 1999. [12](#)
- [61] H. Zhou, G. Schaefer, A. Sadka, and M. Celebi, “Anisotropic mean shift based fuzzy c-means segmentation of deroscopy images,” *IEEE J. Sel. Top. Signal Process.*, vol. 3, no. 1, pp. 26–34, 2009. [13](#)
- [62] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, “Multiscale vessel enhancement filtering,” in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, 1998, pp. 130–137. [13](#), [27](#)

- [63] X. Wang, X. Jiang, and J. Ren, “Blood vessel segmentation from fundus image by a cascade classification framework,” *Pattern Recognit.*, vol. 88, pp. 331–341, 2019. [14](#)
- [64] X. Wang and X. Jiang, “Enhancing retinal vessel segmentation by color fusion,” in *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing*, 2017, pp. 891–895. [14](#), [43](#), [58](#)
- [65] L. Tang, M. Niemeijer, J. M. Reinhardt, M. K. Garvin, and M. D. Abràmoff, “Splat feature classification with application to retinal hemorrhage detection in fundus images,” *IEEE Trans. Med. Imag.*, vol. 32, no. 2, pp. 364–375, 2013. [14](#), [27](#)
- [66] 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 Trans. Biomed. Eng.*, vol. 59, no. 9, pp. 2538–2548, 2012. [14](#), [16](#), [27](#), [38](#), [43](#), [44](#), [55](#), [66](#)
- [67] M. Jahanifar, N. Z. Tajeddin, B. M. Asl, and A. Gooya, “Supervised saliency map driven segmentation of lesions in dermoscopic images,” *IEEE J. Biomed. Health Inform.*, 2018. [14](#)
- [68] D. Fradkin and I. Muchnik, “Support vector machines for classification,” *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol. 70, pp. 13–20, 2006. [14](#)
- [69] Y. He and F. Xie, “Automatic skin lesion segmentation based on texture analysis and supervised learning,” in *Asian Conference on Computer Vision*. Springer, 2012, pp. 330–341. [14](#)
- [70] X. You, Q. Peng, Y. Yuan, Y. M. Cheung, and J. Lei, “Segmentation of retinal blood vessels using the radial projection and semi-supervised approach,” *Pattern Recognit.*, vol. 44, no. 10, pp. 2314–2324, 2011. [15](#), [23](#), [44](#), [58](#), [66](#)

- [71] E. Ricci and R. Perfetti, “Retinal blood vessel segmentation using line operators and support vector classification,” *IEEE Trans. Med. Imag.*, vol. 26, no. 10, pp. 1357–1365, 2007. [15](#)
- [72] E. Tuba, L. Mrkela, and M. Tuba, “Retinal blood vessel segmentation by support vector machine classification,” in *Proceedings of the International Conference Radioelektronika*, 2017, pp. 1–6. [15](#)
- [73] X. Wang and X. Jiang, “Nonlinear retinal image enhancement for vessel detection,” in *Proceedings of the International Conference on Digital Image Processing*, 2017. [25](#), [55](#)
- [74] X. Wang and X. Jiang, “Post-processing for retinal vessel detection,” in *Proceedings of the International Conference on Digital Image Processing*, 2018. [15](#), [36](#)
- [75] N. Altman, “An introduction to kernel and nearest-neighbor nonparametric regression,” *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992. [15](#)
- [76] M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, and M. D. Abramoff, “Comparative study of retinal vessel segmentation methods on a new publicly available database,” in *Proceedings of the Medical Imaging*, 2004, pp. 648–656. [15](#)
- [77] J. Staal, M. D. Abramoff, M. Niemeijer, M. Viergever, and B. Van Ginneken, “Ridge-based vessel segmentation in color images of the retina,” *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 501–509, 2004. [15](#), [36](#), [38](#), [44](#), [66](#)
- [78] R. Vega, G. Sanchez-Ante, L. E. Falcon-Morales, H. Sossa, and E. Guevara, “Retinal vessel extraction using lattice neural networks with dendritic processing,” *Comput. Biol. Med.*, vol. 58, pp. 20–30, 2015. [15](#), [23](#), [40](#), [44](#), [46](#), [67](#)
- [79] C. A. Lupascu, D. Tegolo, and E. Trucco, “Fabrc: retinal vessel segmentation using adaboost,” *IEEE Trans. Inform. Technol. Biomed.*, vol. 14, no. 5, pp. 1267–1274, 2010. [16](#)

