Extending Convolutional Neural Network based Models for Data Augmentation and Segmentation of Fundus Images

Thesis submitted by

Supratim Ghosh

Doctor of Philosophy(Engineering)

Department of Computer Science and Engineering
Faculty Council of Engineering & Technology
Jadavpur University
Kolkata-700032, India
2025

JADAVPUR UNIVERSITY
KOLKATA 700032, INDIA

Index No: 40/18/E

1. Title of the Thesis:

   Extending Convolutional Neural Network based Models for Data Augmentation and Segmentation of Fundus Images

2. Name, Designation & Institute of the supervisor:

   Dr. Mahantapas Kundu
   Professor,
   Department of Computer Science and Engineering,
   Jadavpur University, Kolkata-700032

3. List of Publications:

   a) **JOURNAL**

      i) **Supratim Ghosh**, Sourav Pramanik, Anoop K. Tiwari, Kottakkaran S. Nisar, Mahantapas Kundu and Mita Nasipuri, "2pCePd-Net: Two-Path Cross-Context Encoder With Probability Map-Based Bandpass Decoder for Retinal Vessel Segmentation," in **IEEE Transactions on Instrumentation and Measurement, vol. 74, pp. 1-14, 2025, Art no. 5031314**, doi: 10.1109/TIM.2025.3569005.

      ii) **Supratim Ghosh**, Sourav Pramanik, Anoop K. Tiwari, Kottakkaran S. Nisar, Mahantapas Kundu and Mita Nasipuri, "A Structural Asymmetric Sampling-based Probability-Map Choquet-Fusion Prediction Network (PrSASfNet) for Retinal Blood Vessel Segmentation," in **Springer SN Computer Science** (Communicated).

      iii) **Supratim Ghosh**, Sourav Pramanik, Mahantapas Kundu and Mita Nasipuri, "Iterative Pixel-Performance Loss Projection (αN^{++}) for Improved Fundus Image Segmentation," in **Pattern Recognition Letters** (Communicated).

   b) **CONFERENCE**

i) **Supratim Ghosh**, Mahantapas Kundu and Mita Nasipuri, ”Retinal Blood Vessel Segmentation using a Multi-Scale Layer in Deep Learning,” **2021 IEEE 18th India Council International Conference (INDICON)**, Guwahati, India, 2021, pp. 1-6, doi: 10.1109/INDICON52576.2021.9691545.

ii) **Supratim Ghosh**, Mahantapas Kundu, Mita Nasipuri, “Retinal Vessel Segmentation in Fundus Image Using Low-Cost Multiple U-Net Architecture,” In: Gupta, M., Ghatak, S., Gupta, A., Mukherjee, A.L. (eds) **Artificial Intelligence on Medical Data. Lecture Notes in Computational Vision and Biomechanics, vol 37. Springer, Singapore**. https://doi.org/10.1007/978-981-19-0151-5_12

iii) **Supratim Ghosh**, Mahantapas Kundu, Mita Nasipuri, “Region Separated Vessel Segmentation in Fundus Image Using Multi-scale Layer-Based Convolutional Neural Network,” In: Mandal, J.K., De, D. (eds) **Frontiers of ICT in Healthcare . Lecture Notes in Networks and Systems, vol 519. Springer, Singapore**. https://doi.org/10.1007/978-981-19-5191-6_56

iv) **Supratim Ghosh**, Sourav Pramanik, Mahantapas Kundu, Mita Nasipuri, “Vessel Curvature-Based Data Augmentation Technique for Retinal Fundus Images,” In: Kole, D.K., Roy Chowdhury, S., Basu, S., Plewczynski, D., Bhattacharjee, D. (eds) **Proceedings of 4th International Conference on Frontiers in Computing and Systems. COMSYS 2023. Lecture Notes in Networks and Systems, vol 975. Springer, Singapore**. https://doi.org/10.1007/978-981-97-2614-1_41

4. List of Patents: None

5. List of Presentations in National/International/Conference/Workshops:

i) **Supratim Ghosh**, Mahantapas Kundu and Mita Nasipuri, ”Retinal Blood Vessel Segmentation using a Multi-Scale Layer in Deep Learning,” **2021 IEEE 18th India Council International Conference (INDICON)**, Guwahati, India, 2021, pp. 1-6, doi: 10.1109/INDICON52576.2021.9691545.

ii) **Supratim Ghosh**, Mahantapas Kundu, Mita Nasipuri, “Retinal Vessel Segmentation in Fundus Image Using Low-Cost Multiple U-Net Architecture,” In: Gupta, M., Ghatak, S., Gupta, A., Mukherjee, A.L. (eds) **Artificial Intelligence on Medical Data. Lecture Notes in Com-**

putational Vision and Biomechanics, vol 37. Springer, Singapore. https://doi.org/10.1007/978-981-19-0151-5_12

iii) **Supratim Ghosh**, Mahantapas Kundu, Mita Nasipuri, "Region Separated Vessel Segmentation in Fundus Image Using Multi-scale Layer-Based Convolutional Neural Network," In: Mandal, J.K., De, D. (eds) **Frontiers of ICT in Healthcare . Lecture Notes in Networks and Systems, vol 519. Springer, Singapore**. https://doi.org/10.1007/978-981-19-5191-6_56

iv) **Supratim Ghosh**, Sourav Pramanik, Mahantapas Kundu, Mita Nasipuri, "Vessel Curvature-Based Data Augmentation Technique for Retinal Fundus Images," In: Kole, D.K., Roy Chowdhury, S., Basu, S., Plewczynski, D., Bhattacharjee, D. (eds) **Proceedings of 4th International Conference on Frontiers in Computing and Systems. COMSYS 2023. Lecture Notes in Networks and Systems, vol 975. Springer, Singapore**. https://doi.org/10.1007/978-981-97-2614-1_41

“Statement of Originality”

I, **Supratim Ghosh**, registered on **18/06/2018** do hereby declare that this thesis entitled **“Extending Convolutional Neural Network based Models for Data Augmentation and Segmentation of Fundus Images”** contains original research work done by the undersigned candidate as part of Doctoral studies.

All information in this thesis have been obtained and presented in accordance with existing academic rules and ethical conduct. I declare that, as required by these rules and conduct, I have fully cited and referred all materials and results that are not original to this work.

I also declare that I have checked this thesis as per the “Policy on Anti Plagiarism, Jadavpur University, 2019”, and the level of similarity as checked by iThenticate software is 4%.

Signature of Candidate: Supratim Ghosh

Date: 09-June 2025

M. Kundu 09 June 2025

Certified by Supervisor
(Signature with date, seal)

Professor
Computer Sc & Engg Department
Jadavpur University
Kolkata-700032

CERTIFICATE FROM THE SUPERVISOR

This is certify that the thesis entitled ***"Extending Convolutional Neural Network based Models for Data Augmentation and Segmentation of Fundus Images"*** *submitted by* ***Supratim Ghosh****, who got his name registered on* ***18/06/2018****, for the award of Ph.D. (Engg.) degree of Jadavpur University is absolutely based upon his own work under the supervision of* ***Dr. Mahantapas Kundu****, Professor, Department of Computer Science and Engineering, Jadavpur University, Kolkata and that neither his thesis nor any part of it has been submitted for any degree/diploma or any other academic award anywhere before.*

M. Kundu 09 June 2025

Dr. Mahantapas Kundu
Professor
Department of CSE
Jadavpur University, Kol-32

Professor
Computer Sc & Engg Department
Jadavpur University
Kolkata-700032

Dedicated to my family

ACKNOWLEDGMENTS

A Doctorate degree. A journey to earn a distinct honor to add the ”Dr.” in front of a name. It is said that the last name is the one you earn by virtue of birthright and the first name is the one given lovingly by your parents. But the title of ”Dr.” in front is the one which speaks of the journey you traveled and the only part of your name that you *earned*. And the doctorate degree would not have been possible to achieve without the support and guidance that I received from many people.

Firstly, it is my distinct honor and privilege to show gratitude and respect to my supervisor **Dr. Mahantapas Kundu**, Professor, Department of Computer Science & Engineering, Jadavpur University for his continuous support during my PhD study and related research, his patience, motivation, and knowledge. Without his guidance and constant feedback this PhD research work would not have been successful.

Secondly, I would like to thank **Dr. Mita Nasipuri**, Retired Professor, Department of Computer Science & Engineering, Jadavpur University, for her constant inspiration. Without her persistent help and optimism this dissertation would not have been possible.

Next, I would like to thank **Dr. Sourav Pramanik**, Department of Computer Science, New Alipore College, for his valuable input, suggestions and constant feedback which have made an invaluable contribution towards my PhD work.

I express my sincere thanks to **Dr. Ram Sarkar**, Professor, Department of Computer Science & Engineering, Jadavpur University, for the valuable suggestions that shaped the final dissertation. I would also like to thank **Dr. Nandini Mukherjee**, Professor, Department of Computer Science & Engineering, Jadavpur University, and **Dr. Nirmalya Chowdhury**, Professor, Department of Computer Science & Engineering, Jadavpur University, for their continuous support in their respective tenures as HOD, Department of Computer Science and Engineering, Jadavpur University.

I would like to extend my gratitude to all of my extended family at **CMATER** lab, Jadavpur University for their continuous support and encouragement. A special thank you to **Dr. Kaushiki Roy**, **Dr. Nirmal Das** and **Dr. Kalpita Dutta** for their continuous support and encouragement in the doctoral journey.

I am extending my thanks to all my co-authors for their hard work and substantial contribution to uplift the studies presented in this thesis. Thank you to **Dr. Anoop Tiwari**, Central University of Haryana and **Prof. (Dr.) Kottakkaran Sooppy Nisar**, Prince Sattam bin Abdulaziz University. It has been a true honor to have had the opportunity to collaborate with you and I hope to continue this collaboration in future as well.

No amount of achievement in my life could ever be complete without the blessings of my parents, **Mrs. Piyali Ghosh** and **Mr. Samir Kumar Ghosh**. I would also like to thank **Sourav Das** and **Bhuvnesh Chaturvedi** for their constant support in this journey.

Finally, and most importantly, I would like to thank the ***One*** without whose will not a blade of grass can move.

Supratim Ghosh

Supratim Ghosh

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 An Overview

Ocular diseases, such as diabetic retinopathy, glaucoma, age-related macular degeneration (AMD), and cataracts, are among the leading causes of visual impairment and blindness across the globe [1]. The challenges of these conditions are increasing primarily due to the aging population and the increased incidence of diabetes, hypertension, and other chronic ailments. In this regard. traditional diagnostic practices [2], although effective, heavily rely on the availability of skilled ophthalmologists and subjective interpretation of ocular images, which introduces challenges related to scalability, accessibility, and diagnostic variability. In regards to the limitations, the development and deployment of automated ocular disease detection systems have emerged as a crucial advancement in contemporary ophthalmology [3].

Automated detection systems employ advanced algorithms, particularly based on machine learning and deep learning, to interpret ocular images and recognize pathological traits with unprecedented accuracy [4]. Furthermore, the detection systems offer efficient analysis that is important in high-throughput screening situations [5]. Additionally, the development of the technologies present the potential to provide uniform access to eye care by addressing geographical and resource-based inequalities. The following overview explores the holistic importance of automated eye disease detection systems, including the technological, clinical, economic, and societal aspects.

- **Addressing the Global Burden of Ocular Diseases :** According to the World Health Organization (WHO), nearly 2.2 billion people worldwide live with some degree of vision impairment, among which at least one billion have preventable or uncorrected vision impairment as a result of insufficient access to care [6]. In low and middle-income nations, inadequate quantity of trained eye care personnel further accelerate the issue, especially in rural and remote areas. Instead, computerized ocular disease diagnosis systems offer a budget solution to the emergency. By analyzing for the early detection of ocular pathologies through AI-enlisted diagnostics, these systems can provide pathways for timely referral and treatment.

Additionally, in screening programs, conducted on a community basis, automation can be implemented in primary healthcare facilities such that non-specialists can take ocular photos to be interpreted by intelligent systems [7]. Such a model increases the coverage of eye care services, particularly in remote areas, and has the potential to significantly decrease the prevalence of avoidable blindness. Additionally, early detection through automation helps to improve long-term results by avoiding the advancement of disease towards advanced stages.

- **Enhancing Diagnostic Accuracy and Objectivity :** One of the primary advantages of automated ocular diagnostic systems is the ability to improve diagnostic accuracy and reduce subjectivity [7]. Ocular image interpretation by humans is prone to variability, as it is affected by the subjective experience of the evaluator, mental state and analytical judgment of the clinician. Automated systems, having learned on large annotated datasets, perform according to stable algorithms with reduced intra as well as inter-observer variability [8].

  In this regard, for example, deep convolutional neural networks (CNNs) have demonstrated a considerably superior performance in identifying tiny retinal structures such as microaneurysms, hemorrhages, and exudates [9]— all of which are diagnostic hallmarks of diabetic retinopathy. Similarly, algorithms can precisely calculate cup-to-disc ratios in optic nerve head analysis to assess and diagnose glaucoma. Such systems create uniformity in image interpretation, thereby facilitating standard screening protocols as well as the overall quality of care [10].

- **Reducing Clinical Workload and Improving Efficiency :** Demographic changes and epidemiological upward demand trends for ophthalmic services are increasing the workload of clinicians [11]. In order to limit the burden on clinicians, automated detection systems can be extremely helpful. These systems can automatically categorize images as normal or pathological, acting as the initial screening level, thus alleviating the workload on clinicians to concentrate on cases that need more investigation.

  As a result, patient throughput is increased, waiting times are decreased, and clinical workflows are greatly optimized by this triaging capability [12]. Furthermore, automated systems can work continuously without getting tired, which increases the ability of eye care services to handle extensive screenings like those carried out in national eye care missions, workplace health checks, and school health programs.

- **Facilitating Teleophthalmology and Remote Diagnosis :** Tele-medicine has emerged as a competitive alternative to in-person consultations, especially during public health emergencies such as the COVID-19 pandemic [13]. The success of teleophthalmology initiatives depends on automated systems for detecting ocular diseases. According to this model, ocular images taken in mobile clinics or outlying health centers are electronically sent to centralized facilities for automated analysis. This makes remote diagnosis possible by allowing patients with conditions that could endanger their vision to be referred to tertiary care facilities in a timely manner. The method lessens the need for patient travel and improves access to specialized care for populations in remote locations. Additionally, by making routine monitoring and follow-up easier, teleophthalmology and automation promote continuity of care.

- **Enabling Personalized Medicine :** The shift from generalized to precision medicine can be aided by automated ocular analysis systems. AI models can identify individual disease trajectories by longitudinally monitoring retinal changes and comparing them with lifestyle, genetic, and systemic health factors. This makes it possible to create individualized treatment programs that consider each patient's particular profile.

  For instance, even with glycemic control, some patients with diabetic retinopathy show rapid disease progression. Automated systems can direct more intensive monitoring or intervention strategies and assist in the early identification of such high-risk individuals. Results and resource use are optimized at this level of personalization.

- **Cost-Effectiveness and Resource Optimization :** Healthcare systems have adopted approaches for significant monetary benefits by putting automated detection systems into practice [14]. The long-term advantages of lower diagnostic errors, early detection, and more efficient workflows outweigh the initial costs of deploying AI infrastructure, which include training and technology investments. Automated systems minimize needless referrals and lessen reliance on limited specialist consultations.

  Furthermore, these systems are particularly useful for mass screening programs in public health settings because they can process large volumes of images with little recurring cost. They contribute significantly to preventing the high expenses of advanced disease management, surgery, and vision rehabilitation by identifying conditions early [14].

However, automated ocular disease detection systems present a plethora of obstacles in spite of their potential. These include addressing algorithmic biases, securing regulatory approvals, protecting data privacy, and making model decisions interpretable. Furthermore, in clinical settings, the "black-box" nature of certain deep learning models raises questions regarding trust and accountability.

The present state of research literature is conducted to create explainable AI (XAI) systems that provide clear justification for predictions. Furthermore, for large-scale adoption to be sustained, imaging protocols need to be standardized, validated across a range of populations, and developed collaboratively by clinicians, engineers, and policymakers. It is anticipated that these systems will develop into essential instruments for comprehensive, just, and intelligent eye care. In the following section, a historical overview of the evolution of ocular diagnostic modalities is presented.

1.2 Imaging Modalities for Ocular Disease Detection

The imaging modality used to gather comprehensive data about the anatomy and health of the eye is an essential part of automated ocular disease detection systems [15]. The sensitivity, specificity, and diagnostic performance of automated algorithms are influenced by the imaging modality selection. Numerous imaging methods have been developed over time, each with special benefits and to meet varying diagnostic needs. Among the ophthalmic imaging modality includes, but is not limited to, confocal microscopy, fluorescein angiography, scanning laser ophthalmoscopy (SLO), color fundus photography, optical coherence tomography (OCT), and ultrasound imaging. An extensive review of these modalities is given in the subsequent sub-sections, with particular attention paid to their technical features, clinical uses, and applicability in the context of automated disease detection.

1.2.1 Color Fundus Photography

One of the most popular imaging methods for obtaining close-up imaging of the retina is color fundus photography. A full-color image of the fundus, including the optic disc, macula, blood vessels, and peripheral retina, can be obtained using the non-invasive technique. Fundus photography is perfect for extensive screening programs because it is very accessible, affordable, and simple to use. To find indications of Age-related Macular Degeneration (AMD), hypertensive retinopathy, diabetic retinopathy, and other retinal conditions, the generated images can be examined manually or by AI-based systems. A high-intensity light source and a low-powered microscope coupled to a camera form the basis for a standard fundus camera. The fundamental optical system consists of:

Objective lens: Focuses light onto the retina.

Illumination system: Typically a xenon flash or LED source, co-axially aligned to reduce corneal reflections.

Beam-splitting mirrors: Direct the illumination and imaging paths.

Imaging sensor: High-resolution CCD or CMOS sensor capable of capturing color images.

The primary advancement in contemporary fundus cameras is the incorporation of non-mydriatic capabilities, which enable image acquisition without the need for pharmaceutical pupil dilation. AI-based diagnostic tools and telemedicine software are also integrated with digital fundus cameras. Many deep learning models have been trained to detect microaneurysms, hemorrhages, exudates, and neovascularization and rely on fundus images as their main source of data [16]. Additionally, because of the superior contrast, the green channel of fundus images is especially instructive for vessel segmentation. The position of Fundus photography as a mainstay in ocular diagnostics is further solidified by its widespread availability and compatibility with machine learning models. Figure 1.1 represents a set of standard *fundus* images captured using fundus camera [17].

Figure 1.1: An illustration of a set of Fundus images.

1.2.2 Optical Coherence Tomography (OCT)

Optical Coherence Tomography (OCT), a non-invasive imaging technique, creates high-resolution cross-sectional retinal images using low-coherence interferometry. It provides fine detail of structures like the macula, retinal nerve fiber layer, and photoreceptor integrity, enabling accurate visualization of the retina's internal microarchitecture. It uses near-infrared light (usually around 840 nm for retinal imaging) in low-coherence interferometry. Among the crucial elements are:

Broadband light source: Superluminescent diodes or swept-source lasers.

Michelson interferometer: Splits the light into a reference arm and a sample arm.

Scanning system: Galvanometric mirrors control lateral scanning.

Detector and spectrometer: Capture the interference signal and reconstruct cross-sectional images using Fourier-domain processing.

Two main types of OCT systems exist: Time-Domain OCT (TD-OCT) and Spectral-Domain OCT (SD-OCT), the latter offering higher resolution and faster acquisition rates. Swept-source OCT (SS-OCT) further improves imaging depth and penetration. Clinically, OCT is essential for the early detection and continued treatment of many retinal and optic nerve conditions, such as glaucoma, central serous chorioretinopathy, diabetic macular edema, and age-related macular degeneration (AMD). OCT facilitates the quantitative and objective evaluation of disease progression, which is particularly important for treatment planning and therapeutic efficacy assessment. OCT has promoted the creation of sophisticated AI systems that can perform layer-wise segmentation, automated classification, and change detection from the viewpoint of clinical informatics. OCT devices are costly and frequently need specific training for operation and interpretation, despite their diagnostic advantages. These obstacles prevent them from being integrated into primary care or rural clinics, which bear the burden of the undiagnosed ocular disease. Figure 1.2 represents a set of standard *oct* images captured using interferometry [18].

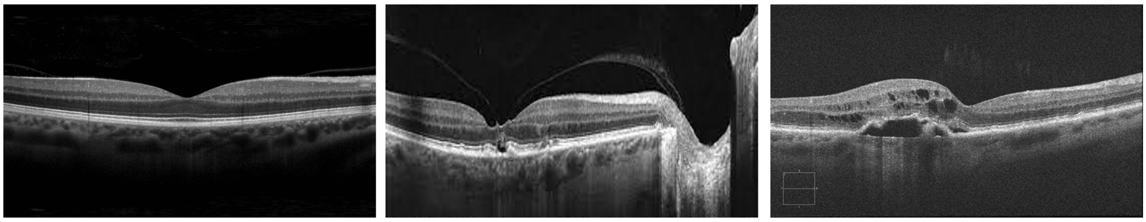


Figure 1.2: An illustration of a set of Optical Coherence Tomography images.

1.2.3 Fluorescein Angiography (FA)

In order to assess retinal and choroidal blood flow, *Fluorescein Angiography* entails injecting sodium fluorescein dye intravenously and then taking serial photos of the retina with a specialized camera. For the detection and tracking of retinal vascular disorders, including diabetic retinopathy, retinal vein occlusions, and choroidal neovascularization in AMD, FA is still the clinical gold standard. A fundus camera with extra optical filters frequently incorporates the camera setup as follows:

Excitation filter: Selects the wavelength (around 490 nm) for fluorescein excitation.

Barrier filter: Allows emitted light (520–530 nm) while blocking reflected excitation light.

High-sensitivity detector: Captures sequential images as the dye circulates.

Timely acquisition of images that is coordinated with the injection protocol is essential. To track the dye's progression, the system needs to be able to process data in real time and at high frame rates. FA is used in clinical settings to evaluate vascular leakage, neovascular complexes, microaneurysms, and capillary non-perfusion. It is extremely useful in complicated diagnostic situations due to its high specificity and dynamic imaging capability. FA is intrusive, though, and may cause adverse effects that range from mild nausea to severe allergic reactions. Its use is limited to tertiary centers due to the need for intravenous access, dye preparation, and skilled interpretation in real time. Due to a lack of extensive, well-annotated datasets and variations in image quality, automated systems for FA interpretation are still in their infancy. Figure 1.3 represents a set of standard *Fluorescein Angiography* images [19].

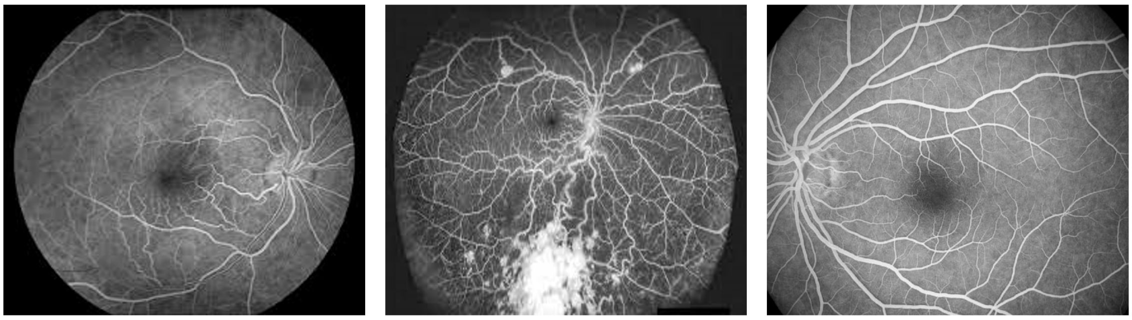


Figure 1.3: An illustration of a set of Fluorescein Angiography images.

1.2.4 Scanning Laser Ophthalmoscopy (SLO)

Confocal laser scanning techniques are used in Scanning Laser Ophthalmoscopy to produce high-contrast retinal images. Through autofluorescence imaging, it is especially useful for observing optic nerve structures and detecting pathologic alterations in the retinal pigment epithelium. The SLO employs a laser beam to scan the retina point-by-point, which is then reconstructed into an image. Key components include:

Laser source: Multiple lasers for different imaging modes (e.g., autofluorescence, infrared reflectance).

Beam deflectors: Galvo mirrors scan the beam horizontally and vertically.

Confocal pinhole: Ensures that only in-focus light reaches the detector, enhancing contrast.

Photodetectors: Avalanche photodiodes or PMTs collect the reflected signal.

The precision optics and alignment are critical in SLO. Integration with other modalities like OCT enables multimodal platforms. To enable multimodal imaging in a single device, SLO-based systems frequently integrate with other modalities, such as OCT. Clinically, SLO is used to assess lipofuscin accumulation in hereditary retinal diseases, optic nerve head cupping in glaucoma, and retinal thinning in degenerative disorders. Early detection of choroidal neovascularization and geographic atrophy has shown promise in regards to AI models trained on SLO data. However, SLO is not widely accessible outside of advanced clinical centers due to its high cost and maintenance needs. Future accessibility may be expanded by additional advancements in portable and affordable SLO systems. Figure 1.4 represents a set of standard *Scanning Laser Ophthalmoscopy* images [20].

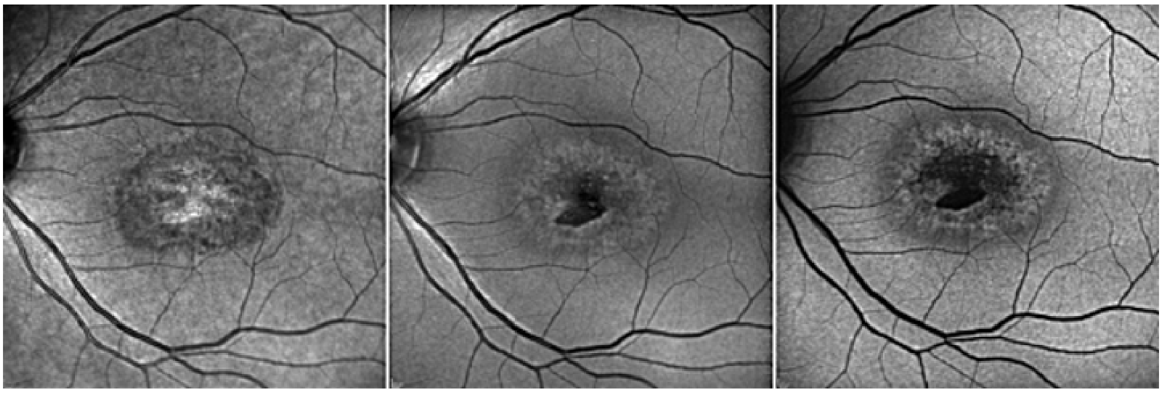


Figure 1.4: An illustration of a set of Scanning Laser Ophthalmoscopy images.

1.2.5 Confocal Microscopy and Adaptive Optics

Retinal tissues can be perceived at cellular resolution using confocal microscopy and adaptive optics. Confocal systems enhance axial resolution by rejecting out-of-focus light. In retinal imaging, adaptive optics (AO) systems further correct optical aberrations in real time.

Wavefront sensor: Hartmann–Shack sensor detects optical aberrations.

Deformable mirror: Adjusts the optical path to compensate for eye-induced distortions.

Confocal scanning system: Combines with AO to achieve cellular-level resolution.

These systems are complex and require active feedback loops, making them more suited to research and advanced clinical centers. These modalities are mainly employed in specialized diagnostics and research to examine the density and morphology of microglia, photoreceptors and retinal pigment epithelium. Adaptive optics corrects optical aberrations to improve image resolution, while confocal techniques improve contrast by rejecting out-of-focus light.

The contemporary technology has made it easier to identify photoreceptor loss in retinitis pigmentosa and cone dysfunction in color vision disorders early in clinical research. In order to interpret ultra-high-resolution imagery, automated image analysis in this field is still in its infancy and requires complex algorithms. Their use is restricted to specialized centers due to the infrastructure and technical know-how required. Figure 1.5 represents a set of standard *Confocal Microscopy* images [21].

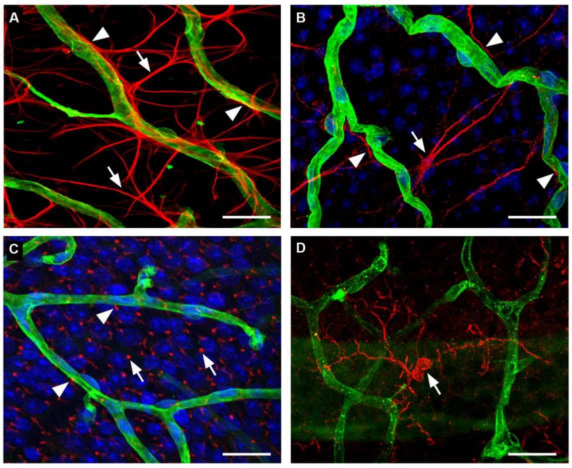


Figure 1.5: An illustration of the set of Confocal Microscopy images denoting retinal blood vessels.

1.2.6 Ultrasound B-Scan Imaging

In the event of the ocular media turning opaque as a result of trauma, dense cataracts, or vitreous hemorrhage, ultrasound B-scan is a useful diagnostic technique. It makes the posterior segment visible for the purpose of identifying foreign bodies, intraocular tumors, and retinal detachment. Ultrasound is essential in emergency and low-vision situations because it does not rely on clear visual pathways unlike optical methods. Ocular ultrasound uses high-frequency probes (10–20 MHz) to image the posterior segment when direct visualization is obstructed. Key components include:

Piezoelectric transducer: Emits and receives ultrasound waves.

Display console: Digitally processes echoes to generate B-mode images.

Manual positioning arm: Allows dynamic manipulation to target specific regions.

The major challenges include speckle noise and operator dependence. Nevertheless, it remains critical in trauma and opaque media diagnosis. The low resolution and high speckle noise of B-scan images make automated interpretation challenging. However, to categorize diseases and segment anatomical features, machine learning methods such as convolutional neural networks have been used. Ultrasound imaging may increase diagnostic capacity when integrated into automated detection systems, especially in areas without access to cutting-edge optical equipment. Figure 1.6 represents a set of standard *Ultrasound B-Scan* images [22].

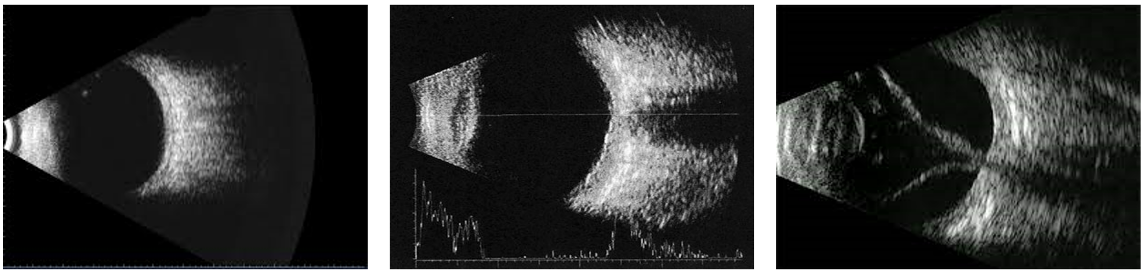


Figure 1.6: An illustration of a set of Ultrasound B-Scan images.

1.2.7 Comparative Evaluation of Imaging Modalities

Every imaging modality presents a unique set of clinical advantages and disadvantages. OCT lacks color detail but is excellent at resolving intra-retinal architecture. FA is invasive and resource-intensive, but it offers unparalleled vascular mapping. SLO provides more contrast, but it comes at a price. Although it lacks structural detail, ultrasound aids in the evaluation of opaque media. High resolution imaging is possible with confocal and adaptive optics, but they are not practical for everyday use.

In contrast, color fundus photography provides the best possible balance of cost-effectiveness, usability, non-invasiveness, and AI system compatibility. It is the preferred modality for screening and teleophthalmology applications due to its capacity to capture an entire retinal overview in a single image. Thus, fundus imaging plays a crucial part in the creation and implementation of automated systems for detecting ocular diseases.

1.3 Importance of Fundus Imaging in Automated Detection of Ocular Diseases

In both clinical and non-clinical contexts, such as primary care units, optometry clinics, and mobile screening units, fundus imaging is extensively accessible. Its usefulness in outreach initiatives and settings with limited resources is further increased by portable fundus cameras. High-throughput screening is made possible by fundus imaging's non-invasiveness and fast acquisition time, which also helps in patient compliance. Some useful characteristics of fundus images are listed below:

- **Rich Pathological Bio-Markers [23]:** Fundus images are perfect for training AI models because of their rich pathological markers, organized layout and consistent resolution. Fundus imaging broadens the diagnostic scope of AI-based models by allowing the detection of several conditions from a single image, such as glaucoma, AMD, diabetic retinopathy, and hypertensive retinopathy. A sample fundus image with marked bio-markers is shown in Figure 1.7.

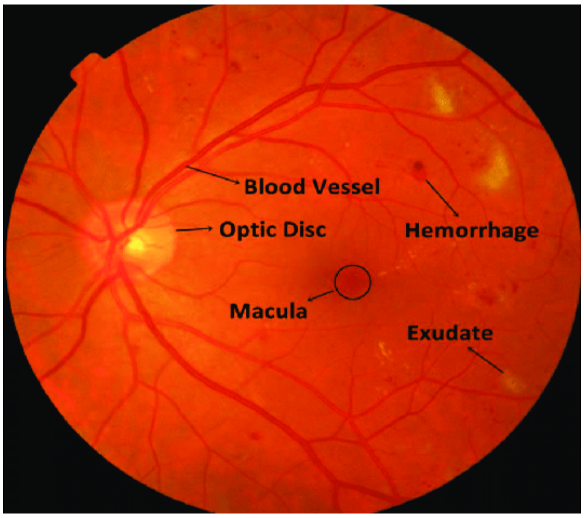


Figure 1.7: A sample fundus image with marked bio-markers [24].

- **Versatility Across Disease Conditions [16]:** Fundus images can be used to diagnose a variety of eye conditions because they can show the optic nerve head, macula, retinal vessels, and peripheral retina. A thorough evaluation of ocular health can be obtained by using automated systems trained on fundus images to identify microaneurysms, hemorrhages, vessel tortuosity, drusen, and disc cupping.

- **Integration with Teleophthalmology [25]:** The digital format of Fundus imaging, ease of transmission, and low hardware requirements make it a good fit for tele-ophthalmology. It serves as the foundation for remote screening programs, which use cloud-based AI platforms to analyze images collected in peripheral centers. In under-served and remote areas, this integration improves access to care.

- **Cost-Effectiveness and Scalability [26]:** Fundus imaging is substantially less expensive than OCT, SLO, and FA in terms of both equipment and operating costs. Because of this, it is very scalable for public health campaigns that try to lower avoidable blindness. Adoption barriers are further decreased by the broad availability of fundus cameras that run on smartphones.

In conclusion, because of the accessibility, adaptability, and compatibility with AI techniques, fundus imaging has become the most appropriate modality for automated ocular disease detection. Other modalities have special diagnostic benefits, but their use is frequently constrained by their expense, intricacy and invasiveness. In the following section, a brief discussion on the ethical and legal aspects associated with fundus imaging is studied.

1.4 Ethical and Legal Considerations in Automated Fundus Image Analysis

Important moral and legal questions do arise by the creation and application of automated systems in fundus image analysis. Stakeholders must consider the effects on patient rights, data protection, bias, accountability, and equitable access as artificial intelligence (AI) is progressively incorporated into clinical workflows for the detection of ocular diseases. The complex legal and ethical frameworks governing AI-based fundus analysis are covered in this section, particularly as they relate to low-resource environments such as rural India.

- **Patient Privacy and Data Protection [27]:** Large retinal image datasets, frequently containing private medical data, are the foundation of automated fundus imaging systems. Strict adherence to privacy and data protection laws is thus necessary for the ethical use of such data. The General Data Protection Regulation (GDPR) establishes strict guidelines for gathering, storing, and processing of personal health data in countries such as the European Union. In a similar vein, India's proposed Digital Personal Data Protection Act (DPDP Act), which prioritizes informed consent, purpose limitation, and data minimization, is anticipated to conform to international standards.

  Data anonymization and secure encryption are essential when it comes to fundus imaging. Data needs to be stored in safe, access-controlled environments, and researchers and developers need to ensure that identifying markers are eliminated. Furthermore, to control the use of images for clinical trials, algorithm training, or commercial purposes, clear data-sharing guidelines and data governance frameworks must be in effect.

- **Informed Consent and Autonomy:** A fundamental component of ethical medical practice is informed consent. Patients need to know exactly how their photos will be used, whether it will be analyzed automatically, and who will be able to view the results. In automated systems, where human oversight may be minimal, this is especially crucial.

- **Accountability and Liability [28]:** Determining responsibility for incorrect decisions becomes difficult when automated systems are used for screening or diagnosis. Depending on the deployment context, the developer, healthcare provider, or regulatory body may be held liable if diseased fundus images is mistakenly classified as normal and treatment is missed.

  In these situations, legal frameworks must specify the person responsible. This entails defining the function of human oversight, formalizing redress procedures, and setting clinical validation standards. The fact that AI tools are assistive rather than autonomous and that the clinician has the final say over diagnosis must be emphasized in professional guidelines.

- **Transparency and Explainability:** Transparency in the system is directly related to trust in AI-driven diagnostics. If the reasoning behind the system's output can be understood, both patients and clinicians are more likely to accept automated fundus analysis. Adoption in high risk environment such as healthcare is hampered by black-box models, which are prevalent in deep learning.

  Explainable AI (XAI) methods that visually represent the areas of the retina influencing a given decision, like saliency maps, class activation maps (CAMs), or attention mechanisms, must be used. These characteristics help with clinical validation and quality assurance in addition to enhancing trust.

- **Regulatory Compliance and Standards:** Globally, laws governing AI-powered medical devices are still being developed. Software as a medical device (SaMD) standards are being outlined by the Central Drugs Standard Control Organization (CDSCO) in India. Internationally, clinical evaluation, risk classification, and post-market surveillance are necessary for the European CE marking process and the U.S. FDA.

  Obtaining regulatory approval for automated fundus analysis tools requires proving their efficacy, safety, and conformity to clinical guidelines. Additionally, regulatory agencies might request proof of cyber-security, robustness, and bias reduction. To comply with these requirements, developers need to keep up-to-date documentation, version control, and audit trails.

- **Intellectual Property and Commercialization Ethics:** There are ethical concerns with the commercialization of AI systems that were trained using patient-contributed images or publicly funded datasets regarding the rightful ownership of the algorithms.

Transparency regarding funding sources, benefit-sharing arrangements, and intellectual property rights is required by ethical commercialization. Public-private partnerships and open-source frameworks can assist in striking a balance between innovation, accessibility, and equity. Furthermore, it is ideal for the money made from AI products to be put back into public health research and infrastructure.

Abiding by the ethical and legal guidelines listed, a few gold standard and medically relevant annotated fundus image datasets have been designed in past literature. In the following section, a subset of different annotated retinal fundus images is presented for a detailed study on respective uses cases.

1.5 Datasets

Fundus image segmentation is a fundamental aspect of automatic retinal image analysis, which is the key to early diagnosis and monitoring of eye diseases like diabetic retinopathy, glaucoma, and hypertensive retinopathy. Segmentation facilitates the structural outline of significant anatomical features such as blood vessels, optic disc, and pathological findings such as exudates and microaneurysms. To develop and test segmentation algorithms, particularly deep learning-based ones, high-quality annotated datasets are essential. In this regard, an exhaustive description of some of the key publicly available fundus image segmentation datasets is provided, including DRIVE, STARE, CHASEDB1, HRF, and IDRiD among the most used in scientific and clinical studies.

These datasets not only differ in the number of images and annotations types offered but also in image resolutions, disease profiles, and imaging devices. Each dataset possesses distinct features that qualify it for different segmentation tasks such as blood vessel segmentation, optic disc segmentation, or lesion detection. A study of these datasets is a key aspect to choose the appropriate benchmark for training and testing deep learning models.

1.5.1 DRIVE: Digital Retinal Images for Vessel Extraction [29]

Prior to the advent of large-scale medical imaging databases, the DRIVE dataset was one of the first and most prominent datasets used for the segmentation of blood vessels in retinal images. It was published alongside a retinal vessel segmentation competition and has since been the de-facto standard for testing segmentation models. Figure 1.8 illustrates a set of DRIVE dataset images.

- **Dataset Overview:**

  The DRIVE dataset contains 40 color fundus images of size 768×584 pixels. The images were captured with a Canon CR5 non-mydriatic 3CCD camera with a 45-degree field of view (FOV). The data was obtained from a diabetic retinopathy screening program in the Netherlands and covers normal and pathological cases.

- **Annotations:**

  Manual segmentation of the vasculature were created by two observers: one was available for training/testing and the other was available for comparative studies. The dataset is segregated into 20 training and 20 test images. Each image includes a binary mask delimiting the vessel structures, as well as the Field of View (FOV) mask for the purpose of evaluation.

- **Significance:**

  DRIVE has become the de-facto standard dataset for designing and evaluating retinal vessel segmentation algorithms. The compact size promotes efficient model development and rapid experimentation. But the limited variability makes supplementary use with other datasets necessary for generalized performance evaluation.

Figure 1.8: Some sample images from of the DRIVE dataset.

1.5.2 STARE: Structured Analysis of the Retina [30]

With the increase in retinal image analysis, a requirement for databases that encompassed a larger variety of pathologies was identified. In this regard, STARE dataset came to address this requirement by offering an image dataset with a diverse range of retinal diseases, thus being particularly useful for model training in pathological environments. Figure 1.9 illustrates a set of STARE dataset images.

- **Dataset Overview:**

  The STARE dataset consists of 20 fundus images obtained at a resolution of 700 × 605 pixels. The images were recorded through a TopCon TRV-50 fundus camera with a 35-degree FOV. In contrast to DRIVE, STARE has a greater number of images with pathological conditions, such as lesions, hemorrhages, and irregular vessels.

- **Annotations:**

  The images had been annotated manually by two skilled observers. Annotations consisted of binary vessel masks and are used extensively for comparative assessment. A subset of the dataset also contains optic nerve annotations.

- **Significance:**

  STARE is usually combined with DRIVE to further improve model training and validation through increased pathological diversity. The dataset plays a critical role in evaluating the robustness of segmentation models in more challenging clinical cases, particularly for disease-specific vessel pathology abnormalities.

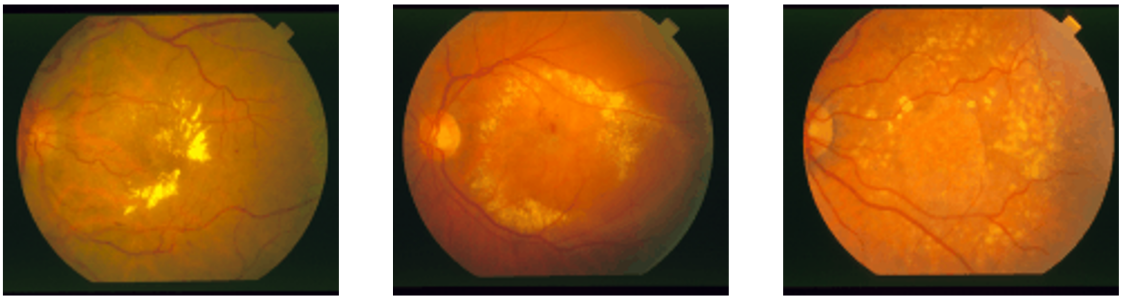


Figure 1.9: Some sample images from the STARE dataset

1.5.3 CHASEDB1: Child Heart and Health Study in England Database [31]

Most retinal image analysis datasets target adult population, however, pediatric analysis remained largely unaddressed. CHASEDB1 acts as a bridge by presenting vessel annotated retinal fundus images of a child for pediatric study. This renders it exceptionally rich for the creation of algorithms specific to pediatric ophthalmology. Figure 1.10 illustrates a set of CHASEDB1 dataset images.

- **Dataset Overview:**

  CHASEDB1 has 28 color fundus photographs (14 pairs) from 14 children aged between 7 and 9 years. These photographs were captured with a Nidek NM-200-D fundus camera of 1280 × 960 pixels resolution and 30-degree field of view.

- **Annotations:**

  Each of the images is annotated manually by two professional graders for the segmentation of blood vessels. The vessel masks contain both veins and arteries, and the dataset has been utilized by some studies for vessel classification as well. There is also a FOV mask available.

- **Significance:**

  Because of the increased resolution and distinct subject population, CHASEDB1 is best suited for high-accuracy pediatric-specific vessel segmentation applications. The dataset ensures models generalize across age ranges and are not skewed toward adult fundus features.

Figure 1.10: Some sample images from the CHASEDB1 dataset

1.5.4 HRF: High-Resolution Fundus Image Database [32]

With the improvement of deep learning algorithms, high-resolution fundus datasets have been increasing in demand. The HRF dataset was proposed to satisfy this requirement, and images can be used to provide detailed analysis for fine structures such as microvasculature and boundaries. Figure 1.11 illustrates a set of HRF dataset images.

- **Dataset Overview:**

  HRF dataset contains 45 fundus images, each of which has a very high resolution of 3504 × 2336 pixels. There are three categories of images: healthy (15 images), diabetic retinopathy (15 images), and glaucomatous eyes (15 images). This segmentation into diseases supports disease-specific segmentation analysis.

- **Annotations:**

  Manual annotations of vessels are given for all the images. Optic disc and cup segmentations are included in some versions of the dataset. High resolution of the dataset makes it appropriate for testing models that need fine spatial accuracy.

- **Significance:**

  The high-quality annotations and large image sizes make it a good choice for fine detail demanding tasks and high diagnostic accuracy. It is especially useful for training models that will be used for early disease detection and to test performance at various spatial scales.

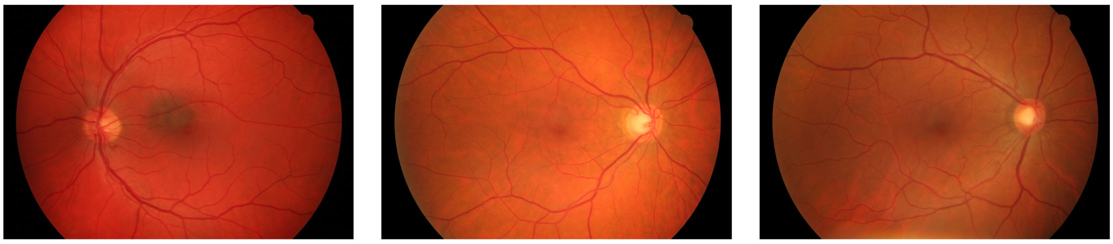


Figure 1.11: Some sample images from the HRF dataset.

1.5.5 IDRiD: Indian Diabetic Retinopathy Image Dataset [33]

With an increase in emphasis on deep learning as the fundamental aspect of diabetic retinopathy screening, datasets with multimodal annotations were required. The IDRiD dataset is noteworthy as a rich resource for lesion segmentation, diabetic retinopathy grading, and anatomical structure localization. Figure 1.12 illustrates a set of IDRiD dataset images.

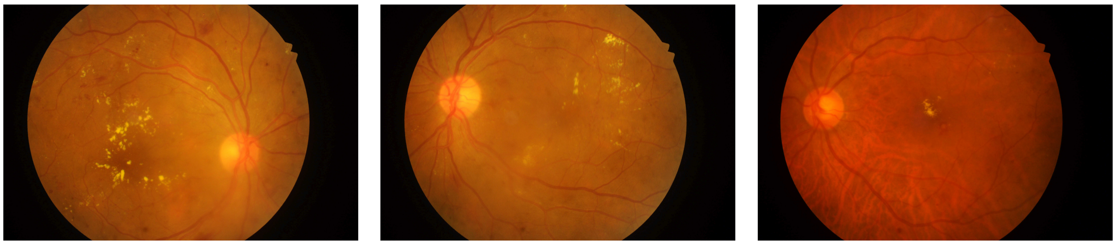


Figure 1.12: Some sample images from the IDRiD dataset

- **Dataset Overview:**

  IDRiD comprises of 516 retinal fundus images taken at a resolution of 4288 × 2848 pixels with a Kowa VX-10a camera. The dataset is separated into training and test sets, which cover a range of diabetic retinopathy grades.

- **Annotations:**

  The dataset offers precise pixel-level annotations for a range of lesion types such as microaneurysms, hemorrhages, hard exudates, and soft exudates. It also offers optic disc segmentation and fovea localization. The multi-label aspect of the annotations enables a broad range of research tasks.

- **Significance:**

  The extensive and multilayered annotations in the dataset render it precious for the development of a multi-task learning model. It allows researchers to come up with algorithms that not only do lesion segmentation but also classification and localization. Therefore, it facilitates end-to-end diagnosis pipelines for diabetic retinopathy screening.

1.6 Literature Survey

Biomarker segmentation within fundus images [34], including blood vessels, optic disc, fovea, and lesions of microaneurysms or exudates, plays a pivotal role in the diagnosis of retinal disorders such as diabetic retinopathy, glaucoma, and age-related macular degeneration. The direction of research has evolved significantly from traditional image processing approaches towards deep learning-based approaches. The following sub-sections provide a brief overview of these approaches as found in literature.

1.6.1 Data Augmentation for Fundus Image Analysis:

Data augmentation is an essential approach for enhancing model robustness, particularly in light of the limited number of annotated fundus images and the high cost of professional annotations. Basic augmentation approaches such as rotation, flipping, scaling, and brightness [35] changes are also commonly used to enhance data variability. However, these might not be sufficient in replicating clinical variations or uncommon pathological patterns.

Consecutively, advanced augmentation techniques have been utilized in recent studies, such as elastic distortions and local intensity perturbations approaches [36]. Generative models such as GANs and VAEs [37] have also been utilized to attempt generating photorealistic lesions or vessel morphology, which might help to generalize better way and avoid overfitting. However, pathology-specific augmentation, including in a disease progression modeling or the generation of artifacts such as blur or occlusion and vessel region variability, is still an underutilized area of research exploration [38].

1.6.2 Handcrafted Feature-Based Approaches:

Early approaches for segmentation were based on handcrafted features and image processing pipelines [39]. It comprised of preprocessing approaches such as contrast enhancement (e.g., CLAHE), noise filtering, and morphological processing to emphasize important structures in Fundus Images. For vessel segmentation of blood vessels, techniques such as matched filtering, multiscale line detection (Frangi filter), and Gabor filters were commonly employed to detect vessel-like structures. Additionally, thresholding and region growing approaches were employed for segmentation of the optic disc, commonly with intensity and color cues in the red or green channels of RGB fundus images. For segmentation of lesions, specifically microaneurysms and exudates, past literature comprises of works utilizing color features (e.g., intensity within the green channel), shape descriptors, and texture features (e.g., LBP, GLCM). SVMs, decision trees, and k-NN classifiers were employed to classify lesion and non-lesion pixels on the basis of such features. These methods, although providing interpretability and decent performance, were usually not robust against imaging quality variation, illumination, and pathology appearance [40].

1.6.3 Deep Learning-Based Methods:

The advent of deep learning significantly boosted the automatic biomarker segmentation in Fundus Images [41]. Fully Convolutional Networks (FCNs), particularly U-Net [42] and its variants, have become the de-facto standard due to their ability to learn hierarchical features from raw pixels without manual engineering. These models can effectively segment vessels, optic discs, and lesions using end-to-end training. Additionally, tasks such as computer vision [43–45] and natural language processing [46], heavily rely on attention mechanisms. By updating the weights at each point, the attention mechanism improves the model potential for classification. The dynamic weights, which are determined from the feature maps, further enhance the model's adaptability. Furthermore, the issue of extracting region-specific information while disregarding unnecessary regions could be resolved by an attention mechanism. Specifically, Self-Attention approach, initially suggested for machine translation tasks [46], has demonstrated superiority in natural language understanding. Self-attention mechanism was introduced into the image segmentation task by non-local neural networks [47]. In order to augment the feature representation for semantic understanding, DANet [48] enhances self-attention in both spatial and channel dimensions for semantic segmentation. This allows for the modeling of long-range interdependence. Self-attention modules, however, require significant memory and processing resources. Hence, a number of studies in the literature have focused on making them less complex [43].

For vessel segmentation, models such as U-Net [42], DeepVessel [49], and Dense U-Net [50] proved to be more accurate on public benchmarks (e.g., DRIVE, STARE). Attention mechanisms and multiscale fusion blocks [51] further enhanced performance by paying special attention to thin vessels and difficult-to-segment regions.

In optic disc and optic cup segmentation, CNNs such as Mask R-CNN and encoder-decoder models [52] performed well in terms of accuracy, particularly when structural priors or boundary-aware loss functions were added. Lesion segmentation for microaneurysms and hemorrhages was also well-handled with hybrid methods that integrated CNNs with recurrent or transformer models [53] to handle spatial dependencies and contextual features. Present stream of research investigates domain adaptation, self-supervised learning, and few-shot learning to combat data shortage and domain drift, which are prevalent in clinical datasets. In spite of the advancements, there are limitations in generalization between datasets, disease artifacts management, and segmenting small or low-contrast biomarkers.

It is thus to be noted, although handcrafted approaches provided the foundation for segmentation of fundus biomarkers, however, deep learning based approaches have extended the limits of performance and automation. In spite of this, there is room for future work to harvest the benefits of both paradigms, particularly in interpretable and low-resource scenarios with more adaptability to diseased cases.

The highlighted literature survey provides a broad glimpse into the existing works in both the fields of augmentation as well as segmentation in Fundus Images. Based on the presented literature survey, a detailed gap analysis for research exploration is discussed in the following section.

1.7 Research Gap Analysis

In spite of research advancements in biomarker segmentation from deep learning, there are a number of research gaps yet to be addressed in literature. Firstly, the unavailability of varied and high-quality annotated datasets is a major limitation of contemporary models, which impede the generalization across different populations and imaging conditions. Most current models are prone to overfitting because they are trained on limited data, which illustrates the need for robust data augmentation techniques [54]. However, current state-of-the-art augmentation techniques (e.g., basic geometric transformations or color jittering) might fail to reflect the intricate variations found in actual fundus images. Furthermore, basic geometric transformations lack providence when used in rotation invariant model designs. Thus, there is an essential need for more task-focused and pathology-conscientious augmentation methods, especially simulating unusual lesions or vasculature abnormalities.

Blood vessel segmentation is still very challenging because of the existence of thin, low-contrast vessels and overlapping structures. While attention-based and multi-scale architectures have enhanced accuracy [55], they are still poor in detecting fine capillaries or vessels in pathological areas. Additionally, the majority of deep models use conventional pixel-wise losses [56], which fail to capture clinical importance. In segmentation of biomarkers, particularly when targets are small or sparse, as in the case of exudates or bifurcations of vessels, such losses can result in inferior recall. Loss optimization methods is thus needed that more accurately capture spatial structure, class imbalance, and clinical priorities. The incorporation of such approaches can encourage networks to pay attention to clinically relevant but hard regions, thus enhancing the overall utility of segmentation in real-world screening and diagnosis applications.

In the following section, the scope of research work addressed in this thesis is discussed in detail followed by an outline of the chapter organization respectively.

1.8 Scope of the Thesis

The aim of the thesis is to tackle ongoing challenges in biomarker segmentation from fundus images, with specific emphasis on enhancing the accuracy, robustness, and clinical validity of automated approaches. While deep learning approaches have made substantial progress, data heterogeneity limitations, partitioning of intricate structures such as blood vessels, and inadequate loss function engineering still constraints the application of such models in real-world clinical environments. In regards to the observations of research gaps identified, the thesis addresses the following specific research paradigms:

Importance of pathology-aware data augmentation in enhancing generalization and robustness of biomarker segmentation models for fundus images

Fundus image segmentation annotated datasets tend to have limited availability and suffer from under-representation of rare pathological cases, contributing to poor model generalization. Current augmentation methodologies are generally restricted to generic geometric or color-scale aberrations, which are not representative of the disease-based pathological variations. The thesis presents an exploration of structure-based augmentation approaches to generate functional datasets for retinal blood vessel segmentation. Furthermore, augmentation techniques that mimic acquisition artifacts (i.e., illumination variability) is also investigated. Through better diversity and representativeness in training data, a study of performance of the model across different clinical conditions is also undertaken.

Consideration of the topological characteristics of retinal blood vessels to enhance blood vessel segmentation, particularly for thin and interrupted vessels

Segmenting blood vessels, especially thin capillaries and bifurcations, is an open-ended problem because of their intricate, sparse, and tree-like structure. Traditional CNN-based approaches tend to lose fine vessels or sacrifice anatomical continuity. Novel architectures proposed in this thesis combines multiscale feature fusion and attention representation to recognize both local and global connectivity of vessels. Additionally, topological priors and vessel connectivity constraints is studied in the network design and training for better emphasis.

Modification of loss functional computation to better capture clinically relevant anatomy and deal with class imbalance in biomarker segmentation problems

Loss functions determine the behavior of models during training. Standard losses such as binary cross-entropy and Dice tend to perform below par in regards to segmentation of small and sparsely located biomarkers. This can be attributed to the fact that biomarkers such as Exudates are highly sensitive to class imbalance and boundary quality. This work attempts to create and test a recomputed projection of loss functions that are specific to fundus biomarker segmentation. The objective of the presented work is to optimize model improvements in harmony with clinical objectives such as low count of false negatives.

1.9 Thesis organization

In this thesis, the problem of enhancing Convolution-based Neural Networks in Augmentation and Segmentation of Bio-Markers in Retina using the Fundus Imaging Modality has been studied. The study presented in this thesis has been sub-divided into three distinct stages, namely, Data Augmentation, Deep Convolutional Neural Network-based Architecture design and Computation of Loss Projection to improve Segmentation output. The thesis is organized and represented in a structure as highlighted below:

- **Chapter 1: Introduction-** This chapter introduces the importance of the study undertaken in the thesis.
- **Chapter 2: Data Augmentation for Fundus Images-** This chapter discusses the techniques and recent approaches in the field of fundus image augmentation. Subsequently, *three* different novel approaches are presented to improve the quality of medically coherent fundus image augmentation.
- **Chapter 3: Retinal Blood Vessel Segmentation from Fundus Images-** This chapter discusses the novel architectures proposed in this thesis and a comparative study with the existing literature is presented to highlight the respective efficacies.
- **Chapter 4: Bio-Marker Segmentation from Fundus Images-** This chapter studies the implications of the novel architectures proposed in Chapter 2 in a broader scale for bio-marker segmentation. Furthermore, to improve upon the

achieved performance, a novel loss projection computation approach is highlighted in this chapter as well.

- **Chapter 5: Conclusion-** This chapter concludes the study presented in this thesis and presents directions of research that can be undertaken in future to improve upon the work.

Chapter 2

Data Augmentation for Fundus Images

2.1 Introduction

The application of Artificial intelligence (AI) and Deep Learning (DL)-based approaches is witnessing a renewed development and research in the field of ophthalmology, especially in the automated diagnosis of diseases in fundus images. Figure 2.1 highlights the impact of AI-based models in the automatic detection of three ophthalmic diseases, namely, Macular Degeneration, Diabetic Retinopathy and Retinopathy of Prematurity.

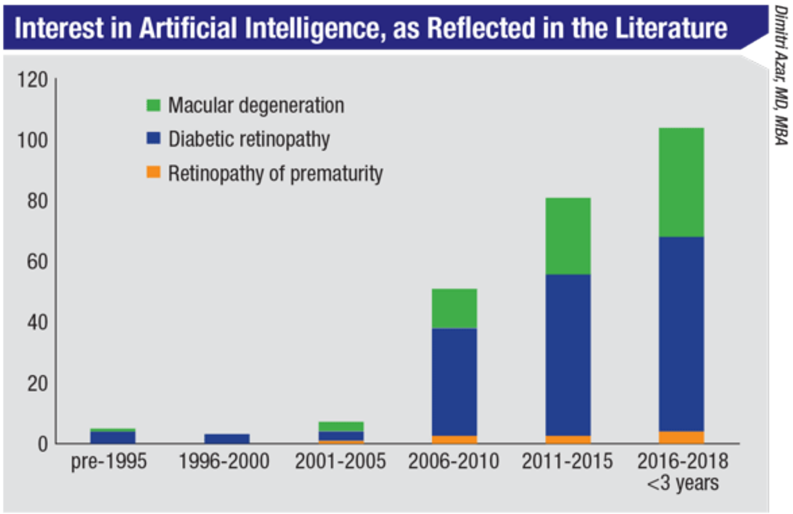


Figure 2.1: Statistical illustration depicting the interest trend in adopting Artificial Intelligence in Ophthalmology [57]

It is noted, from Fig. 2.1, that the trend of AI-based diagnostics increased by almost 20× folds within a span of approximately 10 years. In this regard, Fundus Imaging modalities play an important role in the widespread adoption of AI-based approaches for diagnosis. In spite of the trend, the potential of AI-based diagnostic systems, to be widely adopted, in this domain is limited by two key challenges: data scarcity and data divergence across imaging modalities [58]. These problems inherently impede the reliability, accuracy, and generalizability of diagnostic algorithms, hence limiting their real-world clinical value.

2.1.1 Annotated Data Scarcity in Fundus Imaging

Fundus image datasets for disease diagnosis with *annotated* visual features are crucial for designing an automated system for identifying eye diseases or monitoring disease progression. However, the scarcity of such annotated data [59] is a challenge in training as well as validation of deep learning-based models. In contrast to common image recognition applications with millions of labeled samples (e.g., ImageNet [60]), annotated fundus imaging data are usually small and, at most, comprise of only a few hundred samples combined. The lack of data availability can be attributed to an array of issues as follows:

- **Privacy and Ethical Restrictions:** Medical information is privately owned by each individual as per the law and its collection and dissemination are stringently controlled by legal and ethical regulations [61]. Thus, it becomes challenging to share large volumes of data between institutions.

- **Annotation Cost:** Annotating fundus images requires skilled ophthalmologists to label the existence, absence and/or stage of disease accurately. Furthermore, the visual characteristics of a Fundus Image are equally important and act as a biomarker for monitoring and detecting disease progression. However, labeling is a time and cost-consuming process, which restricts the amount of available annotated data [62].

- **Unavailability of Samples of Varying Retinal Conditions:** Certain retinal conditions are fairly uncommon, so it is challenging to obtain a sufficient number of examples to adequately represent all these classes [63]. This creates class imbalance where frequent conditions get overrepresented, and infrequent ones get underrepresented, making model performance biased.

Due to these constraints, models trained on small or biased datasets tend to overfit [64], where the model learns patterns that are specific to the training set but does not generalize to new, unseen data of varying modality. This is especially risky in ophthalmic applications, where a misdiagnosis can potentially tend to major complications [65]. Furthermore, limited availability of training data also restricts the model capacity to learn the finer visual biomarkers of early stages of disease, which makes early detection less trustworthy.

2.1.2 Data Divergence Across Fundus Imaging Modalities

For fundus image-based disease diagnosis, data divergence [66] presents a significant challenge that arises not only from differences in imaging modality but also from variations introduced by differences in image acquisition protocols. In this regard, the heterogeneity of image acquisition can potentially have a marked influence on the reliability and consistency of AI-driven diagnostic systems.

Fundus images are acquired with various imaging systems, ranging from high-end, hospital-grade equipments to handheld, smartphone-based acquisition units [67]. The imaging systems vary with respect to the sensor resolutions, field of view (FOV), color rendering, quality of illumination, and compression algorithms used in images. Figure 2.2 illustrates the variability among different fundus imaging modalities.

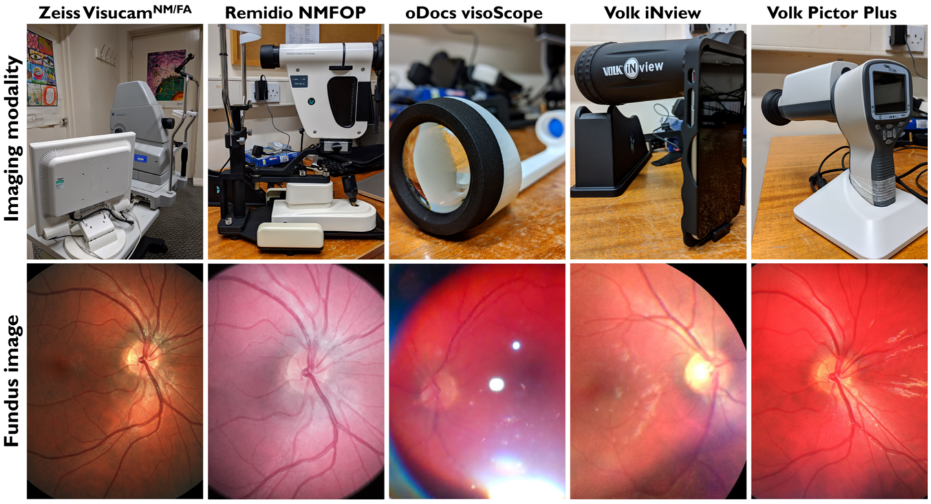


Figure 2.2: An illustration of the fundus imaging modalities using varying equipments. [68]

Subsequently, the same anatomical ocular structure may appear considerably different based on the imaging system utilized. For example, a hemorrhage or microaneurysm could be readily seen in high-resolution scans but become masked or misrepresented on lower-graded images [69]. The subtle variations in the acquisition approaches presents inter-device heterogeneity, which tend to misguide the learning ability of the deep learning models for a particular class of image data. This impacts the generalization ability of a trained model upon deployment in varying clinical settings. A model trained on high-resolution Fundus images of one imaging system alone might perform poorly on another imaging system. This is especially challenging for real-world deployment, where health facilities might use heterogeneous equipments, primarily due to economic, regional, or logistical considerations. Furthermore, publicly available datasets tend to be idealistic in experimental settings, which results in a lack of sufficient quality-based heterogeneous data during training phase [70].

2.1.3 Literature Survey

In regards to the overall challenge of fundus image analysis, retinal blood vessel segmentation is a pre-requisite task that aids in the detection and monitoring of ocular diseases [71]. Reliable segmentation of the vascular structure facilitates quantitative measurement of blood vessel parameters such as vessel tortuosity, branching patterns, and arteriovenous ratio [72]. These characteristics are the clinically relevant markers for medical diagnosis of ocular diseases such as diabetic retinopathy [73], hypertensive retinopathy [74], retinal vein occlusion [75], and glaucoma [76] as well as cardiovascular and neurological well-being. In addition, retinal blood vessel segmentation also serves as a significant biomarker for more challenging diagnosis such as anatomical landmark localization (the optic disc and macula) [77], improves lesion detection by outlining normal anatomy from a diseased one [78] and supports the reduction of false positives in classification pipelines through geometric and spatial constraints [79]. Retinal Blood Vessel segmentation, therefore, is not an independent task but rather a vital pre-processing stage for diagnostic challenges [71].

However, high-performance retinal blood vessel segmentation models are difficult to train due to dataset related challenges [80]. Firstly, the retinal blood vessels constitute a miniscule area of Region of Interest in comparison to the overall fundus image area [81]. This results in a severe class imbalance between retinal blood vessels and background pixels. Furthermore, manual annotation of vessel maps has been well-established in literature but found to be time-consuming and prone to observer bias, traditionally demanding pixel-level accuracy from skilled experts [82]. Therefore, publicly available vessel segmentation datasets are both smaller in size and poorer in imaging diversity, often causing trained models to be prone to overfitting and lack of generalization abilities between different imaging conditions and modalities [83]. To overcome these limitations, data augmentation [84] is necessary. Through artificial modifications of the existing images of the training dataset, data augmentation adds a controlled variability that can mimic real-world imaging variations within the context of existing publicly available datasets. In this regard, popular approaches such as rotation [85], flipping [85], scaling [85], contrast adjustment [86], and gamma correction [87] generates different vessel orientations and lighting modalities for model training. Augmentations using elastic deformation [88] and Gaussian noise [89] mimic minor anatomical or acquisition-related distortions as well.

Furthermore, advanced methodologies target vessel-aware fundus image augmentations, wherein the image transformation is informed of the retinal vessel structure to maintain anatomical topology [90]. Similarly, generative adversarial networks (GANs) based fundus image augmentation approaches [91–93] have been studied in the context of synthetic fundus image generation to mimic realistic fundus images. The generated synthetic datasets can be used to train existing models to produce more robust outcomes with respect to domain shift and low-contrast or occluded vessels. In order to better train models capable of handling the diversified distributions spread across different sensor modality, past works have highlighted the need to obtain a vast volume of data comprising of varying distributions for the purpose of model learning [94]. However, obtaining data from such a large set of possible fundus imaging sensors along with subsequent labelling of the data is a massive task, both in terms of cost and required man-hours. Previous works [95] have emphasized on the use of transfer learning as a possible measure for addressing the issue of unavailability of diverse data as well as to handle probable bias in the available data. Transfer learning works by training the model on a similar source domain and subsequently, fine tuning the training of the model on the target data domain. Although providing superior performances on most image processing and segmentation tasks, the transfer learning approach is yet to achieve superior performance in the domain of medical image segmentation primarily due to the need for sufficiently large number of training data and their associated ground truths during fine tuning of the model in the target domain. To overcome these shortcomings of the supervised learning models for medical image segmentation tasks, the Domain Generalization (DG) approach has been studied extensively as an active topic of research in the medical research community [96]. The DG approach works by active learning from a diversified set of training data which act as a measure to improve the network performance in unseen target domains. Domain Generalization problems can be further separated into two active areas of research, namely, Data Manipulation [97,98] and Representation Learning [99,100] approaches. In recent times, Data Manipulation has been observed as a suitable option for mitigating the issue of unavailability of diversified training data, primarily, by modifying the existing training data distribution of a dataset and generating new samples to better align with real world test distribution. Data Augmentation [101] is one such approach of data manipulation that has been used extensively for application of DG approach in medical image segmentation previously [97].

Transformation based image modification approaches is a popular technique for image augmentation, which not only preserves anatomical but as well as structural characteristics of the images. One such field of research in transformation based augmentation is the Manual Transformation (MT) based approaches which generates variational image characteristic models in the target domain with desired properties [94]. However, designing of an effective augmentation based on MT approach is both laborious and often requires requisite expertise along with a thorough knowledge-base in medical imaging. As an example, image deformation can act as an effective measure for data augmentation in case of prostate segmentation [102], however, in the field of Fundus image analysis, it might be ineffective. The nerve vessels in the Fundus images originate from the optic disc region which is an important property of the vessel network and thus needs to be preserved across all augmented dataset. Automated augmentation policies have been proposed in the past to address this drawback of MT approaches [102], however, the learning capabilities of the model are often times limited by the scope of the training set distribution and thus, inherently assume the similarity of the test set distribution to the training set distribution which can be erroneous. Approaches have been studied in past literature to convert Low Resolution Fundus Images into High Resolution Fundus Images as an initial attempt at image generation [103]. In this aspect, ESRGAN had been proposed in literature for generating realistic fundus images [104]. However, the model suffers from structural distortion anomaly in retinal fundus images. A similar approach was undertaken in Med-SRGAN by Gu et al. [105] as well. Kim et al. proposed an image turing test to generate synthetic retinal images [106]. However, generative models are difficult to adopt for bio-marker segmentation in Fundus Images owing to the need for human ground truth annotation and/or machine generated accurate and gold standard annotation.

To address the bottleneck of generating annotated synthetic retinal fundus image dataset, three different sets of Fundus image augmentation approaches are presented in the subsequent sections.

2.2 Vessel Curvature-based Fundus Image Augmentation

The highlighted research gaps are identified in the domain of Retinal Fundus Image augmentation and a solution is designed to generate a vessel structure modified retinal fundus images along with the respective ground truths. The contribution in the work can be briefly summarized as follows:

1. The representation of a Retinal Vessel Fundus Image ($\mathbb{I}$) as a decomposable MT-based function $\mathbb{F}$ is designed which is parameterized on a modifiable set of parameters. The functional formulation of the image space as a decomposable function permits the design of a controllable set of parameters that can be modulated to generate new data samples which conform to the anatomical and

structural principles of Retinal Fundus Images. Additionally, the functional approach provides greater flexibility in Data Generation as opposed to traditional approaches such as translation, skew, rotation and cropping.

2. A framework for the modulation of the Background Region properties is designed based upon the functional definition F, which separates an input image space into three uniform regions, namely, Optic Disc, High Contrast and Low Contrast regions.
3. A vessel curvature modulation algorithm is designed that can be used to modify the existing vessel structure and curvature to generate new vessel network for data augmentation.

An overview of the *Vessel Curvature-based* Augmentation approach is shown in Figure 2.3.

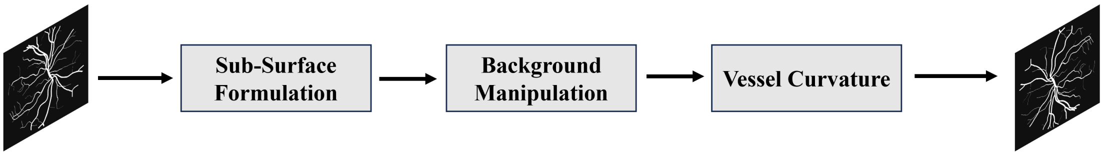


Figure 2.3: Block Diagram of the Vessel Curvature-based Fundus Image Augmentation

An input fundus image $\mathbb{I} \in \Re^2$ as an intensity surface model which can be represented mathematically using Equation 2.2.1 as

$$\mathbb{I}(x, y) = z \tag{2.2.1}$$

where x and y are the positional coordinates of $\mathbb{I}$ and z is the pixel intensity at the position (x,y).

We define a function f, such that

$$\mathbb{F} = f(\mathbb{B}, \mathbb{V}) \tag{2.2.2}$$

where $\mathbb{B}$ is a surface that adheres to the Euclidean properties and $\mathbb{V}$ is a set of curvature surfaces conjoined on the planar surface $\mathbb{B}$.

From Equation 2.2.2, mathematically, we define f as a function that operates on the two surfaces $\mathbb{B}$ and $\mathbb{V}$ such that

$$\mathbb{F} = \int_{(x,y)} [\mathbb{B}(x, y) - m] * \mathbb{V}(x, y)\, d(x, y) \tag{2.2.3}$$

where m refers to a bias constant. The visual representation of $\mathbb{F}$ is shown in Figure 2.4.

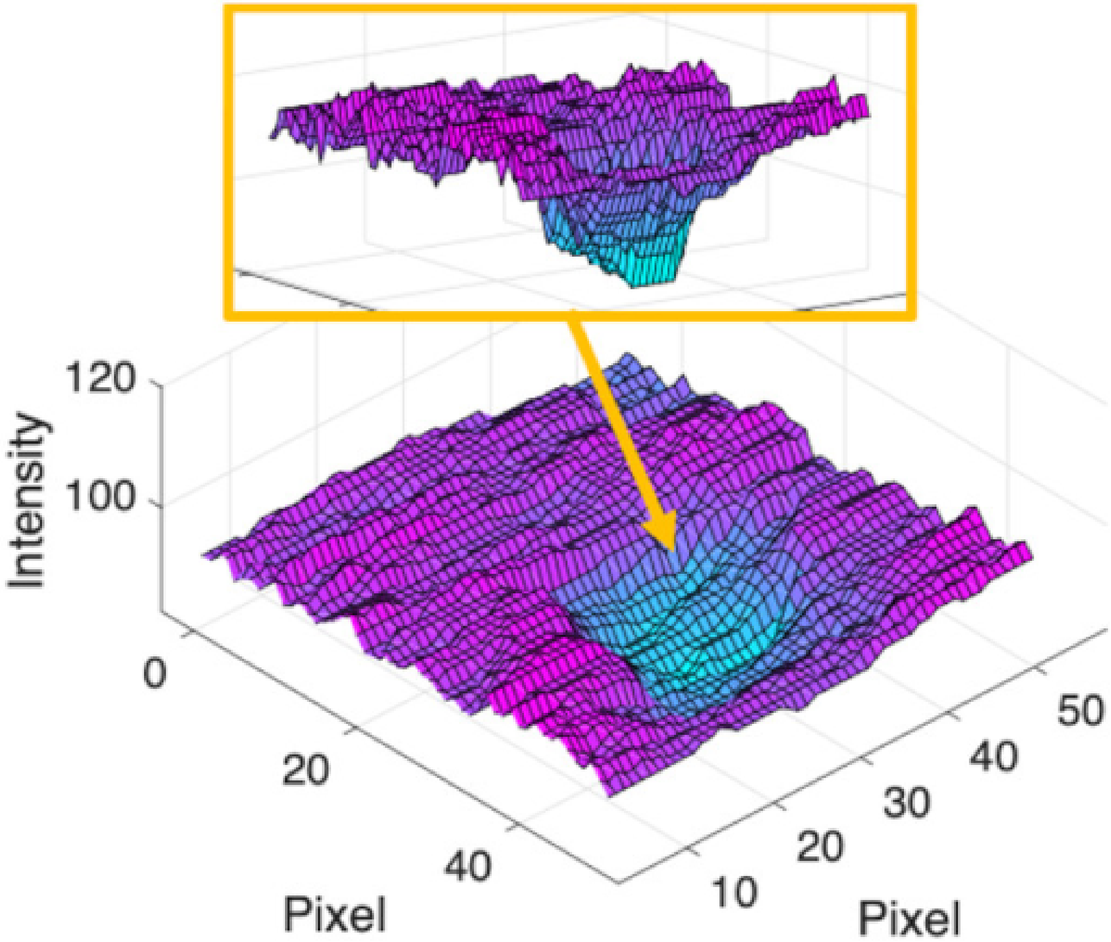


Figure 2.4: A Surface heatmap representing the spatial intensity ($\mathbb{F}$ in Equation 2.2.3) for a fundus image [107].

Equation 2.2.3 highlights an integration over the surface product of $\mathbb{B}$ and $\mathbb{V}$ for each positional coordinate (x,y). As evidenced from Equation 2.2.3, the surfaces of $\mathbb{B}$ and $\mathbb{V}$ dictate the overall surface model of $\mathbb{F}$. Thus, input image $\mathbb{I}$ can be formulated as a surface, defined based on two sub-surfaces, namely, Background Non-Vessel Region ($\mathbb{B}$) and Foreground Vessel Region ($\mathbb{V}$). The modification of the sub-surface properties of $\mathbb{B}$ and $\mathbb{V}$ would result in the modification of the function $\mathbb{F}$ which would result in a new image being generated.

2.2.1 Sub-Surface Formulation

The surface properties of $\mathbb{B}$ and $\mathbb{V}$ can be defined with respect to an input fundus image $\mathbb{I}$. It is postulated that the image surface $\mathbb{B}$ is a superposition of k image surfaces having varying levels of *stability*. In this regard, *stability* of a surface can

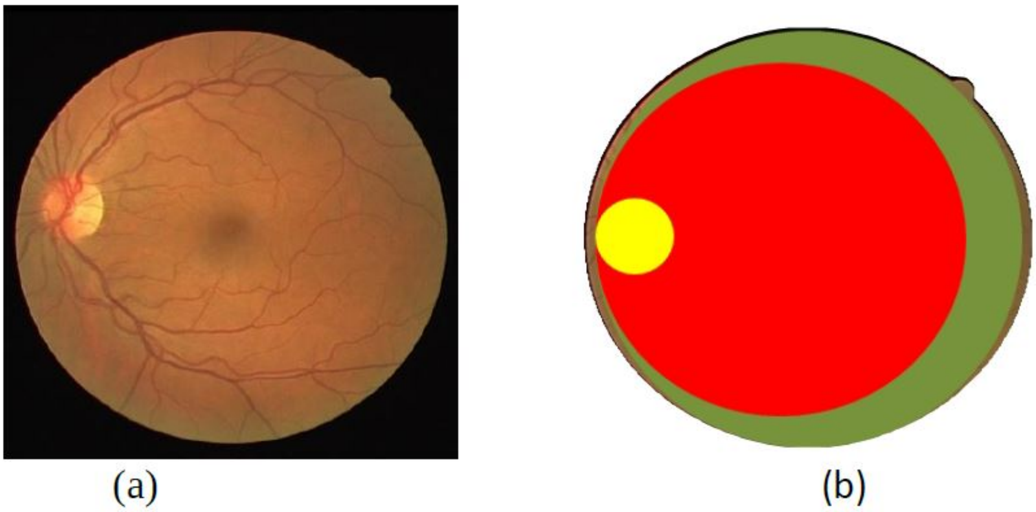


Figure 2.5: An illustration of the image surface ($B_k, \forall k \in [0, 1, 2]$) homogeneity. (a) The original RGB Fundus Image. (b) Sub-Surface Regions B_k super-imposed on (a). Each sub-region is marked with a separate visual aid, namely, **Yellow**: Optic Disc Region, **Red**: High Contrast Region, **Green**: Low Contrast Region.

be constituted as a state of an image surface where all the points possess the same intensity. Mathematically, it can be represented as

$$\forall_{(x,y)} \mathbb{B}_k(x, y) - z_k = 0 \tag{2.2.4}$$

such that z_k is the stability level for image surface B_k at surface level k. A visual representation of the image surface B_k is shown in Figure 2.5.

Image $\mathbb{I}$ can be postulated to comprise of three stable regions based on anatomical and visual characteristics, namely, Optic Disc, High Contrast and Low Contrast regions respectively. The optic disc is a bright, roundish region through which the optic nerve exits the retina; it is highly textured and bright and can therefore be easily identified in a fundus image. By comparison, high contrast regions usually contain notable retinal structures such as vessels or lesions (e.g., hemorrhages, exudates), where intensity variations from the background are clear and diagnostically significant. Low contrast regions, in contrast, depict fairly homogeneous regions such as the background retina or minimal pathological changes without prominent intensity variation but possibly clinically useful information. Thus, it is safe to postulate that $k = 3$ corresponding to the three stable regions. Additionally, the vessel surface $\mathbb{V}$ can be defined as a set of B-Spline Curves with an individual set of control parameters for each curve. By modulating the control parameters, its respective curvature can be modified structurally to generate new vessel structures.

2.2.2 Background Region Manipulation

A background region ($\mathbb{B}$) in a Fundus image can be defined as the set of all pixels which do not belong to the vessel set. For an input green channel Fundus image $\mathbb{I}_G$, region $\mathbb{B}$ is sub-partitioned into three homogeneous regions, namely, Optic Disc, High Contrast and Low Contrast regions respectively. The algorithm to partition the region $\mathbb{B}$ into sub-regions can be formulated as follows. A block diagram of the method is demonstrated in Figure 2.6.

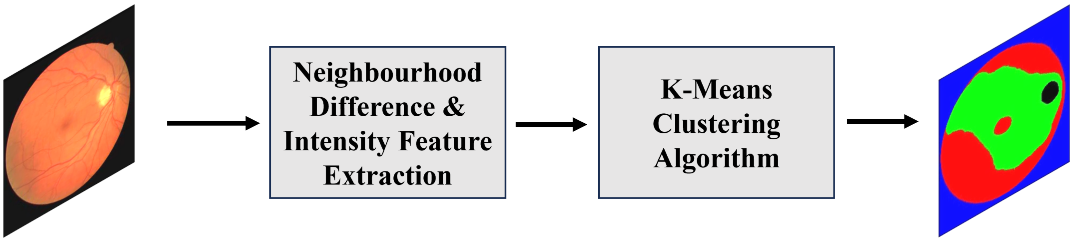


Figure 2.6: Steps for Background Region Partitioning

Firstly, the pixels belonging to the vessel map annotated image $\mathbb{I}_{gt}$ are removed from the corresponding input green channel fundus image $\mathbb{I}_G$. The choice of the green channel of image $\mathbb{I}$ for analysis is based on an observation made by Liskowski et al. [108]. Let the vessel removed image be termed as $\mathbb{I}'_G$. For each pixel in $\mathbb{I}'_G$, two features, namely, Pixel Intensity (f_1) and Neighborhood Difference (f_2) are to be computed for separating the background region into a set of three homogeneous regions, namely, Optic Disc ($\mathbb{I}_{OD}$), High Contrast ($\mathbb{I}_{high}$) and Low Contrast regions ($\mathbb{I}_{low}$) respectively. It is noted that the image domain sub-regions $\mathbb{I}_{OD}$, $\mathbb{I}_{high}$ and $\mathbb{I}_{low}$ correspond to real-world domain of $\mathbb{B}_0$, $\mathbb{B}_1$ and $\mathbb{B}_2$ respectively and can be used interchangeably based on the domain of study.

The pixel intensity value of the vessel removed image $\mathbb{I}'_G$ for each pixel P is considered as the first feature, namely, f_1. For each pixel P in $\mathbb{I}'_G$, the second feature, namely, *Neighborhood Difference* (f_2), is calculated based on the intensity of pixel P and the intensities of the pixels present in the 8-neighborhood of the pixel P. The intensity value of pixel P is subtracted from the 8-neighborhood pixel intensities and the average intensity of the neighborhood is considered as value of feature f_2. The neighborhood pixels, which have minimal intensity difference with the central pixel P, correspond to regions having low contrast in the Fundus Image. Similarly, the neighborhood pixels having high intensity difference with the central pixel P correspond to regions having high contrast in the Fundus Image. Figure 2.7 illustrates the computation of the *Neighborhood Difference* (f_2).

The two features, namely, f_1 and f_2, are collectively used to compute the region

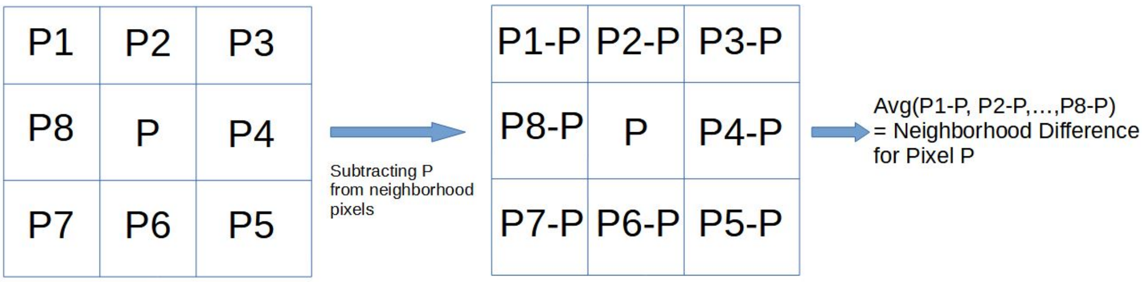


Figure 2.7: An illustration of the computation of the *Neighborhood Difference* (f_2) feature for Background Region Separation.

separation using a k-Means clustering algorithm, having a total of three different clusters, each corresponding to one region in the image, namely, Optic Disc, High Contrast Region, and Low Contrast Region. To identify the clusters and associate the same with the respective regions in the obtained output, the feature values of f_1 and f_2 of each cluster is used. The cluster having the highest pixel intensity (f_1) data points correspond to the Optic Disc region, $\mathbb{I}_{OD}$. The cluster having the highest neighborhood difference intensity (f_2) data points correspond to the high contrast region $\mathbb{I}_{high}$ and the remaining data cluster corresponds to the low contrast region $\mathbb{I}_{low}$ in the input image $\mathbb{I}_G'$. An illustration of the partitioned background region masks is demonstrated in Figure 2.8.

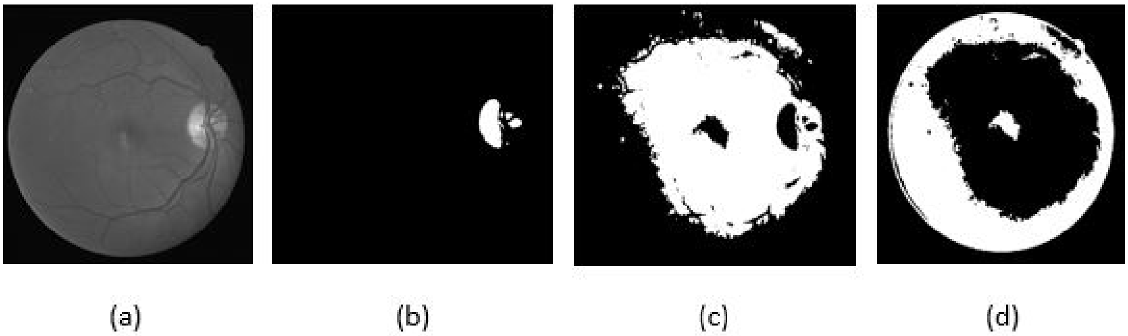


Figure 2.8: Obtained region identifying masks for an input fundus image I (a) Input Fundus Image I (b) Optic Disc Region C_O (c) High Contrast Region C_H (d) Low Contrast Region C_L

The separation of the background region $\mathbb{B}$ into three separate sub-regions provides an opportunity to individually control the characteristic properties of the respective sub-region. As per the definition of image sub-surface B_k, it is observed that the surface B_k is a function of the properties of the region. In other words, functional properties of the region defines the shape and structure of the sub-surface B_k. For background region manipulation, two such functional properties of the sub-surface B, viz., Brightness (Br) and Contrast (Cr) are considered. The properties of Br and Cr for the region B_k can be depicted, mathematically, using Equation 2.2.5 and Equation 2.2.6 as.

$$Br_t^k(x,y) = ((255 - Br_{(t-1)}^k(x,y))/255) * \mathbb{B}_k(x,y) + Br_{(t-1)}^k(x,y) \tag{2.2.5}$$

$$Cr_t^k(x,y) = \alpha * \mathbb{B}_k(x,y) + \gamma \tag{2.2.6}$$

where $\alpha = ((131 * (Cr_{(t-1)}^k(x,y) + 127))/(127 * (131 - Cr_{(t-1)}^k(x,y))))$ and $\gamma = 1 - \alpha$. $Br_t^k(x,y)$ and $Cr_t^k(x,y)$ represents the brightness and contrast respectively at a time period t for point (x,y).

Equations 2.2.5 and 2.2.6 reflects two important parameters for modulation of sub-surface B_k, i.e., Br factor and Cr factor. The modulation of the Br factor varies the brightness quotient of the region with values greater than zero increase the brightness of the region whereas values lesser than zero decrease the brightness of the region. Similarly, the modulation of the Cr factor varies the contrast quotient of the region B_k accordingly.

It should be noted that modulation of the background sub-surface region can cause Noise to be introduced into the image due to variations across different imaging modalities. To mitigate the issue, two constraints are imposed on the Br and Cr factors respectively. Firstly, the overall brightness factor Br being an image-centric measure instead of being region specific, the modulation across the various background sub-surface regions is performed maintaining the same Brightness factor Br for all regions B_k. Secondly, the contrast factor Cr is a region specific measure. However, in Fundus imaging, the measure of contrast for a sub-region B_k is defined as a difference in the visual clarity between the background sub-surface B_k and the set of vessel pixels contained within the sub-regional space of B_k. As a result, the contrast factor Cr needs to be limited by an empirically adopted threshold λ as otherwise the variation of the factor Cr beyond λ would violate the inherent visual clarity of the vessel pixels near the vessel boundaries which might lead to poor learning results during segmentation.

2.2.3 Vessel Curvature Modification

The vessel sub-surface $\mathbb{V}$ can be defined as a set of intersecting B-Spline curves having origin points inside the Optic Disc Region ($\mathbb{I}_{OD}$). Each vessel can be considered as a B-Spline curve that adheres to the curve form given in Equation 2.2.7.

$$C(u) = \Sigma_{i=0}^{n} N_{(i,p)}(u).P_i \tag{2.2.7}$$

However, the B-Spline curve shown in Equation 2.2.7 represents a single curve, however, in a Fundus image, there are multiple vessels with a number of branches. The vessel surface $\mathbb{V}$ can thus be summarized as a superimposition of a set of B-Spline curves. In order to modulate an individual vessel, firstly, isolated and independent B-Spline curves need to be extracted from the vessel set. The vessel surface $\mathbb{V}$ can thus be mathematically formulated as

$$\mathbb{V} = \cup_i \{\mathbb{V}_i\} \tag{2.2.8}$$

As evidenced from Equation 2.2.7, it can be observed that modification of individual vessel curvatures directly result in the change of $\mathbb{V}$ surface. To isolate the individual vessel surfaces $\mathbb{V}_i$, firstly, the Optic Disc Region $\mathbb{I}_{OD}$ is extracted from image $\mathbb{I}$. The centroid (m_o) of $\mathbb{I}_{OD}$ is computed using a standard region centroid detection algorithm [109] and a circular region ($\mathbb{I}^b_{OD}$) centred at m_o that encompasses all pixels that belong to $\mathbb{I}_{OD}$ is delineated. $\mathbb{I}^b_{OD}$ is represented mathematically as

$$\mathbb{I}^b_{OD} = \exists \epsilon \in [1,\infty)\{(x,y) \mid (x,y) - m_o \leq \epsilon, \forall (i,j) \in \mathbb{I}_{OD} \rightarrow (i,j) \in \mathbb{I}^b_{OD}\} \tag{2.2.9}$$

The threshold of ε in Equation 2.2.9 is a dynamic radius for the circular region of $\mathbb{I}^b_{OD}$ and is inherently dependent on the pixels that belong to $\mathbb{I}_{OD}$. In other words, it can be stated that $\mathbb{I}^b_{OD}$ is an approximation of the optic disc region $\mathbb{I}_{OD}$. The approximation of the optic disc region by $\mathbb{I}^b_{OD}$ is necessary to incorporate probable pixels, which are supposed to be part of $\mathbb{I}_{OD}$, which might have been missed by the Background Extraction algorithm. The optic disc region $\mathbb{I}_{OD}$ is approximated by a circular surface as it is evident that in a Fundus Image, Optic Disc is circular in nature mostly.

Subsequently, we remove all vessel pixels in $\mathbb{V}$ which are part of the approximated optic disc region $\mathbb{I}^b_{OD}$ to form an approximately disjoint set of vessels. This approach for vessel manipulation is motivated by a biological aspect observable in Fundus Images: all vessels originate and are inter-twined inside the Optic Disc Region. The removal of the vessel region inside the Optic Disc region creates a disjoint vessel set such that there exists no such vessel $\mathbb{V}_i$ having end points overlapping with end points of another vessel. A modified set of vessels $\mathbb{V}'$ is defined, mathematically as,

$$\mathbb{V}' = \{\cup_i \mathbb{V}_i \in \mathbb{V} \mid end\_points(\mathbb{V}_i) \neq end\_points(\mathbb{V}_j)\} \tag{2.2.10}$$

Although the definition of V' in Equation 2.2.10 is meant to create an approximate disjoint set of vessels, it fails to take into consideration the incident of vessel overlap. To mitigate the issue, the definition of V' is further modified such that

$$\mathbb{V}\text{”} = \cup_{i,j} \mathbb{V}_i, \mathbb{V}_j \in \mathbb{V} \tag{2.2.11}$$

such that

$$\forall \mathbb{V}_j \in \mathbb{V}, end\_points(\mathbb{V}_i) \neq end\_points(\mathbb{V}_j), \forall (x,y) \in \mathbb{V}_i \rightarrow (x,y) \notin \mathbb{V}_j \tag{2.2.12}$$

To implement the approach and remove the vessel overlap points, firstly, a skeletonization algorithm [110] is applied on $\mathbb{V}'$ to generate a vessel spine image $\mathbb{V}'_S$. To identify the vessel overlap points in $\mathbb{V}'_S$, we define a pixel set $\mathfrak{P}$, mathematically, as

$$\mathfrak{P} = \{\mathfrak{P}_i \mid \forall \mathfrak{P}_i \in \mathbb{V}'_S, connectivity(\mathfrak{P}_i) \geq 2 \in \text{8-connected}(\mathfrak{P}_i)\} \tag{2.2.13}$$

Where $connectivity(\mathfrak{P}_i)$ is a function that returns the number of pixels connected to $\mathfrak{P}_i$ in its 8-connected neighborhood.

From the definition of $\mathfrak{P}$, it can be observed that the pixels $\mathfrak{P}_i$ occur at the intersection points in vessel spine image $\mathbb{V}'_S$. The pixels in set $\mathfrak{P}$ are removed from the vessel spine image $\mathbb{V}'_S$ which forms a complete vessel disjoint set, in accordance of the definition of $\mathbb{V}$”. The set of vessels present in $\mathbb{V}$” forms the basis for the vessel curve manipulation algorithm. Firstly, a function $f^{mod}(.)$ is formed which is applied on the set $\mathbb{V}$” to generate new vessel spines to augment the vessel set $\mathbb{V}$”. The function $f^{mod}(.)$ is defined as a rotation operation only on vessels belonging to vessel spine set $\mathbb{V}$”. A set of n vessels from the vessel set $\mathbb{V}$” are selected and the rotation function $f^{mod}(.)$ is applied to generate augmented vessels and substituted in the vessel set $\mathbb{V}$”. To maintain uniformity, the vessels are selected from different positions, distributed uniformly across the fundus image. The rotational limit of $f^{mod}(.)$ is randomly selected within a limit of $\mid \lambda' \mid$. Finally, the vessel spines in $\mathbb{V}$” are region-filled with a Region Growing algorithm [111] to regenerate original vessel structures from the vessel spine to reform vessel set V. Additionally, the vessels that were removed from within $\mathbb{I}^b_{OD}$

are also substituted in the vessel set $\mathbb{V}$ to complete the vessel surface.

The modified vessel network $\mathbb{V}$ is to be added back to the background region based on Equation 2.2.14.

$$\mathbb{I}_{recon} = (\cup_k \mathbb{B}_k) \cup \mathbb{V} \tag{2.2.14}$$

to obtain the final augmented fundus image. It should be noted that $(\cup_k \mathbb{B}_k)$ would henceforth be referred to as $\mathbb{B}'$ for better understanding.

However, upon augmentation to form $\mathbb{I}_{recon}$, the original position of the modified vessels would contain no data which would be erroneous as the empty region would constitute neither a background region nor a vessel region. To mitigate the issue, a raster scan approach is used to fill the region gaps present in the background region image $\mathbb{B}'$ prior to appending the vessel set $\mathbb{V}$ in Equation 2.2.14. Two independent images $\mathbb{B}_t$ and $\mathbb{B}_l$ are formed from $\mathbb{B}'$. The background image $\mathbb{B}'$ is raster scanned from left to right and a gap-fill algorithm $\mathbb{G}$ is applied to form an image $\mathbb{B}_l$. Similarly, the background image $\mathbb{B}'$ is raster scanned from top to bottom and a gap-fill algorithm $\mathbb{G}$ is applied to form an image $\mathbb{B}_t$. An element-wise maximum operation, between pixels at each location (x, y), of images $\mathbb{B}_l$ and $\mathbb{B}_t$ is performed and a cumulative image $\mathbb{B}_m$ is obtained. Finally, the modified background image $\mathbb{B}_m$ is merged and the vessel image $\mathbb{V}$ to form the augmented image $\mathbb{I}'_{recon}$.

The ground truth associated with the augmented data is also generated conjointly with $\mathbb{I}'_{recon}$. Figure 2.9 highlights a detailed overview of the set of steps to generate the augmented images.

2.2.4 Experimental Results and Discussions

The augmented set of data is used to train a vanilla U-Net model [42] for testing the applicability of the augmentation approach. The augmentation approach is validated on the DRIVE [112] dataset for Retinal Vessel Segmentation challenge, which inherently implies all generated data is a derivation of the data samples in the DRIVE dataset. However, the approach is motivated to augment the sample set to form a distribution agnostic dataset. To check the validity of the augmented dataset distribution, the trained model is cross-tested on the STARE [30] dataset to validate the augmentation approach.

Firstly, the trained model is tested on the DRIVE dataset that comprises of 20 test images along with their respective ground truths. Similarly, the trained model is additionally tested on the STARE dataset as well. However, as the STARE dataset does not provide any bifurcation between the training and testing dataset, all the

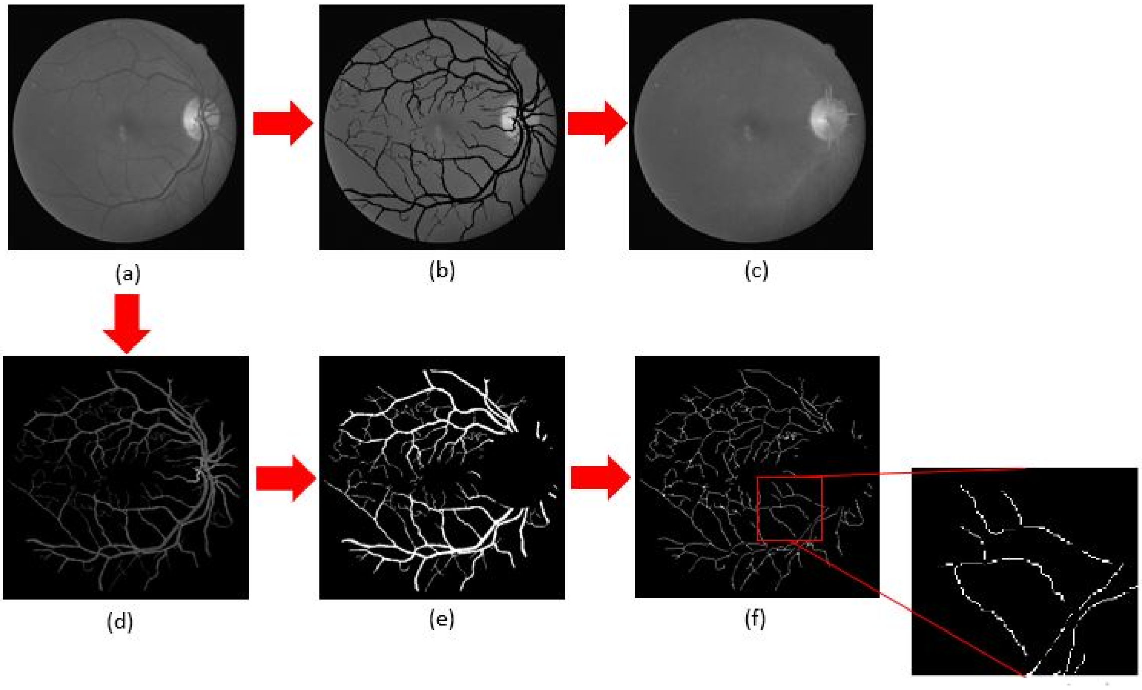


Figure 2.9: Overall Architecture of the Proposed Approach. (a) Input Green Channel image $\mathbb{I}_G$ (b) Background Region $\mathbb{B}$ (c) Gap-Filled Maxima Background Image (d) Vessel Surface Image $\mathbb{V}$ (e) Vessel Image $\mathbb{V}'$ with all pixels within Optic Region $\mathbb{I}^b_{OD}$ removed (f) Disjointed Vessel Set $\mathbb{V}''$

images present in the dataset can be used for testing purposes. Table 2.1 highlights the Accuracy, Sensitivity and Specificity results obtained for DRIVE dataset. A similar comparative analysis is provided for STARE dataset in Table 2.2. Two forms of segmentation approach have been recorded in Table 2.1 and Table 2.2, namely, patch-based and full-scale image. Patch-based approaches split a given input image into a set of patches and individually perform segmentation on each patch. In contrast, full-scale approaches segment the entirety of the image in a single pass.

It can be observed from Table 2.1 and Table 2.2, that the vessel curvature approach has achieved comparable results on Sensitivity in the DRIVE dataset as well as STARE dataset when compared with both patch-based as well as full image segmentation approaches of the past. Additionally, the approach has achieved appreciable results in both Accuracy as well as Specificity measures with most benchmark approaches in literature. The performance of the augmentation approach is especially noticeable for the Sensitivity measure. The performance of a model on Sensitivity dictates the ability of the network to detect the vessel pixels more accurately. In medical image analysis, especially Fundus Imaging, the detection of true vessel pixels is of utmost importance to detect probable ocular abnormalities. Accuracy, although an important metric of performance, fails to truly dictate the strength of a model in

Table 2.1: Performance Comparison on DRIVE dataset

Methods	**Accuracy**	**Specificity**	**Sensitivity**
Jiang et al. [113]	0.891	0.900	0.830
Staal et al. [29]	0.944	0.977	0.719
Ricci et al. [114]	0.959	0.972	0.775
Roychowdhury et al. [115]	0.949	0.978	0.739
Liskowski et al. [108]	0.925	0.924	0.916
Yan et al. [56]	0.954	0.981	0.765
CE-Net [116]	0.954	0.975	0.821
IterNet [117]	0.957	0.983	0.779
HAnet [118]	0.958	0.983	0.797
Genetic U-Net [119]	0.957	0.976	0.830
TP-Net [120]	0.963	0.976	0.875
Vessel Curvature	**0.940**	**0.947**	**0.917**

Table 2.2: Performance Comparison on STARE dataset

Methods	**Accuracy**	**Specificity**	**Sensitivity**
Jiang et al. [113]	0.901	0.900	0.857
Staal et al. [29]	0.952	0.981	0.697
Ricci et al. [114]	0.965	0.939	0.903
Roychowdhury et al. [115]	0.956	0.984	0.732
Liskowski et al. [108]	0.931	0.930	0.931
Yan et al. [56]	0.961	0.985	0.758
CE-Net [116]	0.966	0.984	0.829
IterNet [117]	0.970	0.988	0.771
HAnet [118]	0.967	0.984	0.818
Genetic U-Net [119]	0.971	0.984	0.865
TP-Net [120]	0.972	0.982	0.885
Vessel Curvature	**0.915**	**0.920**	**0.911**

vessel segmentation tasks, primarily, as almost upto 90% of a fundus image is filled with non-vessel pixels. In such a scenario, the performance of the model on Sensitivity becomes of utmost importance and a superior result is achieved while maintaining a balance with the other performance metrics. This highlights the strength of the approach of Vessel Curvature-based Data Augmentation to train Deep Learning models for obtaining superior results in Vessel Segmentation. The slight drop in the performance on STARE dataset can be attributed to a fundamental drawback of the approach. The augmented data samples are derived from the DRIVE dataset which introduces an inherent bias in the data generation process. As a result, the model performs slightly better in DRIVE dataset than on STARE dataset.