- [80] S. Roychowdhury, D. Koozekanani, and K. Parhi, “Blood vessel segmentation of fundus images by major vessel extraction and sub-image classification,” *IEEE J. Biomed. Health Inform.*, vol. 19, no. 3, pp. 1118–1128, 2015. [16](#), [38](#), [44](#)
- [81] B. Barkana, I. Saricicek, and B. Yildirim, “Performance analysis of descriptive statistical features in retinal vessel segmentation via fuzzy logic, ann, svm, and classifier fusion,” *Knowl. Based Syst.*, vol. 118, pp. 165–176, 2017. [16](#), [44](#), [58](#)
- [82] X. Wang, H. Ding, and X. Jiang, “Dermoscopic image segmentation through the enhanced high-level parsing and class weighted loss,” in *IEEE International Conference on Image Processing*, 2019, pp. 245–249. [16](#)
- [83] H. Ding, X. Jiang, B. Shuai, A. Liu, and G. Wang, “Semantic correlation promoted shape-variant context for segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8885–8894. [16](#)
- [84] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778. [16](#), [18](#), [84](#)
- [85] S. Pereira, A. Pinto, V. Alves, and C. Silva, “Brain tumor segmentation using convolutional neural networks in mri images,” *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016. [16](#)
- [86] G. Wang, W. Li, M. Zuluaga, R. Pratt, P. Patel, M. Aertsen, T. Doel, A. David, J. Deprest, S. Ourselin, et al., “Interactive medical image segmentation using deep learning with image-specific fine-tuning,” *IEEE Trans. Med. Imaging*, 2018. [16](#)
- [87] H. Chen, X. Qi, L. Yu, and P. Heng, “Dcan: deep contour-aware networks for accurate gland segmentation,” in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2016, pp. 2487–2496. [16](#)

-
- [88] A. BenTaieb and G. Hamarneh, “Topology aware fully convolutional networks for histology gland segmentation,” in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 460–468. [16](#)
- [89] W. Li, F. Jia, and Q. Hu, “Automatic segmentation of liver tumor in ct images with deep convolutional neural networks,” *Journal of Computer and Communications*, vol. 3, no. 11, pp. 146, 2015. [16](#)
- [90] X. Li, H. Chen, X. Qi, Q. Dou, C. Fu, and P. Heng, “H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes,” *IEEE Trans. Med. Imag.*, vol. 37, no. 12, pp. 2663–2674, 2018. [16](#)
- [91] S. Belal, M. Sadik, R. Kaboteh, O. Enqvist, J. Ulén, M. Poulsen, J. Simonsen, P. Høilund-Carlsen, L. Edenbrandt, and E. Trägårdh, “Deep learning for segmentation of 49 selected bones in ct scans: first step in automated pet/ct-based 3d quantification of skeletal metastases,” *European Journal of Radiology*, vol. 113, pp. 89–95, 2019. [16](#)
- [92] D. Pham, G. Dovletov, S. Warwas, S. Landgraeber, M. Jäger, and J. Pauli, “Deep learning with anatomical priors: imitating enhanced autoencoders in latent space for improved pelvic bone segmentation in mri,” *arXiv preprint arXiv:1903.09263*, 2019. [16](#)
- [93] H. Ding, X. Jiang, B. Shuai, A. Liu, and G. Wang, “Semantic segmentation with context encoding and multi-path decoding,” *IEEE Trans. Image Process.*, 2020. [16](#)
- [94] B. Shuai, Z. Zuo, B. Wang, and G. Wang, “Scene segmentation with dag-recurrent neural networks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 6, pp. 1480–1493, 2018. [17](#), [77](#), [85](#)
- [95] Y. Yuan, M. Chao, and Y. Lo, “Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance,” *IEEE Trans. Med. Imaging*, vol. 36, no. 9, pp. 1876–1886, 2017. [19](#), [76](#), [93](#)

- [96] B. Lin, K. Michael, S. Kalra, and H. Tizhoosh, “Skin lesion segmentation: U-nets versus clustering,” in *Proceedings of the IEEE Symposium Series on Computational Intelligence*, 2017, pp. 1–7. [19](#), [93](#)
- [97] Y. Yuan and Y. Lo, “Improving dermoscopic image segmentation with enhanced convolutional-deconvolutional networks,” *IEEE J. Biomed. Health Inform*, 2017. [19](#), [93](#), [94](#), [96](#)
- [98] M. Al-masni, M. Al-antari, M. Choi, S. Han, and T. Kim, “Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks,” *Comput. Meth. Prog. Bio.*, vol. 162, pp. 221–231, 2018. [20](#)
- [99] G. Lin, A. Milan, C. Shen, and I. Reid, “Refinenet: multi-path refinement networks for high-resolution semantic segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, vol. 1, p. 5. [20](#)
- [100] H. Li, X. He, F. Zhou, Z. Yu, D. Ni, S. Chen, T. Wang, and B. Lei, “Dense deconvolutional network for skin lesion segmentation,” *IEEE J. Biomed. Health Inform*, 2018. [20](#)
- [101] 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 Trans. Med. Imag.*, vol. 35, no. 1, pp. 109–118, 2016. [20](#), [38](#), [43](#), [44](#), [66](#)
- [102] P. Liskowski and K. Krawiec, “Segmenting retinal blood vessels with deep neural networks,” *IEEE Trans. Med. Imag.*, vol. 35, no. 11, pp. 2369–2380, 2016. [20](#), [38](#), [43](#), [44](#)
- [103] Z. Yan, X. Yang, and K. Cheng, “Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation,” *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1912–1923, 2018. [20](#), [66](#)
- [104] A. Oliveira, S. Pereira, and C. Silva, “Retinal vessel segmentation based on fully convolutional neural networks,” *Expert Syst. Appl.*, vol. 112, pp. 229–242, 2018. [20](#), [65](#), [66](#)