Furthermore, two commonly used metrics: Frechet Inception Distance (FID) and Multi-scale Structural Similarity Index Measure (MS-SSIM) [121] are additionally used to evaluate the augmented images in comparison to real images. An overall FID score of 1.8342 is obtained compared to a score of 3.5583 for traditional augmentation (TA) [16]. Similarly, an MS-SSIM score of 0.6431 is obtained compared to 0.6119 for traditional augmentation approaches. The lower FID scores of *Vessel Curvature* data augmentation indicates that the images it generates are of higher quality when compared to the TA method. In addition, the augmentation method produced more coherent images with respect to real dataset than TA method due to the higher average MS-SSIM value.

2.2.5 Ablation Study

Data Augmentation forms the crux of the vessel curvature approach. To highlight the efficacy of the augmentation approach, a comparative analysis is required between the segmentation performance of the pre-augmented dataset and the post-augmented dataset. However, the pre-augmented dataset comprises of only 20 annotations for DRIVE and STARE dataset respectively. To propose a solution on such a small-scale dataset, hand-crafted featured approaches had been proposed in the past in literature [29, 113–115]. Alternatively, Deep Learning approaches had been attempted on the pre-augmented dataset [56, 108], in which the input images have been sub-divided into a set of patches for enhancing the volume of data available for experimentation. As highlighted in Table 2.1 and Table 2.2, the vessel curvature approach of augmentation surpasses the performance obtained in the past literature. Additionally, the vessel curvature augmentation approach has a superior performance in comparison to traditional augmentation approaches, as evidenced in Table 2.1 and 2.2.

2.3 Vessel Curvature-based Fundus Image Augmentation v2.0

A primary drawback of the earlier version of the Vessel Curvature-based Segmentation approach exists in the incoherent structure generated. Figure 2.10 demonstrates the shortcomings of the augmentation approach. It is observed that the generated fundus images are medically and structurally incoherent with present disconnected vessel network. Although suitable for the task of vessel segmentation, however, the v1.0 of the approach does not add to the congruence of the dataset in terms of medical relevance. In this regard, the vessel curvature approach presented earlier is modified and v2.0 of the approach is highlighted next.

Figure 2.10: An illustration of the drawback of *Vessel Curvature* approach shown in Section 2.2. The RED circles highlight the incongruent regions in the augmented images.

2.3.1 Proposed Methodology

Let us consider a fragment of a vessel skeleton, as shown in Figure 2.11. To elaborate on the vessel augmentation approach, firstly, a $terminal_point$ is defined, in the context of a vessel skeleton, as the last end point of the vessel skeleton. It is observed from the vessel skeletal structure, that two cases can arise:

Case 1: A vessel skeletal strand is connected at both points by joint points, i.e., $end\_points(\mathbb{V}_i) \neq terminal\_point$.

Case 2: A vessel skeletal strand is connected at one end point by joint point and other end by terminal point, i.e., $terminal\_point \in end\_points(\mathbb{V}_i)$.

Sub-figure (a) in Figure 2.11 demonstrates Case 1 whereas Sub-figure (b) demonstrates Case 2 respectively.

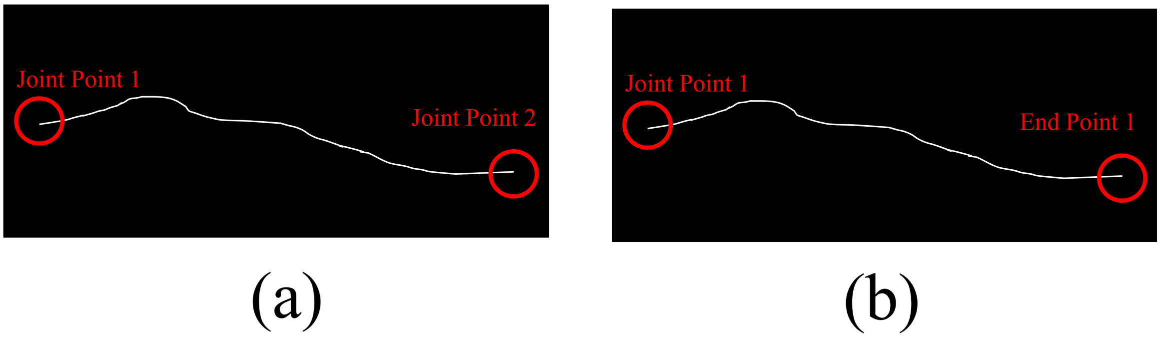


Figure 2.11: An illustration of the conditions for *Vessel Curvature v2.0* approach (a) Case 1 (b) Case 2.

To bend the respective vessels, a displacement force is applied along the positive y-axis of the respective vessel skeleton. The directionality and intensity of displacement applied is implicated in Figure 2.12. The larger arrow depicts a larger displacement as per Equation shown in Figure 2.12. By using the given approach, the augmented dataset obtained is shown in Figure 2.13.

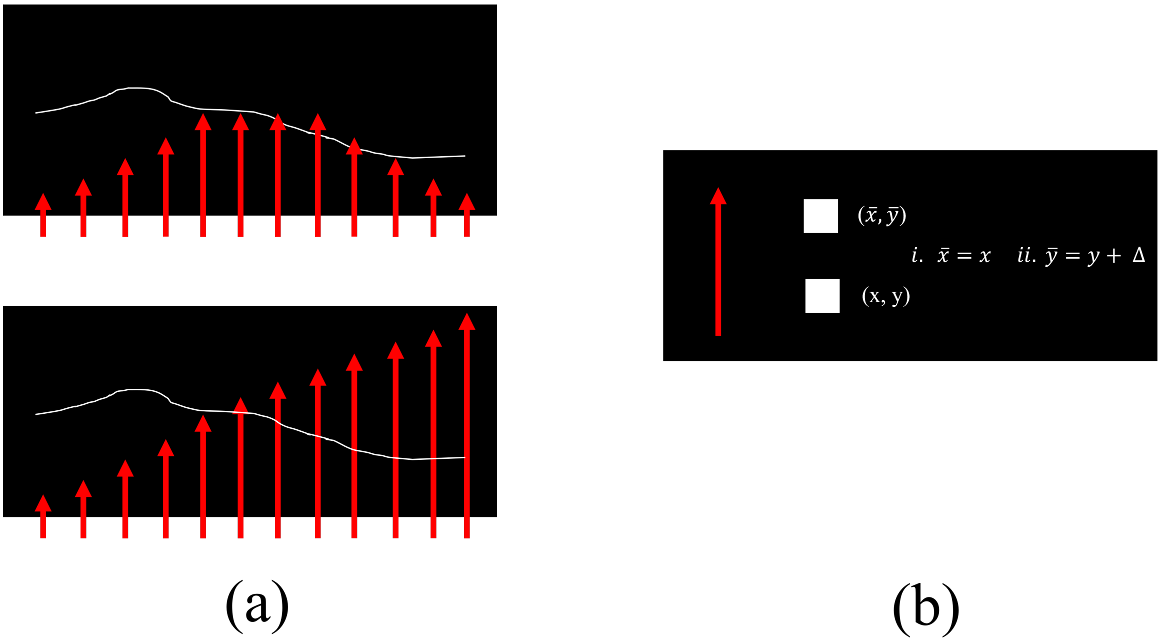


Figure 2.12: An illustration of the force displacement of vessel skeletal strand. (a) Direction and Force of displacement along the skeletal strand (b) Equations for displacement of (x, y) to new points $(\tilde{x}, \tilde{y})$ respectively.

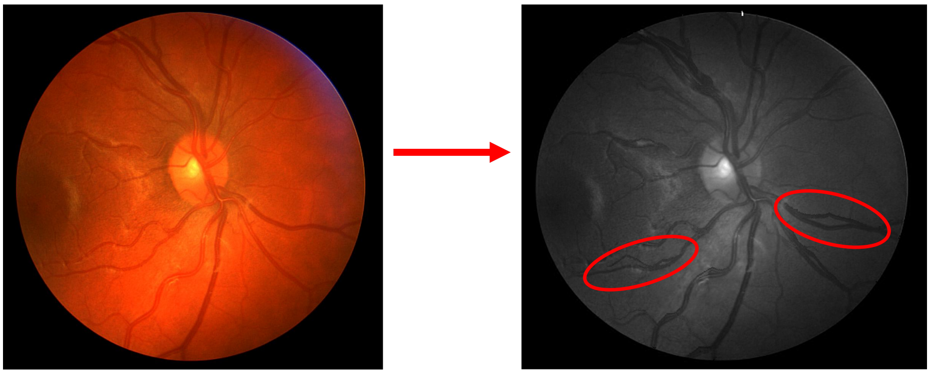


Figure 2.13: An illustration of the modified green channel Fundus Image. The red circle marks the regions with vessel modifications.

2.3.2 Colorization Model Architecture

The *Vessel Curvature v2.0* approach generates images in Green color channel. However, for realistic use-case scenarios, an RGB color-scale model is necessary. In this regard, a U-Net backbone based *Colorization* model is used to convert Green channel Fundus Images into RGB color-scale images. Figure 2.14 outlines the architecture of the colorization model.

The architecture of the Encoder block used in Figure 2.14 is shown in Figure 2.15. Self-Attention mechanism [122] is used in the design to enable the model to place additional emphasis on learning the critical regions of the image. The model is trained by a combination of the standard loss functions *Dice Loss* and *MSE Loss* as shown in Equation 2.3.1

$$\mathcal{L}_{color} = \mathcal{L}_{Dice} + \mathcal{L}_{mse} \tag{2.3.1}$$

2.3.3 Experimental Results and Discussion

A vanilla U-Net model [42] is trained for testing the applicability of the *v2.0* augmentation approach. The augmentation approach is validated on the DRIVE [112] dataset for Retinal Vessel Segmentation challenge. Furthermore, the trained model is cross-tested on the STARE [30] dataset to validate the augmentation approach. Table 2.3 highlights the Accuracy, Sensitivity and Specificity results obtained for DRIVE dataset. A similar comparative analysis is provided for STARE dataset in Table 2.4.

It can be observed from Table 2.3 and Table 2.4, that the vessel curvature v2.0 approach has achieved better performance than earlier approach in the DRIVE dataset as well as STARE dataset. However, the performance of the updated augmentation approach is reduced for the Sensitivity measure. The performance of the approach on

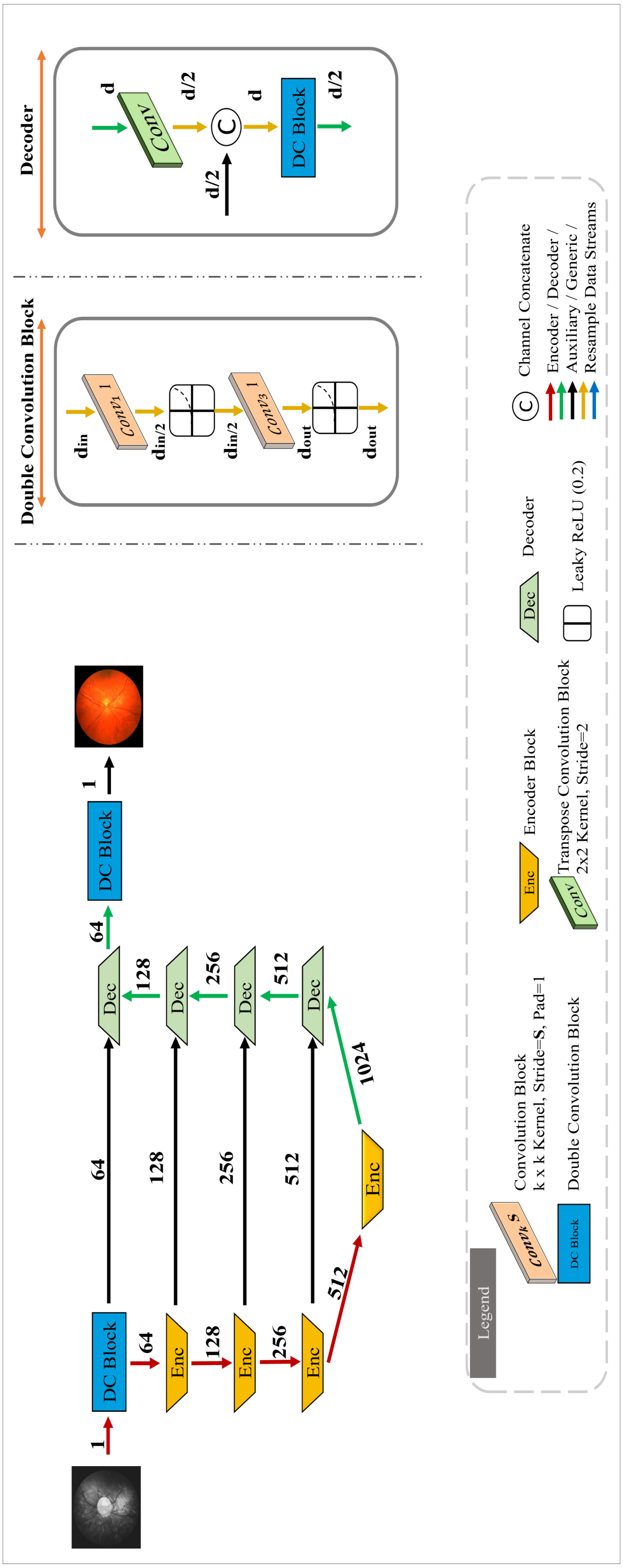


Figure 2.14: An illustration of the U-Net backbone based Colorization model architecture.

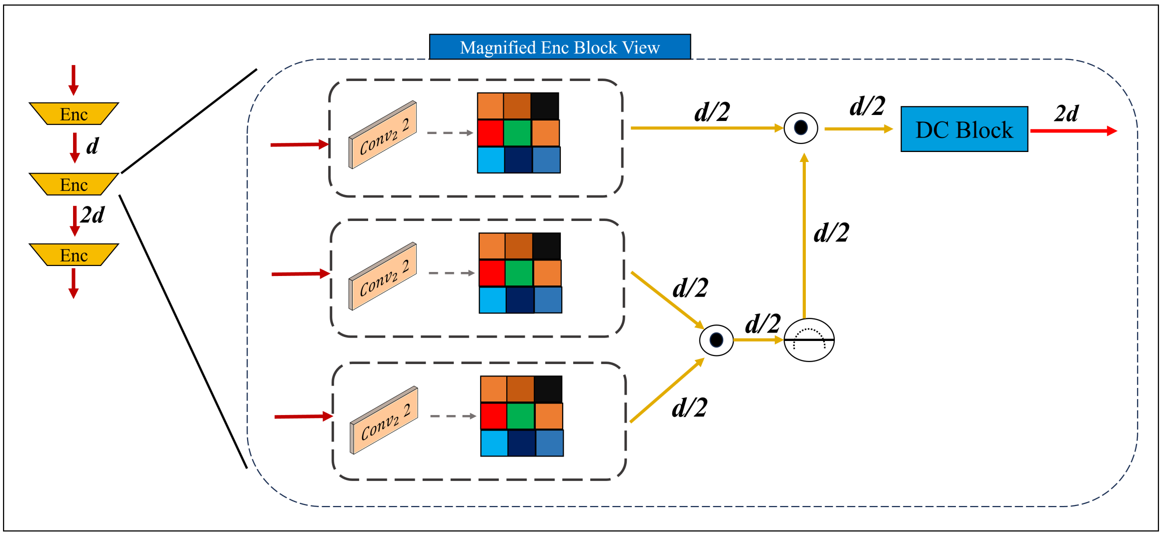


Figure 2.15: An illustration of the Encoder block used in the design of the colorization model.

Table 2.3: Performance Comparison on DRIVE dataset

Methods	**Accuracy**	**Specificity**	**Sensitivity**
Jiang et al. [113]	0.891	0.900	0.830
Staal et al. [29]	0.944	0.977	0.719
Ricci et al. [114]	0.959	0.972	0.775
Roychowdhury et al. [115]	0.949	0.978	0.739
Liskowski et al. [108]	0.925	0.924	0.916
Yan et al. [56]	0.954	0.981	0.765
CE-Net [116]	0.954	0.975	0.821
IterNet [117]	0.957	0.983	0.779
HAnet [118]	0.958	0.983	0.797
Genetic U-Net [119]	0.957	0.976	0.830
TP-Net [120]	0.963	0.976	0.875
Vessel Curvature	0.940	0.947	**0.917**
Vessel Curvature v2.0	**0.960**	**0.980**	0.800

Table 2.4: Performance Comparison on STARE dataset

Methods	**Accuracy**	**Specificity**	**Sensitivity**
Jiang et al. [113]	0.901	0.900	0.857
Staal et al. [29]	0.952	0.981	0.697
Ricci et al. [114]	0.965	0.939	0.903
Roychowdhury et al. [115]	0.956	0.984	0.732
Liskowski et al. [108]	0.931	0.930	0.931
Yan et al. [56]	0.961	0.985	0.758
CE-Net [116]	0.966	0.984	0.829
IterNet [117]	0.970	0.988	0.771
HAnet [118]	0.967	0.984	0.818
Genetic U-Net [119]	0.971	0.984	0.865
TP-Net [120]	0.972	0.982	0.885
Vessel Curvature	0.915	0.920	**0.911**
Vessel Curvature v2.0	**0.960**	**0.980**	0.810

Sensitivity can be attributed to a similarity of data space between the original images and the augmented images. To address the shortcoming of the approach in terms of data variability, the Lambda-Coherence Measure-guided Cartesian-Square (λCMgC^2) Augmentation is demonstrated in the following section.

2.4 Lambda-Coherence Measure-guided Cartesian-Square (λCMgC^2) Augmentation

Conventional data augmentation techniques such as scaling, rotation, and flipping are ineffective for small-scale datasets [123]. To mitigate the issue, effective data augmentation strategy had been proposed in [124], which generates an augmented image by combining training samples. However, it is challenging to directly adopt this model in medical imaging, specifically for fundus images, which requires a strict adherence to structural coherence. Thus, the Lambda-Coherence Measure-guided Cartesian-square ($\lambda CMgC^2$) model for fundus image augmentation is better suited for contemporary approaches.

2.4.1 Problem Formulation

Let $\mathbb{I}_G : \mathbb{I} \rightarrow \Re^2$ be the green channel image plane of $\mathbb{I}$. Each pixel $p_{x,y,v}$ in an image plane($\subset \Re^2$) is characterized by a triplet $<x, y, v>$, where the index pair $(x, y), x = 1, 2, \ldots, h$ and $y = 1, 2, \ldots, w$, represents the pixel location in the image domain Θ, $v(\in \Re^+)$ signifies the pixel value, and $h, w(\in \Re^+)$ are the height and width of the image. Let $\mathbb{I}_G^m$ and $\mathbb{I}_G^n$, $m \neq n$, be two retinal fundus images from a dataset $\mathbb{D}$, and

$\mathbb{I}_{gt}^m, \mathbb{I}_{gt}^n$ are their respective binary pixel-wise ground-truth images, where 0 denotes non-mask pixels and 1 denotes mask pixels. A standard notation of $q_{x,y,0}$ for all non-mask pixels and $q_{x,y,1}$ for all mask pixels in any binary image is used. Inspired from [124], the augmented image $\mathbb{I}^{aug}$ can be defined as the output of an injective function $g(.)$ on a subset of the Cartesian Product on dataset $\mathbb{D}$ as,

$$\mathbb{I}^{aug} = g(\{\mathbb{I}_G^m, \mathbb{I}_G^n\}) \mid \{\mathbb{I}_G^m, \mathbb{I}_G^n\} \subset \mathbb{D}^2 \tag{2.4.1}$$

In Equation 2.4.1, $g : \mathbb{D}^2 \rightarrow \mathfrak{R}^2$ maps the Cartesian-Square set $\mathbb{D}^2$ to the domain of green-channel image plane. The injective function is used to combine the two images $\mathbb{I}_G^m, \mathbb{I}_G^n$ into a single structurally sound fundus image. In most retinal image datasets, it is often times difficult to define such an injective function $g(.)$ while preserving structural anatomy of a retinal fundus image. For example, as shown in Figure 2.16, using a simple linear function $g(\{\mathbb{I}_G^m, \mathbb{I}_G^n\}) = 0.5.\mathbb{I}_G^m + 0.5.\mathbb{I}_G^n$ on the DRIVE [112] dataset can lead to the formation of medically incoherent and impossible images.

Lambda-Coherence Measure-guided Cartesian-Square Sampling (λCMgC^2) augmentation model performs superior in such scenarios to generate fundus images. The Coherence measure denoted by $\widehat{C}$, evaluates the structural coherence between two input images before generating augmented image using the Injective function $g(.)$. Formally, we define $\widehat{C}(\mathbb{I}_G^m, \mathbb{I}_G^n)$ as the degree of positional overlap of the optic disk regions, $\mathbb{I}_{OD}^m$ and $\mathbb{I}_{OD}^n$, of images $\mathbb{I}_G^m$ and $\mathbb{I}_G^n$ respectively. The λCMgC^2 is thus primarily a two-stepped model. Firstly, the OD region from both selected fundus images $\mathbb{I}_G^m$ and $\mathbb{I}_G^n$ and computed the coherence ($\widehat{C}$) between them. Subsequently, the augmented images using injective function $g(.)$ are generated after the optic disc alignment.

2.4.2 Optic Disc Region Isolation

Definition 1 (Optic Disc Region)**.** *Optic disc region* $\mathbb{I}_{OD}(\subseteq \mathbb{I}_G)$ *in a retinal fundus image* $\mathbb{I}_G$ *is a connected and cohesive region such that* $\frac{1}{N_{OD}} \sum\{v|p_{x,y,v} \in \mathbb{I}_{OD}\} > \frac{1}{N-N_{OD}} \sum\{v|p_{x,y,v} \in (\mathbb{I}_G - \mathbb{I}_{OD}))\}$*, where* N_{OD} *represents the number of pixels in* $\mathbb{I}_{OD}$ *and* $N = h \times w$*.*

Here, OD region of a fundus image is considered as a marker to measure structural coherence between a pair of images. Thus, identification of OD regions in fundus images is a primary step to the λCMgC^2 method. An iterative method is used to quantitatively depict and identify the OD region from a retinal fundus image which is discussed as follows. Firstly, vessel regions are removed from the pair of images, i.e., $\mathbb{I}_G^m$ and $\mathbb{I}_G^n$, using Equation 2.4.2.

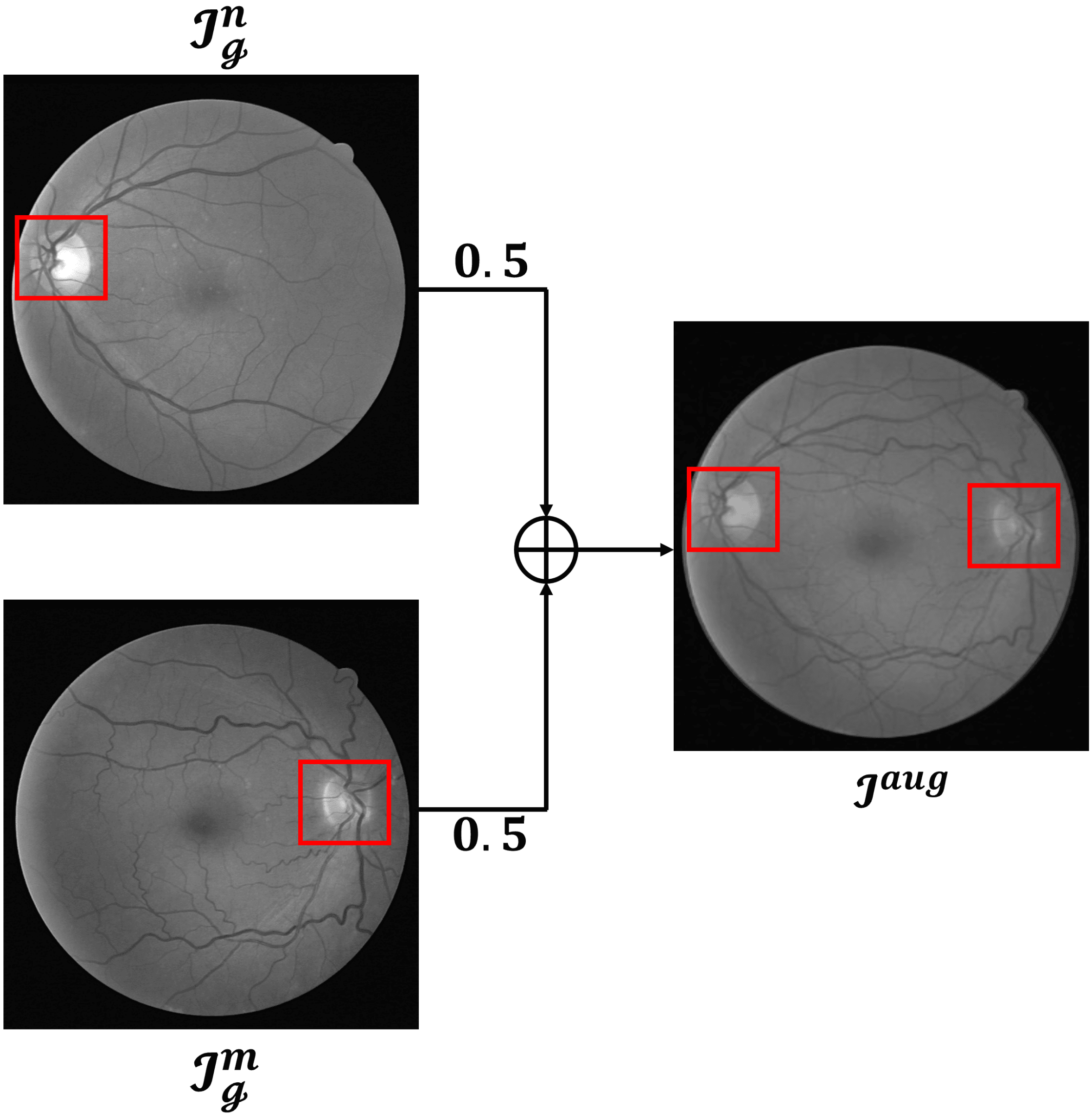


Figure 2.16: Retinal fundus images having different optic disc position (red square), and corresponding Linear Injective Function $g(\{\mathbb{I}_G^m, \mathbb{I}_G^n\})$ augmented image having two optic discs.

$$\mathbb{I}_{VR}^{m} = \{p_{x,y,v} | \forall (x,y), (p_{x,y,v} \in \mathbb{I}_{G}^{m}) \wedge (q_{x,y,0} \in \mathbb{I}_{gt}^{m})\} \tag{2.4.2}$$

In Equation 2.4.2, $\mathbb{I}_{VR}^{m}$ signifies the vessels removed set of pixels of $\mathbb{I}_{G}^{m}$ and $\mathbb{I}_{gt}^{m}$ is the ground truth of $\mathbb{I}_{G}^{m}$. Similar method is used to generate $\mathbb{I}_{VR}^{n}$. It is noted that vessel pixels are disregarded when computing $\widehat{C}$ as OD coherence is measured on the region as a whole which is mostly concerned with background region rather than vessel pixels. For simplicity, $\mathbb{I}_{VR}^{m}$ and $\mathbb{I}_{VR}^{n}$ are represented together as $\mathbb{I}_{VR}^{\{l|l=m,n\}}$. Similar naming convention is also used for $\mathbb{I}_{G}^{m}$ and $\mathbb{I}_{G}^{n}$, i.e., $\mathbb{I}_{G}^{\{l|l=m,n\}}$. It is assumed that $\mathbb{I}_{B}^{\{l|l=m,n\}}$ and $\mathbb{I}_{OD}^{\{l|l=m,n\}}$ are two sets representing background region and optic disc region respectively. At the beginning, $\mathbb{I}_{B}^{\{l|l=m,n\}}$ and $\mathbb{I}_{OD}^{\{l|l=m,n\}}$ are initialized as empty set, i.e., $\mathbb{I}_{B}^{\{l|l=m,n\}} = \Phi$ and $\mathbb{I}_{OD}^{\{l|l=m,n\}} = \Phi$. Now, at every iteration, the pixel having minimum intensity is computed from $\mathbb{I}_{VR}^{\{l|l=m,n\}}$ for further processing as,

$$\hat{m} = min(\{v | \forall p_{x,y,v} \in (\mathbb{I}_{VR}^{\{l|l=m,n\}} - \mathbb{I}_{B}^{\{l|l=m,n\}})\}) \tag{2.4.3}$$

Subsequently, based on the value of $\hat{m}$, all pixels in $\mathbb{I}_{VR}^{\{l|l=m,n\}}$ are partitioned into two groups, namely background pixel set $\mathbb{I}_{B}^{\{l|l=m,n\}}$ and OD pixel set $\mathbb{I}_{OD}^{\{l|l=m,n\}}$, using Equation 2.4.4 and 2.4.5 in order, as

$$\mathbb{I}_{B}^{\{l|l=m,n\}} = \{p_{x,y,v} | \forall p_{x,y,v} \in \mathbb{I}_{VR}^{\{l|l=m,n\}}, v \leqslant \hat{m}\} \tag{2.4.4}$$

$$\mathbb{I}_{OD}^{\{l|l=m,n\}} = \{p_{x,y,v} | \forall p_{x,y,v} \in \{\mathbb{I}_{VR}^{\{l|l=m,n\}} - \mathbb{I}_{B}^{\{l|l=m,n\}}\}\} \tag{2.4.5}$$

After which, the average intensity of pixel elements in $\mathbb{I}_{B}^{\{l|l=m,n\}}$ is computed and compared with the average intensity of $\mathbb{I}_{OD}^{\{l|l=m,n\}}$, which can be defined in Equation 2.4.6.

$$\left(\frac{\sum_{x=1,y=1}^{h,w} \{v | \forall p_{x,y,v} \in \mathbb{I}_{OD}^{\{l|l=m,n\}}\}}{N_{OD}} - \frac{\sum_{x=1,y=1}^{h,w} \{v | \forall p_{x,y,v} \in \mathbb{I}_{B}^{\{l|l=m,n\}}\}}{N_{B}} \right) > t \tag{2.4.6}$$

In Equation 2.4.6, N_{OD} and N_B signify the number of pixels in $\mathbb{I}_{OD}^{\{l|l=m,n\}}$ and $\mathbb{I}_{B}^{\{l|l=m,n\}}$ respectively, and t is a threshold whose value is selected empirically set as 150. If the difference is less than t, the process will stop, and the resulting set $\mathbb{I}_{OD}^{\{l|l=m,n\}}$ will only contain the approximate OD region pixels. Otherwise, the earlier steps are repeated till conclusion of the algorithm (as shown in Figure 2.17). After the completion, the set $\mathbb{I}_{OD}^{\{l|l=m,n\}}$ is treated as binary mask set for OD region, denoted by $q_{x,y,0}, q_{x,y,1}$ notations.

2.4.3 Coherence Measure

After OD region identification, the coherence measure $\widehat{C}(\mathbb{I}_G^m, \mathbb{I}_G^n)$ can be defined formally as:

$$\widehat{C}(\mathbb{I}_G^m, \mathbb{I}_G^n) = n(\{q_{x,y,1} \mid \forall(x,y), q_{x,y,1} \in (\mathbb{I}_{OD}^m \cap \mathbb{I}_{OD}^n)\}) \tag{2.4.7}$$

In Equation 2.4.7, $n(.)$ is the cardinality of a set. It is noted that the coherence $\widehat{C}$ needs to be measured until the OD region set $\mathbb{I}_{OD}^n$ of $\mathbb{I}_G^n$, has maximal intersection with the OD region set $\mathbb{I}_{OD}^m$ of $\mathbb{I}_G^m$. A rotation operation is thus performed on the respective green channel images $\mathbb{I}_G^m, \mathbb{I}_G^n$ until $\widehat{C}$ is maximized. For simplicity, the image $\mathbb{I}_G^m$ is kept unchanged and rotated $\mathbb{I}_G^n$ through an angle of 10^o at every iteration and subsequently measured the coherence.

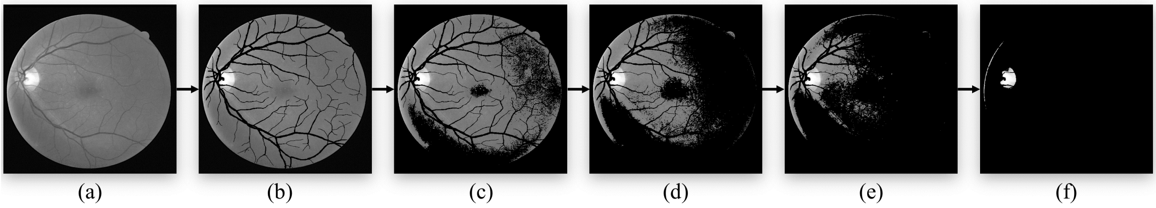


Figure 2.17: Computation of the OD region for $t = 150$: (a) green channel image, (b) vessel removed image using eq. (3), (c-f) represents OD region based on the computation of eq. (7) having output values 109.04, 83.15, 126.13, and 153.78.

Let $\mathcal{A}_{\widehat{C}} = \{(\widehat{C}(\mathbb{I}_G^m, \widehat{\mathbb{I}_G^n}), \theta) | \forall\theta, \theta \in [0^o, 360^o], \widehat{\mathbb{I}_G^n} = ROTATE(\mathbb{I}_G^n, \theta)\}$ be a set comprising of elements of the pair form $< \widehat{C}, \theta >$. Here, $\widehat{C}$ denotes the Coherence Measure computed between images $\mathbb{I}_G^m$ and output of function $ROTATE(\mathbb{I}_G^n, \theta)$. $ROTATE(., \theta)$ denotes a function which rotates an input Image by θ^o and returns the output image respectively. Finally, the element pair $< \widehat{C}, \theta >$ is selected from $\mathcal{A}_{\widehat{C}}$ for which $\widehat{C}$ is highest and rotate the image $\mathbb{I}_G^n$ by the corresponding pair value of θ, as shown in Equation 2.4.8.

$$\theta = arg_\theta(max_{\widehat{C}}(\mathcal{A}_{\widehat{C}})) \tag{2.4.8}$$

Let $\widehat{\mathbb{I}_G^n}$ denote the final θ^o rotated image of $\mathbb{I}_G^n$. It is noted that the ground-truth $\mathbb{I}_{gt}^n$ of $\mathbb{I}_G^n$ is also rotated with the same angle of θ^o and denoted as $\widehat{\mathbb{I}_{gt}^n}$. Finally, the λ-weighted linear injective function $g(.)$ is defined to map the set $\{\mathbb{I}_G^m, \mathbb{I}_G^n\}$ to a green-channel image plane to generate the augmented image $\mathbb{I}^{aug}$ along with ground truth $\mathbb{I}_{gt}^{aug}$ as

$$\begin{aligned} \mathbb{I}^{aug} &= \lambda \mathbb{I}_G^m + (1-\lambda)\widehat{\mathbb{I}}_G^n \\ \mathbb{I}_{gt}^{aug} &= \mathbb{I}_{gt}^m \cup \widehat{\mathbb{I}}_{gt}^n \end{aligned} \tag{2.4.9}$$

In Equation 2.4.9, $\lambda \in (0, 1)$. Figure 2.18 shows the image pairs and corresponding augmented image and respective ground-truth.

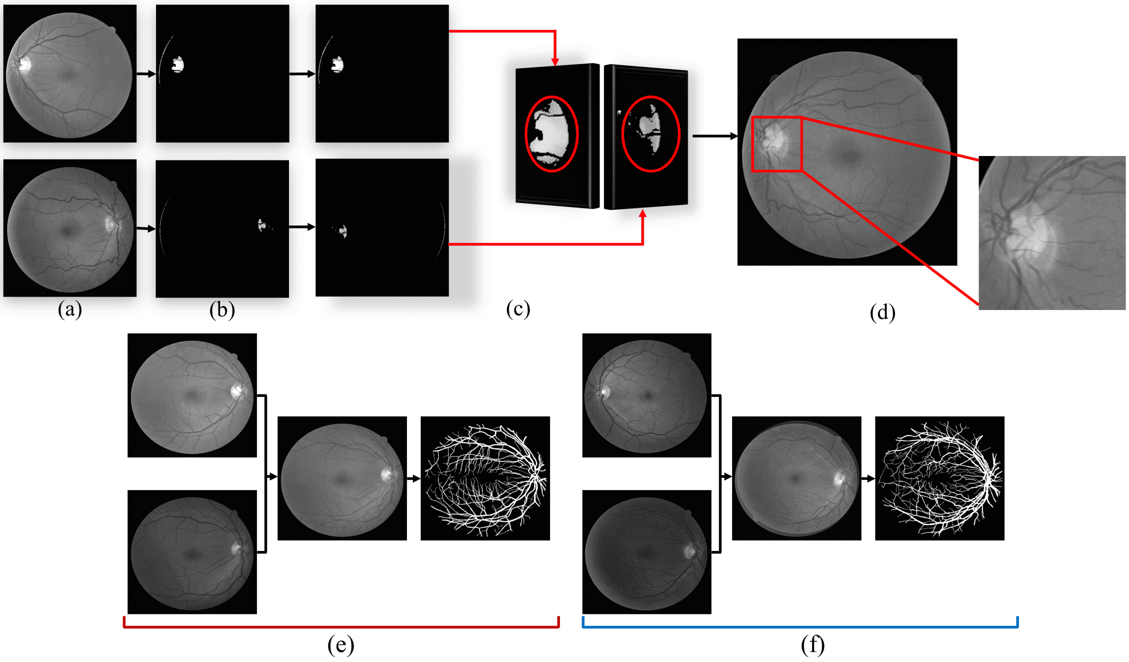


Figure 2.18: Computation of the augmented image: (a) green channel images, (b) isolation of the OD region, (c) OD alignment with $\widehat{C}$ equal to 98.43%, (d) $\lambda CMgC^2$ image with enlarge optic disc region, and (e-f) image pairs and corresponding augmented image and respective ground-truth.

2.4.4 Proof of Structural Integrity in Augmented Image

In this regard, to prove the structural integrity of the augmented image $\mathcal{I}^{aug}$, the following Theorem is needed.

Theorem 1. *Given* λ*-weighted linear injective function* $g : \mathbb{D}^2 \rightarrow \Re^2$*, any vessel region perceptible in* $\mathbb{I}_G^{\{l|l=m,n\}}$ *remains perceptible in* $\mathbb{I}^{aug}$*.*

Proof. Let $\mathbb{I}_{VR}^{aug}$ be defined as a vessel removed augmented image. Let neighborhood intensity difference set be defined as $N_8^-(p_{x,y,v}) = \{\tilde{v} - v \mid \forall p_{\tilde{x},\tilde{y},\tilde{v}} \in N_8(p_{x,y,v})\}$ for any pixel $p_{x,y,v}$, where N_8 is the 8-neighborhood of a pixel. Thus, from this formulation, it is trivial to state that for any pixel $p_{x,y,v} \in (\mathbb{I}^{aug} - \mathbb{I}_{VR}^{aug})$

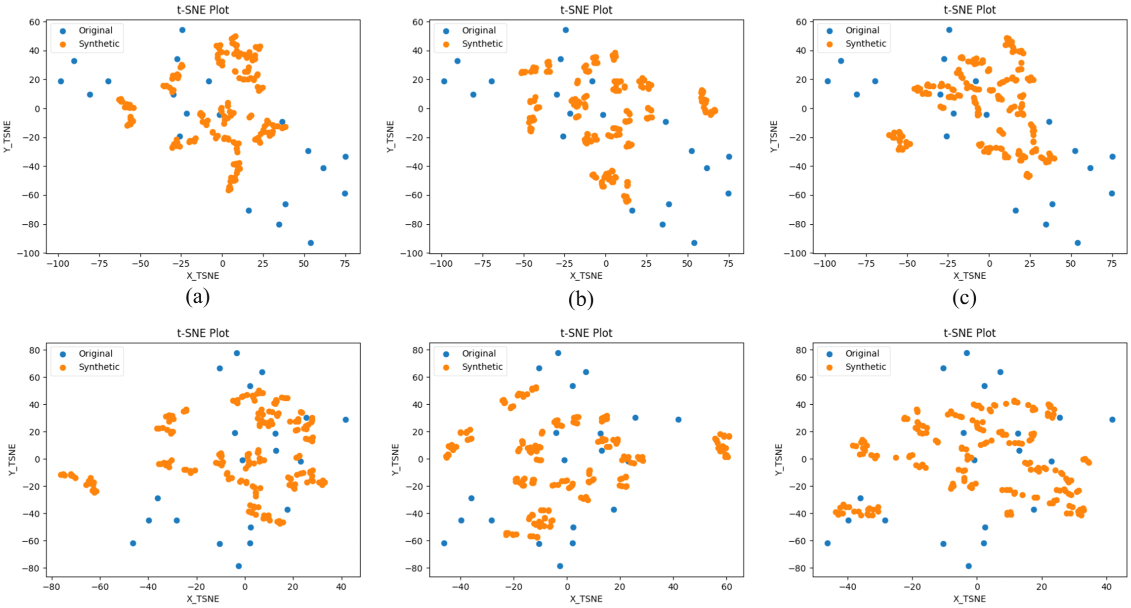
$(N_8^-(p_{x,y,v}) \subset \mathcal{O}) \implies (\nexists q_{\tilde{x},\tilde{y},0} \in \mathbb{I}_{gt}^{aug})$, where $\tilde{x} = x \pm 1, \tilde{y} = y \pm 1$ and $\mathcal{O}$ represents set of infinite zero elements.

Let a hypothetical pixel $p_{x,y,v} \in (\mathbb{I}^{aug} - \mathbb{I}_{VR}^{aug})$ be considered such that $N_8^-(p_{x,y,v}) \subset \mathcal{O}$ and $\exists q_{\tilde{x},\tilde{y},0} \in \mathbb{I}_{gt}^{aug}$. That implies $\exists p_{\tilde{x},\tilde{y},\tilde{v}} \in N_8(p_{x,y,v})$ such that $\tilde{v} = v$. We consider $\tilde{v}^{\{l|l=m,n\}}, v^{\{l|l=m,n\}}$ as pixel values at $(\tilde{x}, \tilde{y})$ and (x, y) in $\mathbb{I}_G^{\{l|l=m,n\}}$ respectively. It follows thus from the formulation of the associated injective function $g(.)$,$[\lambda.\tilde{v}^m + (1-\lambda).\tilde{v}^n] - [\lambda.v^m + (1-\lambda).v^n] = 0$. Rearranging, $\lambda.[\tilde{v}^m - v^m] + (1-\lambda).[\tilde{v}^n - v^n] = 0$ is obtained. As the theorem is applicable for all $0 < \lambda < 1$, it is concluded that $\tilde{v}^m = v^m$ and $\tilde{v}^n = v^n$ respectively, thus implicating that the respective background pixel and vessel pixel in images $\mathbb{I}_G^m, \mathbb{I}_G^n$ are imperceptible. This is a fallacy, as otherwise the original annotations would not have been possible due to imperceptibility. Hence all vessel pixels perceptible in $\mathbb{I}_G^m, \mathbb{I}_G^n$ remain perceptible in $\mathbb{I}^{aug}$ and the theorem is proven by law of contradiction. $\square$

2.4.5 Experimental Results and Discussions

The ability of the $\lambda CMgC^2$ method in generating structurally coherent fundus images is evaluated. Firstly, the $\lambda CMgC^2$ model is evaluated using t-distributed stochastic neighbor embedding (t-SNE) and principal component analysis (PCA) plots. A robust visualization of real images and $\lambda CMgC^2$ images is plotted for the DRIVE and STARE datasets (Figure 2.19-2.20). Three sets of 200 unique synthetic images have been selected from the $\lambda CMgC^2$ generated synthetic image dataset for visual comparison with real images respectively. All synthetic images have been selected from $\lambda CMgC^2$ dataset having $\lambda \in 0.4, 0.5, 0.6$ to demonstrate better visualization between real images and synthetic images which are most dissimilar to original datasets. From Figure 2.19 and 2.20, it can be seen that the generated images closely match the distribution of original images. Most of the synthetic images are positioned close to real images, however, Figure 2.19 depicts that the synthetic image distribution is not centered in one region. This demonstrates the proposed method's ability to generate realistic data even with limited training data.

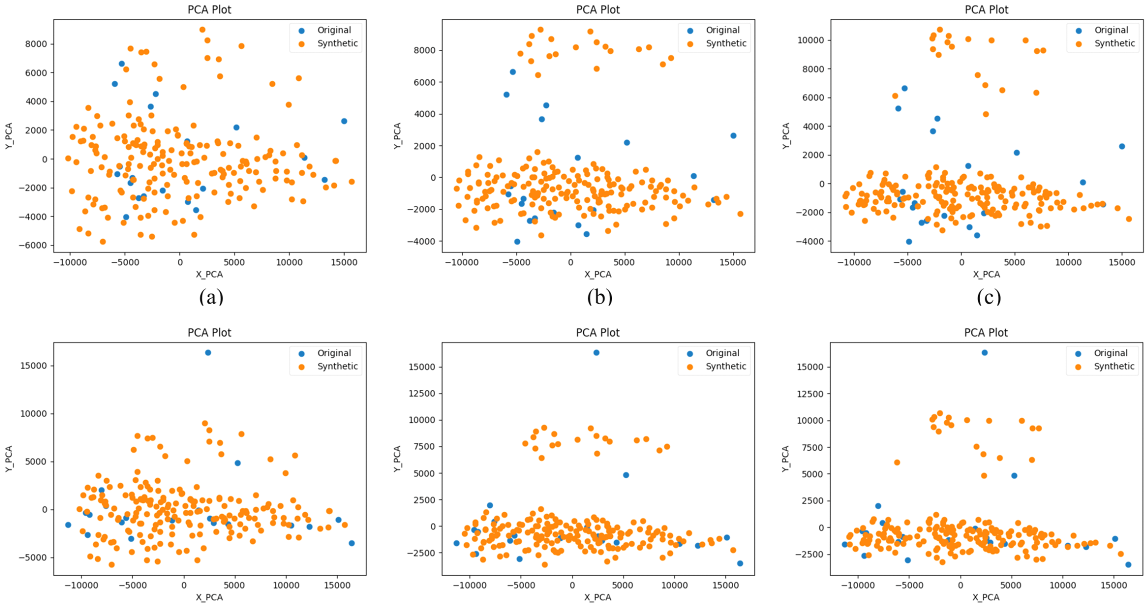
Furthermore, two commonly used metrics: Frechet Inception Distance (FID) and Multi-scale Structural Similarity Index Measure (MS-SSIM) [121] are computed to assess the discriminative features of the $\lambda CMgC^2$-generated images in comparison to real images. FID is highly correlated with image visual quality. A low FID score is usually indicative of the visual authenticity of synthetic data. In addition, MS-SSIM scores of 600 randomly selected pairs of synthetic images are computed to assess the diversity of the images in the real and synthetic datasets. The higher mean MS-SSIM score correlates with higher similarity between the real and synthetic datasets. An



(a) (b) (c)

(d) (e) (f)

Figure 2.19: t-SNE plots to illustrate the distribution of real and $\lambda CMgC^2$-generated images: (a-c) for DRIVE dataset, and (d-f) for STARE dataset.



(a) (b) (c)

(d) (e) (f)

Figure 2.20: PCA plots to illustrate the distribution of real and $\lambda CMgC^2$-generated images: (a-c) for DRIVE dataset, and (d-f) for STARE dataset

overall FID score of 1.5418 is obtained compared to a score of 3.5583 for traditional augmentation (TA) [16]. Similarly, an MS-SSIM score of 0.6948 is obtained compared to 0.6119 for traditional augmentation approaches. The lower FID scores of $\lambda CMgC^2$ indicate that the images it generates are of higher quality when compared to the TA method. In addition, the $\lambda CMgC^2$ method produced more coherent images with respect to real dataset than TA method due to the higher average MS-SSIM value. In medical imaging, it is important to not only generate new images but also medically sound and coherent image. Thus, a balance in similarity with existing data is more important along with diversity in medical domain which is achieved by the model.

Quantitative analysis is performed in two stages. Firstly, the consistency of the $\lambda CMgC^2$ vessel segmentation results is evaluated with the ground-truths annotated by experts. Subsequently, the augmentation model is trained and tested on two baseline networks having similar architectures, viz., the U-Net [125], and the FC-Densenet [126]. Four performance measures have been considered for evaluation, namely, sensitivity (Se), specificity (Sp), and accuracy (Ac) [127]. For a better comparative analysis, the baseline model performances have been evaluated on both Traditional Augmentation [16] as well as $\lambda CMgC^2$ Augmentation (Table-2.5).

Table 2.5: Comparison of the proposed method with the baseline models.

		Traditional			$\lambda CMgC^2$		
DB	**Methods**	**Se**	**Sp**	**Ac**	**Se**	**Sp**	**Ac**
[112]	FC-DNet [126]	0.801	0.974	0.960	0.826	0.976	0.965
	U-Net [125]	0.790	0.972	0.959	0.812	0.976	0.962
[30]	FC-DNet [126]	0.799	0.980	0.970	0.830	0.981	0.973
	U-Net [125]	0.774	0.978	0.965	0.839	0.979	0.970
[128]	FC-DNet [126]	0.810	0.979	0.962	0.835	0.982	0.966
	U-Net [125]	0.799	0.982	0.965	0.841	0.985	0.972
[129]	FC-DNet [126]	0.799	0.981	0.965	0.811	0.981	0.969
	U-Net [125]	0.784	0.979	0.964	0.815	0.980	0.968

A significant improvement in baseline model performance is observed when trained upon $\lambda CMgC^2$ augmentation. This empirically validates the importance of the $\lambda CMgC^2$-based training paradigm.

2.5 Summary

This chapter delved into the function and effect of data augmentation methods in improving the performance of deep learning models in fundus image analysis. Since there is limited access to and variability of annotated medical fundus images, data augmentation is essential in model generalization improvement, overfitting reduction, and modeling real-world imaging variability.

Three strategies of augmentation were studied in detail, viz., *Vessel Curvature-based Fundus Image Augmentation*, *Vessel Curvature-based Fundus Image Augmentation v2.0* and *Lambda-Coherence Measure-guided Cartesian-Square* ($\lambda CMgC^2$) *Augmentation. Vessel Curvature-based Fundus Image Augmentation* and *Vessel Curvature-based Fundus Image Augmentation v2.0* dealt with conventional geometric and structural transformations. The augmentations were used to artificially increase the dataset and make the model aware of variations usually seen in real-world clinical practice. This approach was straightforward yet effective in enhancing robustness without compromising the anatomical integrity of retinal structures.

Lambda-Coherence Measure-guided Cartesian-Square ($\lambda CMgC^2$) *Augmentation*, however, utilized a more sophisticated approach utilizing domain-specific augmentations that combined two existing fundus images into a single fundus image with aligned optic discs. This approach sought to increase not just the volume but also diversity and pathological integrity of the training data.

All augmentation approaches demonstrate improvement in segmentation accuracy and diagnostic validity with $\lambda\mathrm{CMg}C^2$ Augmentation empirically evaluated to be superior. The chapter emphasizes that the augmentation should be clinically relevant, and that it is possible to achieve an efficient training pipeline.

Chapter 3

Retinal Blood Vessel Segmentation from Fundus Images

3.1 Introduction

The retina is a light-sensitive tissue layer at the posterior section of the eye that is responsible for visual perception [130]. Among the innumerable anatomical structures, the retinal blood vessel network contains essential diagnostic information, essential for detecting ocular and systemic illness [131]. In this regard, Fundus imaging, a two-dimensional representation of the retina, is utilized extensively in ophthalmology because it is non-invasive and highly diagnostic. In these fundus images, retinal blood vessel segmentation — the identification and outlining of the vascular structures — has become a key task in medical image analysis.

Segmentation of the retinal vasculature enables clinicians and computer systems to examine the morphology, distribution, and structural integrity of blood vessels, which are key biomarkers for ophthalmic diseases [131]. In addition, new developments in artificial intelligence have placed an even greater emphasis on the importance of accurate segmentation in constructing robust diagnostic and screening devices [25]. The importance of retinal blood vessel segmentation in fundus images and its wider implications for clinical diagnosis, public health, and technological development is discussed henceforth.

3.1.1 Clinical Significance of Retinal Vessel Analysis

One of the most important reasons for segmenting blood vessels in fundus images is the early diagnosis and tracking of retinal diseases which present through typical alterations in the vasculature. These are microaneurysms [132], neovascularization [133], hemorrhages [134], and vascular tortuosity [135] — all of which can be quantified more accurately once the vessels are properly segmented. Additionally, the retinal vasculature is a non-invasive biomarker for systemic diseases such as hypertension [136], diabetes [137], and even cardiovascular diseases [138]. Alterations in the caliber [139], branching pattern [140], and/or density of retinal vessels [141] usually occur before symptoms or observable signs in other organs. For instance, narrowing of arterioles [142] and elevated arteriolar-to-venular ratio [142] may indicate hypertension. Abnormalities of vessel walls and leakage imply diabetic changes [143]. Retinal vessel segmentation, therefore, aids systemic health monitoring via routine eye exams. Segmented retinal vessels can facilitate treatment planning, e.g., in laser photocoagulation [144] or intravitreal injections [145], by displaying a precise map of vessel positions and pathologies. In addition, monitoring throughout the time duration of disease helps evaluate therapeutic responses in conditions such as Diabetic Retinopathy [146].

3.1.2 Automated Screening in Computer-Aided Diagnosis (CAD) Systems

As the retinal diseases, particularly in elderly and low-resource populations, continue to progress, automated screening systems are becoming increasingly important [147]. In this regard, blood vessel segmentation serves as a central building block in such systems, providing critical features to machine learning models for classification or anomaly detection. Also, segmentation maps of blood vessels in deep learning models can serve as attention masks or features to enhance the performance of disease classifiers [148]. The models are aided by explicit anatomical data, resulting in more interpretability and robustness. Vessels segmented can also be utilized in cross-modality image registration, where Optical Coherence Tomography (OCT) or fluorescein angiography images are aligned with fundus images [149]. It facilitates multi-modal analysis, which is essential in advanced diagnostic systems. Thus, automatic segmentation of blood vessels in fundus images is necessary for quick and effective detection of pathological cases. The most recent trend found in solving this problem is to apply Convolutional Neural Networks (CNNs) under the framework of Deep Learning(DL) [108].

3.2 Literature Review

The vessel segmentation algorithms can be broadly categorized into unsupervised and supervised algorithms. The unsupervised algorithms comprise of techniques such as line detectors [150], matched filtering [151], morphological transformations [152–154], model-based methods [78, 155, 156] and multi-scale segmentation methods [32, 115, 157, 158]. However, these algorithms are often computationally complex [159] and perform poorly on images with pathology.

Supervised algorithms, in contrast, rely on the use of classifiers that learn to segment from labeled ground truth images. Algorithms such as the k-Nearest Neighbor [160], Gaussian Mixture Model (GMM) [161], Support Vector Machine (SVM) [114], Neural Networks [162], Decision Trees [31], AdaBoost [163] and Conditional Random Fields [164] have been proposed in the literature as classifiers for vessel and non-vessel pixels. In 2016, Liskowski et al. [108] published a paper highlighting the use of Convolutional Neural Networks (CNNs) for automated self-learning of features and classification using ground truth images. DL based approach for blood vessel segmentation in Fundus images is further pursued by Yan et al. [56]. In the corresponding work, they introduced a joint loss function for the CNN by taking care of traditional pixel wise losses treating all vessel pixels, thick or thin, with equal importance and the segment level loss giving more importance on thin vessel pixels compared to thick ones. Table 3.1 highlights a brief outline of the state-of-the-work deep learning based approaches in the domain of retinal blood vessel segmentation.

Table 3.1: Overview of the State Of The Art Deep Learning Methodologies (SOTA) for Retinal Vessel (RV) Segmentation Methods.

Authors	**Methodology**	**Backbone**	**Loss**
Z.Yan et al. (2018) [56]	Segment-level loss emphasizing more on vessel thickness	Fully Convolutional +Skip-Connection	Segment Loss +Pixel-Loss
Q. Jin et al. (2019) [165]	U-Net shaped architecture with deformable convolution kernel	U-Net	Softmax Loss
Z. Gu et al. (2019) [166]	Encoder-Decoder network with a context extractor module.	U-Net	Dice Loss+ Regularization Loss
L. Li et al. (2020) [167]	Multiple iterations of separated miniaturized U-Net architectures	U-Net	Sigmoid-Cross Entropy Loss
D. Wang et al. (2020) [168]	End-to-End Deep Attention Network with dual decoder for Hard Region segmentation	U-Net	Binary Cross Entropy + Jaccard Loss
Y. Wu et al. (2020) [169]	Inter-network skip connection-based multi-scale deep learning model.	Fully Convolutional +Skip-Connection	Binary Cross Entropy Loss
Z. Lin et al. (2021) [170]	A multi-path scale based High-resolution representation network	Fully Convolutional + Skip-Connection	Hard-focused Cross Entropy Loss
J. Wei et al. (2021) [171]	Genetic algorithm inspired search space optimized U-shaped CNN model	U-Net	Focal Loss
H. Wu et al. (2021) [172]	Scale-aware feature aggregation model with dynamic receptive field	Fully Convolutional +Skip-Connection	Binary Dice Loss
Y. Li et al. (2022) [173]	Combined the Global Transformer model with a dual attention model for segmentation	U-Net +Transformer	Cross Entropy Loss
J. Li et al. (2023) [55]	Dual-path progressive fusion network	Convolutional Model	Binary Cross Entropy Loss

However, despite significant advancements in both model architecture based designs [55, 166, 167, 169–171, 173] as well as training paradigms [56, 165, 174], several challenges remain unaddressed [175]. This includes the automated analysis of vessel structure complexity, image intensity variations, marginal contrast difference between vessels and the background region, and limited availability of blood-vessel annotated datasets for training and evaluation. In the subsequent sections, three different approaches of retinal blood vessel segmentation is demonstrated and an empirical evaluation is conducted to comparatively assess the respective performances.

3.3 Multi-Scale Layer Pixel-based Segmentation Network

Multi-Scale Layer Pixel-based Segmentation Network is inspired from the work of Chaudhuri et al. [176] and Hart et al. [177]. A multi-scale patch size is adopted for each image region in order to better isolate the blood vessel pixels in the image. Since, it is difficult to select or incorporate the multi-scale patch sizes for each region, a convolutional based neural network model is used to combine the different patch sizes into a single feature vector which is finally classified into vessel or non-vessel pixels using a softmax fully connected ANN classifier.

3.3.1 Patch Transformation

A broad outline of the proposed approach is highlighted in Figure 3.1. The first stage of the approach involves the extraction of the three different patch sizes for each pixel present in the Fundus image. Based on the principle highlighted by Hart et al., the green channel of the image is only considered for the algorithm. As reported by Hart et al., the green channel of the Fundus image provides a better contrast between blood vessel pixels and non-blood vessel pixels than other channels or a grayscale image. From the green channel image, three different patch sizes are extracted for each pixel. The patch sizes selected for this work are 9×9, 11×11 and 13×13. The patch sizes are selected based on the knowledge that the blood vessel thickness in the Fundus images usually varies between 1 to 22 pixels. Thus the average pixel width is computed to be approximately 11 pixels. Thus, the three patch sizes are identified for the work. After the extraction of the patches, the system applies different Convolutional Neural Networks on each one of the patches followed by a flatten layer to transform the patches into feature vectors.

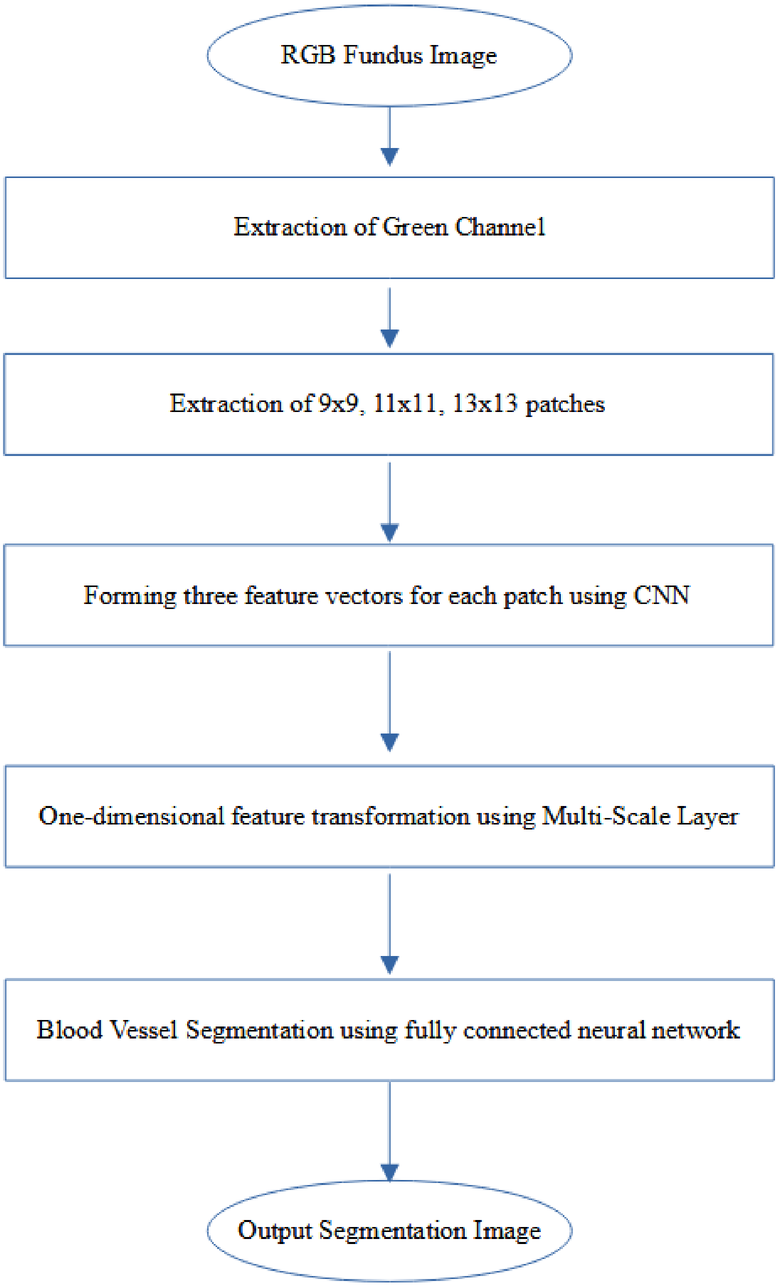


Figure 3.1: An illustration of the flowchart depicting the *Multi-Scale Layer Pixel-based Segmentation Network*

3.3.2 Convolutional Neural Network

The patches extracted from the image pixel are transformed into a one-dimensional feature vector with the help of three different uniform Convolutional Neural Network (CNN). Each of the CNN models comprise of an input layer which takes as input the patches of the image as input. The input layer of the model is followed by four consecutive Convolutional layers comprising of the structure in Table 3.2.

Table 3.2: Convolutional Neural Network Structure

Layer #	Layer Name	# Filters	Filter Size	Activation
Layer 1	Convolution [178]	4	(3,3)	ReLU [179]
Layer 2	Convolution [178]	4	(3,3)	ReLU [179]
Layer 3	Convolution [178]	4	(5,5)	ReLU [179]
Layer 4	Convolution [178]	4	(5,5)	ReLU [179]

The output obtained from the final Convolutional Layers is then consecutively passed through separate Flatten layers which converts the features into a one-dimensional feature vector for each patch type. The symbolic representation of the three different feature vectors are x_1, x_2 and x_3. As the obtained feature vector shape is different for each patch type, with 9×9 having least number of elements in the vector and 13×13 having the highest number of elements, the feature vectors for 9×9 and 11×11 are padded with zeros for uniformity of their shape with 13×13 patch feature vector. The feature vectors finally obtained from this segment of the model is then provided as an input to the *Multi-Scale Layer* for feature combination of the model. An illustration of the neural architecture is shown in Figure 3.2.

3.3.3 Multi-Scale Layer

The *Multi-Scale layer* takes as input three separate feature vectors x_1, x_2 and x_3 as input and combines them on the basis of their importance for classification of the pixel based on its context or adjacent pixel properties. The philosophy behind the layer is that each pixel classification can be modeled as combination of classification decisions based on the pixel adjacent contexts. A low contrast vessel can be better classified if a larger patch size can be used. The use of a smaller patch size for the low contrast pixels will result in failure as sufficient context will not be present for its proper classification. Most of the low contrast pixels are better classified by taking into consideration a larger patch size for a more informed classification. However, the use of a larger patch size might lead to the detection of spurious vessel pixels around

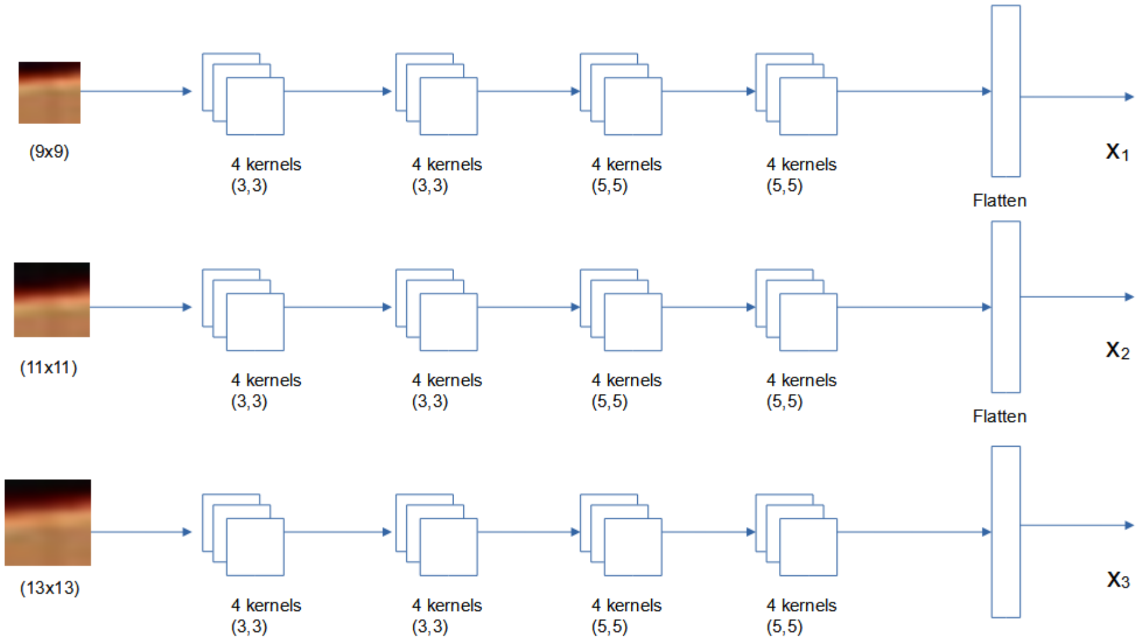


Figure 3.2: An illustration of the applied Convolutional Neural Network

the vascular structures of a thick blood vessel. A smaller patch size is better suited in such a scenario for a more accurate segmentation. In the *Multi-Scale layer*, the features from all three of the patch types are combined in a weighted combination to arrive at the final feature vector for the pixel in the image. The representation of the *Multi-Scale Layer* is provided in Figure 3.3.

The equation used for the combination of the feature vector is shown below.

$$\sigma(x_1 \otimes W_{r_1} \oplus B_{r_1}) = R_1 \tag{3.3.1}$$

$$\sigma(x_2 \otimes W_{r_2} \oplus B_{r_2}) = R_2 \tag{3.3.2}$$

$$\sigma(x_3 \otimes W_{r_3} \oplus B_{r_3}) = R_3 \tag{3.3.3}$$

$$R_1 \oplus R_2 \oplus R_3 = z \tag{3.3.4}$$

$$W_{h_1} \otimes R_1 = h_1 \tag{3.3.5}$$

$$W_{h_2} \otimes R_2 = h_2 \tag{3.3.6}$$

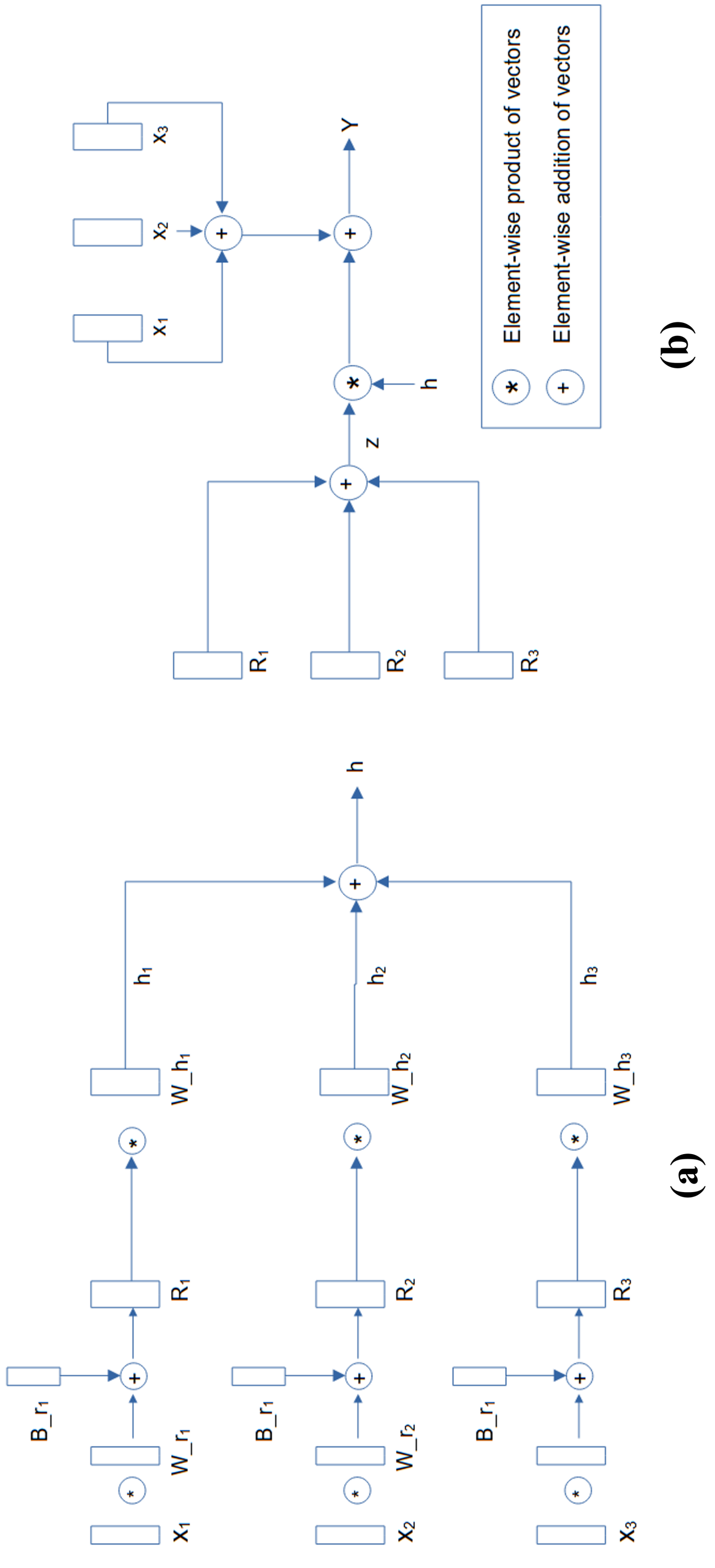


Figure 3.3: An illustration of the *Multi-Scale Layer*. The illustration is partitioned into two parts demonstrating the complete architecture of the layer.

$$W_{h_3} \otimes R_3 = h_3 \tag{3.3.7}$$

$$(x_1 \oplus x_2 \oplus x_3) \oplus z \otimes (h_1 \oplus h_2 \oplus h_3) = Y \tag{3.3.8}$$

The values of $W_{r_1}, W_{r_2}, W_{r_3}, W_{h_1}, W_{h_2}, W_{h_3}, B_{r_1}, B_{r_2}, B_{r_3}$ of the layer are trainable parameters which are learned by the network based on the training provided. The σ operator is used to denote a sigmoid function through which the final set of values for R_1, R_2 and R_3 are computed. Lastly, the output feature vector is provided as an input to the fully connected layer for final classification.

3.3.4 Classification

The combined feature vector Y is taken as an input to a fully connected model of the neural network for final classification. The structure of the fully connected network is shown as below in Table 3.3.

Table 3.3: Classifier Structure

Layer #	**Network Structure**		
	Layer Name	**# Neurons**	**Activation**
Layer 1	Dense Layer	50	ReLU
Layer 2	Dense Layer	10	ReLU
Layer 3	Dense Layer (Softmax)	2	–

The classification network is trained using the ground truth images provided with the Fundus image dataset. The softmax function is used in the final layer of the model for predicting the final classification of the pixel as vessel or non-vessel. An illustration of the classification layer is demonstrated in Figure 3.4.

3.3.5 Experimental Results

The performance of the model is evaluated using the DRIVE [112] dataset. The model is trained with all the training images of the dataset and its performance on the test images are recorded using Accuracy (Acc), Specificity (Sp) and Sensitivity (Se) as the performance indicators. Accuracy of the model computes the percentage of total number of pixels which are correctly classified as Vessel and Non-Vessel pixels. Specificity computes the total number of Non-Vessel Pixels that were correctly classified while Sensitivity calculates the total number of Vessel pixels that are correctly

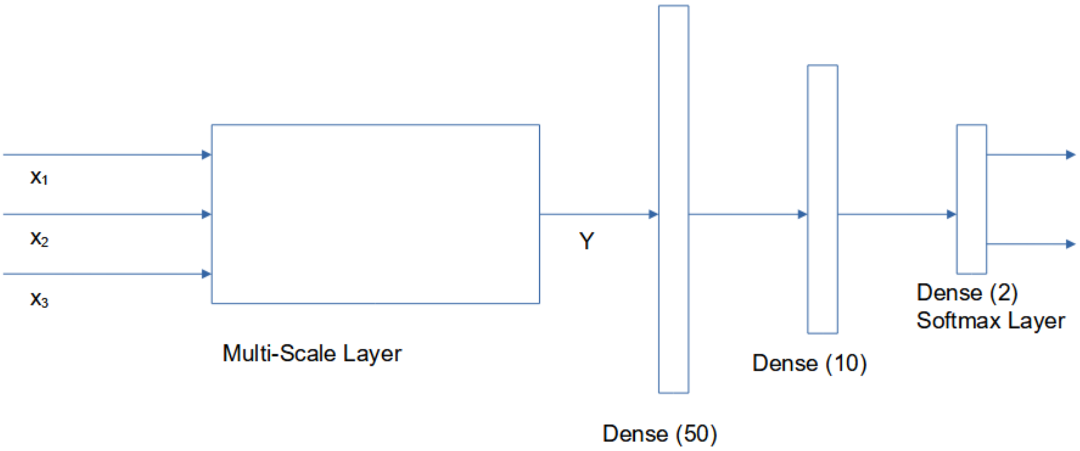


Figure 3.4: An illustration of the Multi-Scale Layer-based Classifier Network.

classified. Table 3.4 presented below highlights the performance comparison of our approach with some of the existing works in the literature.

In addition to the training and testing data, the DRIVE dataset also provides two sets of Ground Truths, annotated by two different Human Observers. For the purpose of evaluation, existing researches have utilized the first set of ground truth as the gold standard and the second set of ground truths as a means of measuring Human performance for the segmentation task. The performance has been computed using the DRIVE dataset images and compared with the first set of ground truth images provided along with the dataset. The Human performance is also computed using the second set of ground truths and has been presented in Table 3.4.

From Table 3.4, it is observed that the presented work performs comparably better compared to existing state of the art works in this field. The current state of the art accuracy obtained is by Wankhede et al. [183] with an accuracy of 97.2% whereas the given model has obtained an accuracy of 94% which is comparable to the other past works. However, the model produced a better Sensitivity value of 78% compared to 74.2% by Wankhede et al. [183]. Overall, the measure of Sensitivity is seen to be better than most of the state of the art systems, bested only by the works done by Jiang et al. [113] and Liskowski et al. [108]. However, the model achieved a Specificity of 95.6% compared to 90% obtained by Jiang et al. [113]. Also, the work achieves a better Accuracy and Specificity when compared to the respective metrics achieved by Liskowski et al. [108]. The value of Sensitivity plays an equally important role as Accuracy in the evaluation of our system as Sensitivity correlates with the number of vessel pixels correctly segmented. It can be seen that Sensitivity achieved is comparable with almost all the state-of-the-art systems while achieving a standard

Table 3.4: Comparison of different research approaches on performance metrics

Performance Comparison

#	Research Work	Type	Accuracy	Specificity	Sensitivity
1	[113]	Supervised	0.891	0.900	0.830
2	[180]	Supervised	0.952	0.983	0.725
3	[114]	Supervised	0.959	0.972	0.775
4	[181]	Unsupervised	0.954	0.982	0.742
5	[160]	Unsupervised	0.942	0.969	0.689
6	[182]	Unsupervised	0.951	0.972	0.731
7	[161]	Unsupervised	0.946	0.978	0.733
8	[183]	Unsupervised	0.972	0.971	0.742
9	[16]	Deep Learning	0.925	0.924	0.916
10	[184]	Deep Learning	0.965	0.988	0.731
11	[185]	Deep Learning	0.954	0.981	0.765
12	Human Observer	–	0.947	0.968	0.789
13	Proposed Approach	Deep Learning	0.940	0.956	0.780

performance in other metrics. The visual output of the system on DRIVE test images is presented in Figure 3.5.

3.4 Structural Asymmetric Sampling-based Probability-Map Choquet-Fusion Prediction Network

This section presents the *Structural Asymmetric Sampling-based Probability-Map Choquet-Fusion Prediction Network (PrSASfNet)* RV segmentation approach. Each of the steps is discussed in detail in the following subsections. An overview of the proposed method is shown in Fig 3.6.

The PrSASfNet architecture is inspired from the design of an encoder-decoder structure with skip connection as the backbone. It mainly comprises of two units, namely, Structural Asymmetric Sampling based Encoder-Decoder Network (SASNet) and Probability-Map Choquet-Fusion based Prediction Network (PrfNet). Figure 3.6 depicts the architecture of the proposed PrSASfNet. The motivation for the operation of the PrSASfNet is founded on two primary features used in delineation of the vessel structure in fundus images, namely, intensities of vessel regions and contrasting difference at vessel margins between vessel region and background region. In the following sub-sections, a detailed overview of each individual unit of the network

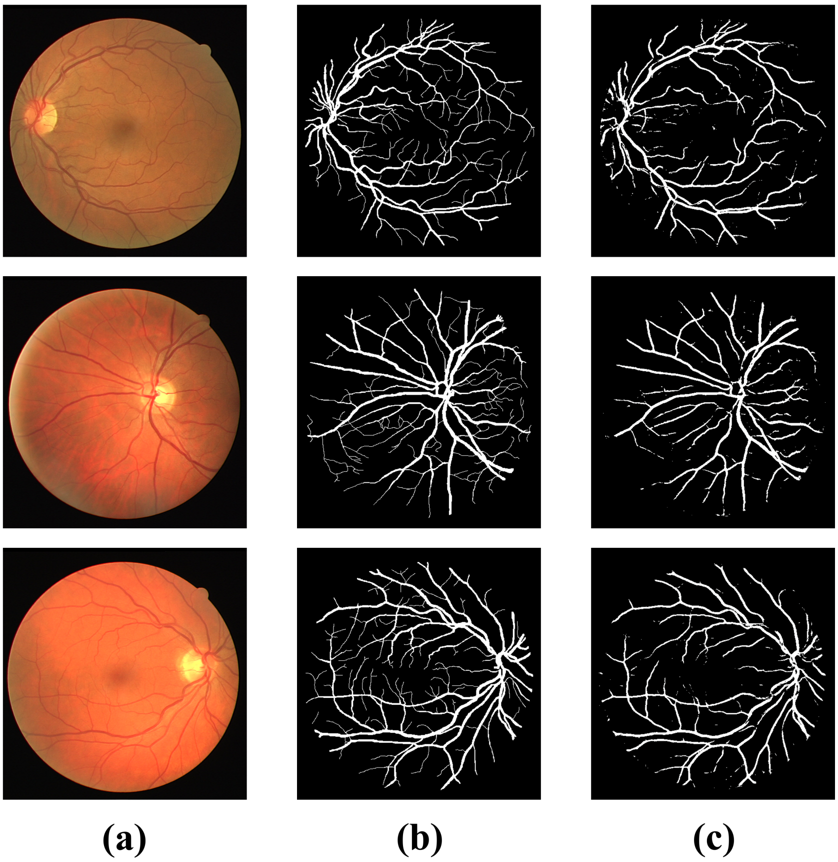


Figure 3.5: Qualitative depiction of the model performance on the DRIVE dataset. (a) Original RGB Fundus Image (b) Human annotated Vessel Ground Truth (c) *Multi-Scale Layer Pixel-based Segmentation Network* Achieved Output

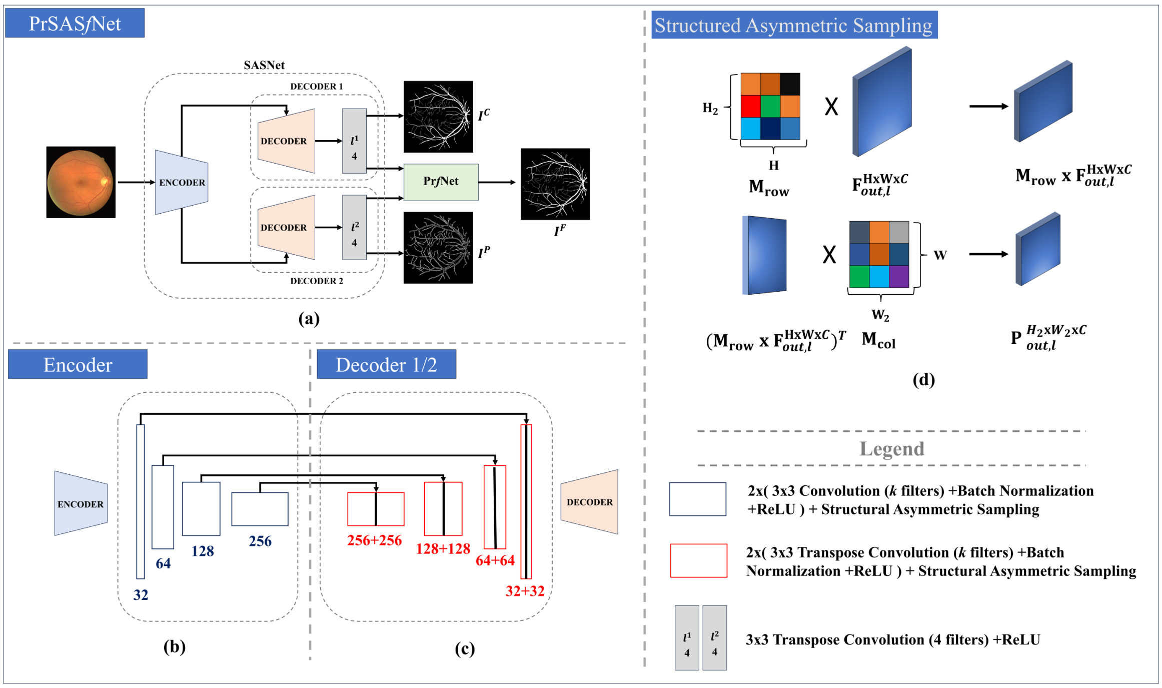


Figure 3.6: An overview of the proposed convolution-based PrSASfNet architecture. (a) Block Diagram of the proposed model. Here, l^1/l^2 represents the penultimate layer of the respective decoder. (b) Architectural Design of the proposed encoder. (c) Architectural Design of Decoder n, where $n \in \{1, 2\}$. (d) Design of the Structural Asymmetric Sampling (SAS) Layer. The output obtained from $l^{1|2}$ comprises of 4 channels which is further passed through a 3×3 Convolution to generate $\mathcal{I}^C$ and $\mathcal{I}^P$ respectively. Simultaneously, the output from $l^{1|2}$ is passed through the PrfNet block, as depicted in (a), to obtain final segmented image $\mathcal{I}^F$.

is presented.

3.4.1 Identification of the approximate structure of RVs using SASNet

This section presents the *Structural Asymmetric Sampling based Encoder-Decoder Network (SASNet)* for identifying approximate vessel regions in fundus images. SASNet is structurally based on a standard Encoder-Decoder architecture with skip connections (as shown in Figure 3.6), with a fundamental change made to the network's pooling and upsampling module for both the encoder and decoder arms respectively. A Structural Asymmetric Sampling (SAS) Layer in used, inspired from [186], instead of a standard pooling and upsampling operation. It employs two trainable asymmetric matrices to perform a sampling of features, which ensures feature preservation and representation. Figure 3.6(d) illustrates the functioning of our proposed SAS layer. The mathematical formulation of SAS Layer is detailed below.

Let $\mathcal{F}_{out,l}^{H\times W\times C}$ be denoted as the convolution feature map output at layer l having dimension of $H \times W \times C$, where H, W and C signify height, width, and channels, respectively. The feature maps generated post convolution operation can be formally defined as

$$\mathcal{F}_{out,l}^{H\times W\times C} = concat_i(\forall i, max(0, \mathcal{F}_{in,l}^{H\times W\times C_1} * \mathcal{K}_{i,l}^{3\times 3\times C_1})) \tag{3.4.1}$$

where $\mathcal{F}_{in,l}^{H\times W\times C_1}$ is the feature map input at layer l of the model having a dimension of $H \times W \times C_1$, $\mathcal{K}_{i,l}^{3\times 3\times C_1}$ is the i^{th} kernel at layer l having dimension $3 \times 3 \times C_1$, $*$ signifies the convolution operation. The feature map $\mathcal{F}_{out,l}^{H\times W\times C}$, generated from a specific convolution layer l, is taken as input to the SAS layer.

$$\mathcal{P}_{out,l}^{H_2\times W_2\times C} = (M_{col} \times (M_{row} \times \mathcal{F}_{out,l}^{H\times W\times C})^T)^T \tag{3.4.2}$$

where, $\times$ represents matrix multiplication, M_{col} and M_{row} are two trainable asymmetric matrices with dimensions $H_2 \times H$ and $W_2 \times W$ respectively. $(.)^T$ signifies the matrix transpose operation for each filter along the channel dimension and $\mathcal{P}_{out,l}^{H_2\times W_2\times C}$ is the final output of the SAS layer. Here, the value of H_2 and W_2 is considered to be equal to $\frac{H}{2}$ and $\frac{W}{2}$ respectively for the encoder arm of the network. Similarly, the values of $2H$ and $2W$ is calculated respectively for H_2 and W_2 in the decoder arm. Thus, SASNet acts as a Classification model for RV segmentation to generate a coarse segmentation map ($\mathcal{I}^C$) with respect to the vessel annotated ground truth ($\mathcal{I}^{GT}$). The SASNet Classification model is trained using the generalized Dice-Loss function (L_{Dice}) [187].

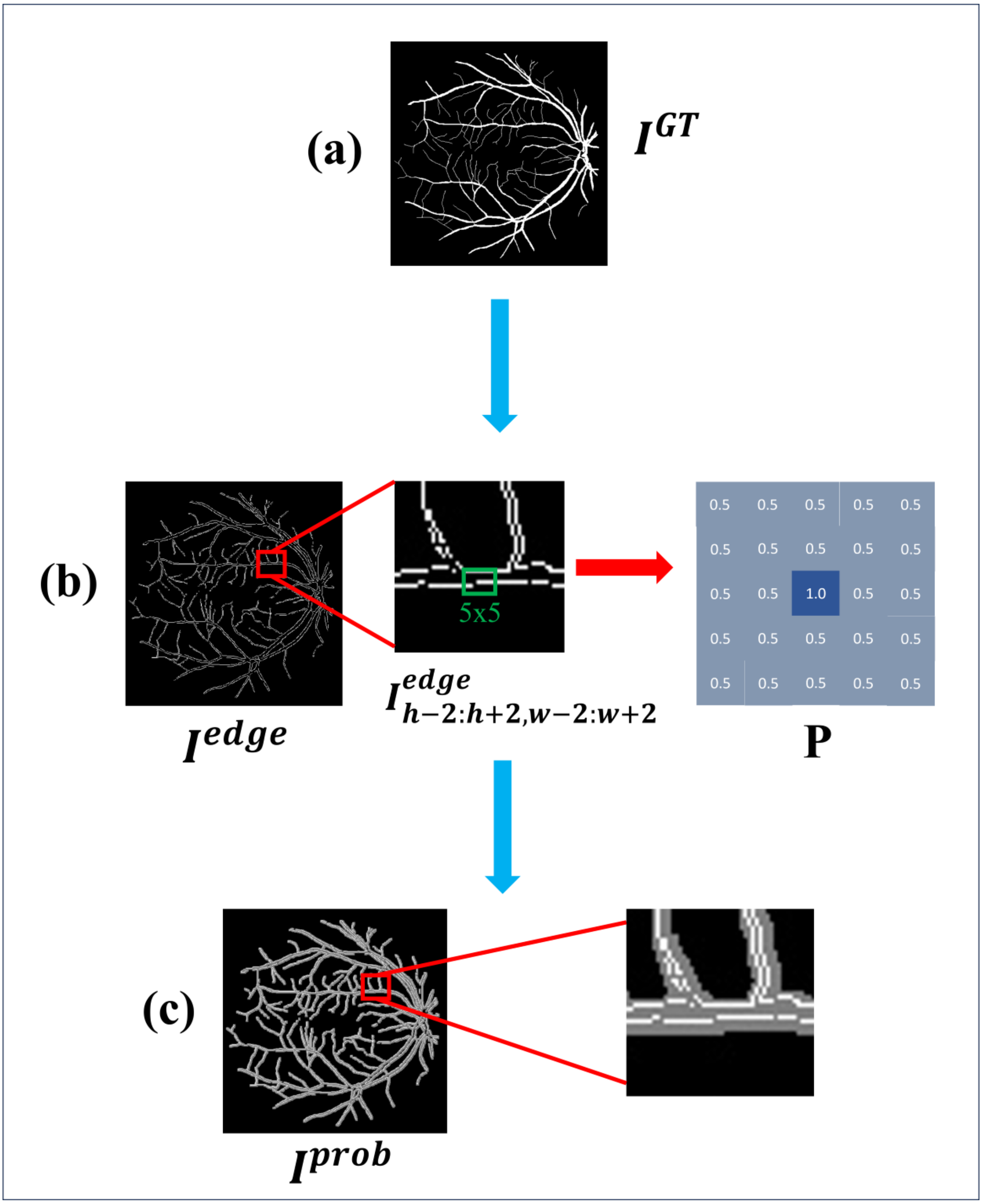
In parallel to the primary Decoder arm (shown as *Decoder 1* in Fig. 3.6(a)), a secondary structurally identical Decoder arm (shown as *Decoder 2*) is also designed. The secondary Decoder is trained to generate the probability of a vessel pixel to be a vessel boundary. To train *Decoder 2*, a boundary-probability map is prepared separately as ground truth ($\mathcal{I}^{Prob}$) to train the regression model. The probability-map based ground truth is prepared derived from vessel-annotated ground truth as explained next.

Firstly, a Canny Edge Detection algorithm [188] is applied on the vessel annotation $\mathcal{I}^{GT}$ to obtain the boundary pixels (as shown in Fig. 3.7(b)). However, most annotated ground truths suffer from annotation bias [189] which may lead to drop in model performance. To achieve a balance between annotation bias and model performance, a *Boundary Region* in this work. A *Boundary Region* is defined as a set of all pixels with proximity near Canny Edge generated boundary pixels. A proximity distance of 2 is considered. In consequence, a Boundary Region based Probability-Map is thus generated by iterating over all edge pixels obtained from Canny Edge detection algorithm. At each edge pixel, a 5 × 5 kernel (as proximity is 2) is placed with the center on the edge pixel. The kernel center point is assigned a probability of 1. The remaining points within the kernel are assigned a probability of 0.5. The output image is considered as the Boundary Region based Probability-Map ground truth image ($\mathcal{I}^{prob}$) of image $\mathcal{I}^{GT}$. The generated $\mathcal{I}^{prob}$ is finally used to train the *Decoder 2* arm of SASNet as a regression model using the Mean-Squared Error function (L_{MSE}) [190] and predict the boundary probability image ($\mathcal{I}^{P}$).

A detailed algorithm for Region-based Probability Map computation ($\mathcal{I}^{prob}$) is presented in Algorithm 1.

Algorithm 1 Computation of Region based Probability-Map Ground Truth $\mathcal{I}^{prob}$ for PrfNet model

```
Input: $\mathcal{I}^{GT}$
Output: $\mathcal{I}^{prob}$
Initialize: $\mathcal{I}^{edge} \leftarrow f^{Canny}(\mathcal{I}^{GT})$
for $h$=2 to $H-2$ do                                  ▷ As kernel is of size 5 × 5
    for $w$=2 to $W-2$ do                              ▷ As kernel is of size 5 × 5
        if $\mathcal{I}^{edge}_{h,w} \in \{1\}$ then
            $P \leftarrow \mathcal{I}^{edge}_{h-2:h+2,w-2:w+2}$     ▷ Extract 5 × 5 patch centered at $(h, w)$
            Set $P_{2,2} \leftarrow 1$ and $\forall_{i,j \notin \{2\}} P_{i,j} \leftarrow 0.5$
            Set $\mathcal{I}^{prob}_{h-2:h+2,w-2:w+2} \leftarrow P$
```



0.5	0.5	0.5	0.5	0.5
0.5	0.5	0.5	0.5	0.5
0.5	0.5	1.0	0.5	0.5
0.5	0.5	0.5	0.5	0.5
0.5	0.5	0.5	0.5	0.5

Figure 3.7: An illustration of the Probability Map generation ($\mathcal{I}^{Prob}$). (a) An input vessel annotated image $\mathcal{I}^{GT}$ (b) A Canny Edge Detector is applied on $\mathcal{I}^{GT}$. At each edge coordinate, a 5×5 kernel is applied. (c) Probability Map image ($\mathcal{I}^{Prob}$) for the input image $\mathcal{I}^{GT}$.

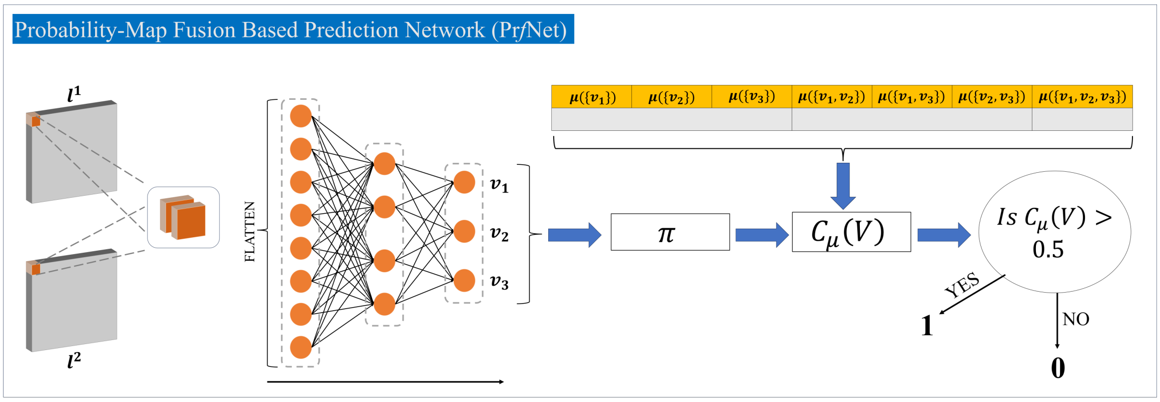


Figure 3.8: An illustration of the Probability-Map Choquet-Fusion based Prediction Network (PrfNet). μ set represents the learnable Fuzzy Measure.

3.4.2 Identification of the boundary of RVs using PrfNet

The difference in the contrast between vessel boundaries and background region varies across different regions of the Fundus Image. This poses an additional challenge in the design of any automated RV segmentation model. The vessel-boundary surface is often misclassified owing to the complexities [191]. To mitigate the issue, a Regression-based model is designed to predict a Boundary Probability-Map ($\mathcal{I}^P$). The coarse segmentation image $\mathcal{I}^C$ and the predicted Boundary Probability-Map $\mathcal{I}^P$ are combined using a Choquet-Fusion guided unit PrfNet to generate the final vessel segmented image $\mathcal{I}^F$. The design of the PrfNet is detailed as follows.

Feature set from the penultimate layers of each decoder arm (namely, l^1 and l^2 respectively) is extracted and channel-wise concatenated to form an 8-channel feature map (shown in Figure 3.8). At each pixel coordinate, a $3 \times 3 \times 8$ patch is extracted, as shown in the figure. The extracted patch is spatially squashed [192] per channel to a shape of $1 \times 1 \times 8$ and, finally, flattened to form a single 8-dimensional feature vector. The flattened vector is passed through an ANN (Fig. 3.8) to extract a vector V of length three, viz., $V = \{\sigma(v_1), \sigma(v_2), \sigma(v_3)\}$, (shown in Figure 3.8), where σ is the Sigmoid function. Set V is considered as a feature set, in this work, to classify each pixel coordinate of the final $\mathcal{I}^F$ image as vessel pixel or background pixel. To classify set V, we take inspiration from the Choquet Integral [193] for data fusion in fuzzy systems. Firstly, a feature correlation is determined in set V by defining a fuzzy measure set ($\mu : 2^V \rightarrow [0, 1]$) as a trainable parameter in this work. Mathematically, μ is formulated as

$$\mu = \{w_p \mid \forall p \in \mathrm{P}(V) - \varphi, 0 \leq w_p \leq 1\} \tag{3.4.3}$$

where $\mathrm{P}(V)$ is defined as the Power Set [194] of V, φ is the null element and w_p

represents an element of the trainable parameter set μ. Two additional constraints on the fuzzy set μ to simplify the optimization, viz., $\mu(\{v_1\}) = \mu(\{v_2\}) = \mu(\{v_3\})$ and $\mu(\{v_1, v_2\}) = \mu(\{v_1, v_3\}) = \mu(\{v_2, v_3\})$. A function $\pi : [0, 1]^n \to \mathbb{N}$ is defined in this work which selects the smallest value in a set of cardinality n and returns the position of the element in the set. The formulations are combined of μ and π and thus formulate a prediction score $C_\mu(V)$ based on principles of Choquet Integral as shown in Algorithm 2.

Algorithm 2 Computation of Fuzzy Prediction Score for PrfNet model

```
Input: $V$, $\mu$, $\pi$
Output: $C_\mu(V)$
Initialize: $A_\pi = \{\}$, $C_\mu(V) \leftarrow 0$, $e \leftarrow 0$
for steps=1 to 3 do
    $C_\mu(V) \leftarrow C_\mu(V) + |V_{\pi(V-A_\pi)} - e|.\mu(V - A_\pi)$    ▷ Compute the prediction score
    $e \leftarrow V_{\pi(V-A_\pi)}$    ▷ Extracting smallest element from Set $V - A_\pi$
    $A_\pi \leftarrow A_\pi \cup \{e\}$    ▷ Adding the extracted element to set $A_\pi$
```

All pixel coordinates having $C_\mu(V) > 0.5$ are finally classified as Vessel Pixel. To train the fuzzy measure set μ, the Binary-Cross Entropy function (L_{BCE}) [195] has been used in this work to update the fuzzy measure set across iterations. The fuzzy measure updation is defined as

$$\mu_{iter+1} = \mu_{iter} - \eta \frac{\partial L_{BCE}}{\partial \mu} \tag{3.4.4}$$

where $iter$ signifies the iteration and η is the learning rate. The final obtained image is denoted by $\mathcal{I}^F$. It is, however, noted that boundary conditions and monotonicity constraints have not been imposed on the fuzzy set μ in this case to limit the computational overhead. Additionally, the SASNet unit and the PrfNet unit of the proposed work are trained separately, in sequence of mention, to achieve overall performance.

The proposed convolutional architecture is coded in the PyTorch environment [196] having 32GB RAM with a 15GB T4 GPU for faster processing. Each individual module of the platform is trained using the Adam optimizer [197]. The training is performed for a maximum iteration of 1000 and, for each iteration, the learning rate is fixed at $3e^{-4}$, constrained on an Early Stopping criterion based on the Loss per iteration. A batch size of 4 is considered for each iteration to limit the need for excess memory.

3.4.3 Experimental Results

In this section, a detailed qualitative as well as quantitative analysis of the proposed PrSASfNet architecture is presented on four public benchmark datasets, namely, DRIVE [112], STARE [198], CHASEDB1 [199] and HRF [200] respectively. The DRIVE dataset comprises of RGB fundus images captured at a resolution of 768 × 584 pixels and segregated into 20 training and 20 testing vessel annotated sets of images for retinal vessel segmentation. The STARE dataset includes a retinal vessel annotated gold standard of 20 RGB fundus images. In this approach, the dataset is split at random into a set of 10 images for training and a set of 10 images for testing as the dataset does not provide any segregation. CHASEDB1 dataset comprises of 28 color fundus images captured with a resolution of 1280 × 960 pixels each. The dataset is also partitioned at random into a set of 14 images for training and 14 images for testing respectively. The HRF dataset is a high-resolution fundus image dataset consisting of 45 fundus images with annotated vessel segmentations. Each image is captured at a resolution of 3504 × 2336 pixels. The dataset is split at random into a set of 23 images for training and 22 images for testing respectively. To increase the volume of data for model training, standard data augmentation approaches had been applied on the training set and the performance achieved on the test set is highlighted in subsequent sections. All images, both training and testing, have been resolved to a dimension of 512 × 512 in this work.

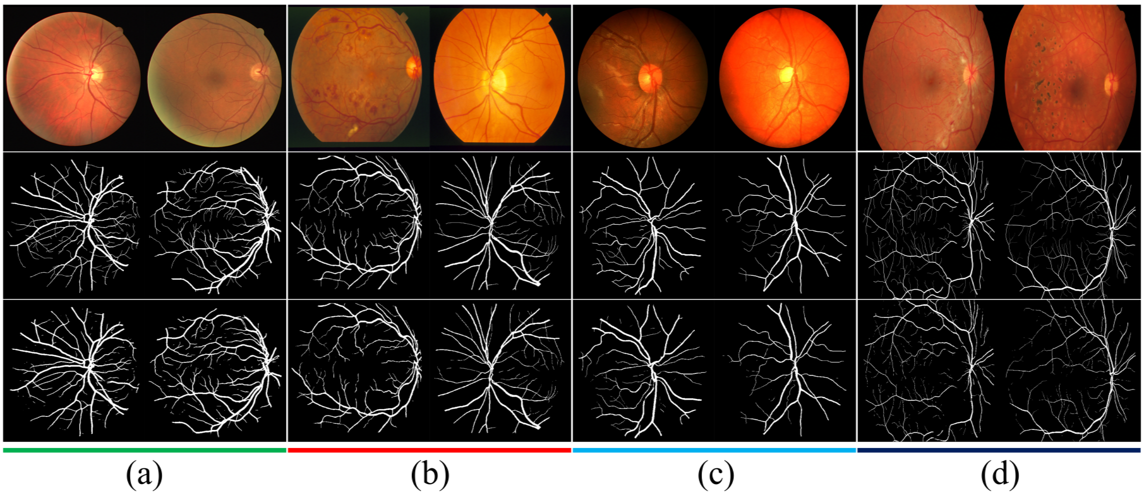
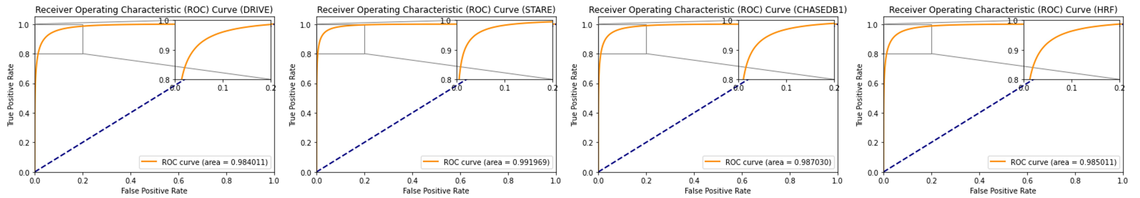


Figure 3.9: An illustration for the qualitative analysis of the segmented output obtained on (a) DRIVE (b) STARE (c) CHASEDB1 (d) HRF datasets respectively. The first row corresponds to the original color fundus images. The second row corresponds to the human-annotated ground truth for vessel segmentation. The last row corresponds to the segmented output obtained from the proposed PrSASfNet model.