- [105] K. Hu, Z. Zhang, X. Niu, Y. Zhang, C. Cao, F. Xiao, and X. Gao, “Retinal vessel segmentation of color fundus images using multiscale convolutional neural network with an improved cross-entropy loss function,” *Neurocomputing*, vol. 309, pp. 179–191, 2018. [20](#), [66](#)
- [106] Z. Yan, Xi. Yang, and K. Cheng, “A three-stage deep learning model for accurate retinal vessel segmentation,” *IEEE J. Biomed. Health Inf.*, vol. 23, no. 4, pp. 1427–1436, 2018. [20](#), [66](#)
- [107] L. Yu, Z. Qin, T. Zhuang, Y. Ding, Z. Qin, and K. Choo, “A framework for hierarchical division of retinal vascular networks,” *Neurocomputing*. [20](#), [66](#)
- [108] S. Feng, Z. Zhuo, D. Pan, and Q. Tian, “Ccnet: a cross-connected convolutional network for segmenting retinal vessels using multi-scale features,” *Neurocomputing*. [20](#), [66](#)
- [109] B. Zhang, L. Zhang, L. Zhang, and F. Karay, “Retinal vessel extraction by matched filter with first-order derivative of gaussian,” *Comput. Biol. Med.*, vol. 40, no. 4, pp. 438–445, 2010. [23](#)
- [110] X. Jiang, “Image detail-preserving filter for impulsive noise attenuation,” *IEE Proc.-Vis. Image Signal Process.*, vol. 150, no. 3, pp. 179–185, 2003. [26](#), [27](#)
- [111] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, “Detection of blood vessels in retinal images using two-dimensional matched filters,” *IEEE Trans. Med. Imag.*, vol. 8, no. 3, pp. 263–269, 1989. [27](#), [52](#)
- [112] Z. Lu, X. Jiang, and A. Kot, “Color space construction by optimizing luminance and chrominance components for face recognition,” *Pattern Recognit.*, vol. 83, pp. 456–568, 2018. [28](#)
- [113] J. Ren and X. Jiang, “Regularized 2-d complex-log spectral analysis and subspace reliability analysis of micro-doppler signature for uav detection,” *Pattern Recognit.*, vol. 69, pp. 225–237, 2017. [28](#)

- [114] J. Zhang, Y. Chen, E. Bekkers, M. Wang, B. Dashtbozorg, and B. ter Haar Romeny, “Retinal vessel delineation using a brain-inspired wavelet transform and random forest,” *Pattern Recognit.*, vol. 69, pp. 107–123, 2017. [28](#), [36](#), [38](#), [40](#), [44](#), [46](#), [66](#), [67](#)
- [115] X. Jiang, “Linear subspace learning-based dimensionality reduction,” *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 16–26, 2011. [28](#), [30](#), [31](#)
- [116] X. Jiang, “Asymmetric principal component and discriminant analyses for pattern classification,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 5, pp. 931–937, 2009. [28](#)
- [117] X. Jiang, B. Mandal, and A. Kot, “Eigenfeature regularization and extraction in face recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 3, pp. 383–394, 2008. [31](#)
- [118] A. Hoover, V. Kouznetsova, and M. Goldbaum, “Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response,” *IEEE Trans. Med. Imag.*, vol. 19, no. 3, pp. 203–210, 2000. [37](#), [43](#), [44](#), [46](#), [47](#), [63](#), [66](#), [67](#), [69](#), [71](#), [72](#)
- [119] L. Gu, X. Zhang, H. Zhang, H. Li, and L. Cheng, “Segment 2d and 3d filaments by learning structured and contextual features,” *IEEE Trans. Med. Imag.*, vol. 36, no. 2, pp. 596–606, 2017. [40](#), [46](#), [67](#), [70](#)
- [120] H. Ding, X. Jiang, B. Shuai, A. Liu, and G. Wang, “Context contrasted feature and gated multi-scale aggregation for scene segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2393–2402. [43](#), [88](#)
- [121] A. M. Mendonca and A. Campilho, “Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction,” *IEEE Trans. Med. Imag.*, vol. 25, no. 9, pp. 1200–1213, 2006. [44](#)