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

Figure 3.10: An illustration of the ROC Curve for computaion of AUC on (a) DRIVE (b) STARE (c) CHASEDB1 (d) HRF dataset

3.4.4 Performance Evaluation of Vessel Segmentation

Fig. 3.9 highlights the segmentation outputs obtained for each set of fundus images from DRIVE, STARE, CHASEDB1 and HRF datasets respectively for a qualitative analysis. The first two rows correspond to the color fundus images and their corresponding vessel annotated ground truths respectively. The last row presents the obtained vessel segmentation output. It is observed from the figure that the segmented output bears significant resemblance to the human annotated ground truth. However, apart from qualitative aspects, quantitative assessment is also necessary for complete evaluation of the approach. To perform a quantitative analysis, firstly, PrSASfNet model is compared with a vanilla U-Net model [125]. The U-Net model is selected for comparative analysis as it follows a similar skip-connection based encoder-decoder architecture. Also, U-Net model is a popular model in literature and is most commonly used for biomedical image segmentation. Table 3.5 highlights the statistical performance between a vanilla U-Net model and our proposed architecture in Retinal Blood Vessel segmentation task. It is observed that the architecture achieves a better performance overall in comparison to vanilla U-Net model. For a comparative analysis, four standard benchmarks for biomedical segmentation evaluation have been used in this work, namely, Sensitivity (Se), Specificity (Sp), Accuracy (Acc) and Area Under Curve (AUC) respectively [108]. Figure 3.10 highlights the obtained ROC Curve for the model across all four benchmark datasets. It is noted from Table-3.5 achieves almost 5.1%, 3.7%, 1.5% and 0.2% increase in Sensitivity performance in contrast to baseline U-Net model. Simultaneously, an increase of 0.7%, 1.1% and 2.0% in Accuracy is observable. The approach has a minimal impact on the HRF dataset however, as noted from Table-3.5. The dimensionality reduction of the HRF data to a resolution of 512×512 might have attributed to loss of information which reduces the impact. Nonetheless, the significant boost in performance across the other datasets establishes the efficacy of the approach.

To establish the achieved performance, the PrSASfNet model is additionally com-

Table 3.5: Comparison of the PrSASfNet architecture with a vanilla U-Net model.

Databases	Methods	Sensitivity	Specificity	Accuracy	AUC
DRIVE	U-Net [125]	0.809	0.972	0.955	0.979
	This Work	**0.860**	**0.972**	**0.962**	**0.984**
STARE	U-Net [125]	0.818	0.979	0.965	0.986
	This Work	**0.855**	**0.988**	**0.976**	**0.992**
CHASEDB1	U-Net [125]	0.829	0.983	0.965	0.986
	This Work	**0.844**	**0.988**	**0.985**	**0.987**
HRF	U-Net [125]	0.821	0.979	0.964	0.985
	This Work	**0.823**	**0.980**	**0.964**	**0.985**

pared against other SOTA methods in literature. Table 3.6 highlights the comparative analysis across all four benchmark datasets. Additionally, the scores obtained are compared beyond three decimal places in the event of a tie. It is noted that the model has achieved highest Sensitivity (0.860) and second-best AUC (0.984) score amongst contemporary works listed on the DRIVE dataset. For the STARE dataset, the model performs superior across three parameters in comparison to other methods, namely, Specificity (0.988), Accuracy (0.976) and AUC (0.992) respectively. With context to CHASEDB1 dataset, the model achieves better performance in terms of Specificity (0.988) and Accuracy (0.985) scores. On the HRF dataset, the architecture achieves a superior performance in terms of Sensitivity (0.824). Additionally, the model also achieves a second-best performance in terms of AUC (0.984) on DRIVE dataset, Sensitivity (0.844) on CHASEDB1 dataset and AUC (0.985) on HRF dataset respectively. In the following section, an ablation study is presented to investigate the individual unit performance independently.

3.4.5 Ablation Study

To better gauge the efficacy of the approach, a series of Ablation Studies have been conducted and the results obtained are reported in this section. The following subsections indepedently highlight the noted observations.

3.4.5.1 Efficacy of SAS-Layer

Firstly, the efficacy of the proposed *Structural Asymmetric Sampling* layer is evaluated. To gauge the efficacy of the layer, the structural architecture of *Encoder-Decoder 1* skip connected architecture is replicated. The SAS layer is replaced by standard MaxPooling layer in Encoder and UpSampling Layer in Decoder-1. The model, thus

Table 3.6: Quantitative Performance Analysis of the PrSASfNet architecture on DRIVE, STARE, CHASEDB1 and HRF datasets. (Color Coding top performances: RED-1st, GREEN-2nd and BLUE-3rd)

Databases/	**DRIVE**				**STARE**				**CHASEDB1**				**HRF**			
Methods	**Se**	**Sp**	**Acc**	**AUC**	**Se**	**Sp**	**Acc**	**AUC**	**Se**	**Sp**	**Acc**	**AUC**	**Se**	**Sp**	**Acc**	**AUC**
[56]	0.765	0.981	0.954	0.975	0.758	0.984	0.961	0.980	0.763	0.980	0.961	0.978	0.808	0.941	0.943	–
[165]	0.796	0.980	0.956	0.980	0.759	0.987	0.964	0.983	0.815	0.975	0.961	0.980	0.746	0.987	0.965	0.983
[166]	0.820	0.975	0.954	0.983	0.829	0.984	0.966	0.987	-	-	-	-	-	-	-	-
[167]	0.779	0.983	0.954	0.984	0.771	0.988	0.970	0.987	-	-	-	-	-	-	-	-
[168]	0.799	0.981	0.958	0.982	0.818	0.984	0.967	0.988	0.824	0.981	0.967	0.987	0.780	0.984	0.965	0.983
[170]	0.836	0.974	0.956	0.979	0.856	0.982	0.968	0.987	-	-	-	-	-	-	-	-
[172]	0.828	0.983	0.969	0.983	0.820	0.983	0.973	0.984	0.836	0.984	0.974	0.986	0.811	0.982	0.968	0.984
[171]	0.830	0.975	0.957	0.982	0.865	0.984	0.972	0.992	0.846	0.984	0.966	0.988	0.822	0.981	0.966	0.987
[173]	0.835	0.982	0.970	0.986	0.848	0.986	0.976	0.990	0.844	0.985	0.976	0.989	0.817	0.982	0.969	0.985
[55]	0.828	0.977	0.957	–	0.828	0.985	0.965	–	-	-	-	-	-	-	-	-
This Work	0.860	0.972	0.963	0.984	0.855	0.988	0.976	0.992	0.844	0.988	0.985	0.987	0.824	0.973	0.964	0.985

formed, is similar in structure as a U-Net model but with lesser parameters. The term *LwU-Net* is coined to address this architecture. In parallel, a similar architecture is also designed but with the SAS Layer in both the Encoder and Decoder-1 arms (coined as *LwU-Net*+SAS). Table-3.7 is a brief highlight of the proposed architectures respectively. Each layer highlighted in Table comprises of a series of dual successive blocks of operations with each block comprising of 3 × 3 Convolution with stride as 1 and zero-padding of depth 1. Each convolution operation is followed by Batch Normalization and ReLU Activation layers respectively. The *LwU-Net* model as well as the *LwU-Net*+SAS Layer model is trained using Dice Loss and used to predict the vessel segmentation output. Table-3.8 highlights the achieved performance across Sensitivity (Se), Specificity (Sp) and Accuracy (Acc) respectively, across DRIVE and STARE datasets. It is noted that the U-Net+SAS layer model performs better than the base U-Net model with an improvement of almost 3% and 4% in Accuracy on the respective datasets. Furthermore, an improvement of almost 8% and 9% is also noted on the respective datasets. This establishes the efficacy of SAS Layer integration.

Table 3.7: Structural Comparison between *LwU-Net* and *LwU-Net*+SAS.

	***LwUNet* (#Channels)**		***LwU-Net*+SAS (#Channels)**	
	Encoder	**Decoder**	**Encoder**	**Decoder**
I/O	3	1	3	1
Layer1	32	32	32	32
Layer2	64	64	64	64
Layer3	128	128	128	128
Layer4	256	256	256	256

Table 3.8: Statistical Results of Ablation Study between *LwU-Net* and *LwU-Net*+SAS model.

Databases	**LwU-Net**			**LwU-Net+SAS**		
	Se	**Sp**	**Acc**	**Se**	**Sp**	**Acc**
DRIVE	0.750	0.940	0.930	**0.830**	**0.980**	**0.960**
STARE	0.730	0.930	0.910	**0.820**	**0.970**	**0.950**

3.4.5.2 Efficacy of PrfNet module

The efficacy of the *Probability-Map Fusion Based Prediction Network* is evaluated in this sub-section. As the PrfNet module is computed in adjunct with the Probability

Map $\mathcal{I}^{prob}$, the subsequent modules are removed and the *LwU-Net*+SAS model, shown in Table-3.7, is considered. The comparative analysis between the *LwU-Net*+SAS model and our proposed PrSAS*f*Net model is presented under Table-3.9, as shown. It is noted that a significant impact is observed on Sensitivity across both datasets, with an improvement of almost 3% noted. The change in Sensitivity is also reflected in the achieved Accuracy, as a boost of almost 1% and 2%, respectively, is noted across both datasets. However, a small aberration 1% in Specificity is noted on the DRIVE dataset which might be attributed to a drop in performance across low contrast regions.

Table 3.9: Statistical Results of Ablation Study between *LwU-Net*+SAS and PrSAS*f*Net model.

Databases	**LwU-Net+SAS**			**Our Work**		
	Se	**Sp**	**Acc**	**Se**	**Sp**	**Acc**
DRIVE	0.830	**0.980**	0.960	**0.860**	0.970	**0.960**
STARE	0.820	0.970	0.950	**0.850**	**0.980**	**0.970**

3.4.5.3 Performance in Cross-Validation

To gauge the generalization potential of the approach, *Cross-Validation* is used to evaluate the performance of the model. Two standard datasets, namely, DRIVE and STARE, is used to conduct the cross-validation studies. Initially, the PrSAS*f*Net model is trained on the augmented DRIVE dataset and tested on the STARE dataset. Subsequently, the model is trained on the augmented STARE dataset and tested on the DRIVE dataset. Table highlights the achieved performance on cross-validation experiments. A comparative analysis is also presented with the state-of-the-art DPF-Net model [55] for elucidating the capabilities of the work. It is noted that the approach achieves an overall average increase of 0.5% in Specificity and 0.1% in Accuracy across both cross-validation studies with respect to benchmark performances. However, an overall drop of 2.55% is noted in Sensitivity across both studies.

3.4.5.4 Performance in Diseased Samples

To better evaluate the practicality of the proposed approach, the model performance on Diseased data samples is studied in this section. Seven diseased images from DRIVE dataset and ten diseased images from STARE dataset is selected for testing the pre-trained model performance in this work. Table-3.11 highlights the achieved performance across disease data samples. It is noted that the proposed approach

Table 3.10: Statistical Results of Cross-Validation on proposed PrSASfNet model.

Train	Test	Model	Sensitivity	Specificity	Accuracy
DRIVE	STARE	U-Net [125]	0.613	0.947	0.904
		DPF-Net [55]	**0.749**	0.983	0.953
		Our Work	0.711	**0.988**	**0.953**
STARE	DRIVE	U-Net [125]	0.655	0.931	0.889
		DPF-Net [55]	**0.746**	0.984	0.948
		Our Work	0.733	**0.989**	**0.950**

achieves an overall superior performance in comparison to past literature. This establishes the practical applications of the proposed model in real-life scenarios.

Table 3.11: Statistical Results of Diseased Samples on proposed PrSASfNet model.

Model	DRIVE			STARE		
	Se	**Sp**	**Acc**	**Se**	**Sp**	**Acc**
U-Net [125]	0.755	0.979	0.957	0.778	0.978	0.963
CE-Net [166]	0.815	0.979	0.965	0.794	**0.983**	0.969
IterNet [167]	0.823	**0.982**	0.968	0.809	0.979	0.966
Our Work	**0.847**	0.981	**0.968**	**0.822**	0.982	**0.970**

3.5 2pCePd-Net: Two-Path Cross-context Encoder with Probability Map-based Band-pass Decoder for Retinal Vessel Segmentation

Let $\mathfrak{R}^3$ signify a 3D continuous space, and $\mathcal{I} : \Theta \rightarrow \mathfrak{R}^3$ be the RGB fundus image, where $\Theta(\subset \mathfrak{R}^3) = (x, y, z)|x, y, z \in \mathbb{Z}$ is defined as the image domain. Typically, $\mathcal{I}$ consists of three distinguishable sub-regions, namely, retinal vessels ($\mathcal{I}_{rv}$), optic disc ($\mathcal{I}_{od}$), and background ($\mathcal{I}_b$).

The 2pCePd-Net architecture is inspired from the design of an encoder-decoder structure as the backbone with three distinct sections, namely, Two-path Cross-Dictionary Residual ($2pCd^+$) Encoder, Context-Aware Encoder Resampler Gate ($CERg$) and Probabilistic Band-Pass Fusion ($\hat{p}BPf$) Module. Figure 3.11 illustrates the structure of the proposed model. As can be observed in the figure, each horizontal segment of the network is referred to as a *level*, denoted by l. This notation is used extensively across subsequent model discussion. The following sub-sections provide detailed explanation of each individual section of the network.

3.5.0.1 Two-Path Cross-Dictionary Residual ($2pCd^{+}$) Encoder

The model starts from the Encoder section as highlighted in Figure 3.11. The Encoder module is structured in alignment with an Encoder-Decoder module with skip connections, as is common in literature [125]. However, such a structure also introduces challenges in the Vessel Segmentation tasks in Retinal Fundus images [55]. Notably, the continuous downsampling of features using a generic MaxPooling layer does not preserve the subtle features necessary for decoding thin vessel regions [173]. To address this issue, a $2pCd^{+}$ Encoder is introduced which preserves the important features for segmentation by introducing a Cross-Dictionary based Attention module, as depicted in Figure 3.12.

Firstly, a chain of Strided-Convolutions (stride=2) is deployed, which runs parallel to the Encoder stream and acts as an auxiliary stream of features. It is noted that the MaxPooling operation in Encoder stream enhances the translation invariance capabilities in CNN models whereas the Strided Convolution in auxiliary stream preserves better feature information [201]. To combine the advantages of both modules, a Cross-Dictionary guided Encoder block, named Cd^{+}, is designed in the work. Let the auxiliary stream input for a given **Enc** block at level l be denoted as f_{in}^{aux} and the encoder stream input from **Enc** block at level *l-1* as f_{in}^{enc} respectively. To enhance the feature representation in the downsampled encoder output, inspiration is taken from Cross-Attention [46] mechanism and compute the dictionary value matrices D_v^{enc} and D_v^{aux}, along with dictionary key matrix D_k^{enc} and dictionary query matrix D_q^{aux} respectively, defined as follows:

$$\begin{aligned}
D_v^{enc} &= \hat{f}_{in}^{enc} * \Phi_v^{enc} \\
D_k^{enc} &= \hat{f}_{in}^{enc} * \Phi_k^{enc} \\
D_v^{aux} &= f_{in}^{aux} * \Phi_v^{aux} \\
D_q^{aux} &= f_{in}^{aux} * \Phi_q^{aux}
\end{aligned} \tag{3.5.1}$$

In Equation 3.5.1, $\hat{f}_{in}^{enc}$ represents downsampled feature vector of f_{in}^{enc} and Φ_v^{enc}, Φ_k^{enc}, Φ_v^{aux}, Φ_q^{aux} represent the Convolutional Weight matrices respectively. Subsequently, the salient features are merged from the two dictionaries and cross-attention is com-

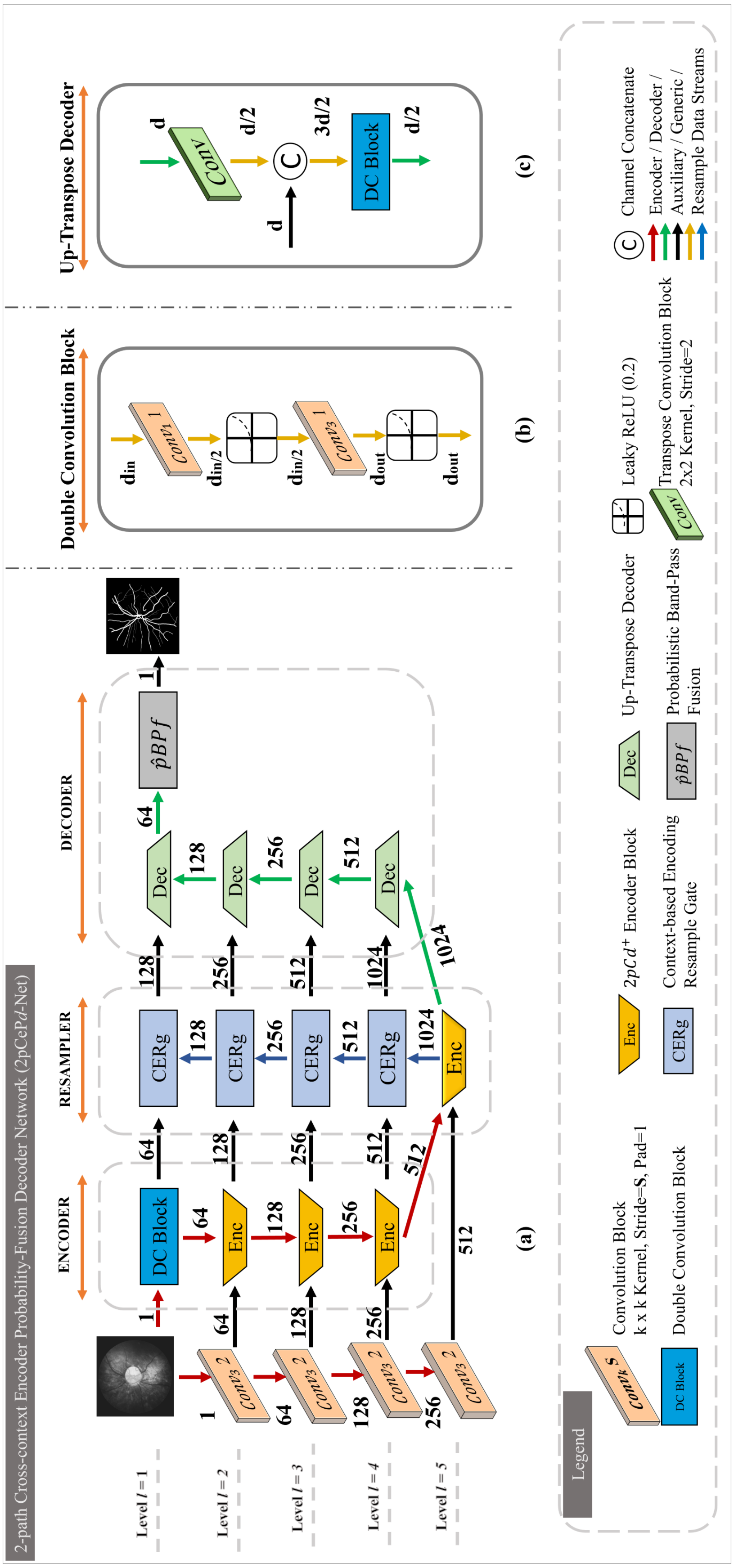


Figure 3.11: An illustration of the 2pCePd-Net model. (a) The symbolic block based depiction of the proposed 2pCePd-Net. The symbolic interpretations of the network blocks are highlighted within the *Legend*. (b) The detailed symbolic design of the Double Convolution Block (marked as *DC Block* in (a)). (c) The detailed symbolic design of the Up-Transpose Decoder Block (marked as *Dec* in (a)).

puted on the downsampled feature space $\hat{f}_{in}^{enc}$ as follows:

$$f^{temp} = softmax(\frac{(D_k^{enc})^T \odot D_q^{aux}}{\sqrt{d}}) \odot D_v^{enc} \tag{3.5.2}$$

In Equation 3.5.2, f^{temp} represents cross-attention feature vector, d represents the channel dimension, $(D_k^{enc})^T$ represents the transposed matrix and $\odot$ represents the Hadamard Product. Finally, the output of the Cd^+ Encoder block ($f_{out}^{Cd^+}$) is computed as follows:

$$f_{out}^{Cd^+} = f^{temp} + \hat{f}_{in}^{enc} + \sigma(D_v^{aux}) \tag{3.5.3}$$

In Equation 3.5.3, $\sigma(.)$ denotes the Sigmoid function. The addendum of the $\sigma(D_v^{aux})$ in Equation 3.5.3 incorporates an additional pathway for f^{temp} to recover any depreciated feature of importance due to the initial MaxPooling operation. Furthermore, the resultant feature ($f^{temp}+\sigma(D_v^{aux})$) is added to the Encoder stream input f_{in}^{enc}, similar to a Residual block [202]. This enforces the network to learn the Difference transformation [202] function between input and output instead of the complete transformation, which leads to better encoding. However, it is noted that although the Cd^+ Encoder block is capable of capturing relevant features, it might get biased towards certain patterns and dependencies, a common phenomenon in single-head attention mechanisms as discussed in [203]. To mitigate such issues, a multi-block $n - Cd^+$ Encoder module is deployed to avoid local fitting and preferential dependencies, where n represents the number of Cd^+ blocks. The value of $n = 2$ is empirically set for maintaining a balance between quality of output with computational resources and thus design the proposed $2 - Cd^+$ Encoder block as shown in Figure 6. The outputs obtained from each encoder are combined as a Channel concatenation followed by subsequent convolution operations, as shown in figure, to obtain the final combined encoder output f_{out}^{enc}. The combined encoder architecture with multiple Cd^+ blocks and combinatorial layers is denoted as $2pCd^+$ block in this work.

3.5.0.2 Context-Aware Encoder Resampler Gate (CERg)

Similar to standard CNN architectures, a skip-connection bridge is employed between the Encoder section and Decoder section respectively for better segmentation. However, the usage of skip-connections present an alternate pathway to the network to bypass the deeper layers, leading to dependence of the network on shallower features [16]. Additionally, traditional skip-connections duplicate the less-informed features of the model from shallower layers which often lead to incorrect target boundary detection in medical image segmentation [204]. To simultaneously mitigate both the issues in

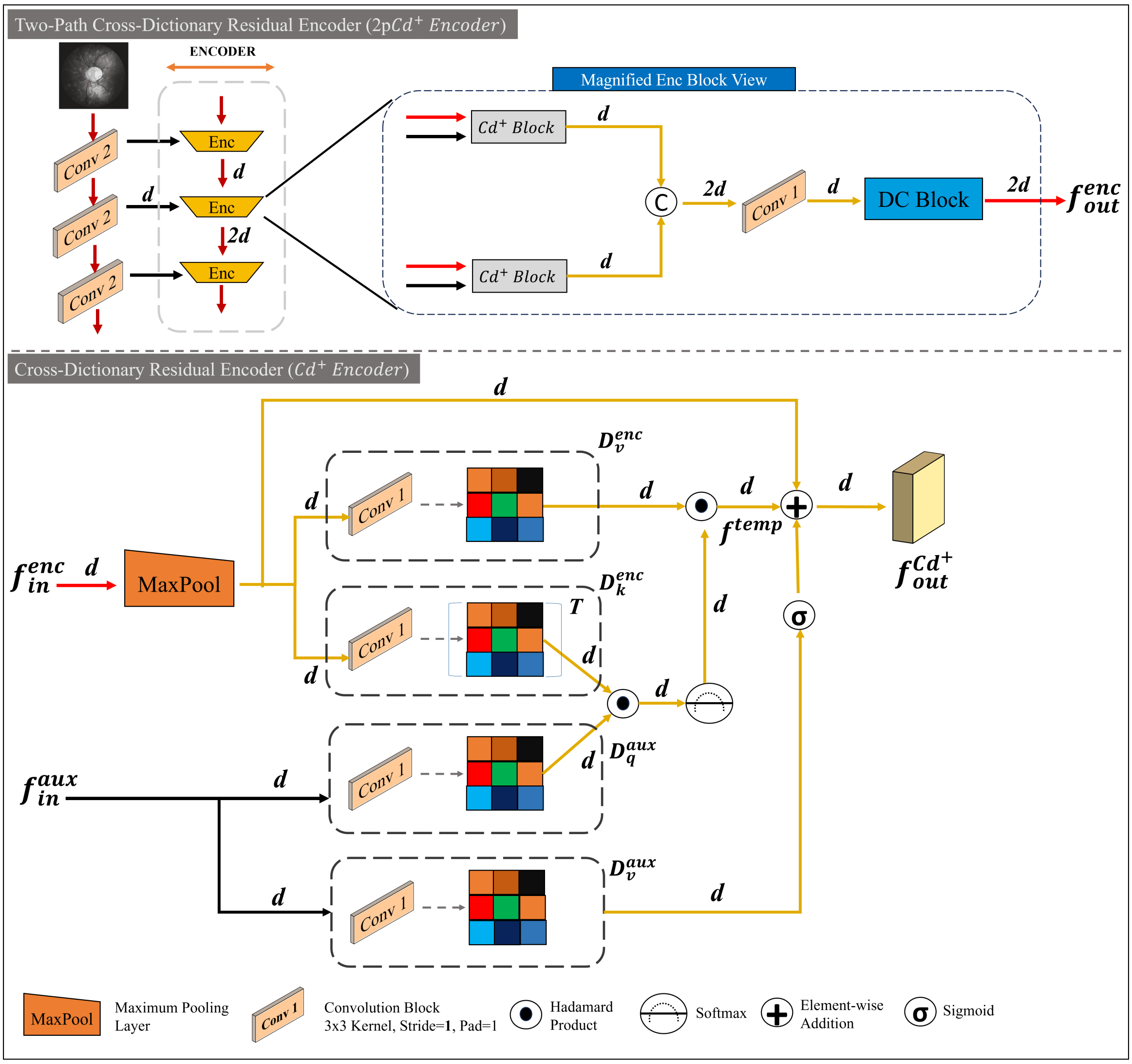


Figure 3.12: A depiction of the $2pCd^{+}$ Encoder Block.

traditional skip-connections, a Context-Aware Encoder Resampler Gate is employed. Firstly, three axioms are presented relevant to all learned encoded feature space in a CNN model for retinal vessel segmentation.

Axiom 1: A well-learned feature space $\mathcal{F}$ would represent all vessel regions separate from background regions irrespective of local vessel properties.

Axiom 2: A well-learned feature space $\mathcal{F}$ would be devoid of any influence from noise.

Axiom 3: A feature space $\hat{\mathcal{F}}$ derived from $\mathcal{F}$ would preserve all information present in $\mathcal{F}$.

To incorporate the axioms in the network, an inspiration is taken from past context-sensitive networks [205]. Thus, a Context-Aware Skip connection bridge is designed to incorporate information from deeper layers. This introduces a framework to modulate an over-dependence of the network on less-informed features on skip-connections, shown in Figure 3.13. As can be observed in the figure, the CERg block comprises of three distinct sub-modules, namely, Feature Modulation (f^{FM}), Outlier Filter (f^{OF}) and Vector Resampler (f^{VR}) respectively.

Let us consider a CERg module on skip-connection at level l. It is well-documented in literature that Encoder outputs at level $\hat{l} < l$ are more susceptible to noisy inputs [125] than deeper layers. A simple skip-connection at shallower layers would transmit a noise-infused feature onto the Decoder, inducing an additional difficulty on the Decoder to adhere to Axiom 1. In such scenarios, an emphasis on Axiom 2 might aid to mitigate an additional workload on Decoders. Furthermore, it is also important to preserve the relevant features on the skip connection for improving segmentation. These observations lead to the inspiration of the CERg Skip connection module. Firstly, a Feature Modulation unit takes an encoder output f^{enc}_{out} at level l as input as well as the CERg block output at level $l + 1$ and computes f^{FM}. In this context, it is noted that Encoder outputs at deeper layers are less prone to noise [170] while preserving the relevant features necessary for segmentation (Axiom 3). Thus, f^{FM} is equated mathematically as

$$f^{FM} = f^{enc}_{out} \odot \sigma(\hat{f}^{CERg}_{l+1} * \Phi^{CERg}_{1,l+1}) \tag{3.5.4}$$

In Equation 3.5.4, $\hat{f}^{CERg}_{l+1}$ represents De-Convoluted matrix and $\Phi^{CERg}_{1,l+1}$ represents the convolutional weights respectively. Equation 3.5.4 highlights a modulation where less represented features from deeper layers are suppressed in f^{FM} (due to $\sigma(.)$), thus putting emphasis on useful information. An Outlier Filter unit is appended to the Feature Modulation unit to further boost informative signal in the output f^{FM} while depreciating the less-represented features. Mathematically, it is represented as

$$\begin{aligned}\boldsymbol{f^{OF}} &= \boldsymbol{f^{FM} + (\sigma(\hat{f}^{CERg}_{l+1} * \Phi^{CERg}_{2,l+1}))} \\ &\quad \boldsymbol{\odot(\tanh(\hat{f}^{CERg}_{l+1} * \Phi^{CERg}_{3,l+1}))}\end{aligned} \tag{3.5.5}$$

In Equation 3.5.5, $\Phi^{CERg}_{2,l+1}, \Phi^{CERg}_{3,l+1}$ represents the convolutional weights respectively. Equation 3.5.5 highlights that a less represented feature in $\hat{f}^{CERg}_{l+1}$ would generate a tangent value closer to 0 or even negative whereas a well-represented feature would generate a value closer to 1. This property of hyperbolic tangent is used as an additive input to boost or depreciate the features respectively. Due to the additive nature of the design, a sigmoid function is multiplied with the tangent output to potentially limit the range of feature fluctuation for a stabilized learning. Finally, to limit the feature values within range [-1,1] (a zero-centered distribution [201]) and prevent gradient abnormality [206], a Vector Resampler unit is introduced. It is desirable that informative features have a higher representation in the feature vector in comparison to less-informative features in the final output. Thus, the feature space is resampled in the Vector Resampler unit and compute the final output as

$$f^{VR} = \tanh\{(f^{OF} * \Phi^{OF}) \odot (\hat{f}^{CERg}_{l+1} * \Phi^{CERg}_{4,l+1})\} \tag{3.5.6}$$

In Equation 3.5.6, $\Phi^{CERg}_{4,l+1}, \Phi^{OF}$ represents the convolutional weights respectively. The obtained feature map f^{VR} is channel-concatenated with the input f^{enc}_{out} and sent as an input to the decoder module as represented in Figure 3.13.

3.5.0.3 Probabilistic Band-Pass Fusion ($\hat{p}BPf$)

An Up-Transpose Decoder module is incorporated in the Decoder Section of the network, as shown in Figure 3.11. An Up-Transpose Decoder at level l takes inputs from CERg block at level l and Up-Transpose Decoder block from level $l+1$ respectively. It is noted that a model performance not only depends on model architecture but also on the training methodology which directs the gradients in the network to adjust the weights optimally. In this work, the model is directed to optimally segment a challenging region, defined herewith as Margin-of-Error (MoE) region (shown in Figure 3.14). The MoE region is defined as a vessel region having poor contrast with the adjacent background region. This enhances the difficulty in identifying the boundary of the vessels as well as proper delineation of thin vessels in RV segmentation task. It is trivial to state that all thin vessel regions are subset of MoE regions as well. To better augment the learning process, especially in MoE region, the use of a probability map is proposed to direct the model learning as discussed next.

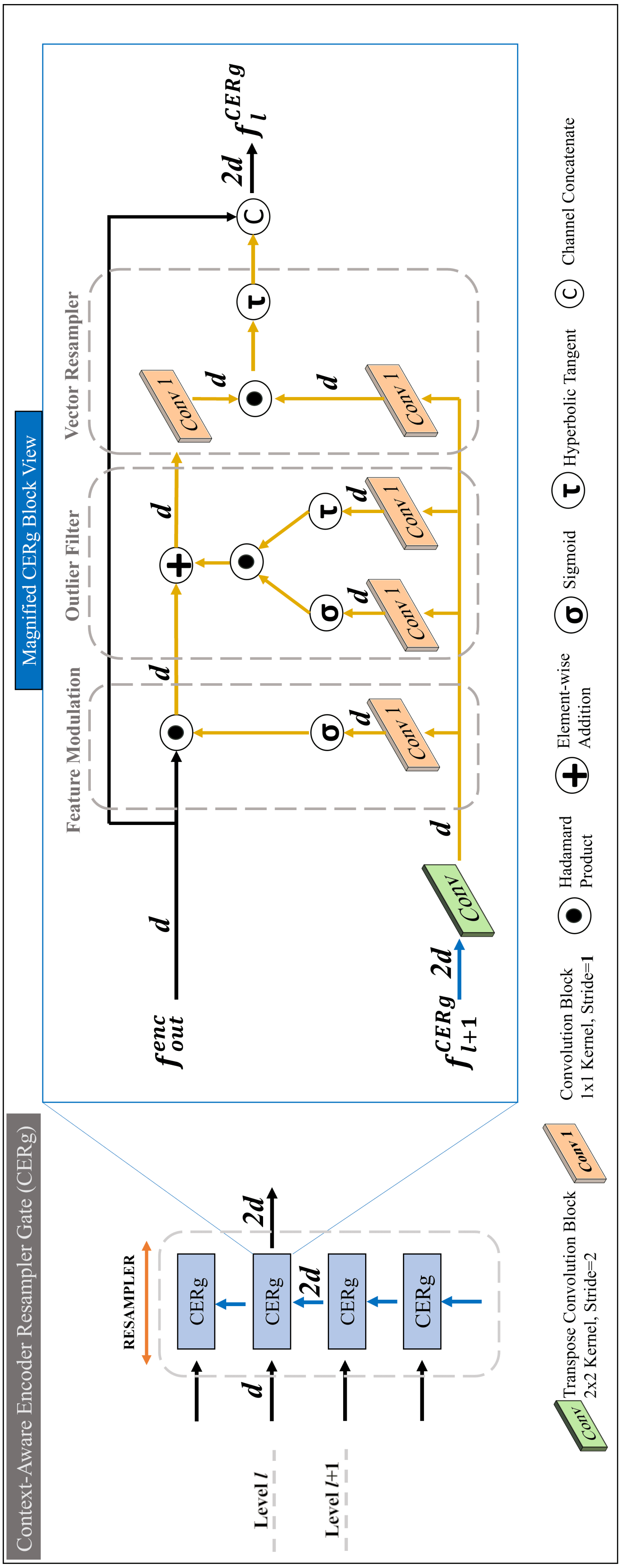


Figure 3.13: A depiction of the $CERg$ Resampler Block.

Let us consider $\mathcal{I}_{gt}^{aug}$ be the ground-truth image of the fundus image $\mathcal{I}^{aug}$. Initially, a Canny edge detector [207] is applied on the image $\mathcal{I}_{gt}^{aug}$ to obtain a binary vessel edge map $\mathcal{I}_{gt}^{C}$, which highlights the vessel edges in the ground-truth, as shown in Figure 3.15. The local pixel direction $d_{x,y}$ is calculated for $q_{x,y,1} \in \mathcal{I}_{gt}^{C}$ as highlighted in [207]. Additionally, $d_{x,y}^{\perp}$ is denoted as the orthogonal direction to $d_{x,y}$ which is directed towards the vessel. To illustrate the formulation, a fragment of a vessel i is considered comprising of two vessel edge fragments E_1^i and E_2^i which form the boundary of the vessel i, as shown in Figure 3.15. For each pixel $q_{x_1,y_1,1} \in E_1^i$, a traversal is conducted along $d_{x_1,y_1}^{\perp}$ till the pixel $q_{x_2,y_2,1} \in E_2^i$ is reached. A hypothetical edge E_{CV}^i is defined, parallel to d_{x_1,y_1} which passes through the mid-point between pixels $q_{x_1,y_1,1}$ and $q_{x_2,y_2,1}$, as shown in Figure 3.15. A window W_k is defined of size $k \times 1$, positioned on E_{CV}^i and oriented along $d_{x_1,y_1}^{\perp}$, intersecting $q_{x_1,y_1,1}$ and $q_{x_2,y_2,1}$ (shown in Figure 3.16). Here. k is equal to the Euclidean distance between pixels $q_{x_1,y_1,1}$ and $q_{x_2,y_2,1}$. All pixels are denoted, which lie within window W_k, as u_j with $j = 0, \ldots, k-1$, where $u_0 = q_{x_1,y_1,1}$ and $u_{k-1} = q_{x_2,y_2,1}$. The probability of a pixel u_j to be the edge pixel is thus computed as

$$\mathcal{P}(q_j) = \delta^j (1-\delta)^{|(k/2)-j|} \tag{3.5.7}$$

In Equation 3.5.7, $\delta = 0.25$ is empirically set in this work. Eq. 16 assigns the highest probability to the edge pixels $q_{x_1,y_1,1}$, $q_{x_2,y_2,1}$ and gradually decreases as j approaches the edge E_{CV}^i, as shown in Figure 3.16. The process is repeated across all pixels $q_{x,y,1}$ in $\mathcal{I}_{gt}^{C}$ to generate the final vessel boundary map I^{Pr}. Finally, the generated probability map $\mathcal{I}^{Pr}$ is used as an auxiliary ground truth to train the model.

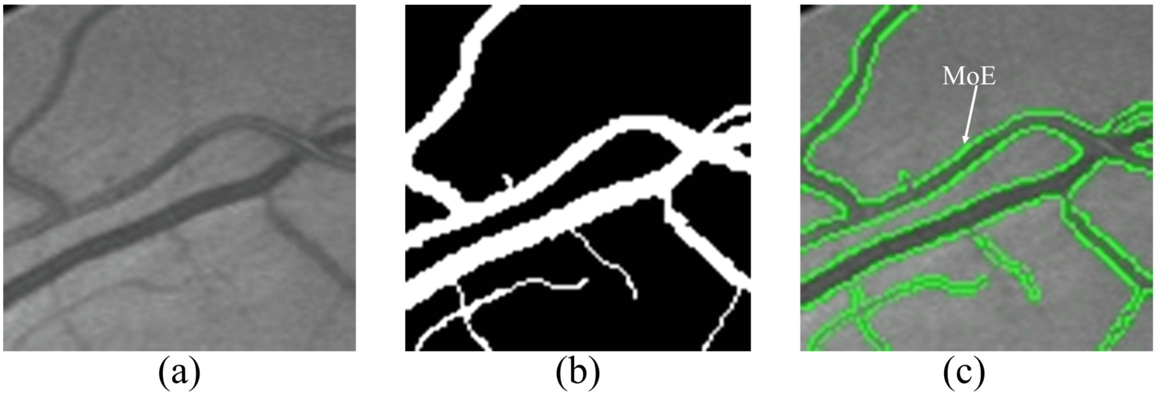
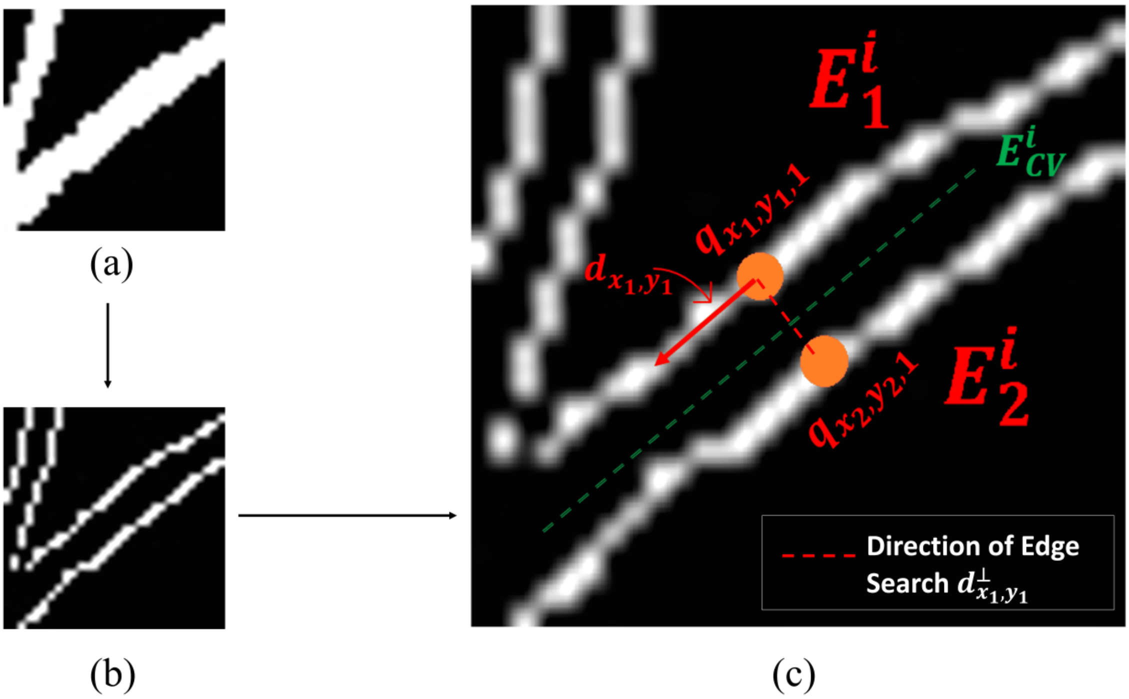


Figure 3.14: illustrates the MoE. (a) fundus image patch, (b) corresponding ground-truth patch, and (c) denotes the MoE (green line).

A Probabilistic Band-Pass Fusion module is proposed to incorporate the probability map in the training process of the network, as shown in Figure 3.17 (Part-1).

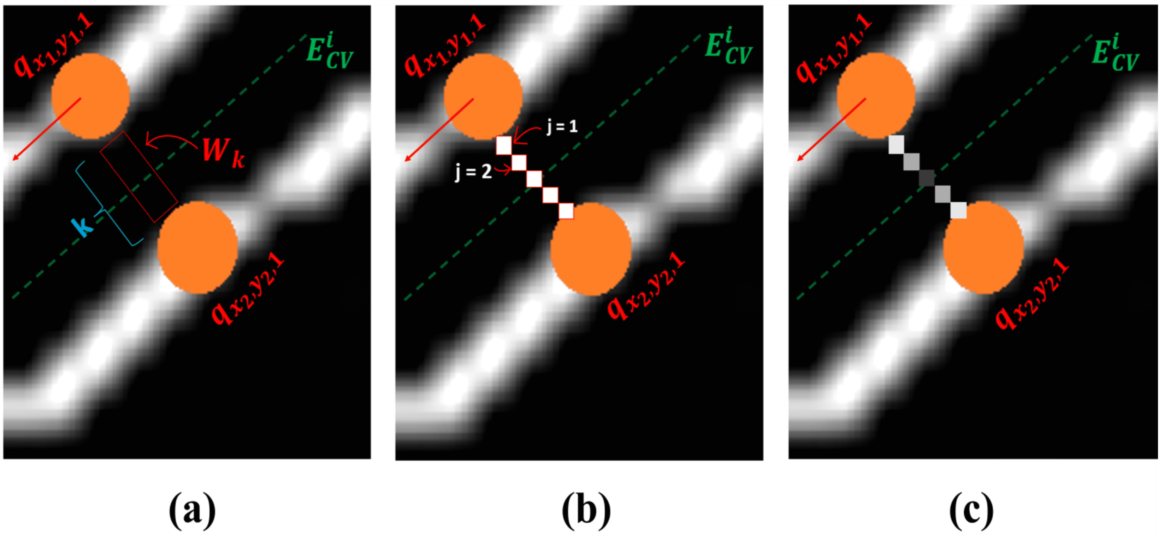


(a)

(b)

(c)

Figure 3.15: (a) vessel segment from a ground-truth, (b) vessel boundary after applying Canny-edge detector, and (c) edge pair searching in context of point $p_{x_1,y_1,1}$ of edge E_1^i.



(a)

(b)

(c)

Figure 3.16: Computation of the probability map for vessel i at point $p_{x_1,y_1,1}$. (a) represents the window W_k of length $k \times 1$ at point $p_{x_1,y_1,1}$, (b) computation of Probability at point q_j (Eq. 13), and (c) probability map generated for vessel i at point $p_{x_1,y_1,1}$.

As depicted in figure, the probability map is convoluted and concatenated, at lower as well as higher-feature level in order to generate two feature maps, namely, m_1 and m_2 respectively. The model is thus further tasked to predict the probability map along with the vessel segmentation, putting emphasis on the MoE region specifically. However, this emphasis on MoE region introduces a susceptibility in the model to noise [208] which in turn gets encoded into the m_1 feature vector. To address the issue, a *Self-Rectified High Pass Gate* (sRHP) is introduced which enables the feature map m_1 to rectify the encoded space and reduce the influence of low-frequency background noise (low contrast regions). In addition, a *Self-Rectified Low Pass Gate* (sRLP) is proposed which enables the higher-order m_2 feature map to remove high frequency spikes in intensities which manifest as speckle noise in segmented outputs of standard CNN models [209]. The sRHP and sRLP gates can be represented mathematically as

$$\begin{aligned} sRHP &= m_1 + \sigma(m_1) \\ sRLP &= m_2 - \sigma(m_2) \end{aligned} \tag{3.5.8}$$

In Equation 3.5.8, it is noted that sRHP module effectively enhances the prominence of dominant features due to the additive operation. Additionally, it enhances the bandwidth between low-frequency noise and information in the feature space, thus making noise more noticeable. In contrast, sRLP depreciates dominant features which leads to better representation of less-prominent features. Simultaneously, a higher frequency noise in the feature space gets better suppressed as well due to the subtractive operation. The combined *sRHP+sRLP Gates* (coined as Band-Pass Gates) thus possess the ability to counter-act the effect of noise induced due to MoE detection. The final segmented output $\mathcal{I}^{RV}$ is thus obtained from the $\hat{p}BPf$ fusion module along with the predicted probability map $\mathcal{I}_{LPr}$, as shown in Figure 3.17 (Part-2). In the subsequent section, the training paradigm used in this work is discussed to train the model.

3.5.0.4 Training and Implementation Details

The proposed $2pCePd$-Net is trained using a generalized Dice Loss function [126], denoted as $L_{\mathcal{I}^{RV}}$, for a better overall segmentation. In addition, the network is trained with an auxiliary objective to optimize the predicted probability map, which is trained using a Mean Squared Error (MSE) function [125], termed as $L_{\mathcal{I}^{LPr}}$.

The loss functions are combined and the final loss function is presented as:

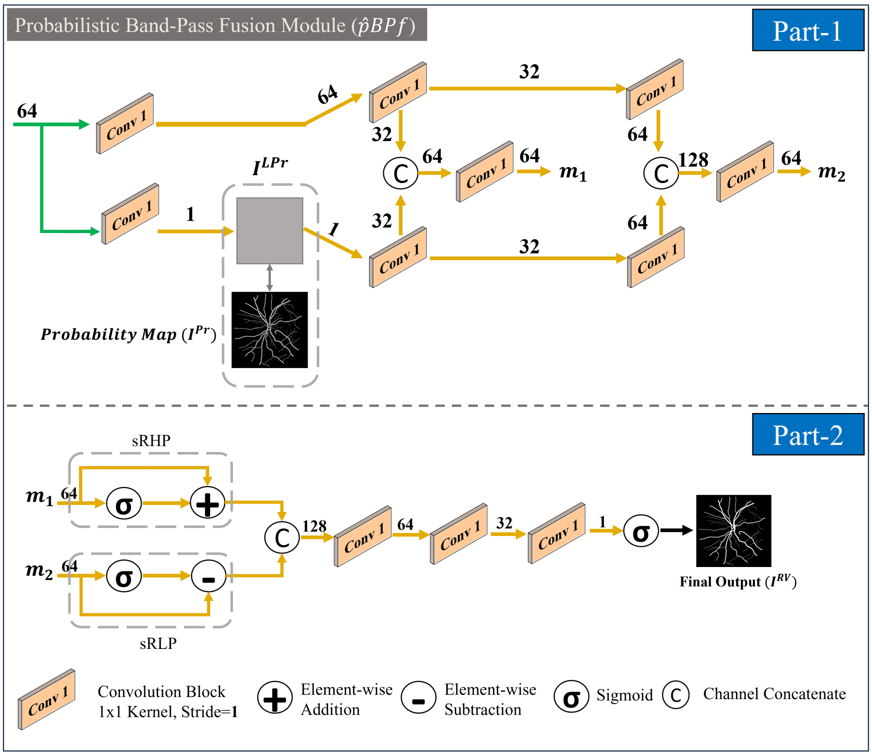


Figure 3.17: A depiction of the $\hat{p}BPf$ Fusion Block.

$$L = \alpha * L_{\mathcal{I}^{RV}} + \beta * L_{\mathcal{I}^{LPr}} \tag{3.5.9}$$

In Eq. 18, α, β are loss coefficient weights, empirically set to 0.6 and 0.4 respectively.

3.5.1 Experimental Results and Discussion

In this section, the segmentation performance of the method is systematically assessed on four public benchmark datasets. Firstly, the fundus images dataset used for experimentation is described. Secondly, the efficacy of the proposed data augmentation method $\lambda CMgC^2$ is evaluated. Finally, the efficacy of the presented segmentation results with respect to individual blocks of the proposed 2pCePd-Net model is evaluated. The model is coded in the PyTorch environment having 64GB of RAM with 2×RTX 3060 GPUs for faster processing. Each individual module of the platform is trained using the Adam optimizer with a maximum iteration of 1000 and a learning rate of $3e^{-4}$, constrained on an Early Stopping criterion based on the Loss per iteration. To prevent overfitting, an L2 Regularization is used in the training as well. All modules are trained with a batch size of 3 to limit the need for excess memory.

3.5.1.1 Qualitative Evaluation of the segmentation Result

Figure 3.18 illustrates examples of fundus images from the DRIVE, STARE, CHASEDB1, and HRF databases and the corresponding segmentation results obtained using the proposed model. First row highlights the original fundus images from DRIVE, STARE, CHASEDB1, and HRF. The second and third rows depict corresponding green channel images and ground-truths. The fourth row highlights the segmentation results from the proposed model. Comparing the segmentation results of the proposed model with the respective ground-truths, it can be observed that the proposed model holds an excellent agreement with the ground-truths.

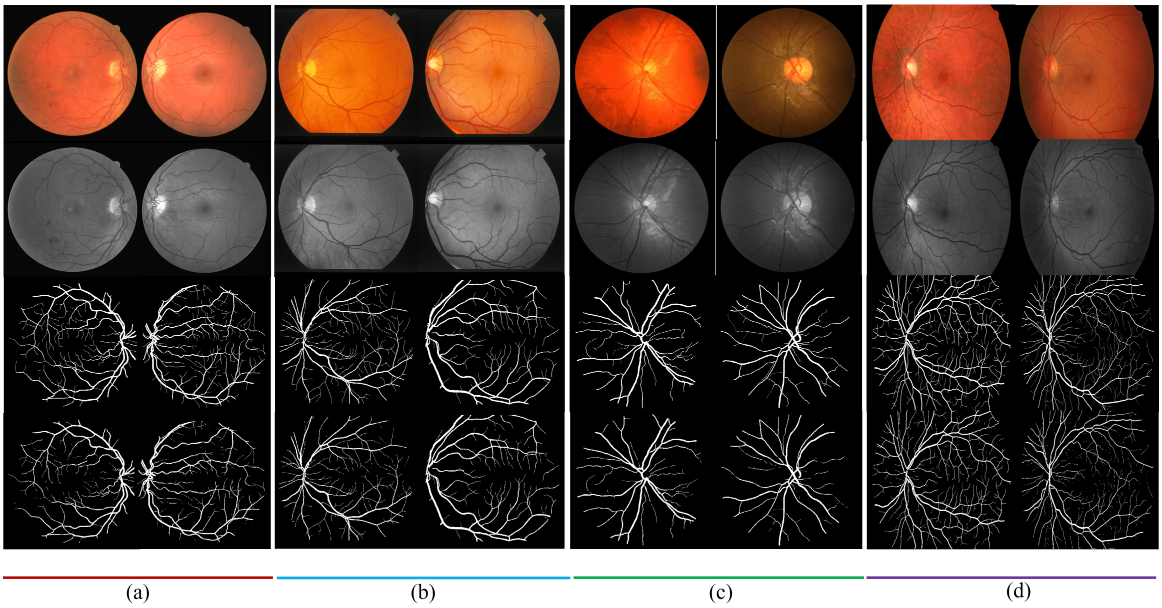


Figure 3.18: Comparison of the method's segmentation results and ground-truths. (a) DRIVE, (b) STARE, (c) CHASEDB, (d) HRF

3.5.1.2 Quantitative Evaluation of the segmentation Result

Quantitative analysis is performed in two stages. Firstly, the consistency of the $2pCePd$-Net segmentation results is evaluated with the ground-truths annotated by experts. In this context, the linear regression plots and the Bland-Altman plots are constructed, shown in Figure 3.19. It can be observed that there is a high correlation (R^2) between the proposed model's segmentation results and the ground-truths for all the four databases, with values of 0.9335 (DRIVE), 0.9212 (STARE), 0.9691 (CHASEDB1), and 0.9502 (HRF), respectively. It is also seen that there are very few vessel regions outside of the 95% confidence interval for all four databases [see lower row of Figure 3.19]. Furthermore, the Cohen's kappa coefficient (κ) [16] is calculated to evaluate the agreement of the proposed method's segmentation results with the ground-truths. In general, the value of k varies in between 0 and 1, and $\kappa \geq 0.8$ signify high consistency between proposed model's results and that of the ground-truths. The obtained kappa (κ) values for the four datasets are: 0.86 (DRIVE), 0.91 (STARE), 0.89 (CHASEDB1), and 0.85 (HRF) respectively. These results ensure that the proposed model is highly consistent with the ground truths.

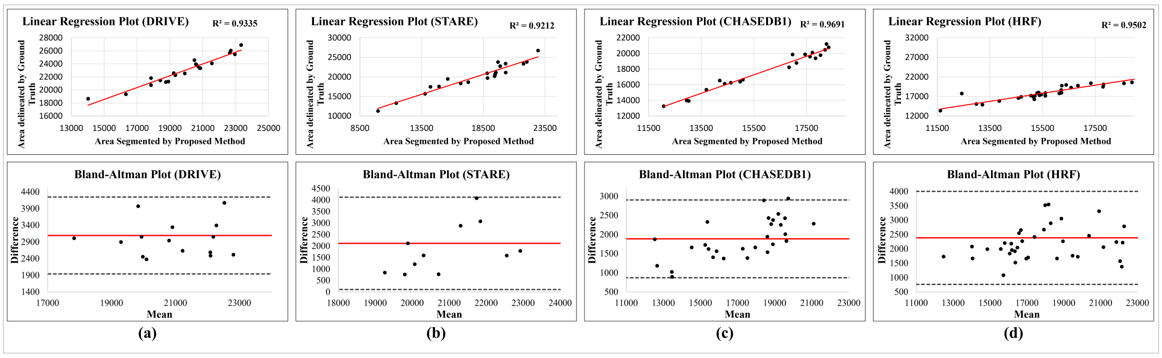


Figure 3.19: Linear Regression and Bland-Altman plots of 2$pCePd$-Net against the ground-truths. (a) DRIVE, (b) STARE, (c) CHASEDB1, (d) HRF

3.5.1.3 Comparative Analysis with State-of-the-Art

Table-3.12 highlights that the model achieves a significantly higher Se (0.832) and at par AUC (0.987) score compared to the other methods listed on the DRIVE dataset. Furthermore, the overall Ac (0.976), Sp (0.989) and F1 (0.857) scores achieved by 2$pCePd$-Net are highest compared to those of the other methods. Furthermore, the model outperforms the existing methods in terms of Ac (0.981), Sp (0.989), Se (0.867), AUC (0.993) and F1 (0.869) scores on the STARE dataset. It is also noted that the achieved Se (0.867) is far superior in comparison to other models. For the CHASEDB1 dataset, 2$pCePd$-Net outperforms the other competing methods in terms of Sp (0.990), Se (0.884), Ac (0.982), F1 (0.871) and AUC (0.992) scores. Table-3.13 highlights that the achieved performance on HRF dataset. It is observed that the proposed model outperforms existing literature with an overall score of Se (0.8259), Sp (0.9896), Ac (0.9770), F1 (0.8468) and AUC (0.9869) respectively. To further investigate the performances, a series of assessment have been conducted on the proposed work and the observations are reported next.

Table 3.12: Performance Comparison on DRIVE, STARE, and CHASEDB1 datasets.

Methods/	**DRIVE**					**STARE**					**CHASEDB1**				
	Se	**Sp**	**Acc**	**F1**	**AUC**	**Se**	**Sp**	**Acc**	**F1**	**AUC**	**Se**	**Sp**	**Acc**	**F1**	**AUC**
[16]	0.781	0.980	0.953	–	–	–	–	–	–	–	–	–	–	–	–
[56]	0.765	0.981	0.954	–	0.975	0.758	0.984	0.961	–	0.980	0.763	0.980	0.961	–	0.978
[165]	0.796	0.980	0.956	0.823	0.980	0.759	0.987	0.964	0.814	0.983	0.815	0.975	0.961	0.788	0.980
[210]	0.813	0.978	0.960	–	–	–	–	–	–	–	–	–	–	–	–
[166]	0.820	0.975	0.954	0.824	0.983	0.829	0.984	0.966	0.836	0.987	0.802	0.981	0.965	0.805	0.986
[167]	0.779	0.983	0.954	0.820	0.984	0.771	0.988	0.970	0.814	0.987	0.797	0.982	0.965	0.807	0.986
[169]	0.799	0.981	0.958	0.829	0.983	–	–	–	–	–	–	–	–	–	–
[168]	0.799	0.981	0.958	0.829	0.982	0.818	0.984	0.967	0.838	0.988	0.824	0.981	0.967	0.819	0.987
[170]	0.836	0.974	0.956	0.827	0.979	0.856	0.982	0.968	0.846	0.987	0.848	0.979	0.966	0.830	0.986
[172]	0.828	0.983	0.969	–	0.983	0.820	0.983	0.973	–	0.984	0.836	0.984	0.974	–	0.986
[171]	0.830	0.975	0.957	0.831	0.982	0.865	0.984	0.972	0.863	0.992	0.846	0.984	0.966	0.822	0.988
[174]	0.832	0.986	0.946	–	–	–	–	–	–	–	0.781	0.979	0.956	–	–
[211]	0.821	0.983	0.968	0.819	0.981	0.819	0.985	0.973	0.819	0.982	0.808	0.984	0.973	0.792	0.983
[212]	0.826	0.983	0.969	0.823	0.983	0.833	0.986	0.975	0.832	0.987	0.822	0.986	0.975	0.809	0.987
[213]	0.819	0.982	0.968	0.816	0.983	0.831	0.984	0.973	0.822	0.987	0.819	0.985	0.974	0.800	0.986
[173]	0.835	0.982	0.970	–	0.986	0.848	0.986	0.976	–	0.990	0.844	0.985	0.976	–	0.989
[214]	0.816	0.976	0.956	–	–	0.790	0.983	0.964	–	–	–	–	–	–	–
[55]	0.828	0.977	0.957	0.830	–	0.828	0.985	0.965	0.836	–	0.830	0.984	0.967	0.830	–
[127]	**0.836**	0.983	0.970	0.832	0.987	0.853	0.987	0.977	0.847	0.992	0.796	0.989	0.977	0.814	0.990
2pCeP*d*	0.832	**0.989**	**0.976**	**0.857**	**0.987**	**0.867**	**0.989**	**0.981**	**0.869**	**0.993**	**0.884**	**0.990**	**0.982**	**0.871**	**0.992**

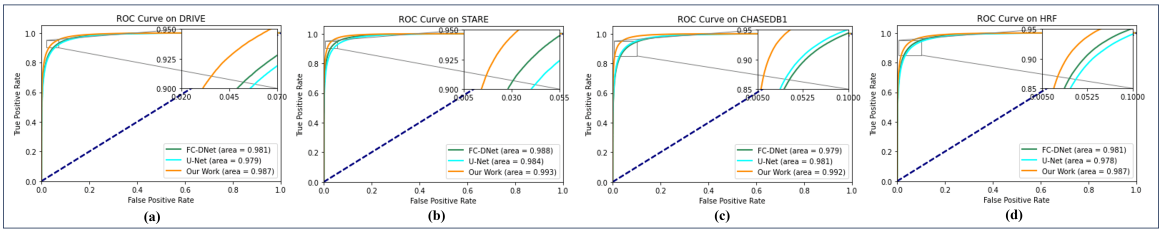


Figure 3.20: The ROC Plot and the AUC scores of Table-I. (a) DRIVE, (b) STARE, (c) CHASEDB1, and (d) HRF

Table 3.13: Performance Comparison on HRF dataset.

Methods	**Se**	**Sp**	**Acc**	**F1**	**AUC**
JSL-PWL [56]	0.8087	0.9417	0.9437	–	–
DU-Net [165]	0.7464	0.9874	0.9651	–	0.9831
HA-Net [168]	0.7803	0.9843	0.9654	0.8074	0.9837
SCS-Net [172]	0.8114	0.9823	0.9687	–	0.9842
Genetic U-Net [171]	0.8220	0.9818	0.9667	0.8179	**0.9872**
LA-Net [211]	0.8022	0.9814	0.9677	0.7921	0.9807
TransUNet [212]	0.8116	0.9834	0.9702	0.8068	0.9841
PLVS-Net [213]	0.8059	0.9819	0.9684	0.7970	0.9827
GT-DLA-dsHFF [173]	0.8178	0.9823	0.9698	–	0.9853
DPL-GTF-EFA [127]	0.8225	0.9840	0.9716	0.8165	0.9867
This Work	**0.8259**	**0.9896**	**0.9770**	**0.8468**	0.9869

Table 3.14: Comparison of the proposed method with the baseline models.

		Traditional			$\lambda CMgC^2$		
Database	**Methods**	**Se**	**Sp**	**Acc**	**Se**	**Sp**	**Acc**
DRIVE [112]	2pCePd	0.825	0.986	0.972	0.832	**0.989**	**0.976**
STARE [30]	2pCePd	0.859	0.987	0.978	**0.867**	**0.989**	**0.981**
CHASEDB1 [128]	2pCePd	0.851	0.989	0.979	**0.884**	**0.990**	**0.982**
HRF [129]	2pCePd	0.822	0.988	0.972	**0.826**	**0.989**	**0.977**

Additionally, the improvement in performance of the $2pCePd$-Net is assessed on the $\lambda CMgC^2$ augmented dataset. The choice of $\lambda CMgC^2$ augmentation approach for the performance analysis hinges on the superior performance of the chosen approach in comparison to other approaches. It is thus noted that the use of augmented data further improves the model performance and establishes a new state-of-the-art in this regard.

3.5.2 Ablation Study

To evaluate the utility of the modules designed for the $2pCePd$-Net model, a series of comparative assessment have been conducted with respect to the DRIVE and STARE datasets and the obtained results are as follows.

Table 3.15: Comparison of Computational Complexity

Methods	**Params. ($1 \mathrm{x} 10^6$)**	**FLOPS ($1 \mathrm{x} 10^9$)**	**Mem. (MB)**	**Time (s)**	**Acc**
U-Net [125]	8.03	51.04	30.67	0.071	0.962
CE-Net [166]	13.40	124.48	110.77	0.076	0.954
Iter-Net [167]	13.60	194.65	52.05	0.107	0.954
TransUNet [212]	**36.46**	225.40	142.14	0.092	0.969
LA-Net [211]	0.44	72.19	1.69	0.306	0.968
PLVS-Net [213]	0.54	75.17	1.90	0.069	0.968
DPF-Net [55]	3.63	56.99	40.44	0.124	0.957
GT-DLA-dsHFF [173]	26.08	**473.92**	125.68	**0.445**	0.970
DPL-GTF-EFA [127]	16.86	256.09	**194.73**	0.153	0.970
This Work	25.14	198.32	105.51	0.108	**0.976**

3.5.2.1 Computational Complexity

An assessment of the computational complexity of the proposed $2pCePd$-Net is conducted and relevant observations are reported in Table-3.15. The complexity analysis is evaluated across four major metrics, namely, structural parameters, floating-point operations per second (FLOPS), memory required and inference time needed for each test sample. The respective Accuracy (Ac) is also presented in Table-3.15 for ease of readability. It is observed that the proposed model has a greater number of structural parameters in comparison to baseline models (such as U-Net [125]). However, the number of parameters is slightly lesser than recent GT-DLA-dsHFF [173] model while achieving a higher overall Accuracy. Additionally, the achieved overall performance is significantly higher than TransUNet model [212] (Table-3.12) in spite of having nearly 24% less parameters. Similar trend is also observed with respect to FLOPS metric. The $2pCePd$-Net model achieves a better Accuracy whilst requiring only 50% of FLOPS in comparison to models having comparable number of parameters such as GT-DLA-dsHFF. Furthermore, the work has a far lesser memory requirement of modest 105.51 MB and inferential time of 0.1083 seconds in contrast to recent models such as GT-DLA-dsHFF and DPL-GTF-EFA [127]. Noticeably, the proposed model has a lesser floating point operation in comparison to DPL-GTF-EFA model in spite of having a larger set of parameters. This can be attributed to the design of an End-to-End Convolutional model in our work which reduces the activation memory footprint [206] and enhances speed of execution. It can thus be inferred that the $2pCePd$-Net model comprises of a moderate complexity and better memory utilization while achieving an improved performance compared to other state-of-the-art approaches. The moderate quantity of parameters coupled with low memory and inferential time establishes the

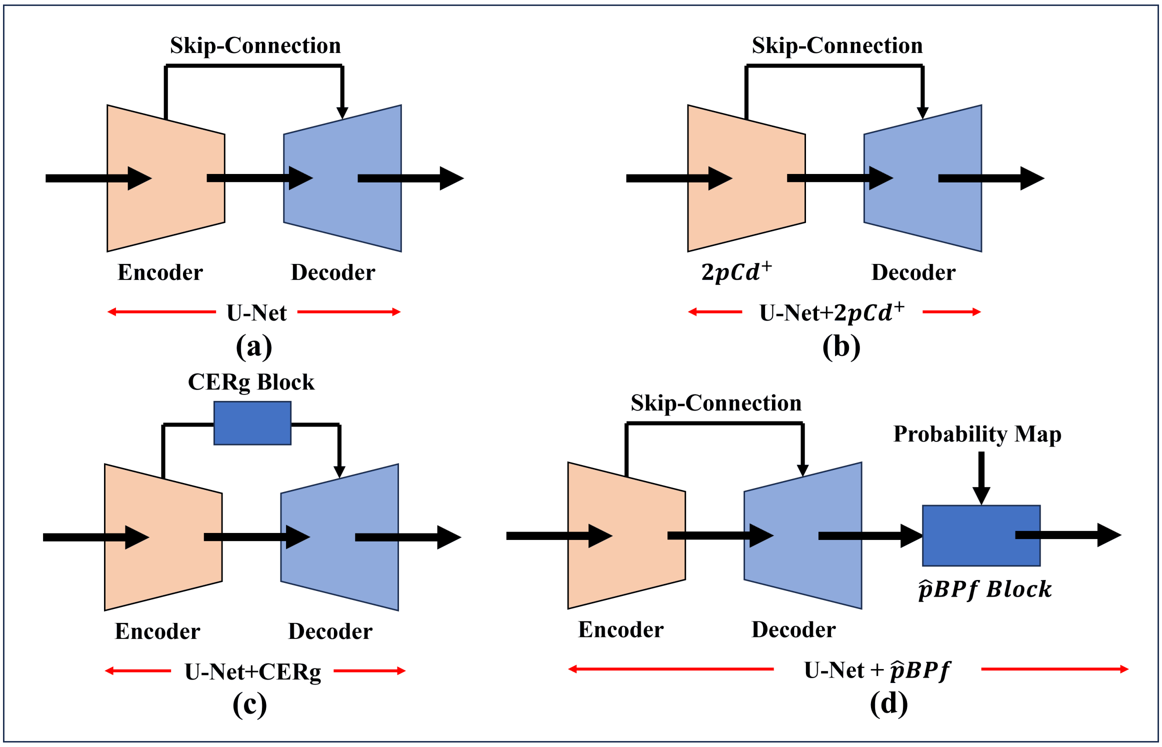


Figure 3.21: Network Block Diagram used for Ablation Study. (a) U-Net (b) U-Net+$2pCd^{+}$ (c) U-Net+CERg (d) U-Net+$\hat{p}BPf$

practicality of the approach.

3.5.2.2 Structural Block-Wise Performance

The individual performances of the proposed $2pCd^{+}$ Encoder, CERg Block and $\hat{p}BPf$ Fusion block, separately, is analyzed by considering a U-Net architecture [125] with level $l = 5$ as a baseline for the ablation study (as shown in Fig. 3.21(a)).

Table 3.16: Ablation Study of the Network Blocks with U-Net.(↑ indicates the increase compared to baseline U-Net.)

DB	Methods	Se (↑x100)	Sp (↑x100)	Acc (↑x100)
DRIVE	U-Net	0.812	0.976	0.962
	U-Net+$2pCd^{+}$	0.815 (↑ 0.3)	**0.981 (↑ 0.5)**	**0.967 (↑ 0.5)**
	U-Net+CERg	0.814 (↑ 0.2)	0.978 (↑ 0.2)	0.964 (↑ 0.2)
	U-Net+(w/o)	0.817 (↑ 0.5)	0.978 (↑ 0.2)	0.963 (↑ 0.1)
	U-Net+$\hat{p}BPf$	**0.821 (↑ 0.9)**	0.978 (↑ 0.2)	0.965 (↑ 0.3)
STARE	U-Net	0.839	0.979	0.970
	U-Net+$2pCd^{+}$	0.842 (↑ 0.3)	0.981 (↑ 0.2)	0.972 (↑ 0.2)
	U-Net+CERg	0.841 (↑ 0.2)	0.979 (– 0.0)	0.970 (– 0.0)
	U-Net+(w/o)	0.843 (↑ 0.4)	0.981 (↑ 0.2)	0.972 (↑ 0.2)
	U-Net+$\hat{p}BPf$	**0.849 (↑ 1.0)**	**0.983 (↑ 0.4)**	**0.974 (↑ 0.4)**

Efficacy with $2pCd^{+}$ Block: To test the capabilities of the **Enc** block, a simple U-Net model is considered and the Encoder arm is replaced with $2pCd^{+}$ Encoder (as shown in Figure 3.21(b)). Table 3.16 highlights the statistical results obtained for both DRIVE and STARE datasets, with and without the proposed block. The obtained results for the U-Net+$2pCd^{+}$ structure demonstrates an overall increase in the performance across different metrics, viz., a mean increase of 0.35% in Accuracy, 0.35% in Specificity and 0.3% in Sensitivity in the presence of the $2pCd^{+}$ block. To gauge the feature representation, the $2pCd^{+}$ block output at $l = 5$ (denoted by f^{enc}) is extracted. A U-Net inspired AutoEncoder is subsequently trained on the annotated vessel images and the bottleneck feature output (f^{auto}) is extracted. f^{auto} is selected for this comparison as it comprises of a compact feature representation on vascular structure [215]. Finally, the correlation measure of Mutual Information ($\mathfrak{I}(.)$) is selected for feature comparison in this work. A correlation measure of $\mathfrak{I}(f^{auto}, f^{enc}) = 0.121$ is achieved for $2pCd^{+}$ block. In contrast, the bottleneck layer of baseline U-Net (f^{unet}) achieved score of $\mathfrak{I}(f^{auto}, f^{unet}) = 0.061$ respectively. This highlights that f^{enc} encodes a better vascular structure information in encoded space, in alignment with the annotated vessel map.

Efficacy with CERg Block: Similar to the above study, the capabilities of the CERg block is tested by designing a U-Net architecture with our proposed CERg skip-connection block (as shown in Fig. 3.21(c)). Table 3.16 highlights the obtained results. Only a marginal improvement is noticeable in the model performance. To further investigate the strength of the CERg block in noise filtration, Gaussian Noise [16] is added into the test dataset and the respective performances are noted. It is observed that baseline U-Net model achieves Se, Sp and Ac mean scores of 0.681, 0.919 and 0.905 respectively on the Gaussian Noise induced data. In contrast, *U-Net+CERg* model achieves a superior 0.741, 0.947 and 0.934 scores respectively in the same task. Thus, the introduction of CERg block enhanced the noise tolerance efficacy in the model, as was the objective.

Efficacy with $\hat{p}BPf$ Fusion Block: To test the $\hat{p}BPf$ Fusion Block, the last output layer of a U-Net architecture (as shown in Fig. 3.21(d)) is replaced and recorded the impact of the proposed block. Two sets of tests are reported in Table 3.16, viz., **a.** without Band-Pass Gate (marked *U-Net+w/o*) and **b.** with both modules (marked *U-Net+*$\hat{p}BPf$). As can be observed in Table 3.16, the fusion block enhances the capabilities of a standard U-Net model with an overall mean increase of 0.35% in Accuracy, 0.3% in Specificity and 0.95% in Sensitivity respectively. It is noted that $\hat{p}BPf$ Fusion block has the highest overall impact on STARE dataset performance. This can be attributed to the use of a secondary objective function (MSE Loss) which leads to improvement. However, the presence of lower contrast regions in DRIVE enforces additional complexity on MoE detection [16]. A well-informed feature space is thus necessary to tackle this challenge. Hence, $2pCd^+$ achieves a better performance on DRIVE. Furthermore, *w/o* set achieves a slightly diminished performance in comparison. This can be attributed to a poor performance on noise filtration, introduced by the MoE objective [208], without Band-Pass Gate. The complete $\hat{p}BPf$ layer however does not suffer from such issues and achieves a better performance. Thus, it is noted that all three modules have individual roles in the network and are co-dependent to achieve the overall performance.

3.5.2.3 Encoder Scalability

It is observed from the approach that a more generalized Encoder system can be designed with an $npCd^+$ architecture, wherein n represents the number of Cd^+ blocks. Table-3.17 represents the performance analysis of such a generalized encoder block with differing values of n. Both the quantitative performance and the computational complexity of the overall model have been reported for each value of n. It is noted that the model is fairly robust in terms of complexity and scales modestly as shown in Table-3.17. Each addition of an encoder block to the model incurred only 3 MB of additional memory. Furthermore, it is observed that the cumulative computational complexity is still at par with current state-of-the-art models (Table-3.15) in spite of additional encoder blocks in the model. However, addition of Encoder blocks results in a dip of Accuracy, as shown in Table-3.17, beyond $n = 2$. This can be attributed to overfitting of features by the Encoder blocks due to excess complexity. Thus, $n = 2$ had been empirically selected in this work which lead to the design of the $2pCePd$-Net model.

Table 3.17: Encoder Scalability Performance Analysis

#Cd^+ Blocks	**Parameters ($1\text{x}10^6$)**	**FLOPS ($1\text{x}10^9$)**	**Memory (MB)**	**Accuracy**
$1 - Cd^+$	26.87	266.37	102.30	0.961
$2 - Cd^+$	27.65	273.82	105.51	**0.976**
$3 - Cd^+$	28.44	281.26	108.19	0.972
$4 - Cd^+$	29.23	288.71	111.97	0.973
$5 - Cd^+$	30.01	296.16	114.50	0.970

Table 3.18: Comparative Assessment on Diseased Images

Database	Method	Sensitivity	Specificity	Accuracy
DRIVE	U-Net [125]	0.7546	0.9790	0.9573
	CE-Net [166]	0.8149	0.9794	0.9650
	Iter-Net [167]	0.8234	0.9818	0.9679
	GT-DLA-dsHFF [173]	0.8259	**0.9821**	0.9684
	Our Work	**0.8309**	0.9811	**0.9692**
STARE	U-Net [125]	0.7783	0.9779	0.9631
	CE-Net [166]	0.7945	0.9829	0.9689
	Iter-Net [167]	0.8087	0.9791	0.9665
	GT-DLA-dsHFF [173]	0.8328	0.9835	0.9724
	Our Work	**0.8451**	**0.9847**	**0.9749**

3.5.2.4 Segmentation in Complex Regions

To evaluate the robustness of the model on diseased images, $2pCePd$-Net performance is evaluated on seven diseased images from DRIVE dataset and ten diseased images from STARE dataset respectively. Table 3.18 presents the comparative analysis of the achieved results. It is observed that the proposed model achieves a better overall performance on the diseased data compared to other contemporary models. In comparison to $2pCePd$-Net performance in Table 3.12, the proposed model experiences a modest drop of 0.3% in Accuracy across both datasets on diseased images. Notably, a marginal drop of only 0.1% in comparison to literature is noted in Specificity on DRIVE dataset. However, the Sensitivity attained is far superior in contrast to existing literature. This establishes the performance attained by the work in diseased dataset. Fig. 3.22 visually represents the model performance on diseased images.

3.5.2.5 Cross-Validation

To evaluate the generalization of $2pCePd$-Net, a cross-validation experiment was performed on the DRIVE and STARE dataset respectively. Table-3.19 highlights the comparative assessment of the statistical results. It is noted from Table 3.19 that the work achieves a better overall performance compared to literature. The model trained on DRIVE and tested on STARE dataset achieves an overall increase of 0.4% and 0.5% with respect to current benchmark approaches across Specificity and Accuracy respectively. A drop of 0.3% in Sensitivity is noted which can probably be attributed to a better generalized vessel feature learned from DRIVE dataset in DPL-GTF-EFA model [127]. Nonetheless, in the approach, a better overall performance is achieved when trained on STARE and tested on DRIVE dataset with an increase of 0.9%, 0.2% and 0.3% respectively across the three parameters of evaluation. This

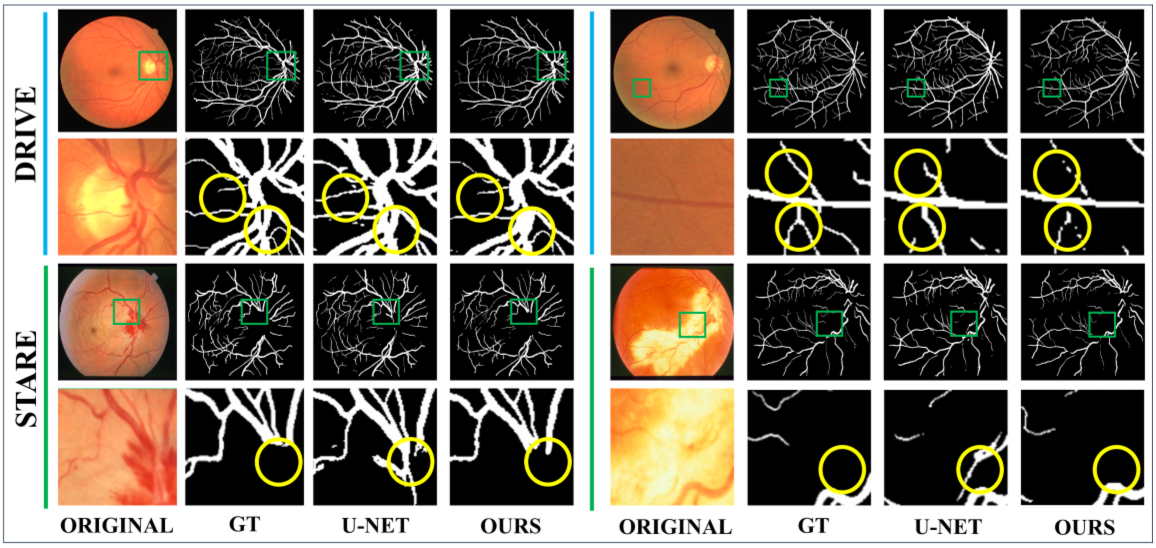


Figure 3.22: Qualitative Analysis of $2pCePd$-Net segmentation Results in Diseased images. Odd rows display the images with critical regions marked in green. Even rows display the magnified view of the marked green regions. Yellow markers highlight the areas of major challenges.

indicates the robustness and generalization potential of the work.

Table 3.19: Comparative Assessment of Cross-Validation Results

Train	Test	Model	Sensitivity	Specificity	Accuracy
DRIVE	STARE	DPF-Net [55]	0.749	0.983	0.953
		DPL-GTF-EFA [127]	**0.825**	0.984	0.972
		Our Work	0.822	**0.988**	**0.977**
STARE	DRIVE	DPF-Net [55]	0.746	0.984	0.948
		DPL-GTF-EFA [127]	0.810	0.982	0.967
		Our Work	**0.819**	**0.984**	**0.970**

3.5.2.6 Extendibility

The $2pCePd$-Net model with $\lambda CMgC^2$ augmentation is shown to improve upon existing works in literature. However, one point of study is the scalability. As shown in Fig.3.23(a), the number of parameters rises in cubic series if the number of kernels is increased in each layer. In Fig.3.23(a), x-axis corresponds to the number of kernels in layer $l = 1$ and denotes an equivalent increase across all model layers overall. A similar trend is, however, also observed in a baseline U-Net model (as shown in Fig. 3.23(a)). It is thus noted that scalability is an architectural challenge inherent to Encoder-Decoder models. Simultaneously, the choice of speed against accuracy is an additional challenge in deep learning (DL) paradigm. The number of floating point operations increases with increasing layer complexity, in alignment with DL paradigm (shown in Fig. 3.23(b)). However, in contrast, Accuracy is observed to increase with layer complexity and reach a peak at 64 (Fig. 3.23(c)). The subsequent increase in complexity leads to overfitting which diminishes the achieved Accuracy. As diagnostic accuracy is crucial in medical image segmentation, providence has been given to Accuracy over Speed and layer complexity of 64 is chosen. It is also noted, from Fig.3.23(d), that an increase in image resolution leads to an exponential demand on FLOPS. This limits the current model to consider image resolutions upto 512×512 dimension for inference in reasonable computing time.

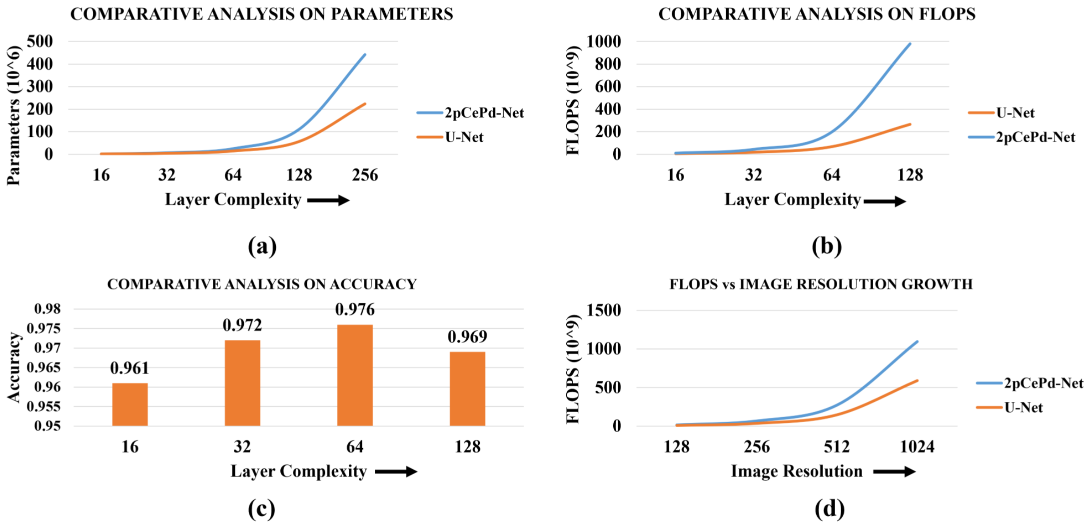


Figure 3.23: An comparative assessment of limitations. (a)-(b) Scalability Analysis of Computational Complexity (c)-(d) Speed vs Accuracy Analysis

Furthermore, the application of $2pCePd$-Net model on other bio-markers of Fundus Images, viz., Optic Disc and Exudates, presents opportunity for further improvement. To establish a baseline, the $2pCePd$-Net+(w/o) model is used to evaluate the potential of the model for biomarker segmentation, apart from retinal blood vessels. The (w/o)- model version is utilized as the probability-map fusion module requires a probability map to be generated which is inconsequential at a minute scale such as Exudates. Table 3.20 highlights the results obtained upon application of the model on the Optic Disc and Hard Exudates segmentation challenges from IDRiD dataset. The model performance with training on *Dice Loss*, *Dice Loss + L2 Regularization* and *Dice Loss + Binary Crossentropy Loss* is highlighted in the tabulation.

Table 3.20: Performance Comparison of $2pCePd$-Net+(w/o) model on IDRiD dataset

Task	**Loss**	**Dice**
Optic Disc	$\mathcal{L}^{DC}$	0.820
	$\mathcal{L}^{DC+L2}$	0.822
	$\mathcal{L}^{DC+BCE}$	0.825
Hard Exudates	$\mathcal{L}^{DC}$	0.671
	$\mathcal{L}^{DC+L2}$	0.592
	$\mathcal{L}^{DC+BCE}$	0.702

It is noted that the *Dice Loss + Binary Crossentropy Loss* presents a better performance in comparison to other losses. In the subsequent Chapter, a study is presented along with the exploration of a loss projection approach for improving the performance of the model in the domain of Bio-Marker segmentation.

3.6 Summary

This chapter introduced the development and evaluation of new deep learning models specific to the analysis of retinal fundus images. Utilizing the latest developments in convolutional and attention architectures, the models were crafted to meet fundamental challenge of fundus image interpretation, viz., vessel segmentation. The models were exhaustively tested on publicly available and augmented data to guarantee an adequate performance evaluation. The findings proved that the methodologies suggested in this study outperformed all the current state-of-the-art solutions on measures, specifically, accuracy, specificity, and sensitivity respectively. These results prove the effectiveness of the architectures designed for identifying intricate visual patterns present within retinal images. Overall, the conclusions present evidence for the growing body of research on AI-supported ophthalmic retinal blood vessel segmentation and prove the usefulness of deep learning in improving automated screening devices for retinal conditions.

Chapter 4

Bio-Marker Segmentation from Fundus Images

4.1 Introduction

The ocular network is a complex biological structure which is examined for diagnosis and possible prevention of blindness-inducing diseases such as Glaucoma and Diabetic Retinopathy [216]. The objective of fundus image bio-marker segmentation is to properly outline anatomical features like the optic disc, vessels, and macula. In this regard, Fundus Imaging is an imaging modality that provides a detailed view of the retinal structures. Successful and correct analysis of the retinal components thus forms an important aspect for the design of an automated system. Past researches with hand-crafted features [30, 112, 217, 218] had made significant progress in this regard in spite of data scarcity and limited technology. However, the vascular complexity of the retina, Optic Oclusion and Irregular Contrast, often increases the challenge of automated segmentation of components, especially in low-contrast regions of images.

Deep learning based models have attempted, in past literature [219], to address the complexity of bio-marker segmentation in Fundus Images. End-to-end Convolutional network models [220], [221], [222], have especially proven their ability to outperform the traditional hand-crafted benchmark models in this regard. The ability to extract localized features and infer complex structures layer by layer have collectively enhanced the prowess of neural models in segmentation tasks. It is however noted that the performance of a neural model is dependent simultaneously on the architecture, training paradigm and loss-based objective functions [223].

4.1.1 Understanding Loss Functions

A loss function measures the disparity between the segmentation output predicted by the model and the ground truth annotation [224]. The deep learning model adjusts its parameters during training to reduce this loss. The design and selection of the loss function have a direct impact on what the model learns from data, especially in domains like medical image segmentation where data can be imbalanced or noisy [225].

4.1.1.1 Segmentation Challenges in Fundus Imaging

- **Class Imbalance [226]:** Class imbalance is a common predicament found in Fundus images. For instance, blood vessels will cover a minimal area of the image in relation to the background. Standard loss functions such as cross-entropy are not necessarily good at such imbalances and hence perform poorly on minority classes.

- **Fine Structures [227]:** Anatomical features of fundus images tend to be slender and complex. Their recording should be with a loss function that penalizes mismatches more severely in the case of small, complex regions than in large uniform areas.

- **Boundary Precision [228]:** Precise boundary delineation is another important aspect in automated segmentation, particularly for boundaries such as optic disc and optic cup. Loss functions used for training a model must be boundary-sensitive to mismatches.

4.1.1.2 Commonly Utilized Loss Functions

- **Binary Cross-Entropy (BCE) Loss [229]:** BCE is commonly used in binary segmentation tasks. Each pixel is independently considered for computation, and the loss between the predicted class labels and ground-truth labels is computed. However, it is susceptible to class imbalance and might fail for thin objects such as blood vessels.

- **Dice Coefficient Loss [230]:** The Dice loss, which is based on the Dice similarity coefficient, measures the overlap between predicted and ground truth masks directly. It is especially helpful for imbalanced data since it concentrates on the region of interest.

- **Jaccard Loss (IoU Loss) [231]:** Similar to Dice loss, Jaccard loss also calculates the intersection over union of the predicted and ground truth masks. It is stronger than Dice and can result in improved boundary alignment.

- **Tversky Loss [232]:** It is an extension of the Dice loss that adds trainable parameters to regulate the balance between false positives and false negatives. It is particularly useful for segmenting small structures in fundus images.

- **Focal Loss [233]:** Initially developed for object detection, focal loss corrects class imbalance by reducing the weight of simplistic training samples and rewarding more attention to difficult samples. It is usually used with Dice or BCE for better segmentation performance.

- **Boundary Loss [234]:** This loss function specifically targets reducing the distance between the boundaries of the predicted and ground truth masks. It is useful in enhancing edge alignment.

4.1.1.3 Regularization Through Loss Functions

Loss functions have the ability to regularize the learning process implicitly by instructing the model to attend to relevant patterns while deterring overfitting. Regularization via loss functions promotes generalizability [235], particularly in fundus image segmentation where labeled data is sparse and highly diverse.

- **Boundary Regularization [235]:** Boundary loss functions prompt the model to take more notice of the exact borders of anatomical structures. It serves as regularization by favoring structural correctness rather than pixel-level correctness. By doing so, the model discourages overfitting to the dominant background class and becomes more sensitive to decisive boundaries, for example, boundaries of the optic disc and optic cup.

- **Attention-Based Losses [236]:** Losses that include spatial or channel attention mechanisms enable the model to pay attention to diagnostically significant regions. These mechanisms act as implicit regularizers by emphasizing informative features and reducing noise and redundancy. In the case of fundus images, this translates to improved attention on lesions, vessels, and the macula.

- **Topological Regularization [237]:** Topology-sensitive loss functions guarantee the model retains the appropriate topological relationships of the segmented structures. For instance, maintaining continuity in blood vessel segmentation ensures clinical interpretability is maintained. Such regularization prevents the model from learning anatomically unrealistic representations.

- **Adversarial Losses as Regularizers [238]:** In segmentation models based on GAN, adversarial loss is a powerful regularizer in terms of inducing global coherence in the predictions. The discriminator forces the segmentation model to produce outputs close to real anatomical masks and avoids unrealistic segmentations that fit training data perfectly but are incorrect on novel images.

- **Uncertainty-Aware Losses [239]:** The incorporation of uncertainty into the loss function enables the model to communicate its confidence in its predictions.

The regularization here avoids overconfident but inaccurate outputs, hence promoting conservative, trustworthy predictions particularly in uncertain regions.

- **Multi-Task Losses [240]:** Loss functions that optimize segmentation concurrently with auxiliary tasks (such as classification or landmark detection) incorporate implicit regularization. The shared task representation avoids overfitting on a specific goal and results in more robust learning.

In this regard, the efficacy of the Dice Loss [241] and its variants in medical image segmentation tasks have been highlighted extensively in recent works [55, 173, 174]. However, in such scenarios, the projection of per pixel loss, solely based on distance between model predicted probability and ground truth, might achieve subpar performance [242]. A dynamic loss computation paradigm for each pixel is hence necessary to account for limitations due to model architecture. Subsequently, the per-pixel loss projection is modified based upon additional parameters, namely, adjacent pixel performance, pixel saturation etc. To address the issue of loss performance, Regularization [243] had been highlighted as a measure in past literature to prevent model overfitting. In addition, the use of a secondary objective function [56, 170] is also often highlighted in related works. However, both measures lack a dynamic framework to address and modify the loss projections based upon individual pixel performances. To address the issues, a dynamic framework of loss computation for individual pixel performances is presented henceforth.

4.2 Methodology

Let us consider a trainable model $\mathcal{M}^{\theta}$ for region segmentation of an image $\mathcal{I}$ of size HxW with model weight configuration θ. Additionally, the segmentation *ground truth* (GT) of image $\mathcal{I}$ is defined as $\mathcal{I}^{gt}$ respectively, comprising of two classes of regions, namely, *Background Pixels* (denoted as 0) and *GT Mask Pixels* (denoted as 1). The prediction loss ($\mathcal{L}$) of model $\mathcal{M}^{\theta}$, is formulated mathematically, as

$$\mathcal{L} = \Phi(\mathcal{M}^{\theta}(\mathcal{I}), \mathcal{I}^{gt}) \tag{4.2.1}$$

where Φ represents the objective function and $\mathcal{M}^{\theta}(\mathcal{I}) \in [0, 1]$. A successful learning is usually attributed to the minimization of the loss $\mathcal{L}$. For a pair of iterations $\hat{i} =< i-k, i >$, the respective loss pair is considered as $< \mathcal{L}_{i-k}, \mathcal{L}_i >$ and define *Iterative Difference of Loss at iteration i* ($\nabla_{\hat{i}}\mathcal{L}$) as

$$\nabla_{\hat{i}}\mathcal{L} = \mathcal{L}_i - \mathcal{L}_{i-k} = \Phi(\mathcal{M}_i^{\theta}(\mathcal{I}), \mathcal{I}^{gt}) - \Phi(\mathcal{M}_{i-k}^{\theta}(\mathcal{I}), \mathcal{I}^{gt}) \tag{4.2.2}$$

where $0 < k \leq i$ is the window of iteration pair. Eq. 4.2.2 refers to *difference of loss* $\nabla_{\hat{i}}\mathcal{L}$ having $\mathcal{M}_i^{\theta}$ denoted as the prediction of model $\mathcal{M}$ at a given iteration i. Equation 4.2.2 is generalized by computing the partial derivative of $\mathcal{L}$ with respect to iterations as

$$\frac{\partial \mathcal{L}}{\partial i} = \left(\frac{\partial \mathcal{L}}{\partial P}\right)\left(\frac{\partial P}{\partial i}\right) \tag{4.2.3}$$

where P corresponds to predicted probability at the output layer of model $\mathcal{M}^{\theta}$. From Equation 4.2.3, it is trivial to note that the rate of change of loss is directly correlated with the rate of change of predicted probabilities (P) with respect to iterations (i). Additionally, it can be observed from Eq. 4.2.3 that the saturation of loss convergence ($\frac{\partial \mathcal{L}}{\partial i} \to 0$) can be attributed to two possible scenarios:

Case 1: Maximal optimization of objective function, i.e., $\lim_{i\to\infty} \mathcal{M}_i^{\theta}(\mathcal{I}) = \mathcal{I}^{gt}$.

Case 2: Negligible update across iterations for a subset of predicted probabilities, i.e., $\frac{\partial P}{\partial i} \to 0$.

Case 2 highlights a critical issue which often results in an unstable training environment for the model leading to a state of impaired learning [244]. In such scenarios, a modification of the computed loss [243] is often times beneficial to mitigate the subpar learning. To address Case 2, an *Iteration-Directed Differential Weighted Kernel with Counter-based Depreciation Normalized* ($\mathcal{L}^{\alpha N^{++}}$) loss projection can be used.

4.2.1 Iteration-Directed Differential Weighted Kernel

Let us consider the model predictions for class $\mathcal{C}$ set of pixels at iteration pair $\hat{i} =< i-k, i >$ as $< \mathcal{M}_{i-k}^{\theta_\mathcal{C}}(\mathcal{I}), \mathcal{M}_i^{\theta_\mathcal{C}}(\mathcal{I}) >$. *Iterative Difference of Probabilities* ($\nabla\mathcal{P}_{\hat{i}}^{\mathcal{C}}$) at iteration pair $\hat{i}$ is defined as

$$\nabla\mathcal{P}_{\hat{i}}^{\mathcal{C}} =\mid \mathcal{M}_i^{\theta_\mathcal{C}}(\mathcal{I}) - \mathcal{M}_{i-k}^{\theta_\mathcal{C}}(\mathcal{I}) \mid \tag{4.2.4}$$

It is noted that pixel having $\nabla\mathcal{P}_{\hat{i}}^{\mathcal{C}} \simeq 0$ exhibit the issue highlighted in Case 2. To emphasize on the pixels exhibiting $\nabla\mathcal{P}_{\hat{i}}^{\mathcal{C}} \simeq 0$, a weighted-kernel is required to increase the loss projection of non-performing pixels based on the difference of predicted probabilities. With this motivation, an *Iteration-Directed Differential Weighted Kernel* $\alpha_{\hat{i}}^{\mathcal{C}}$ for class $\mathcal{C}$ and iteration pair $\hat{i}$ is defined, and equated as

$$\alpha_{\hat{i}}^{\mathcal{C}} = \frac{(\mathcal{M}_i^{\theta_{\mathcal{C}}}(\mathcal{I}) - \mathcal{I}^{gt})}{\mid \mathcal{M}_i^{\theta_{\mathcal{C}}}(\mathcal{I}) - \mathcal{I}^{gt} \mid} \odot \left(1 - \frac{\nabla \mathcal{P}_{\hat{i}}^{\mathcal{C}}}{\mid \sum \nabla \mathcal{P}_{\hat{i}}^{\mathcal{C}} \mid}\right) \tag{4.2.5}$$

where $\odot$ is the Hadamard Product, $\frac{(\mathcal{M}_i^{\theta_{\mathcal{C}}}(\mathcal{I}) - \mathcal{I}^{gt})}{|\mathcal{M}_i^{\theta_{\mathcal{C}}}(\mathcal{I}) - \mathcal{I}^{gt}|}$ introduces direction to the kernel (Positive for Background Pixels and Negative for GT Mask Pixels) as well as attenuate the projected loss for pixels which satisfy Case 1. $\mid \sum \nabla \mathcal{P}_{\hat{i}}^{\mathcal{C}} \mid$ is used for normalization in Equation 4.2.5. The weight-kernel $\alpha_{\hat{i}}^{\mathcal{C}}$ is thus incorporated as an objective function Ω to form the $\mathcal{L}^{\alpha N^{++}}$ loss projection, which is equated as

$$\begin{aligned}
\mathcal{L}_i^{\alpha N^{++}} &= \sum_{\mathcal{C}} \Omega((1 + \alpha_{\hat{i}}^{\mathcal{C}}) \odot \mathcal{M}_i^{\theta_{\mathcal{C}}}(\mathcal{I}), \mathcal{I}^{gt}) \\
\frac{\partial \mathcal{L}_i^{\alpha N^{++}}}{\partial \theta} \\
&= \sum_{\mathcal{C}} (1 + \alpha_{\hat{i}}^{\mathcal{C}})\, \Omega^{'}((1 + \alpha_{\hat{i}}^{\mathcal{C}}) \odot \mathcal{M}_i^{\theta_{\mathcal{C}}}(\mathcal{I}), \mathcal{I}^{gt}) \left(\frac{\partial \mathcal{M}_i^{\theta_{\mathcal{C}}}(\mathcal{I})}{\partial \theta}\right)
\end{aligned} \tag{4.2.6}$$

where $\mathcal{L}_i^{\alpha N^{++}}$ is the proposed projected loss at iteration i. $\mathcal{L}^{\alpha N^{++}}$ is thus added to the loss function, defined by Eq. 4.2.1, to form a dual-loss objective and approach better convergence. It is also highlighted that differentiation with respect to $\alpha_{\hat{i}}^{\mathcal{C}}$ is not accounted for in Equation 4.2.6 as the kernel is recomputed at every iteration and is thus modified with respect to θ already. Equation 4.2.1 is thus modified and the total loss is defined at iteration i as

$$\mathcal{L}_i^{proj} = \beta_1 \Phi(\mathcal{M}^{\theta}(\mathcal{I}), \mathcal{I}^{gt}) + \beta_2 \mathcal{L}_i^{\alpha N^{++}} \tag{4.2.7}$$

where β_1, β_2 are coefficients empirically set to 0.5. As can be observed in Equation 4.2.6, the loss projection term is differentiable *iff* the objective function Ω used is differentiable thus establishing its usability.

It is noted that all formulations have been defined considering a loss minimization framework. Furthermore, background pixels are denoted as **0** and mask pixels as **1** in conformance with literature. The following characteristics are noted from the formulation. Firstly, $\alpha_{\hat{i}}^{\mathcal{C}}$ is bound within limits $[-1, 1]$ as $0 \leq \frac{\nabla \mathcal{P}_{\hat{i}}^{\mathcal{C}}}{|\sum \nabla \mathcal{P}_{\hat{i}}^{\mathcal{C}}|} \leq 1$ and $\frac{(\mathcal{M}_i^{\theta_{\mathcal{C}}}(\mathcal{I}) - \mathcal{I}^{gt})}{|\mathcal{M}_i^{\theta_{\mathcal{C}}}(\mathcal{I}) - \mathcal{I}^{gt}|} \in \{-1, 1\}$. Furthermore, it can be implied from limits of $\alpha_{\hat{i}}^{\mathcal{C}}$ that for all pixels, $0 \leq (1 + \alpha_{\hat{i}}^{\mathcal{C}}) \leq 2$. To gauge the impact of the formulation, three independent observations are noted as follows.

- $\boldsymbol{((1 + \alpha_{\hat{i}}^{\mathcal{C}}) >> 1)}$: It is noted that for all $(1 + \alpha_{\hat{i}}^{\mathcal{C}}) >> 1$, $(1 + \alpha_{\hat{i}}^{\mathcal{C}}) \odot \mathcal{M}_i^{\theta_{\mathcal{C}}}(\mathcal{I}) >> \mathcal{M}_i^{\theta_{\mathcal{C}}}(\mathcal{I})$. This implies that for all background pixels in $\mathcal{I}^{gt} \in \{0\}$, *iff* $\nabla \mathcal{P}_{\hat{i}}^{\mathcal{C}} \simeq 0$, then for an objective function $\Omega(.)$, $\Omega((1+\alpha_{\hat{i}}^{\mathcal{C}})\odot\mathcal{M}_i^{\theta c}(\mathcal{I}),\mathcal{I}^{gt}) >> \Omega(\mathcal{M}_i^{\theta c}(\mathcal{I}),\mathcal{I}^{gt})$ (as $\mathcal{M}_i^{\theta c}(\mathcal{I}) > \mathcal{I}^{gt}$). In other words, the loss projection of a non-performing pixel is elevated for background pixels having $\mathcal{I}^{gt} \in \{0\}$, irrespective of the choice of $\Omega(.)$.

- $\boldsymbol{((1+\alpha_{\hat{i}}^{\mathcal{C}}) << 1)}$: It is noted that for all $(1+\alpha_{\hat{i}}^{\mathcal{C}}) << 1$, $(1+\alpha_{\hat{i}}^{\mathcal{C}})\odot\mathcal{M}_i^{\theta c}(\mathcal{I}) << \mathcal{M}_i^{\theta c}(\mathcal{I})$. This implies that for all mask pixels in $\mathcal{I}^{gt} \in \{1\}$, *iff* $\nabla\mathcal{P}_{\hat{i}}^{\mathcal{C}} \simeq 0$, then $\Omega((1+\alpha_{\hat{i}}^{\mathcal{C}})\odot\mathcal{M}_i^{\theta c}(\mathcal{I}),\mathcal{I}^{gt}) >> \Omega(\mathcal{M}_i^{\theta c}(\mathcal{I}),\mathcal{I}^{gt})$ (as $\mathcal{M}_i^{\theta c}(\mathcal{I}) < \mathcal{I}^{gt}$). In other words, the loss projection of a non-performing pixel is elevated for mask pixels in $\mathcal{I}^{gt} \in \{1\}$, irrespective of the choice of $\Omega(.)$.

- $\boldsymbol{\mid (1+\alpha_{\hat{i}}^{\mathcal{C}}) \mid \simeq 1}$: It is noted that for all pixels in $\mathcal{I}^{gt}$, iff $\nabla\mathcal{P}_{\hat{i}}^{\mathcal{C}} >> 0$, then $\mid 1+\alpha_{\hat{i}}^{\mathcal{C}} \mid \simeq 1$. This implies that $(1+\alpha_{\hat{i}}^{\mathcal{C}})\odot\mathcal{M}_i^{\theta c}(\mathcal{I}) \simeq \mathcal{M}_i^{\theta c}(\mathcal{I})$. Thus, in other words, the impact on remaining pixels, having $\nabla\mathcal{P}_{\hat{i}}^{\mathcal{C}} >> 0$, is considered negligible.

Thus, the *weighted kernel* $\boldsymbol{\alpha_{\hat{i}}^{\mathcal{C}}}$ impacts specifically non-performing pixels in conformance with the Case 2 objective. A graphical illustration of the formulation is depicted in Figure 4.1.