- [122] B. Lam and Y. Gao and A. Liew, “General retinal vessel segmentation using regularization-based multiconcavity modeling,” *IEEE Trans. Med. Imag.*, vol. 29, no. 7, pp. 1369–1381, 2010. [44](#)
- [123] Y. Wang, G. Ji, P. Lin, and E. Trucco, “Retinal vessel segmentation using multiwavelet kernels and multiscale hierarchical decomposition,” *Pattern Recognit.*, vol. 46, no. 8, pp. 2117–2133, 2013. [44](#)
- [124] K. Rezaee, J. Haddadnia, and A. Tashk, “Optimized clinical segmentation of retinal blood vessels by using combination of adaptive filtering, fuzzy entropy and skeletonization,” *Appl. Soft Comput.*, vol. 52, pp. 937–951, 2017. [44](#)
- [125] H. Fu, Y. Xu, D. Wong, and J. Liu, “Retinal vessel segmentation via deep learning network and fully-connected conditional random fields,” in *Proc. 9th Int. Symp. Biomed. Imaging*, 2016, pp. 698–701. [66](#)
- [126] B. Shuai, H. Ding, T. Liu, G. Wang, and X. Jiang, “Toward achieving robust low-level and high-level scene parsing,” *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1378–1390, 2019. [75](#)
- [127] H. Ding, X. Jiang, B. Shuai, A. Liu, and G. Wang, “Semantic correlation promoted shape-variant context for segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019. [75](#)
- [128] X. Wang, X. Jiang, H. Ding, and J. Liu, “Bi-directional dermoscopic feature learning and multi-scale consistent decision fusion for skin lesion segmentation,” *IEEE Trans. Image Process.*, vol. 29, pp. 3039–3051, 2019. [75](#)
- [129] J. Garcia-Arroyo and B. Garcia-Zapirain, “Segmentation of skin lesions in dermoscopy images using fuzzy classification of pixels and histogram thresholding,” *Comput. Meth. Prog. Bio.*, vol. 168, pp. 11–19, 2019. [76](#)
- [130] I. Diaz, “Dermaknet: incorporating the knowledge of dermatologists to convolutional neural networks for skin lesion diagnosis,” *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 547–559, 2018. [76](#)

- [131] V. Singh, M. Abdel-Nasser, H. Rashwan, F. Akram, N. Pandey, A. Lalande, B. Presles, S. Romani, and D. Puig, “Fca-net: adversarial learning for skin lesion segmentation based on multi-scale features and factorized channel attention,” *IEEE Access*, vol. 7, pp. 130552–130565, 2019. [76](#)
- [132] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. Yuille, “Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, 2018. [76](#)
- [133] L. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking atrous convolution for semantic image segmentation,” *arXiv preprint arXiv:1706.05587*, 2017. [76](#)
- [134] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al., “Imagenet large scale visual recognition challenge,” *Int. J. Comput. Vision*, vol. 115, no. 3, pp. 211–252, 2015. [79](#)
- [135] J. Huang, Y. Yang, K. Zhou, X. Zhao, Q. Zhou, H. Zhu, Y. Yang, C. Zhang, Y. Zhou, and W. Zhou, “Rapid processing of a global feature in the on visual pathways of behaving monkeys,” *Front. neurosci.*, vol. 11, pp. 474, 2017. [80](#)
- [136] D. Hubel, *Eye, brain, and vision.*, Scientific American Library/Scientific American Books, 1995. [80](#)
- [137] M. Ahissar and S.I Hochstein, “The reverse hierarchy theory of visual perceptual learning,” *Trends Cogn. Sci.*, vol. 8, no. 10, pp. 457–464, 2004. [80](#)
- [138] D. Zoccolan, “Invariant visual object recognition and shape processing in rats,” *Behav. Brain Res.*, vol. 285, pp. 10–33, 2015. [80](#)
- [139] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014. [84](#)

- [140] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, “Large kernel matters—improve semantic segmentation by global convolutional network,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4353–4361. [85](#)
- [141] ISIC 2016, “Skin lesion analysis towards melanoma detection,” <https://challenge.kitware.com/phase/566744dccad3a56fac786787>. [86](#), [93](#), [94](#), [95](#)
- [142] ISIC 2017, “Skin lesion analysis towards melanoma detection,” <https://challenge.kitware.com/phase/584b0afacad3a51cc66c8e24>. [86](#), [93](#), [94](#)
- [143] M. Lin, Q. Chen, and S. Yan, “Network in network,” *arXiv preprint arXiv:1312.4400*, 2013. [98](#)
- [144] F. Riaz, S. Naeem, R. Nawaz, and M. Coimbra, “Active contours based segmentation and lesion periphery analysis for characterization of skin lesions in dermoscopy images,” *IEEE J. Biomed. Health*, vol. 23, no. 2, pp. 489–500, 2018. [99](#)