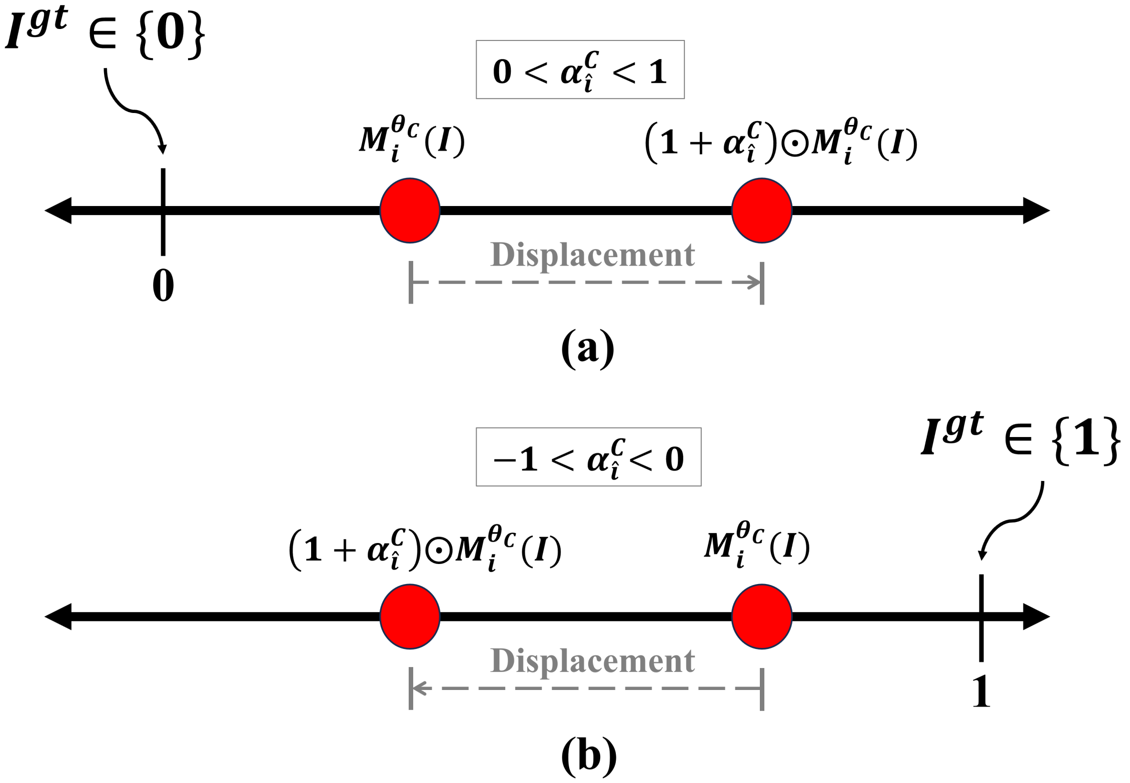


Figure 4.1: A graphical illustration of the $\alpha_{\hat{i}}^{\mathcal{C}}$ formulation. (a) Represents the Background Pixel scenario. (b) Represents the Mask Pixel scenario.

However, Equation 4.2.7 suffers from a major drawback. It is noted that repeated emphasis across iterations on a non-performing pixel having $\nabla\mathcal{P}^{\mathcal{C}}_{\hat{i}} \simeq 0$ might hamper the overall learning of the model. To address this drawback, a *Counter-based Depreciated Normalization* to readjust the computed weight-kernel is defined which is discussed in following section.

4.2.2 Counter-based Depreciated Normalization

Let us consider a spatial location $(h, w) \in \mathcal{I}^{gt}$. Intuitively, if the magnitude of $\alpha^{\mathcal{C}}_{\hat{i}}(h, w)$ is greater than λ (where $0 \leq \lambda \leq 1$), consistently across iterations, then such a point $(\alpha^{\mathcal{C}}_{\hat{i}}(h, w))$ exhibits the drawback highlighted earlier. The emphasis on such a pixel position should depreciate accordingly to overcome the drawback. In this regard, a *Counter Depreciation Matrix* $\tilde{D}_{\mathcal{C}}$ of dimension HxW is defined separately for each class $\mathcal{C}$. The depreciation matrix $\tilde{D}_{\mathcal{C}}$ is initialized to zero and subsequently computed, across (h, w) at every iteration as

$$\begin{aligned} (\mid \alpha^{\mathcal{C}}_{\hat{i}}(h, w) \mid > \lambda) &\implies (\tilde{D}_{\mathcal{C}}(h, w) = \tilde{D}_{\mathcal{C}}(h, w) + 1) \\ (\mid \alpha^{\mathcal{C}}_{\hat{i}}(h, w) \mid \leq \lambda) &\implies (\tilde{D}_{\mathcal{C}}(h, w) = 0) \end{aligned} \tag{4.2.8}$$

The *Counter-based Depreciation Normalized Iterative-Differential Weighted Kernel* $\bar{\alpha}^{\mathcal{C}}_{\hat{i}}$ is thus defined as

$$\bar{\alpha}^{\mathcal{C}}_{\hat{i}} = \frac{\alpha^{\mathcal{C}}_{\hat{i}} \odot e^{-\tilde{D}_{\mathcal{C}}}}{\mid \Sigma\alpha^{\mathcal{C}}_{\hat{i}} \odot e^{-\tilde{D}_{\mathcal{C}}} \mid} \tag{4.2.9}$$

where $\mid \Sigma\alpha^{\mathcal{C}}_{\hat{i}} \odot e^{-\tilde{D}_{\mathcal{C}}} \mid$ is used to normalize the kernel weights. Thus, the final projected loss $\mathcal{L}^{\alpha N^{++}}_{\hat{i}}$ is computed as

$$\mathcal{L}^{\alpha N^{++}}_{\hat{i}} = \sum_{\mathcal{C}} \Omega((1 + \bar{\alpha}^{\mathcal{C}}_{\hat{i}}) \odot \mathcal{M}^{\theta_{\mathcal{C}}}_{i}(\mathcal{I}), \mathcal{I}^{gt}) \tag{4.2.10}$$

The *Counter-based Depreciation Normalized Iterative-Differential Weighted Kernel* $\bar{\alpha}^{\mathcal{C}}_{\hat{i}}$ is defined as $\bar{\alpha}^{\mathcal{C}}_{\hat{i}} = \frac{\alpha^{\mathcal{C}}_{\hat{i}} \odot e^{-\tilde{D}_{\mathcal{C}}}}{|\Sigma\alpha^{\mathcal{C}}_{\hat{i}} \odot e^{-\tilde{D}_{\mathcal{C}}}|}$. It is trivial to note that $e^{-\tilde{D}_{\mathcal{C}}}$ is a depreciating factor used to modulate the value of $\bar{\alpha}^{\mathcal{C}}_{\hat{i}}$. A higher value in $\tilde{D}_{\mathcal{C}}$ accordingly reduces the respective magnitude of $\bar{\alpha}^{\mathcal{C}}_{\hat{i}}$ and vice-versa. To ensure a consistent weightage of each position in $\bar{\alpha}^{\mathcal{C}}_{\hat{i}}$ with respect to all predictions in $\mathcal{M}^{\theta_{\mathcal{C}}}_{i}(\mathcal{I})$, the values in $\bar{\alpha}^{\mathcal{C}}_{\hat{i}}$ have been normalized.

The scope of study is limited within $k = 2$ to reduce the computational over-

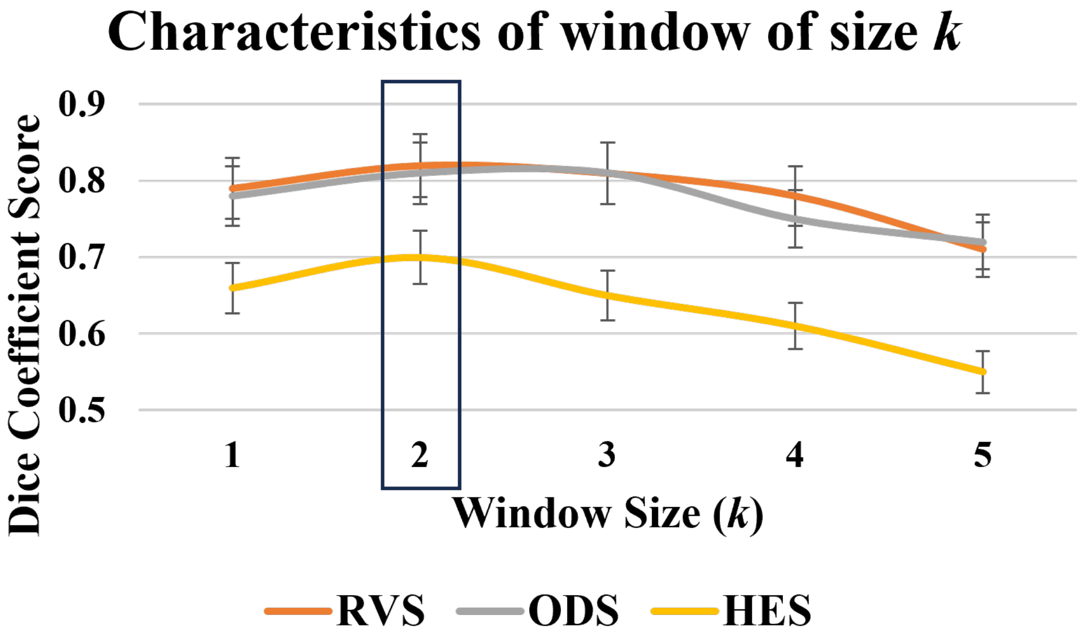


Figure 4.2: A plot depicting the performance of our proposed $\mathcal{L}^{proj}$ based on window size k. The x-axis depicts the choice for the value k and the y-axis depicts the *Dice Coefficient Score*. Here, RVS, ODS and HES respectively imply *Retinal Vessel Segmentation, Optic Disc Segmentation* and *Hard Exudates Segmentation*.

head. The performance of the αN^{++} loss projection is dependent on the choice of the hyper-parameter k, which represents the window size for iteration pair $\hat{i}$, as shown in Figure 4.2. It is noted that the Dice Coefficient drops for all the three segmentation challenges, namely, RVS, ODS and HES, beyond $k > 2$. A larger value for k would thus generate weighted-kernels which are ill-suited to track the performance of each pixel, as evident from Figure 4.2. Thus, k is set to 2, empirically, to obtain better performances. In addition, as $\mathcal{L}^{\alpha N^{++}}$ is not applicable for iterations less than k, only the original loss defined by Equation 4.2.1 is used in these cases.

4.3 Experimental Results & Discussion

To establish the efficacy of our approach, a standard segmentation objective function, viz., Dice Loss [241], had been selected as Φ in Equation 4.2.10 along with U-Net [125] architecture as the model $\mathcal{M}^{\theta}$. In addition, the standard BCE Loss [173] is employed as the objective function Ω in Equation 4.2.10 to maintain the Dice-BCE dual loss parity, popularly observed in medical image segmentation [168].

4.3.1 Dataset

To conduct the experiment, two benchmark fundus datasets, namely, STARE [30] and IDRiD [33] had been earmarked for this work. The STARE dataset comprises of Retinal Vessel Segmentation (RVS) task in Fundus images. It contains 20 annotated RGB Fundus images, randomly separated into 15 training and 5 test data for this work. The IDRiD dataset, in contrast, comprises of 54 Fundus images as training data as well as 27 images as test data and provides lesion annotations for five different tasks, out of which, two tasks, namely, Hard Exudate Segmentation (HES) and Optic Disc Segmentation (ODS), have been addressed in this work for comparative study.

4.3.2 Implementation Details

The PyTorch framework [245] had been used in this work for computation design. For data augmentation, traditional augmentation approach of rotation is used in this work with a rotation interval of 45^o. All computations reported in this work had been performed on a 4.2 GHz processor with $64GB$ RAM and 2×RTX 3060 GPUs, each having 12GB VRAM. The Adam optimizer has been used for all computations in this work with an epoch cycle of 500 with *Early Stopping Criterion* enabled and learning rate set to 3x10^{-4}.

4.3.3 Results

To begin with the experimentation, a value of λ is assigned in Equation 4.2.8 prior to further computation. In order to ascertain an optimal value, an iterative computation with incremental values of λ is conducted as shown in Table-4.1. The obtained results are evaluated on the Dice Coefficient (DC) [241] along with Precision (Pr) and Recall (Re) measures, as highlighted. The proposed $\mathcal{L}^{proj}$ loss projection, with Dice Loss as Φ and BCE Loss as Ω functions respectively, performed better in terms of DC for $\lambda = 0.7$ (from Table-4.1). Thus, all further experimentation in this work is reported with λ empirically set as 0.7.

Table 4.1: Iterative Computation of Counter Threshold (λ) in $\mathcal{L}^{\alpha N^{++}}$

	RVS			HES			ODS		
λ	**Pr**	**Re**	**Dice**	**Pr**	**Re**	**Dice**	**Pr**	**Re**	**Dice**
0.3	0.80	0.80	0.80	0.69	0.66	0.67	0.80	0.77	0.78
0.5	**0.85**	0.77	0.81	**0.72**	0.69	0.70	0.80	**0.81**	0.80
0.7	0.82	**0.82**	**0.82**	0.70	**0.70**	**0.70**	**0.82**	0.80	**0.81**
0.9	0.83	0.79	0.81	0.66	0.58	0.62	0.81	0.80	0.80

Subsequently, the performance of $\mathcal{L}^{proj}$ is compared with respect to two benchmark approaches, namely, Dice Loss with L2 normalization ($\mathcal{L}^{DC+L2}$) and Dice Loss with BCE Loss ($\mathcal{L}^{DC+BCE}$). Furthermore, two additional loss functions, namely, *Distance Map Loss* ($\mathcal{L}^{DM}$) [246] and *Structure Loss* ($\mathcal{L}^{SL}$) [247], are also chosen for comparative analysis. The selected loss functions are also based on a weighted-kernel approach for loss computation.

Table 4.2: Performance Comparison of αN^{++} Loss Projection with baseline model. (↑ indicates the % increase in performance.)

Task	**Loss**	**Precision**	**Recall**	**Dice**
Vessel	$\mathcal{L}^{DC}$	0.720	**0.830**	0.770
	$\mathcal{L}^{DC+L2}$	0.690	0.720	0.700
	$\mathcal{L}^{DC+BCE}$	0.820	0.780	0.800
	$\mathcal{L}^{DM}$	0.750	0.760	0.750
	$\mathcal{L}^{SL}$	0.770	0.810	0.790
	$\mathcal{L}^{proj}$	**0.820**	0.820	**0.820**
Optic Disc	$\mathcal{L}^{DC}$	0.730	0.770	0.750
	$\mathcal{L}^{DC+L2}$	0.680	**0.810**	0.740
	$\mathcal{L}^{DC+BCE}$	0.690	0.750	0.720
	$\mathcal{L}^{DM}$	0.720	0.690	0.700
	$\mathcal{L}^{SL}$	0.760	0.780	0.770
	$\mathcal{L}^{proj}$	**0.820**	0.800	**0.810**
Hard Exudates	$\mathcal{L}^{DC}$	0.580	0.670	0.620
	$\mathcal{L}^{DC+L2}$	0.430	0.510	0.470
	$\mathcal{L}^{DC+BCE}$	0.670	0.640	0.650
	$\mathcal{L}^{DM}$	0.510	0.550	0.530
	$\mathcal{L}^{SL}$	0.650	0.680	0.660
	$\mathcal{L}^{proj}$	**0.700**	**0.700**	**0.700**

Table-4.2 highlights the respective performances obtained across all three data used. To better gauge the performance, all approaches are compared with a baseline U-Net model, trained with Dice Loss, ($\mathcal{L}^{DC}$) and the percentage of improvement is reported as an additional parameter in Table-4.2. The $\mathcal{L}^{proj}$ loss achieves the highest overall increase in performance across all parameters in comparison to other approaches (highlighted in Table-4.2). Additionally, the framework supersedes the performance of contemporary loss functions significantly, as observed in Table-4.2. Figure 4.3 depicts a visual representation of the obtained results on the respective challenges.

To study the overall impact of the approach, a comparison of the $\mathcal{L}^{proj}$ loss is presented in Table-4.3 for Retinal Vessel Segmentation on the STARE dataset with respect to other state of the art models in literature. To maintain a parity in com-

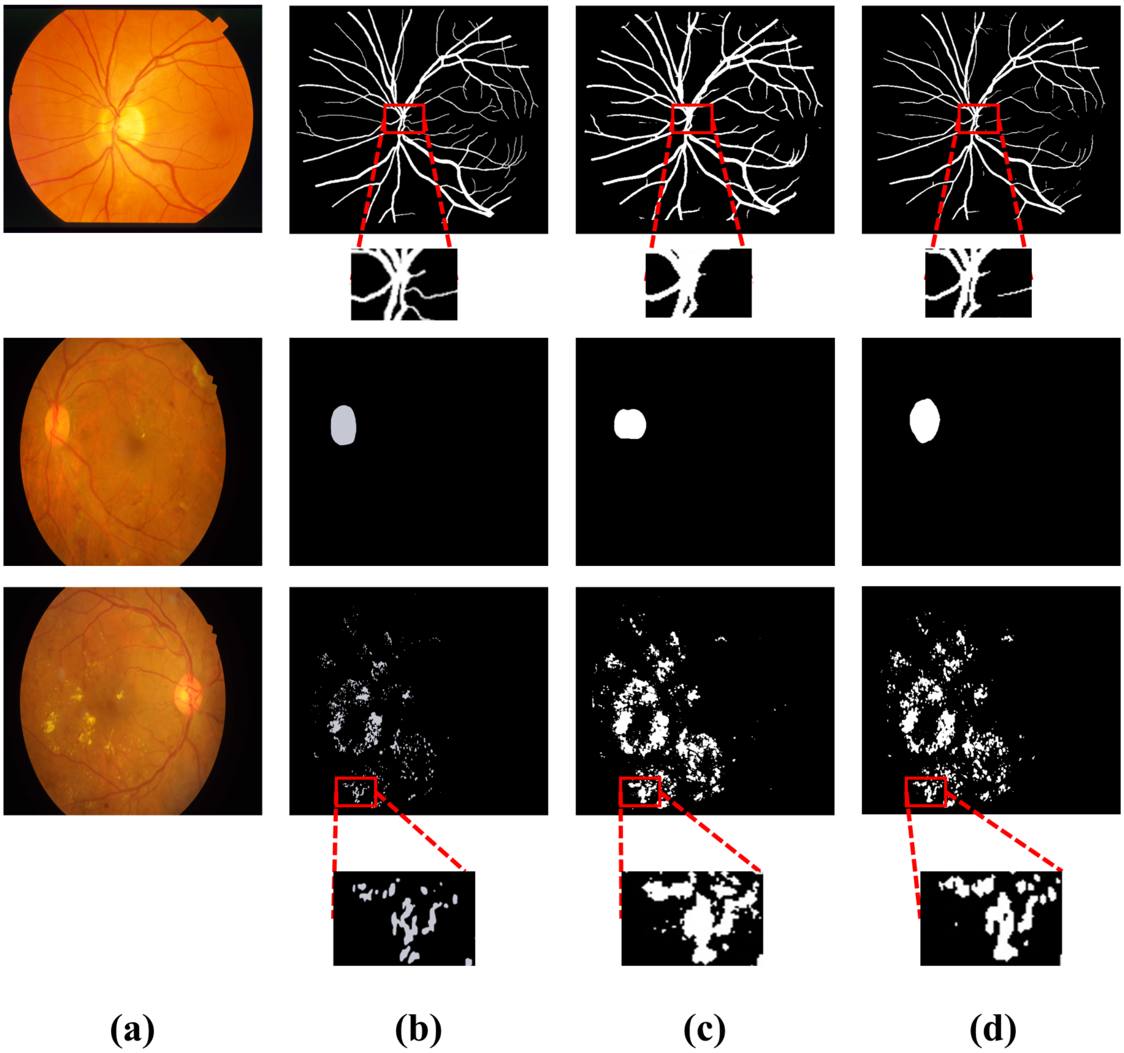


Figure 4.3: A visual depiction of the segmentation achieved by respective U-Net models. Each row depicts Vessel, Optic Disc and Hard Exudates Segmentation respectively. Row 1 corresponds to STARE dataset for RVS. Row 2-3 corresponds to IDRiD dataset for ODS and HES respectively. (a) Original Fundus Images (b) Ground Truths (c) U-Net+$\mathcal{L}^{DC}$ output (d) U-Net+$\mathcal{L}^{proj}$ output.

parison, all the state-of-the-art models are faithfully implemented on both *Dice Loss* ($\mathcal{L}^{DC}$) as well as the $\mathcal{L}^{proj}$ loss. Furthermore, the study also includes the $2pCePd$-Net+(w/o) model performance on STARE dataset, as highlighted in Chapter 3.

Table 4.3: Performance Comparison of αN^{++} Loss Projection with STARE dataset for Retinal Vessel Segmentation.

	Dice ($\mathcal{L}^{DC}$)			**Dice+αN^{++}** ($\mathcal{L}^{proj}$)		
Models	**Pr**	**Re**	**Dice**	**Pr**	**Re**	**Dice**
U-Net [125]	0.721	0.829	0.771	0.822	0.824	0.823
SCS-Net [248]	0.770	0.821	0.795	0.811	0.827	0.819
HA-Net [168]	**0.852**	0.818	0.835	0.860	0.826	0.843
CE-Net [166]	0.842	0.828	0.836	0.845	0.835	0.841
MISSFormer [221]	0.851	0.832	0.842	0.867	0.838	0.852
H2Former [222]	0.852	0.841	0.847	0.855	0.847	0.853
$2pCePd$	**0.871**	**0.867**	**0.869**	**0.874**	**0.870**	**0.872**

It is observed from Table-4.3, the $\mathcal{L}^{proj}$ loss improves the performance of existing models significantly. An average improvement of almost 2% is observed on the Dice Coefficient measure across all methods. It is noted from Table-4.3 that major improvements in performance are observable in simpler baseline models. The use of $\mathcal{L}^{proj}$ improves the Dice Coefficient in U-Net by almost 5%. In contrast, state-of-the-art models such as H2Former registers an increase of 0.6% in Dice Coefficient. The $2pCePd$-Net also achieves a similar performance, as highlighted in Table-4.3. This can be attributed to the design of sophisticated learning blocks in complex models which inherently reduces the dependence on learning paradigms. Hence, the improvement in performance is relatively lower for sophisticated state-of-the-art models. Nonetheless, the addendum of loss projection is observed to improve the respective model performances. As a visual study, Figure 4.4 depicts the training loss curve for $\mathcal{L}^{DC}$ loss against the loss curve of $\mathcal{L}^{proj}$ approach on U-Net model. It is observed that the loss function achieves a higher convergence across all tasks when compared to baseline Dice Loss under similar environmental constraints.

4.3.4 Ablation Study

To conduct an ablation study of the work, the choice of objective functions, namely, Φ and Ω respectively is modified. It is trivial to note that the choice of objective functions in Equations 4.2.7 and 4.2.10 can dictate the performance of model $\mathcal{M}^{\theta}$ that is achievable by the approach on a given task. To study the extent of such an impact, both objective functions Φ and Ω are set to the Dice Loss in Equation 4.2.7 and the observed results are presented as a comparative study across all datasets in

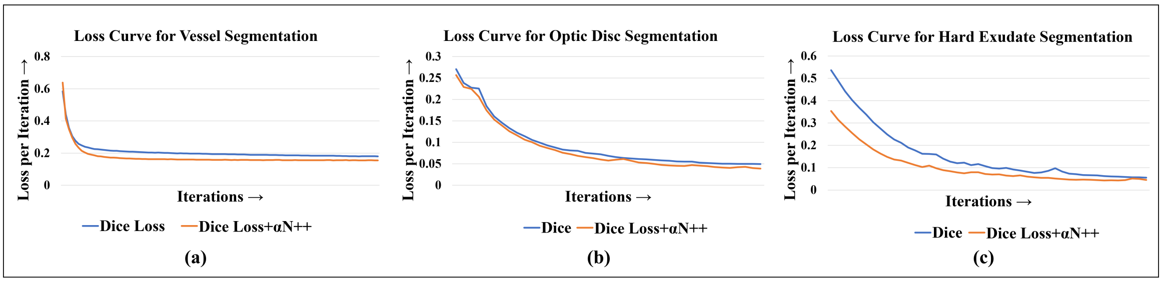


Figure 4.4: Comparative Analysis of the Training Loss Curve between $\mathcal{L}^{DC}$ and $\mathcal{L}^{proj}$ on (a) Vessel (b) Optic Disc and (c) Hard Exudate Segmentation Tasks.

Table-4.5. Table-4.5 highlights that the combination of Dice Loss with BCE Loss as Ω function performs superior for the work, however, $\mathcal{L}^{\alpha N^{++}}$ with Dice Loss as Ω function still has a higher performance when compared to a baseline U-Net model ($\mathcal{L}^{DC}$), as inferred from Table-4.2 and Table-4.5 respectively. Table-4.5 highlights that a cross-domain Φ and Ω functions in Equation 4.2.7 projects a better overall performance in contrast to homogeneous-domain objective functions.

Table 4.4: Comparative Analysis of Dice Loss with different Ω in αN^{++} Loss Projection with U-Net model.

Segmentation Task	**Ω Function**	**Precision**	**Recall**	**Dice Coefficient**
Vessel	Dice	0.731	**0.840**	0.780
	BCE	**0.822**	0.824	**0.823**
Optic Disc	Dice	0.811	0.749	0.779
	BCE	**0.823**	**0.798**	**0.810**
Hard Exudates	Dice	0.652	**0.732**	0.689
	BCE	**0.695**	0.696	**0.695**

Table 4.5: Comparative Analysis of U-Net baseline model with $2pCePd$-Net+(w/o) model using αN^{++} Loss Projection and $\mathcal{L}^{proj}$ Loss.

Segmentation Task	**Model**	**Precision**	**Recall**	**Dice Coefficient**
Optic Disc	U-Net	0.823	0.798	0.810
	$2pCePd$-Net+(w/o)	**0.829**	**0.833**	**0.831**
Hard Exudates	U-Net	0.695	0.696	0.695
	$2pCePd$-Net+(w/o)	**0.730**	**0.738**	**0.734**

4.4 Summary

This chapter introduced the development and evaluation of new training approach specific to the analysis of bio-marker segmentation in retinal fundus images. The αN^{++} Loss Projection is demonstrated in this chapter to leverage the potential of non-performing pixel positions to improve the bio-marker segmentation performance of well-known standard models along with the $2pCePd$-Net+(w/o) model in Fundus Images. The models were exhaustively tested on publicly available STARE and IDRiD datasets retinal blood vessel, optic disc and hard exudate segmentation. The findings proved that the methodology suggested in this study outperformed all the current state-of-the-art solutions on measures, specifically, Precision, Recall, and Dice Score respectively. Overall, the conclusions present evidence for the efficacy of the loss projection approach for retinal bio-marker segmentation in Fundus Images.

Chapter 5

Conclusion

5.1 Introduction

In this dissertation, the goal of study was to investigate three related topics forming a necessary pipeline for retinal fundus image analysis: data augmentation approaches to increase annotated dataset for segmentation, a more precise retinal blood vessel segmentation approach to increase model robustness and a novel loss projection-based biomarker segmentation method. By addressing past bottlenecks in literature for fundus image analysis pipeline, the approaches highlighted in this work conjointly produced a comprehensive framework improving segmentation performance. The primary conclusions, contributions, implications, limitations, and future directions are discussed in this concluding chapter in the following sub-sections.

5.2 Summary of Key Contributions

The first and foremost objective of the dissertation is to address the problem of imbalanced and limited datasets in bio-marker annotated segmentation datasets in Fundus Images, which is a prevalent problem in fundus imaging, especially with the advent of Deep Neural Networks. In this regard, it is observed from literature that, conventional deep neural network models have a tendency to overfit distributions when trained on smaller datasets with high class imbalance.

Chapter 2 in this dissertation introduces a collection of data augmentation approaches, namely, *Vessel Curvature*, *Vessel Curvature v2.0* and $\lambda CMgC^2$, designed in order to address the bottleneck dataset challenges in retinal fundus images. The augmentation approaches emphasized on utilizing retinal blood vessels in fundus images to be used as anchor points for novel vessel structure-based augmentation approaches. In this regard, a strong emphasis had been placed in proposing vessel-preserving structural deformations to maintain structural coherence while increasing the variation of the standard datasets.

An experimental evaluation demonstrated the efficacy of the augmentation approaches and presented a significant performance improvement with respect to retinal blood vessel segmentation in Fundus Images. Particularly, models trained on the proposed augmented datasets consistently performed better than the non-augmented as well as traditional-based augmented approaches in terms of generalization. Additionally, the thesis demonstrated that $\lambda CMgC^2$ augmentation that replicated realistic pathological vessel connectivity had a greater impact than traditional or vessel disconnected augmentations. The results obtained highlight the importance of structural augmentation approaches, designed to address the bottleneck associated with segmentation in retinal imaging.

Chapter 3 in this dissertation introduces a collection of retinal blood vessel segmentation approaches, namely, *Multi-Scale Layer based Network*, *PrSASfNet* and *2pCePd-Net*, designed in order to segment the augmented datasets formed earlier as well as traditional retinal blood vessel segmentation datasets. The existence of thin retinal blood vessel branches as well as the poor vessel contrast sensitivity with background poses a significant challenge to precisely segment the vessel structure. A set of three extended Convolutional Neural Network architectures enhanced with novel feature computation blocks and attention mechanisms is proposed as a solution in this regard. It is observed qualitatively that the attention-based layers assisted the model concentration on prominent vascular patterns whereas the multiscale block allowed the model to capture finer vessel structures.

Standard benchmark datasets, namely, DRIVE, STARE, CHASEDB_1 and HRF, are used to assess the respective model architectures. The results demonstrated that the porposed methodologies considerably improved segmentation performance as well as decreasing noisy artifacts such as broken vessels and false positives. Notably, the vessel probability map boundary-aware fusion module enhanced the segmented vascular continuity significantly by improving capillary and boundary delineation. This dissertation highlighted the congruence of $2pCePd-Net$ approach with $\lambda CMgC^2$ data augmentation to better address the challenging task of retinal blood vessel segmentation.

Chapter 4 in this dissertation introduces the performance of the standard benchmark segmentation models on bio-marker segmentation task such as optic disc and hard exudates segmentation. It is observed that the standard segmentation models face difficulties, primarily due to the structural imbalances, resulting in usually small, sparse, and frequently indistinct vessel structures. In this regard, the performance of $2pCePd-Net+(w/o)$ model is tested on a standard benchmark IDRiD dataset for extended bio-marker segmentation in Fundus Images. To improve the achieved performance, the novel αN^{++} loss projection approach is demonstrated in the dissertation to exceed the model performance beyond standards. Targeted loss recomputed learning is made possible by this projection, which highlights saturated learning localities and introduces an informed readjustment to progress the model convergence. This approach achieved a superior performance, particularly in hard exudates challenge based on standard metrics of evaluation, namely, precision, recall and Dice scores. Compared to conventional loss computation approaches, the proposed approach was better at delineating minute lesions and reducing false negatives, which is critical for early disease detection. This work showcases the potential of saturated pixel-wise loss-aware computation in improving the localization and segmentation of clinically important structures.

The individual components of the pipeline, namely, retinal fundus image augmentation, vessel segmentation and bio-marker segmentation forms an important component, each dealing with a challenging task, to create a coherent pipeline for reliable and accurate fundus image analysis. The $2pCePd$-Net architecture trained for vessel and biomarker segmentation tasks had broader implications primarily due to the data augmentation approaches highlighted. The appearance of major pathological markers near or along retinal blood vessels offers an anatomical context that is utilized to guide biomarker detection. Furthermore, the concentration of learning on diagnostically relevant regions through loss projection demonstrates an approach to enhance weak supervision of benchmark models.

From a clinical point of view, an improved and benchmark segmentation model of retinal bio-markers in fundus images is important for the design of automated retinal diseases grading unit, releasing a significant workload of ophthalmologists, and allow large-scale screening to be conducted in rural places with few resources. Accurate vessel segmentation is additionally important for measuring changes in blood vessels over time, which is used for tracking the progress of a disease and treatment plan.

The approaches that have been developed in this dissertation, presents new areas of research in fundus imaging for attention-guided learning and multi-task supervision. The dissertation presents a better understanding on approaches to combine multi-domain segmentation tasks, as well as augmentation pipelines that demonstrate the use of retinal anatomical priors in deep learning architectures. These contributions presents new frontiers to lead further research in retinal fundus image analysis.

5.3 Limitations and Future Scope

The approaches presented in this dissertation demonstrate an effective end-to-end pipeline for segmenting important retinal bio-markers in fundus images. However, in spite of the progress achieved, there are scopes for further improvement. Firstly, the datasets used, in spite of being accepted benchmarks, might not present the full range of real-world fundus imaging conditions. The data augmentation approaches, being a derivative of existing standard benchmarks, suffer from a similar drawback as a result. The efficacy of the models in context of real-world scenario requires further experimentation and study. Secondly, the design of the presented segmentation architectures, with the use of attention mechanisms and dual-branch processing, require more computing power than standard baseline models. This imposes restrictions on the applicability of the models on edge devices or clinical settings lacking significant hardware infrastructure. Finally, the quality of the annotations itself has an impact on the accuracy of the training and evaluation processes. The levels of accuracy for fine vessels and small lesions is a subjective standard depended upon the inter-observer bias. This difference between observers adds noise that presents additional challenges for both model learning and reported metrics.

Building on this thesis's contributions, a number of interesting avenues for further study become apparent. The investigation of few-shot or semi-supervised learning techniques is one such avenue that can lessen the reliance on sizable, annotated datasets. These methods may allow for efficient training with a large amount of unlabeled data and a small number of labeled samples.

Future research could also focus on temporal analysis of fundus images. Models that can analyze image sequences over time may provide insights into the progression and prognosis of many retinal conditions, many of which are progressive. Current static image analysis methods would gain a dynamic dimension if longitudinal data were integrated.

In future studies, a detailed exploration of the contributions presented in this dissertation can present a number of interesting research domains for further study. The investigation of few-shot or semi-supervised learning approaches for retinal fundus image segmentation is one such possibility that can significantly reduce the dependence on annotated gold-standard fundus imaging datasets. The exploration of such approaches may provide a more suitable interface for efficient training with a large volume of unlabeled data and a small volume of labeled samples.

Future research could also focus on temporal analysis of fundus images to track progression of disease more efficiently. Models that can analyze image sequences and video frames over a sequence of time might provide additional insights into the progression and prognosis of multiple retinal conditions, some of which are progressive in nature. Current static image analysis approaches would add a more dynamic dimension if time-series based data were integrated. It is also crucial to implement the developed models in clinical settings and conduct extensive studies to validate the effectiveness of designed deep neural network models. In addition to evaluating the applicability of models, iterative improvement of current benchmarks is also a domain of study. To support the design of a thorough diagnostic tool, future need to integrate multimodal data sources such as OCT images, patient demographics, and clinical history for a more holistic diagnosis. Finally, improving the explainability and transparency of segmentation models is still a major challenge for contemporary model designs. Interpretable frameworks are required in future research to help clinicians better understand model decisions, boosting confidence and ease practice adoption among general populous.

References

[1] E. Generali, L. Cantarini, and C. Selmi, “Ocular involvement in systemic autoimmune diseases,” *Clin. Rev. Allergy Immunol.*, vol. 49, no. 3, pp. 263–270, Dec. 2015.

[2] H. Eguchi, F. Hotta, T. Kuwahara, H. Imaohji, C. Miyazaki, M. Hirose, S. Kusaka, M. Fukuda, and Y. Shimomura, “Diagnostic approach to ocular infections using various techniques from conventional culture to next-generation sequencing analysis,” *Cornea*, vol. 36, no. 1, pp. S46–S52, Nov. 2017.

[3] T. J. Bennett and C. J. Barry, “Ophthalmic imaging today: an ophthalmic photographer’s viewpoint - a review,” *Clin. Experiment. Ophthalmol.*, vol. 37, no. 1, pp. 2–13, Jan. 2009.

[4] J.-H. Han, “Artificial intelligence in eye disease: Recent developments, applications, and surveys,” *Diagnostics (Basel)*, vol. 12, no. 8, p. 1927, Aug. 2022.

[5] X. He, N. Zhang, W. Cao, Y. Xing, and N. Yang, “Application progress of high-throughput sequencing in ocular diseases,” *J. Clin. Med.*, vol. 11, no. 12, p. 3485, Jun. 2022.

[6] S. Kuthyar, C. L. Anthony, T. Fashina, S. Yeh, and J. G. Shantha, “World health organization high priority pathogens: Ophthalmic disease findings and vision health perspectives,” *Pathogens*, vol. 10, no. 4, p. 442, Apr. 2021.

[7] A. Alhazimi and F. Almarek, “Impact of artificial intelligence on diagnosing eye diseases – a meta-analysis,” *Adv. Hum. Biol.*, vol. 14, no. 1, pp. 42–54, Jan.

[8] T. Makino, S. Jastrzebski, W. Oleszkiewicz, C. Chacko, R. Ehrenpreis, N. Samreen, C. Chhor, E. Kim, J. Lee, K. Pysarenko, B. Reig, H. Toth, D. Awal, L. Du, A. Kim, J. Park, D. K. Sodickson, L. Heacock, L. Moy, K. Cho, and K. J. Geras, “Differences between human and machine perception in medical diagnosis,” *Sci. Rep.*, vol. 12, no. 1, p. 6877, Apr. 2022.

[9] M. Akil, Y. Elloumi, and R. Kachouri, “Detection of retinal abnormalities in fundus image using CNN deep learning networks,” in *State of the Art in Neural Networks and their Applications.* Elsevier, 2021, pp. 19–61.

[10] A. Neto, J. Camera, S. Oliveira, A. Cláudia, and A. Cunha, “Optic disc and cup segmentations for glaucoma assessment using cup-to-disc ratio,” *Procedia Comput. Sci.*, vol. 196, pp. 485–492, 2022.

[11] R. A. Cooper, T. E. Getzen, H. J. McKee, and P. Laud, “Economic and demographic trends signal an impending physician shortage,” *Health Aff. (Millwood)*, vol. 21, no. 1, pp. 140–154, Jan. 2002.

[12] X. Li, D. Tian, W. Li, B. Dong, H. Wang, J. Yuan, B. Li, L. Shi, X. Lin, L. Zhao, and S. Liu, “Artificial intelligence-assisted reduction in patients’ waiting time for outpatient process: a retrospective cohort study,” *BMC Health Serv. Res.*, vol. 21, no. 1, p. 237, Mar. 2021.

[13] D. Kalogeropoulos, C. Kalogeropoulos, M. Stefaniotou, and M. Neofytou, “The role of tele-ophthalmology in diabetic retinopathy screening,” *J. Optom.*, vol. 13, no. 4, pp. 262–268, Oct. 2020.

[14] H. Wu, K. Jin, C. C. Yip, V. Koh, and J. Ye, “A systematic review of economic evaluation of artificial intelligence-based screening for eye diseases: From possibility to reality,” *Surv. Ophthalmol.*, vol. 69, no. 4, pp. 499–507, Jul. 2024.

[15] G. A. Saleh, N. M. Batouty, S. Haggag, A. Elnakib, F. Khalifa, F. Taher, M. A. Mohamed, R. Farag, H. Sandhu, A. Sewelam, and A. El-Baz, “The role of medical image modalities and AI in the early detection, diagnosis and grading of retinal diseases: A survey,” *Bioengineering (Basel)*, vol. 9, no. 8, p. 366, Aug. 2022.

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

[17] R. Bernardes, P. Serranho, and C. Lobo, “Digital ocular fundus imaging: a review,” *Ophthalmologica*, vol. 226, no. 4, pp. 161–181, Sep. 2011.

[18] I. A. Viedma, D. Alonso-Caneiro, S. A. Read, and M. J. Collins, “Deep learning in retinal optical coherence tomography (OCT): A comprehensive survey,” *Neurocomputing*, vol. 507, pp. 247–264, Oct. 2022.

[19] L. A. Yannuzzi, K. T. Rohrer, L. J. Tindel, R. S. Sobel, M. A. Costanza, W. Shields, and E. Zang, “Fluorescein angiography complication survey,” *Ophthalmology*, vol. 93, no. 5, pp. 611–617, May 1986.

[20] J. Fischer, T. Otto, F. Delori, L. Pace, and G. Staurenghi, “Scanning laser ophthalmoscopy (SLO),” in *High Resolution Imaging in Microscopy and Ophthalmology.* Cham: Springer International Publishing, 2019, pp. 35–57.

[21] C. M. St Croix, S. H. Shand, and S. C. Watkins, “Confocal microscopy: comparisons, applications, and problems,” *Biotechniques*, vol. 39, no. 6 Suppl, pp. S2–5, Dec. 2005.

[22] R. H. Silverman, “Principles of ophthalmic ultrasound,” *Expert Rev. Ophthalmol.*, vol. 18, no. 6, pp. 379–389, Nov. 2023.

[23] A. Grzybowski, K. Jin, J. Zhou, X. Pan, M. Wang, J. Ye, and T. Y. Wong, “Retina fundus photograph-based artificial intelligence algorithms in medicine: A systematic review,” *Ophthalmol. Ther.*, vol. 13, no. 8, pp. 2125–2149, Aug. 2024.

[24] L. Godlin Atlas, K. P. Arjun, K. Sampath Kumar, R. K. Dhanaraj, and A. Nayyar, “Deep learning enabled hemorrhage detection in retina with DPFE and splat segmentation in fundus images,” *Biomed. Signal Process. Control*, vol. 88, no. 105692, p. 105692, Feb. 2024.

[25] D. C. DeBuc, “The role of retinal imaging and portable screening devices in tele-ophthalmology applications for diabetic retinopathy management,” *Curr. Diab. Rep.*, vol. 16, no. 12, Dec. 2016.

[26] Y. Wang, C. Liu, W. Hu, L. Luo, D. Shi, J. Zhang, Q. Yin, L. Zhang, X. Han, and M. He, “Economic evaluation for medical artificial intelligence: accuracy vs. cost-effectiveness in a diabetic retinopathy screening case,” *NPJ Digit. Med.*, vol. 7, no. 1, p. 43, Feb. 2024.

[27] L. F. Nakayama, L. Zago Ribeiro, F. K. Malerbi, and C. S. Regatieri, “Artificial intelligence, data sharing, and privacy for retinal imaging under brazilian data protection law,” *Int. J. Retina Vitreous*, vol. 11, no. 1, p. 41, Apr. 2025.

[28] S. T. Padmapriya and S. Parthasarathy, “Ethical data collection for medical image analysis: A structured approach,” *Asian Bioeth. Rev.*, vol. 16, no. 1, pp. 1–14, Apr. 2023.

[29] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, “Ridge-based vessel segmentation in color images of the retina,” *IEEE Trans. Med. Imaging*, vol. 23, no. 4, pp. 501–509, Apr. 2004.

[30] 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. Imaging*, vol. 19, no. 3, pp. 203–210, 2000.

[31] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "An ensemble classification-based approach applied to retinal blood vessel segmentation," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 9, pp. 2538–2548, Sep. 2012.

[32] A. Budai, R. Bock, A. Maier, J. Hornegger, and G. Michelson, "Robust vessel segmentation in fundus images," *Int. J. Biomed. Imaging*, vol. 2013, p. 154860, Dec. 2013.

[33] P. Porwal and et al., "Idrid: Diabetic retinopathy – segmentation and grading challenge," *Med. Image Anal.*, vol. 59, p. 101561, 2020.

[34] R. Li, Y. Hui, X. Zhang, S. Zhang, B. Lv, Y. Ni, X. Li, X. Liang, L. Yang, H. Lv, Z. Yin, H. Li, Y. Yang, G. Liu, J. Li, G. Xie, S. Wu, and Z. Wang, "Ocular biomarkers of cognitive decline based on deep-learning retinal vessel segmentation," *BMC Geriatr.*, vol. 24, no. 1, p. 28, Jan. 2024.

[35] E. Goceri, "Medical image data augmentation: techniques, comparisons and interpretations," *Artif. Intell. Rev.*, vol. 56, no. 11, pp. 1–45, Mar. 2023.

[36] A. Raj, N. A. Shah, and A. K. Tiwari, "A novel approach for fundus image enhancement," *Biomed. Signal Process. Control*, vol. 71, no. 103208, p. 103208, Jan. 2022.

[37] A. Q. Saeed, S. N. H. Sheikh Abdullah, J. Che-Hamzah, A. T. Abdul Ghani, and W. A. K. Abu-ain, "Synthesizing retinal images using end-to-end VAEs-GAN pipeline-based sharpening and varying layer," *Multimed. Tools Appl.*, Oct. 2023.

[38] R. Leonardo, J. Goncalves, A. Carreiro, B. Simoes, T. Oliveira, and F. Soares, "Impact of generative modeling for fundus image augmentation with improved and degraded quality in the classification of glaucoma," *IEEE Access*, vol. 10, pp. 111 636–111 649, 2022.

[39] L. Rundo and C. Militello, "Image biomarkers and explainable AI: handcrafted features versus deep learned features," *Eur. Radiol. Exp.*, vol. 8, no. 1, p. 130, Nov. 2024.

[40] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, “Image matching from handcrafted to deep features: A survey,” *Int. J. Comput. Vis.*, vol. 129, no. 1, pp. 23–79, Jan. 2021.

[41] B. Goutam, M. F. Hashmi, Z. W. Geem, and N. D. Bokde, “A comprehensive review of deep learning strategies in retinal disease diagnosis using fundus images,” *IEEE Access*, vol. 10, pp. 57 796–57 823, 2022.

[42] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in *Lecture Notes in Computer Science*, ser. Lecture notes in computer science. Cham: Springer International Publishing, 2015, pp. 234–241.

[43] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, “Ccnet: Criss-cross attention for semantic segmentation,” in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 603–612.

[44] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141.

[45] M. Ren and R. S. Zemel, “End-to-end instance segmentation with recurrent attention,” 2017. [Online]. Available: https://arxiv.org/abs/1605.09410

[46] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, ser. NIPS’17. Red Hook, NY, USA: Curran Associates Inc., 2017, p. 6000–6010.

[47] X. Wang, R. Girshick, A. Gupta, and K. He, “Non-local neural networks,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7794–7803.

[48] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, “Dual attention network for scene segmentation,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 3141–3149.

[49] H. Fu, Y. Xu, S. Lin, D. W. Kee Wong, and J. Liu, “DeepVessel: Retinal vessel segmentation via deep learning and conditional random field,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, ser. Lecture notes in computer science. Cham: Springer International Publishing, 2016, pp. 132–139.

[50] C. Wang, Z. Zhao, Q. Ren, Y. Xu, and Y. Yu, "Dense u-net based on patch-based learning for retinal vessel segmentation," *Entropy (Basel)*, vol. 21, no. 2, p. 168, Feb. 2019.

[51] Z. Li, M. Xu, X. Yang, and Y. Han, "Multi-label fundus image classification using attention mechanisms and feature fusion," *Micromachines (Basel)*, vol. 13, no. 6, p. 947, Jun. 2022.

[52] P. Bharati and A. Pramanik, "Deep learning techniques—R-CNN to mask R-CNN: A survey," in *Computational Intelligence in Pattern Recognition*, ser. Advances in intelligent systems and computing. Singapore: Springer Singapore, 2020, pp. 657–668.

[53] H. Zhang, W. Ni, Y. Luo, Y. Feng, R. Song, and X. Wang, "TUnet-LBF: Retinal fundus image fine segmentation model based on transformer unet network and LBF," *Comput. Biol. Med.*, vol. 159, no. 106937, p. 106937, Jun. 2023.

[54] J. Han, Y. Wang, and H. Gong, "Fundus retinal vessels image segmentation method based on improved u-net," *IRBM*, vol. 43, no. 6, pp. 628–639, Dec. 2022.

[55] J. Li, G. Gao, L. Yang, G. Bian, and Y. Liu, "Dpf-net: A dual-path progressive fusion network for retinal vessel segmentation," *IEEE Trans. Instrum. Meas.*, 2023.

[56] Z. Yan, X. Yang, and K.-T. 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, Sep. 2018.

[57] S. E. Christopher Kent, "AI & Ophthalmology: The Pros and Cons — reviewofophthalmology.com," https://www.reviewofophthalmology.com/article/ai-and-ophthalmology-the-pros-and-cons, [Accessed May-2025].

[58] K. Jin and J. Ye, "Artificial intelligence and deep learning in ophthalmology: Current status and future perspectives," *Adv. Ophthalmol. Pract. Res.*, vol. 2, no. 3, p. 100078, Nov. 2022.

[59] A. K. Upadhyay and A. K. Bhandari, "Advances in deep learning models for resolving medical image segmentation data scarcity problem: A topical review," *Arch. Comput. Methods Eng.*, vol. 31, no. 3, pp. 1701–1719, Apr. 2024.

[60] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition.* IEEE, Jun. 2009.

[61] H. Jin, Y. Luo, P. Li, and J. Mathew, “A review of secure and privacy-preserving medical data sharing,” *IEEE Access*, vol. 7, pp. 61 656–61 669, 2019.

[62] G. Quellec, M. Lamard, P.-H. Conze, P. Massin, and B. Cochener, “Automatic detection of rare pathologies in fundus photographs using few-shot learning,” *Med. Image Anal.*, vol. 61, no. 101660, p. 101660, Apr. 2020.

[63] T. Krzywicki, P. Brona, A. M. Zbrzezny, and A. E. Grzybowski, “A global review of publicly available datasets containing fundus images: Characteristics, barriers to access, usability, and generalizability,” *J. Clin. Med.*, vol. 12, no. 10, May 2023.

[64] M. Aftab, T. Ahmad, S. Adeel, S. H. Bhatti, and M. Irfan, “Hyper-parameter tuning through innovative designing to avoid over-fitting in machine learning modelling: a case study of small data sets,” *J. Stat. Comput. Simul.*, vol. 95, no. 7, pp. 1595–1609, May 2025.

[65] A. Rashid and F. A. Jakobiec, “Avoiding the major complication of ophthalmic pathology: misdiagnosis. a review of three common diagnostic challenges,” *Semin. Ophthalmol.*, vol. 29, no. 5-6, pp. 468–474, Sep. 2014.

[66] S. Chokuwa and M. H. Khan, “Divergent domains, convergent grading: Enhancing generalization in diabetic retinopathy grading,” in *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2025, pp. 3667–3677.

[67] V. Kumar and K. Paul, “Fundus imaging-based healthcare: Present and future,” *ACM Trans. Comput. Healthc.*, vol. 4, no. 3, pp. 1–34, Jul. 2023.

[68] S. Das, H. J. Kuht, I. De Silva, S. S. Deol, L. Osman, J. Burns, N. Sarvananthan, U. Sarodia, B. Kapoor, T. Islam, R. Sampath, A. Poyser, V. Konidaris, R. Anzidei, F. A. Proudlock, and M. G. Thomas, “Feasibility and clinical utility of handheld fundus cameras for retinal imaging,” *EYE*, vol. 37, no. 2, pp. 274–279, Feb. 2023.

[69] P. Khojasteh, B. Aliahmad, and D. K. Kumar, “Fundus images analysis using deep features for detection of exudates, hemorrhages and microaneurysms,” *BMC Ophthalmol.*, vol. 18, no. 1, Dec. 2018.

[70] A. Raj, A. K. Tiwari, and M. G. Martini, “Fundus image quality assessment: survey, challenges, and future scope,” *IET Image Process.*, vol. 13, no. 8, pp. 1211–1224, Jun. 2019.

[71] K. S. Kumar and N. P. Singh, “Retinal disease prediction through blood vessel segmentation and classification using ensemble-based deep learning approaches,” *Neural Comput. Appl.*, vol. 35, no. 17, pp. 12 495–12 511, Jun. 2023.

[72] S. W. Franklin and S. E. Rajan, “Computerized screening of diabetic retinopathy employing blood vessel segmentation in retinal images,” *Biocybern. Biomed. Eng.*, vol. 34, no. 2, pp. 117–124, 2014.

[73] A. W. Stitt, T. M. Curtis, M. Chen, R. J. Medina, G. J. McKay, A. Jenkins, T. A. Gardiner, T. J. Lyons, H.-P. Hammes, R. Simó, and N. Lois, “The progress in understanding and treatment of diabetic retinopathy,” *Prog. Retin. Eye Res.*, vol. 51, pp. 156–186, Mar. 2016.

[74] S. Erden and E. Bicakci, “Hypertensive retinopathy: incidence, risk factors, and comorbidities,” *Clin. Exp. Hypertens.*, vol. 34, no. 6, pp. 397–401, Apr. 2012.

[75] M. Ip and A. Hendrick, “Retinal vein occlusion review,” *Asia Pac. J. Ophthalmol. (Phila.)*, vol. 7, no. 1, pp. 40–45, Jan. 2018.

[76] R. J. Casson, G. Chidlow, J. P. M. Wood, J. G. Crowston, and I. Goldberg, “Definition of glaucoma: clinical and experimental concepts,” *Clin. Experiment. Ophthalmol.*, vol. 40, no. 4, pp. 341–349, May 2012.

[77] S. Roychowdhury, D. D. Koozekanani, S. N. Kuchinka, and K. K. Parhi, “Optic disc boundary and vessel origin segmentation of fundus images,” *IEEE J. Biomed. Health Inform.*, vol. 20, no. 6, pp. 1562–1574, Nov. 2016.

[78] B. S. Y. Lam, Y. Gao, and A. W.-C. Liew, “General retinal vessel segmentation using regularization-based multiconcavity modeling,” *IEEE Trans. Med. Imaging*, vol. 29, no. 7, pp. 1369–1381, Jul. 2010.

[79] Z. Xiao, M. Adel, and S. Bourennane, “Bayesian method with spatial constraint for retinal vessel segmentation,” *Comput. Math. Methods Med.*, vol. 2013, p. 401413, Jul. 2013.

[80] C. Xiao, E. Choi, and J. Sun, “Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review,” *J. Am. Med. Inform. Assoc.*, vol. 25, no. 10, pp. 1419–1428, Oct. 2018.

[81] 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. Imaging*, vol. 23, no. 10, pp. 1196–1204, Oct. 2004.

[82] A. Lahiri, K. Ayush, P. Kumar Biswas, and P. Mitra, "Generative adversarial learning for reducing manual annotation in semantic segmentation on large scale miscroscopy images: Automated vessel segmentation in retinal fundus image as test case," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.

[83] K. Jin, X. Huang, J. Zhou, Y. Li, Y. Yan, Y. Sun, Q. Zhang, Y. Wang, and J. Ye, "FIVES: A fundus image dataset for artificial intelligence based vessel segmentation," *Sci. Data*, vol. 9, no. 1, p. 475, Aug. 2022.

[84] D. A. van Dyk and X.-L. Meng, "The art of data augmentation," *J. Comput. Graph. Stat.*, vol. 10, no. 1, pp. 1–50, Mar. 2001.

[85] C. Y. Lin, M. Wu, J. A. Bloom, I. J. Cox, M. L. Miller, and Y. M. Lui, "Rotation, scale, and translation resilient watermarking for images," *IEEE Trans. Image Process.*, vol. 10, no. 5, pp. 767–782, 2001.

[86] M. Zhou, K. Jin, S. Wang, J. Ye, and D. Qian, "Color retinal image enhancement based on luminosity and contrast adjustment," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 3, pp. 521–527, Mar. 2018.

[87] S. Rahman, M. M. Rahman, M. Abdullah-Al-Wadud, G. D. Al-Quaderi, and M. Shoyaib, "An adaptive gamma correction for image enhancement," *EURASIP J. Image Video Process.*, vol. 2016, no. 1, Dec. 2016.

[88] E. Castro, J. S. Cardoso, and J. C. Pereira, "Elastic deformations for data augmentation in breast cancer mass detection," in *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. IEEE, Mar. 2018.

[89] F. Luisier, T. Blu, and M. Unser, "Image denoising in mixed Poisson-Gaussian noise," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 696–708, Mar. 2011.

[90] B. Aktas, D. D. Ates, O. Duzyel, and A. Gumus, "Diffusion-based data augmentation methodology for improved performance in ocular disease diagnosis using retinography images," *Int. J. Mach. Learn. Cybern.*, Dec. 2024.

[91] J. Guo, Z. Pang, F. Yang, J. Shen, and J. Zhang, "Study on the method of fundus image generation based on improved GAN," *Math. Probl. Eng.*, vol. 2020, pp. 1–13, Jul. 2020.

[92] D. Shenkut and V. Bhagavatula, "Fundus GAN - GAN-based fundus image synthesis for training retinal image classifiers," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, Jul. 2022.

[93] H.-T. Wu, X. Cao, Y. Gao, K. Zheng, J. Huang, J. Hu, and Z. Tian, "Fundus image enhancement via semi-supervised GAN and anatomical structure preservation," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 8, no. 1, pp. 313–326, Feb. 2024.

[94] L. Zhang, X. Wang, D. Yang, T. Sanford, S. Harmon, B. Turkbey, B. J. Wood, H. Roth, A. Myronenko, D. Xu, and Z. Xu, "Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation," *IEEE Trans. Med. Imaging*, vol. 39, no. 7, pp. 2531–2540, Jul. 2020.

[95] M. R. Shoaib, H. M. Emara, A. S. Mubarak, O. A. Omer, F. E. Abd El-Samie, and H. Esmaiel, "Revolutionizing diabetic retinopathy diagnosis through advanced deep learning techniques: Harnessing the power of gan model with transfer learning and the diagan-cnn model," *Biomedical Signal Processing and Control*, vol. 99, p. 106790, Jan. 2025. [Online]. Available: http://dx.doi.org/10.1016/j.bspc.2024.106790

[96] J. Wang, C. Lan, C. Liu, Y. Ouyang, T. Qin, W. Lu, Y. Chen, W. Zeng, and P. S. Yu, "Generalizing to unseen domains: A survey on domain generalization," 2021.

[97] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," 2017.

[98] S. Yun, D. Han, S. Chun, S. J. Oh, Y. Yoo, and J. Choe, "CutMix: Regularization strategy to train strong classifiers with localizable features," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, Oct. 2019.

[99] H. Li, S. J. Pan, S. Wang, and A. C. Kot, "Domain generalization with adversarial feature learning," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Jun. 2018.

[100] X. Fan, Q. Wang, J. Ke, F. Yang, B. Gong, and M. Zhou, "Adversarially adaptive normalization for single domain generalization," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Jun. 2021.

[101] S. Wang, L. Yu, K. Li, X. Yang, C.-W. Fu, and P.-A. Heng, "DoFE: Domain-oriented feature embedding for generalizable fundus image segmentation on unseen datasets," *IEEE Trans. Med. Imaging*, vol. 39, no. 12, pp. 4237–4248, Dec. 2020.

[102] J. Lyu, Y. Zhang, Y. Huang, L. Lin, P. Cheng, and X. Tang, "AADG: Automatic augmentation for domain generalization on retinal image segmentation," *IEEE Trans. Med. Imaging*, vol. 41, no. 12, pp. 3699–3711, Dec. 2022.

[103] S. Fadnavis, "Image interpolation techniques in digital image processing: an overview," *International Journal of Engineering Research and Applications*, vol. 4, no. 10, pp. 70–73, 2014.

[104] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018, pp. 0–0.

[105] Y. Gu, Z. Zeng, H. Chen, J. Wei, Y. Zhang, B. Chen, Y. Li, Y. Qin, Q. Xie, Z. Jiang *et al.*, "Medsrgan: medical images super-resolution using generative adversarial networks," *Multimedia Tools and Applications*, vol. 79, pp. 21 815–21 840, 2020.

[106] M. Kim, Y. N. Kim, M. Jang, J. Hwang, H.-K. Kim, S. C. Yoon, Y. J. Kim, and N. Kim, "Synthesizing realistic high-resolution retina image by style-based generative adversarial network and its utilization," *Scientific Reports*, vol. 12, no. 1, p. 17307, 2022.

[107] R. Ramli, K. Hasikin, M. Y. I. Idris, N. K. A. Karim, and A. W. A. Wahab, "Fundus image registration technique based on local feature of retinal vessels," *Appl. Sci. (Basel)*, vol. 11, no. 23, p. 11201, Nov. 2021.

[108] P. Liskowski and K. Krawiec, "Segmenting retinal blood vessels with_newline deep neural networks," *IEEE Trans. Med. Imaging*, vol. 35, no. 11, pp. 2369–2380, Nov. 2016.

[109] A. Gahramanova and University of South Florida, "Locating centers of mass with image processing," *Undergrad. J. Math. Model. One Two*, vol. 10, no. 1, May 2019.

[110] P. K. Saha, G. Borgefors, and G. Sanniti di Baja, “A survey on skeletonization algorithms and their applications,” *Pattern Recognit. Lett.*, vol. 76, pp. 3–12, Jun. 2016.

[111] A. Mehnert and P. Jackway, “An improved seeded region growing algorithm,” *Pattern Recognit. Lett.*, vol. 18, no. 10, pp. 1065–1071, Oct. 1997.

[112] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken, “Ridge-based vessel segmentation in color images of the retina,” *IEEE Trans. Med. Imaging*, vol. 23, no. 4, pp. 501–509, 2004.

[113] 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, Jan. 2003.

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

[115] 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, Jul. 2015.

[116] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, “CE-net: Context encoder network for 2D medical image segmentation,” *IEEE Trans. Med. Imaging*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019.

[117] L. Li, M. Verma, Y. Nakashima, H. Nagahara, and R. Kawasaki, “IterNet: Retinal image segmentation utilizing structural redundancy in vessel networks,” in *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, Mar. 2020.

[118] D. Wang, A. Haytham, J. Pottenburgh, O. Saeedi, and Y. Tao, “Hard attention net for automatic retinal vessel segmentation,” *IEEE J. Biomed. Health Inform.*, vol. 24, no. 12, pp. 3384–3396, Dec. 2020.

[119] J. Wei, G. Zhu, Z. Fan, J. Liu, Y. Rong, J. Mo, W. Li, and X. Chen, “Genetic U-Net: Automatically designed deep networks for retinal vessel segmentation using a genetic algorithm,” *IEEE Trans. Med. Imaging*, vol. 41, no. 2, pp. 292–307, Feb. 2022.

[120] Z. Qu, L. Zhuo, J. Cao, X. Li, H. Yin, and Z. Wang, "TP-Net: Two-path network for retinal vessel segmentation," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 4, pp. 1979–1990, Apr. 2023.

[121] A. Borji, "Pros and cons of gan evaluation measures," *Comput. Vis. Image Underst.*, vol. 179, pp. 41–65, 2019.

[122] H. Zhao, J. Jia, and V. Koltun, "Exploring self-attention for image recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[123] X. Sun, H. Fang, Y. Yang, D. Zhu, L. Wang, J. Liu, and Y. Xu, "Robust retinal vessel segmentation from a data augmentation perspective," in *8th Int. Workshop OMIA 2021, Proc.* Springer, 2021, pp. 189–198.

[124] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.

[125] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Med. Image Comput. Comput.-Assist. Interv. MICCAI 2015, Proc. Part III.* Springer, 2015, pp. 234–241.

[126] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 11–19.

[127] Y. Li, L. Xu, Y. Jin, X. Kuang, Y. Zhang, W. Cui, and T. Zhang, "Diffusion probabilistic learning with gate-fusion transformer and edge-frequency attention for retinal vessel segmentation," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–13, 2024.

[128] C. G. Owen, A. R. Rudnicka, R. Mullen, S. A. Barman, D. Monekosso, P. H. Whincup, J. Ng, and C. Paterson, "Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (caiar) program," *Invest. Ophthalmol. Vis. Sci.*, vol. 50, no. 5, pp. 2004–2010, 2009.

[129] T. Köhler, A. Budai, M. F. Kraus, J. Odstrčilik, G. Michelson, and J. Hornegger, "Automatic no-reference quality assessment for retinal fundus images using vessel segmentation," in *Proc. 26th IEEE Int. Symp. Comput.-Based Med. Syst.* IEEE, 2013, pp. 95–100.

[130] H. E. Grossniklaus, E. E. Geisert, and J. M. Nickerson, “Introduction to the retina,” in *Progress in Molecular Biology and Translational Science*, ser. Progress in molecular biology and translational science. Elsevier, 2015, pp. 383–396.

[131] V. Romano, B. Steger, M. Ahmad, G. Coco, L. Pagano, S. Ahmad, Y. Zhao, Y. Zheng, and S. B. Kaye, “Imaging of vascular abnormalities in ocular surface disease,” *Surv. Ophthalmol.*, vol. 67, no. 1, pp. 31–51, Jan. 2022.

[132] V. Mayya, S. Kamath S., and U. Kulkarni, “Automated microaneurysms detection for early diagnosis of diabetic retinopathy: A comprehensive review,” *Comput. Methods Programs Biomed. Update*, vol. 1, no. 100013, p. 100013, 2021.

[133] H. E. Grossniklaus and W. R. Green, “Choroidal neovascularization,” *Am. J. Ophthalmol.*, vol. 137, no. 3, pp. 496–503, Mar. 2004.

[134] R. Srivastava, L. Duan, D. W. K. Wong, J. Liu, and T. Y. Wong, “Detecting retinal microaneurysms and hemorrhages with robustness to the presence of blood vessels,” *Comput. Methods Programs Biomed.*, vol. 138, pp. 83–91, Jan. 2017.

[135] E. Grisan, M. Foracchia, and A. Ruggeri, “A novel method for the automatic grading of retinal vessel tortuosity,” *IEEE Trans. Med. Imaging*, vol. 27, no. 3, pp. 310–319, Mar. 2008.

[136] C. Y.-L. Cheung, M. K. Ikram, C. Sabanayagam, and T. Y. Wong, “Retinal microvasculature as a model to study the manifestations of hypertension,” *Hypertension*, vol. 60, no. 5, pp. 1094–1103, Nov. 2012.

[137] N. Ashton, “Vascular changes in diabetes with particular reference to the retinal vessels; preliminary report,” *Br. J. Ophthalmol.*, vol. 33, no. 7, pp. 407–420, Jul. 1949.

[138] S. Guo, S. Yin, G. Tse, G. Li, L. Su, and T. Liu, “Association between caliber of retinal vessels and cardiovascular disease: A systematic review and meta-analysis,” *Curr. Atheroscler. Rep.*, vol. 22, no. 4, p. 16, May 2020.

[139] K. McGeechan, G. Liew, P. Macaskill, L. Irwig, R. Klein, B. E. K. Klein, J. J. Wang, P. Mitchell, J. R. Vingerling, P. T. V. M. Dejong, J. C. M. Witteman, M. M. B. Breteler, J. Shaw, P. Zimmet, and T. Y. Wong, “Meta-analysis: retinal vessel caliber and risk for coronary heart disease,” *Ann. Intern. Med.*, vol. 151, no. 6, pp. 404–413, Sep. 2009.

[140] B. R. Masters, "Fractal analysis of the vascular tree in the human retina," *Annu. Rev. Biomed. Eng.*, vol. 6, no. 1, pp. 427–452, 2004.

[141] H. Akil, A. S. Huang, B. A. Francis, S. R. Sadda, and V. Chopra, "Retinal vessel density from optical coherence tomography angiography to differentiate early glaucoma, pre-perimetric glaucoma and normal eyes," *PLoS One*, vol. 12, no. 2, p. e0170476, Feb. 2017.

[142] S. Zhang, R. Zheng, Y. Luo, X. Wang, J. Mao, C. J. Roberts, and M. Sun, "Simultaneous arteriole and venule segmentation of dual-modal fundus images using a multi-task cascade network," *IEEE Access*, vol. 7, pp. 57 561–57 573, 2019.

[143] R. Heitmar, G. Y. H. Lip, R. E. Ryder, and A. D. Blann, "Retinal vessel diameters and reactivity in diabetes mellitus and/or cardiovascular disease," *Cardiovasc. Diabetol.*, vol. 16, no. 1, Dec. 2017.

[144] J. R. Evans, M. Michelessi, and G. Virgili, "Laser photocoagulation for proliferative diabetic retinopathy," *Cochrane Database Syst. Rev.*, vol. 2015, no. 11, p. CD011234, Nov. 2014.

[145] T. Y. Y. Lai, S. Liu, S. Das, and D. S. C. Lam, "Intravitreal injection–technique and safety," *Asia Pac. J. Ophthalmol. (Phila.)*, vol. 4, no. 6, pp. 321–328, Nov. 2015.

[146] A. Markan, A. Agarwal, A. Arora, K. Bazgain, V. Rana, and V. Gupta, "Novel imaging biomarkers in diabetic retinopathy and diabetic macular edema," *Ther. Adv. Ophthalmol.*, vol. 12, p. 251584142095051, Jan. 2020.

[147] T. Khan, M. Y. Bertram, R. Jina, B. Mash, N. Levitt, and K. Hofman, "Preventing diabetes blindness: cost effectiveness of a screening programme using digital non-mydriatic fundus photography for diabetic retinopathy in a primary health care setting in south africa," *Diabetes Res. Clin. Pract.*, vol. 101, no. 2, pp. 170–176, Aug. 2013.

[148] M. Goliaš and E. Šikudová, "Retinal blood vessel segmentation and inpainting networks with multi-level self-attention," *Biomed. Signal Process. Control*, vol. 102, no. 107343, p. 107343, Apr. 2025.

[149] J. Zhang, Y. Wang, J. Dai, M. Cavichini, D.-U. G. Bartsch, W. R. Freeman, T. Q. Nguyen, and C. An, "Two-step registration on multi-modal retinal images

via deep neural networks," *IEEE Trans. Image Process.*, vol. 31, pp. 823–838, Jan. 2022.

[150] U. T. V. Nguyen, A. Bhuiyan, L. A. F. 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, Mar. 2013.

[151] 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. Imaging*, vol. 19, no. 3, pp. 203–210, Mar. 2000.

[152] A. M. Mendonça and A. Campilho, "Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction," *IEEE Trans. Med. Imaging*, vol. 25, no. 9, pp. 1200–1213, Sep. 2006.

[153] 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.

[154] M. S. Miri and A. Mahloojifar, "Retinal image analysis using curvelet transform and multistructure elements morphology by reconstruction," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 5, pp. 1183–1192, May 2011.

[155] K. A. Vermeer, F. M. Vos, H. G. Lemij, and A. M. Vossepoel, "A model based method for retinal blood vessel detection," *Comput. Biol. Med.*, vol. 34, no. 3, pp. 209–219, Apr. 2004.

[156] B. Y. Lam and H. Yan, "A novel vessel segmentation algorithm for pathological retina images based on the divergence of vector fields," *IEEE Trans. Med. Imaging*, vol. 27, no. 2, pp. 237–246, Feb. 2008.

[157] A. Sarbasova and M. M. Hasan, "Multiscale blood vessel segmentation in retinal fundus images algorithm implementation and analysis," in *Embracing Global Computing in Emerging Economies*, ser. Communications in computer and information science. Cham: Springer International Publishing, 2015, pp. 113–121.

[158] M. A. Palomera-Pérez, M. E. Martinez-Perez, H. Benítez-Pérez, and J. L. Ortega-Arjona, "Parallel multiscale feature extraction and region growing: application in retinal blood vessel detection," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 500–506, Mar. 2010.

[159] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi, “DREAM: diabetic retinopathy analysis using machine learning,” *IEEE J. Biomed. Health Inform.*, vol. 18, no. 5, pp. 1717–1728, Sep. 2014.

[160] M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, and M. D. Abramoff, “Comparative study of retinal vessel segmentation methods on a new publicly available database,” in *Medical Imaging 2004: Image Processing*, J. M. Fitzpatrick and M. Sonka, Eds. SPIE, May 2004.

[161] J. V. B. Soares, J. J. G. Leandro, R. M. Cesar Júnior, H. F. Jelinek, and M. J. Cree, “Retinal vessel segmentation using the 2-D gabor wavelet and supervised classification,” *IEEE Trans. Med. Imaging*, vol. 25, no. 9, pp. 1214–1222, Sep. 2006.

[162] D. Marin, A. Aquino, M. E. Gegundez-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. Imaging*, vol. 30, no. 1, pp. 146–158, Jan. 2011.

[163] C. A. Lupascu, D. Tegolo, and E. Trucco, “FABC: retinal vessel segmentation using AdaBoost,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 5, pp. 1267–1274, Sep. 2010.

[164] J. I. Orlando, E. Prokofyeva, and M. B. 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, Jan. 2017.

[165] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, “Dunet: A deformable network for retinal vessel segmentation,” *Knowl.-Based Syst.*, vol. 178, pp. 149–162, 2019.

[166] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, “Ce-net: Context encoder network for 2d medical image segmentation,” *IEEE Trans. Med. Imaging*, vol. 38, no. 10, pp. 2281–2292, 2019.

[167] L. Li, M. Verma, Y. Nakashima, H. Nagahara, and R. Kawasaki, “Iternet: Retinal image segmentation utilizing structural redundancy in vessel networks,” in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2020, pp. 3656–3665.

[168] D. Wang, A. Haytham, J. Pottenburgh, O. Saeedi, and Y. Tao, "Hard attention net for automatic retinal vessel segmentation," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 12, pp. 3384–3396, 2020.

[169] Y. Wu, Y. Xia, Y. Song, Y. Zhang, and W. Cai, "Nfn+: A novel network followed network for retinal vessel segmentation," *Neural Netw.*, vol. 126, pp. 153–162, 2020.

[170] Z. Lin, J. Huang, Y. Chen, X. Zhang, W. Zhao, Y. Li, L. Lu, M. Zhan, X. Jiang, and X. Liang, "A high resolution representation network with multi-path scale for retinal vessel segmentation," *Comput. Methods Programs Biomed.*, vol. 208, p. 106206, 2021.

[171] J. Wei, G. Zhu, Z. Fan, J. Liu, Y. Rong, J. Mo, W. Li, and X. Chen, "Genetic u-net: automatically designed deep networks for retinal vessel segmentation using a genetic algorithm," *IEEE Trans. Med. Imaging*, vol. 41, no. 2, pp. 292–307, 2021.

[172] H. Wu, W. Wang, J. Zhong, B. Lei, Z. Wen, and J. Qin, "Scs-net: A scale and context sensitive network for retinal vessel segmentation," *Med. Image Anal.*, vol. 70, p. 102025, 2021.

[173] Y. Li, Y. Zhang, J.-Y. Liu, K. Wang, K. Zhang, G.-S. Zhang, X.-F. Liao, and G. Yang, "Global transformer and dual local attention network via deep-shallow hierarchical feature fusion for retinal vessel segmentation," *IEEE Trans. Cybern.*, 2022.

[174] Y. Tan, K.-F. Yang, S.-X. Zhao, and Y.-J. Li, "Retinal vessel segmentation with skeletal prior and contrastive loss," *IEEE Trans. Med. Imaging*, vol. 41, no. 9, pp. 2238–2251, 2022.

[175] O. O. Sule, "A survey of deep learning for retinal blood vessel segmentation methods: Taxonomy, trends, challenges and future directions," *IEEE Access*, vol. 10, pp. 38 202–38 236, 2022.

[176] 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. Imaging*, vol. 8, no. 3, pp. 263–269, 1989.

[177] W. E. Hart, M. Goldbaum, B. Côté, P. Kube, and M. R. Nelson, "Automated measurement of retinal vascular tortuosity," *Proc. AMIA Annu. Fall Symp.*, pp. 459–463, 1997.

[178] J. Cao, Y. Li, M. Sun, Y. Chen, D. Lischinski, D. Cohen-Or, B. Chen, and C. Tu, "DO-Conv: Depthwise over-parameterized convolutional layer," *IEEE Trans. Image Process.*, vol. 31, pp. 3726–3736, May 2022.

[179] C. Banerjee, T. Mukherjee, and E. Pasiliao, Jr, "An empirical study on generalizations of the ReLU activation function," in *Proceedings of the 2019 ACM Southeast Conference*, vol. 201. New York, NY, USA: ACM, Apr. 2019, pp. 164–167.

[180] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi, "Blood vessel segmentation of fundus images by major vessel extraction and subimage classification," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 3, pp. 1118–1128, May 2015.

[181] Y. Zhao, L. Rada, K. Chen, S. P. Harding, and Y. Zheng, "Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images," *IEEE Trans. Med. Imaging*, vol. 34, no. 9, pp. 1797–1807, Sep. 2015.

[182] T. Mapayi, S. Viriri, and J.-R. Tapamo, "Adaptive thresholding technique for retinal vessel segmentation based on GLCM-energy information," *Comput. Math. Methods Med.*, vol. 2015, p. 597475, Feb. 2015.

[183] P. R. Wankhede and K. B. Khanchandani, "Retinal blood vessel segmentation in fundus images using improved graph cut method," in *2018 International Conference on Smart Systems and Inventive Technology (ICSSIT).* IEEE, Dec. 2018.

[184] M. Zhang, W. Li, and D. Chen, "Blood vessel segmentation in fundus images based on improved loss function," in *2019 Chinese Automation Congress (CAC).* IEEE, Nov. 2019.

[185] C. Kromm and K. Rohr, "Inception capsule network for retinal blood vessel segmentation and centerline extraction," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI).* IEEE, Apr. 2020.

[186] W. Han, X. Dong, Y. Zhang, D. Crandall, C.-Z. Xu, and J. Shen, "Asymmetric convolution: An efficient and generalized method to fuse feature maps in multiple vision tasks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–15, 2024.

[187] Q. Huang, J. Sun, H. Ding, X. Wang, and G. Wang, "Robust liver vessel extraction using 3d u-net with variant dice loss function," *Computers in Biology and Medicine*, vol. 101, pp. 153–162, 2018.

[188] R. Medina-Carnicer, R. Muñoz-Salinas, E. Yeguas-Bolivar, and L. Diaz-Mas, "A novel method to look for the hysteresis thresholds for the canny edge detector," *Pattern Recognition*, vol. 44, no. 6, pp. 1201–1211, 2011.

[189] Z. Liao, S. Hu, Y. Xie, and Y. Xia, "Modeling annotator preference and stochastic annotation error for medical image segmentation," *Medical Image Analysis*, vol. 92, p. 103028, 2024.

[190] L. Xu, "Least mean square error reconstruction principle for self-organizing neural-nets," *Neural Networks*, vol. 6, no. 5, pp. 627–648, 1993.

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

[192] X. Jin, Y. Xie, X.-S. Wei, B.-R. Zhao, Z.-M. Chen, and X. Tan, "Delving deep into spatial pooling for squeeze-and-excitation networks," *Pattern Recognition*, vol. 121, p. 108159, 2022.

[193] C. Tan and X. Chen, "Intuitionistic fuzzy choquet integral operator for multi-criteria decision making," *Expert Systems with Applications*, vol. 37, no. 1, pp. 149–157, 2010.

[194] W. Bandler and L. Kohout, "Fuzzy power sets and fuzzy implication operators," *Fuzzy Sets And Systems*, vol. 4, no. 1, pp. 13–30, Jul. 1980.

[195] M. Yeung, E. Sala, C.-B. Schönlieb, and L. Rundo, "Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation," *Computerized Medical Imaging and Graphics*, vol. 95, p. 102026, 2022.

[196] S. Imambi, K. B. Prakash, and G. R. Kanagachidambaresan, *PyTorch.* Cham: Springer International Publishing, 2021, pp. 87–104.

[197] P. Verma, V. Tripathi, and B. Pant, "Comparison of different optimizers implemented on the deep learning architectures for covid-19 classification," *Materials Today: Proceedings*, vol. 46, pp. 11 098–11 102, 2021, international Conference on Technological Advancements in Materials Science and Manufacturing.

[198] A. Hoover, V. Kouznetsova, and M. Goldbaum, “Stare: Structured analysis of the retina dataset,” 2000.

[199] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, “Chase db1: Retinal vessel reference dataset,” 2012. [Online]. Available: https://researchdata.kingston.ac.uk/96/

[200] T. Köhler, A. Budai, M. F. Kraus, J. Odstrčilik, G. Michelson, and J. Hornegger, “Automatic no-reference quality assessment for retinal fundus images using vessel segmentation,” in *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, 2013, pp. 95–100.

[201] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, “Striving for simplicity: The all convolutional net,” 2015. [Online]. Available: https://arxiv.org/abs/1412.6806

[202] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *2016 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.

[203] J. Tang, B. Zou, C. Li, S. Feng, and H. Peng, “Plane-wave image reconstruction via generative adversarial network and attention mechanism,” *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–15, 2021.

[204] Y. Chen, C. Xu, W. Ding, S. Sun, X. Yue, and H. Fujita, “Target-aware u-net with fuzzy skip connections for refined pancreas segmentation,” *Appl. Soft Comput.*, vol. 131, p. 109818, 2022.

[205] S. Das, D. Jyotishi, and S. Dandapat, “Automated detection of heart valve diseases using stationary wavelet transform and attention-based hierarchical lstm network,” *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–10, 2023.

[206] N. C. Thompson, K. Greenewald, K. Lee, and G. F. Manso, “The computational limits of deep learning,” 2022. [Online]. Available: https://arxiv.org/abs/2007.05558

[207] R. C. Gonzalez, *Digital image processing.* Pearson education india, 2009.

[208] S. Xie and Z. Tu, “Holistically-nested edge detection,” in *2015 IEEE Int. Conf. Comput. Vis. (ICCV)*, 2015, pp. 1395–1403.

[209] S. Das, D. Jyotishi, and S. Dandapat, "Heart valve diseases detection based on feature-fusion and hierarchical lstm network," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.

[210] L. Mou, L. Chen, J. Cheng, Z. Gu, Y. Zhao, and J. Liu, "Dense dilated network with probability regularized walk for vessel detection," *IEEE Trans. Med. Imaging*, vol. 39, no. 5, pp. 1392–1403, 2019.

[211] X. Li, Y. Jiang, M. Li, and S. Yin, "Lightweight attention convolutional neural network for retinal vessel image segmentation," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 1958–1967, 2021.

[212] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," 2021. [Online]. Available: https://arxiv.org/abs/2102.04306

[213] M. Arsalan, T. M. Khan, S. S. Naqvi, M. Nawaz, and I. Razzak, "Prompt deep light-weight vessel segmentation network (plvs-net)," *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, vol. 20, no. 2, p. 1363–1371, Oct. 2022.

[214] Y. Liu, J. Shen, L. Yang, H. Yu, and G. Bian, "Wave-net: A lightweight deep network for retinal vessel segmentation from fundus images," *Comput. Biol. Med.*, vol. 152, p. 106341, 2023.

[215] A. Ashraf, N. M. Nawi, T. Shahzad, M. Aamir, M. A. Khan, and K. Ouahada, "Dimension reduction using dual-featured auto-encoder for the histological classification of human lungs tissues," *IEEE Access*, vol. 12, pp. 104 165–104 176, 2024.

[216] R. Bernardes, P. Serranho, and C. Lobo, "Digital Ocular Fundus Imaging: A Review," *Ophthalmologica*, vol. 226, no. 4, pp. 161–181, 09 2011.

[217] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "Blood vessel segmentation methodologies in retinal images–a survey," *Comput. Methods Programs Biomed.*, vol. 108, no. 1, pp. 407–433, 2012.

[218] 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. Imaging*, vol. 25, no. 9, pp. 1214–1222, 2006.

[219] C. L Srinidhi, P. Aparna, and J. Rajan, “Recent advancements in retinal vessel segmentation,” *J. Med. Syst.*, vol. 41, pp. 1–22, 2017.

[220] W. Li, X. Xiong, P. Xia, L. Ju, and Z. Ge, “ TP-DRSeg: Improving Diabetic Retinopathy Lesion Segmentation with Explicit Text-Prompts Assisted SAM ,” in *Proc. Med. Image Comput. Comput. Assist. Interv. – MICCAI 2024*, vol. LNCS 15008. Springer Nature Switzerland, October 2024.

[221] X. Huang, Z. Deng, D. Li, X. Yuan, and Y. Fu, “Missformer: An effective transformer for 2d medical image segmentation,” *IEEE Trans. Med. Imaging*, vol. 42, no. 5, pp. 1484–1494, 2023.

[222] A. He, K. Wang, T. Li, C. Du, S. Xia, and H. Fu, “H2former: An efficient hierarchical hybrid transformer for medical image segmentation,” *IEEE Trans. Med. Imaging*, vol. 42, no. 9, pp. 2763–2775, 2023.

[223] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *2015 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015, pp. 3431–3440.

[224] Q. Wang, Y. Ma, K. Zhao, and Y. Tian, “A comprehensive survey of loss functions in machine learning,” *Ann. Data Sci.*, vol. 9, no. 2, pp. 187–212, Apr. 2022.

[225] S. Roshan, J. Tanha, M. Zarrin, A. F. Babaei, H. Nikkhah, and Z. Jafari, “A deep ensemble medical image segmentation with novel sampling method and loss function,” *Comput. Biol. Med.*, vol. 172, no. 108305, p. 108305, Apr. 2024.

[226] Z. Li, K. Kamnitsas, and B. Glocker, “Analyzing overfitting under class imbalance in neural networks for image segmentation,” *IEEE Trans. Med. Imaging*, vol. 40, no. 3, pp. 1065–1077, Mar. 2021.

[227] G. Wang, W. Li, M. A. Zuluaga, R. Pratt, P. A. Patel, M. Aertsen, T. Doel, A. L. David, J. Deprest, S. Ourselin, and T. Vercauteren, “Interactive medical image segmentation using deep learning with image-specific fine tuning,” *IEEE Trans. Med. Imaging*, vol. 37, no. 7, pp. 1562–1573, Jul. 2018.

[228] R. Wang, S. Chen, C. Ji, J. Fan, and Y. Li, “Boundary-aware context neural network for medical image segmentation,” *Med. Image Anal.*, vol. 78, no. 102395, p. 102395, May 2022.

[229] P. Hurtik, S. Tomasiello, J. Hula, and D. Hynar, "Binary cross-entropy with dynamical clipping," *Neural Comput. Appl.*, vol. 34, no. 14, pp. 12 029–12 041, Jul. 2022.

[230] R. Zhao, B. Qian, X. Zhang, Y. Li, R. Wei, Y. Liu, and Y. Pan, "Rethinking dice loss for medical image segmentation," in *2020 IEEE International Conference on Data Mining (ICDM)*. IEEE, Nov. 2020.

[231] T. Eelbode, J. Bertels, M. Berman, D. Vandermeulen, F. Maes, R. Bisschops, and M. B. Blaschko, "Optimization for medical image segmentation: Theory and practice when evaluating with dice score or jaccard index," *IEEE Trans. Med. Imaging*, vol. 39, no. 11, pp. 3679–3690, Nov. 2020.

[232] D. S. A. Damit, S. N. Sulaiman, M. K. Osman, N. K. A. Karim, N. F. Razali, and M. I. F. Marzuki, "Navigating tversky loss function hyperparameter spaces using particle swarm optimization for myocardial scar segmentation," in *2024 20th IEEE International Colloquium on Signal Processing & Its Applications (CSPA)*. IEEE, Mar. 2024.

[233] J. Chen, C. Fu, H. Xie, X. Zheng, R. Geng, and C.-W. Sham, "Uncertainty teacher with dense focal loss for semi-supervised medical image segmentation," *Comput. Biol. Med.*, vol. 149, no. 106034, p. 106034, Oct. 2022.

[234] H. Kervadec, J. Bouchtiba, C. Desrosiers, E. Granger, J. Dolz, and I. Ben Ayed, "Boundary loss for highly unbalanced segmentation," in *Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning*, ser. Proceedings of Machine Learning Research, M. J. Cardoso, A. Feragen, B. Glocker, E. Konukoglu, I. Oguz, G. Unal, and T. Vercauteren, Eds., vol. 102. PMLR, 08–10 Jul 2019, pp. 285–296.

[235] M. Tang, F. Perazzi, A. Djelouah, I. Ben Ayed, C. Schroers, and Y. Boykov, "On regularized losses for weakly-supervised cnn segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

[236] X. Shi, F. Xing, K. Xu, P. Chen, Y. Liang, Z. Lu, and Z. Guo, "Loss-based attention for interpreting image-level prediction of convolutional neural networks," *IEEE Trans. Image Process.*, vol. 30, pp. 1662–1675, Jan. 2021.

[237] C. Zhou, H. Meng, M. Li, and Z. Zhou, "On learning label noise robust networks via regularization: A topological view," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. PP, May 2025.

Supratim Ghosh 09-June 2025

M. Kundu

09 June 2025

Professor
Computer Sc. & Engg. Department
Jadavpur University
Kolkata 700032

[238] C. Wang, C. Xu, C. Wanga, and D. Tao, "Perceptual adversarial networks for image-to-image transformation," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4066–4079, Aug. 2018.

[239] L. Chen, Y. Wang, J. Yang, Y. Zheng, T. Han, B. Zhang, and T. Cao, "Uncertainty-aware focal loss for object segmentation," *Eng. Appl. Artif. Intell.*, vol. 149, no. 110599, p. 110599, Jun. 2025.

[240] Y. Jin, H. Li, Q. Dou, H. Chen, J. Qin, C.-W. Fu, and P.-A. Heng, "Multi-task recurrent convolutional network with correlation loss for surgical video analysis," *Med. Image Anal.*, vol. 59, no. 101572, p. 101572, Jan. 2020.

[241] Y.-C. Liu, D. S. Tan, J.-C. Chen, W.-H. Cheng, and K.-L. Hua, "Segmenting hepatic lesions using residual attention u-net with an adaptive weighted dice loss," in *2019 IEEE Int. Conf. Image Process. (ICIP)*, 2019, pp. 3322–3326.

[242] Q. Ming and X. Xiao, "Towards accurate medical image segmentation with gradient-optimized dice loss," *IEEE Signal Process. Lett.*, vol. 31, pp. 191–195, 2024.

[243] Y. Wang, Z.-P. Bian, J. Hou, and L.-P. Chau, "Convolutional neural networks with dynamic regularization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 5, pp. 2299–2304, 2021.

[244] J. Chi, X. Han, C. Wu, H. Wang, and P. Ji, "X-net: Multi-branch unet-like network for liver and tumor segmentation from 3d abdominal ct scans," *Neurocomputing*, vol. 459, pp. 81–96, 2021.

[245] A. Paszke and et al., *PyTorch: an imperative style, high-performance deep learning library.* Red Hook, NY, USA: Curran Associates Inc., 2019.

[246] F. Caliva, C. Iriondo, A. M. Martinez, S. Majumdar, and V. Pedoia, "Distance map loss penalty term for semantic segmentation," 2019. [Online]. Available: https://arxiv.org/abs/1908.03679

[247] D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "Pranet: Parallel reverse attention network for polyp segmentation," in *Med. Image Comput. Comput. Assist. Interv. – MICCAI 2020.* Cham: Springer International Publishing, 2020, pp. 263–273.

[248] H. Wu, W. Wang, J. Zhong, B. Lei, Z. Wen, and J. Qin, "Scs-net: A scale and context sensitive network for retinal vessel segmentation," *Med. Image Anal.*, vol. 70, p. 102025, 2021.

2pCeP*d*-Net: Two-Path Cross-Context Encoder With Probability Map-Based Bandpass Decoder for Retinal Vessel Segmentation

Supratim Ghosh, Sourav Pramanik, Anoop Kumar Tiwari, Kottakkaran Sooppy Nisar,
Mahantapas Kundu, and Mita Nasipuri, *Life Senior Member, IEEE*

***Abstract*—Accurate automatic retinal blood vessel segmentation in fundus images plays an important role in the early diagnosis of any ocular disease detection system. However, most of the past literature has yet to attain a superior result primarily due to a lack of sufficient annotated data and the complexity of the vessel structure under challenging background conditions. In this work, we propose the design of a coherence measure-guided data augmentation model, named lambda-coherence measure-guided Cartesian-square (λCMgC^2), for enriching the existing datasets with synthetic and structurally coherent fundus images thus alleviating the issue of data scarcity. Subsequently, we propose a novel end-to-end convolutional network, called two-path cross-context encoder with probability map-based bandpass decoder (2pCeP*d*-Net) for the segmentation of blood vessels endowed with a novel 2pCd$^+$ encoder block with CERg skip connection and a novel $\hat{p}$BPf-enabled decoder block. The proposed work has been evaluated using four standard datasets, namely, DRIVE, STARE, CHASEDB1, and HRF, and has obtained a benchmark accuracy (Ac) of 97.6%, 98.1%, 98.2%, and 97.7%, respectively. Statistically, our proposed model has further achieved benchmark results across sensitivity (Se), specificity (Sp), *F*1, and AUC measures of evaluation as well.**

***Index Terms*—Cross-attention, data augmentation, data fusion, deep learning (DL), fundus image.**

I. INTRODUCTION

RETINAL fundus images (RFIs) form an important basis for the diagnosis of a variety of ocular diseases, including glaucoma, diabetic retinopathy, macular degeneration, and so on [1], [2], which often lead to blindness. Prevention of such diseases requires early screening. Typically, a fundus image contains a large number of retinal components, such as an optic disk (OD), macula, and retinal vessels (RVs). Morphological characteristics of RVs, such as angles and branching patterns, are important in the diagnosis of ocular disorders [1], [2]. As a result, precise RV segmentation from fundus images is critical for ocular illness diagnosis and monitoring. However, manual RV segmentation in fundus images is an extremely tricky and time-consuming process that requires professionally capable physicians [3]. Thus, automatic segmentation of RVs is now an active field of research. Various automatic RV segmentation methods for fundus images have been developed [1], [4] in the past. Despite substantial progress in this field, the existing approaches still fail to address significant issues, primarily due to the variation of intensity values, morphological structures of vessels, and the poor contrast observed in vessel boundaries. Additionally, existing standard datasets have a limited amount of annotated vessel images. Thus, the limitation of data in RV segmentation often leads to overfitting problems and becomes a bottleneck to modern approaches. Several techniques [1], [4] have been developed in the past to address the aforementioned issues which can be divided into two groups: traditional approaches and deep learning (DL) approaches, which are discussed as follows.

A. Traditional Segmentation Approaches

Traditional approaches considered only the pixel intensity and the morphological features of a vessel for segmentation [5], [6], [7]. Geometric vessel structure detection using a minimum generated super-pixel tree had been explored by Sheng et al. [8] in past literature. Deformable model-based segmentation approaches have also been studied in the literature in this regard [9]. Akbarizadeh et al. [10] proposed the use of statistical-based kurtosis wavelet energy (KWE) and skewness wavelet energy (SWE) [11] for texture recognition in SAR images. The image binarization approaches highlighted in [10] and [11] can be extrapolated to RV segmentation as well, as discussed in Karn et al. [12]. Although traditional approaches often require little resource for execution, however, the sensitivity (Se) to noise and inhomogeneity in retinal images often present a major bottleneck

---

Received 19 July 2024; revised 11 April 2025; accepted 28 April 2025. Date of publication 12 May 2025; date of current version 22 May 2025. This work was supported by the Prince Sattam Bin Abdulaziz University under Grant PSAU/2024/R/1445. The Associate Editor coordinating the review process was Dr. Dan Zhang. *(Corresponding author: Sourav Pramanik.)*

Supratim Ghosh is with the Department of Computer Science and Engineering (AIML), Netaji Subhash Engineering College, Kolkata 700152, India (e-mail: supratimghosh2772@gmail.com).

Sourav Pramanik is with the Department of Computer Science, New Alipore College, Kolkata 700053, India (e-mail: souravpramanik@newaliporecollege.ac.in).

Anoop Kumar Tiwari is with the Department of Computer Science and Information Technology, Central University of Haryana, Mahendragarh 123031, India (e-mail: dranoop@cuh.ac.in).

Kottakkaran Sooppy Nisar is with the Department of Mathematics, College of Science and Humanities in Al-Kharj, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia (e-mail: n.sooppy@psau.edu.sa).

Mahantapas Kundu and Mita Nasipuri are with the Department of Computer Science and Engineering, Jadavpur University, Kolkata 700032, India (e-mail: mahantapas@gmail.com; mitanasipuri@gmail.com).

Digital Object Identifier 10.1109/TIM.2025.3569005

for adopting traditional approaches to segment retinal blood vessels [13].

B. DL-Based Approaches

In contrast to traditional approaches, DL has provided superior performance in the domain of RV segmentation. Liskowski and Krawiec [13] exhibited one of the first research in this domain utilizing patch-based methodology for training a deep neural network for segmentation. U-Net model variants [14], [15], [16], [17], [18], [19] have been explored in the literature previously to train different algorithms for superior segmentation performance. Jin et al. [20] employed a U-Net-shaped architecture with a deformable convolution kernel to exploit local features for final segmentation. Mou et al. [21] proposed the use of a dense and dilated network for obtaining initial detection of seed points followed by a probability-driven regularized walk algorithm for segmentation. Gu et al. [22] proposed the use of an encoder–decoder network with a context extractor module for feature modulation. A dense atrous-convolution module had been proposed for context extraction by the authors which had been integrated with a residual block to complete the network. Li et al. [23] used multiple iterations of separated miniaturized U-Net architectures for segmentation. Tan et al. [24] proposed the use of a skeletal prior structure to segment the vessel skeleton along with a contrastive loss to better identify the segmented vessel morphology. The work proposed by Li et al. [3] combined the global Transformer model with a dual-attention model to better segment the RFIs. Data scarcity has been a major bottleneck for the development of sophisticated algorithms for RV segmentation as well. Derivative augmentation approaches such as contrast tuning and geometric transformations [25] had been studied extensively in past literature to address the issue. In this regard, geometric transformations such as rotated, warped, and slanted [13] are considered the current benchmark in data augmentation. However, geometrically altered images often lack pragmatic consistency in the real world, introducing new challenges in designing a transformation-agnostic segmentation model. Recently, a color space mixing-based data augmentation approach has been introduced [24]. Additionally, generative models such as the R-sGAN model had been proposed by Zhao et al. [26] to generate synthetic fundus images by forming a mixture of different color spaces. However, most GANs suffer from improper data distribution consistency between existing and augmented data [26], which makes their adoption problematic.

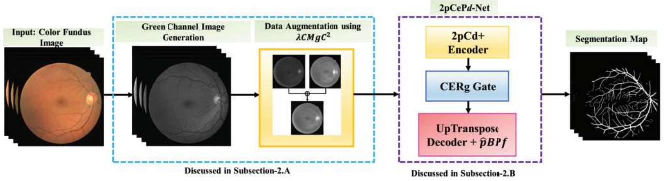


Fig. 1. Block diagram of the proposed RV segmentation method.

To address the aforementioned challenges, we propose in this work a novel end-to-end convolutional neural network-based model, named 2pCeP*d*-Net for RV segmentation. The primary contributions of this work are as follows.

1) A novel lambda-coherence measure-guided Cartesian square (λCMgC^2)-sampling model that leverages both the anatomical as well as the geometrical properties of the retinal images in existing datasets to generate augmented images.
2) A novel two-path cross-dictionary residual (2pCd$^+$) encoder block that incorporates a cross-dictionary attention module with a three-way residual block to incorporate both low-level as well as high-level feature information for learning the encoded space.
3) A novel context-aware encoder resampler gate (CERg) that alleviates skip-connection dependency in the encoder–decoder network while introducing a context-aware feature learning module that bridges the encoder–decoder segment.
4) A novel probability map-based bandpass fusion module ($\hat{p}$BPf), which introduces two distinct sRHP and sRLP noise modulation blocks and a probability map-based margin-of-error (MoE)-guided learning paradigm.

The remainder of this article is organized as follows. Section II introduces and discusses in detail our proposed methodology followed by the experimental results in Section III. The observations from the proposed work are concluded finally in Section IV.

II. PROPOSED METHODOLOGY

Let $\Re^3$ signify a 3-D continuous space, and $\mathcal{I}: \Theta \to \Re^3$ be the RGB fundus image, where $\Theta(\subset \Re^3) = (x, y, z)|x, y, z \in \mathbb{Z}$ is defined as the image domain. Typically, $\mathcal{I}$ consists of three distinguishable subregions, namely, RVs ($\mathcal{I}_{\text{rv}}$), an optic disc ($\mathcal{I}_{\text{od}}$), and a background ($\mathcal{I}_b$). The main objective of this work is to segment $\mathcal{I}_{\text{rv}}$ region from the fundus image. An overview of our proposed RV segmentation method is shown in Fig. 1.

A. Augmentation Using λCMgC^2 Model

Conventional data augmentation techniques such as scaling, rotation, and flipping are ineffective for small-scale datasets [27]. To mitigate the issue, an effective data augmentation strategy has been proposed in [28], which generates an augmented image by combining training samples. However, it is challenging to directly adopt this model in medical imaging, specifically for fundus images, which requires strict adherence to structural coherence. Thus, in this work, we propose a novel λCMgC^2 model for fundus image augmentation.

Let $\mathcal{I}_g: \mathcal{I} \to \Re^2$ be the green channel image plane of $\mathcal{I}$. In this work, each pixel $p_{x,y,v}$ in an image plane ($\subset \Re^2$) is characterized by a triplet x, y, where the index pair $(x, y), x = 1, 2, \ldots, h$ and $y = 1, 2, \ldots, w$, represents the pixel location in the image domain Θ, $v(\in \Re^+)$ signifies the pixel value, and $h, w(\in \Re^+)$ are the height and width of the image, respectively. Let $\mathcal{I}_g^m$ and $\mathcal{I}_g^n$, $m \neq n$, be two RFIs from a dataset $\mathbb{D}$, and $\mathcal{I}_{\text{gt}}^m, \mathcal{I}_{\text{gt}}^n$ are their respective binary pixel-wise ground-truth images, where 0 denotes nonmask pixels and 1 denotes mask pixels. We use the notation of $q_{x,y,0}$ for all nonmask pixels and $q_{x,y,1}$ for all mask pixels in any binary image. Inspired from [28], the augmented image $\mathcal{I}^{\text{aug}}$ can be defined as the output of

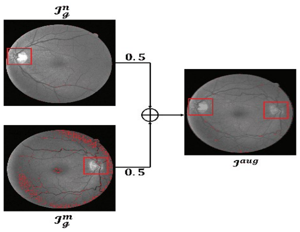


Fig. 2. RFIs having different optic disc position (red square), and the corresponding linear injective function $g(\{\mathcal{I}_g^m, \mathcal{I}_g^n\})$-augmented image having two optic discs.

an injective function $g(.)$ on a subset of the Cartesian product on dataset $\mathbb{D}$ as

$$\mathcal{I}^{\text{aug}} = g\left(\{\mathcal{I}_g^m, \mathcal{I}_g^n\}\right) \mid \{\mathcal{I}_g^m, \mathcal{I}_g^n\} \subset \mathbb{D}^2. \tag{1}$$

In the above equation, $g: \mathbb{D}^2 \rightarrow \Re^2$ maps the Cartesian-square set $\mathbb{D}^2$ to the domain of green-channel image plane. The injective function is used in this work to combine the two images $\mathcal{I}_g^m, \mathcal{I}_g^n$ into a single structurally sound fundus image. In most retinal image datasets, it is oftentimes difficult to define such an injective function $g(.)$ while preserving the structural anatomy of an RFI. For example, as shown in Fig. 2, using a simple linear function $g(\{\mathcal{I}_g^m, \mathcal{I}_g^n\}) = 0.5.\mathcal{I}_g^m + 0.5.\mathcal{I}_g^n$ on the DRIVE dataset can lead to the formation of medically incoherent and impossible images.

Thus, in this work, we have introduced a novel λCMgC^2 sampling augmentation model to generate fundus images. The coherence measure, denoted by $\widehat{C}$, evaluates the structural coherence between two input images before generating an augmented image using the injective function $g(.)$. Formally, we define $\widehat{C}(\mathcal{I}_g^m, \mathcal{I}_g^n)$ as the degree of positional overlap of the OD regions, $\mathcal{I}_{\text{od}}^m$ and $\mathcal{I}_{\text{od}}^n$, of images $\mathcal{I}_g^m$ and $\mathcal{I}_g^n$, respectively. The proposed λCMgC^2 is thus primarily a two-stepped model. First, we have identified the OD region from both selected fundus images $\mathcal{I}_g^m$ and $\mathcal{I}_g^n$ and computed the coherence ($\widehat{C}$) between them. Subsequently, we generated the augmented images using the injective function $g(.)$ after the optic disc alignment.

Definition 1 (Optic Disc Region): Optic disk region $\mathcal{I}_{\text{od}}(\subseteq \mathcal{I}_g)$ in an RFI $\mathcal{I}_g$ is a connected and cohesive region such that $(1/N_{\text{od}})\sum\{v|p_{x,y,v} \in \mathcal{I}_{\text{od}}\} > (1/(N - N_{\text{od}}))\sum\{v|p_{x,y,v} \in (\mathcal{I}_g - \mathcal{I}_{\text{od}}))\}$, where N_{od} represents the number of pixels in $\mathcal{I}_{\text{od}}$ and $N = h \times w$.

Here, the OD region of a fundus image is considered a marker to measure structural coherence between a pair of images. Thus, the identification of OD regions in fundus images is a primary step to our λCMgC^2 method. In this work, we have proposed an iterative method to quantitatively depict and identify the OD region from an RFI which is discussed as follows. First, we have removed vessel regions from the pair of images, that is, $\mathcal{I}_g^m$ and $\mathcal{I}_g^n$, using the following equation:

$$\mathcal{I}_{\text{vr}}^m = \left\{p_{x,y,v} | \forall(x,y), \left(p_{x,y,v} \in \mathcal{I}_g^m\right) \wedge \left(q_{x,y,0} \in \mathcal{I}_{\text{gt}}^m\right)\right\}. \tag{2}$$

In the above equation, I_{vr}^m signifies the vessels removed set of pixels of $\mathcal{I}_g^m$ and $\mathcal{I}_{\text{gt}}^m$ is the ground truth of $\mathcal{I}_g^m$. Similar method is used to generate $\mathcal{I}_{\text{vr}}^n$. It is noted that vessel pixels are disregarded when computing $\widehat{C}$ as OD coherence is measured on the region as a whole which is mostly concerned with background region rather than vessel pixels. For simplicity, we have represented $\mathcal{I}_{\text{vr}}^m$ and $\mathcal{I}_{\text{vr}}^n$ together as $\mathcal{I}_{\text{vr}}^{\{l|l=m,n\}}$. Similar naming convention is also used for $\mathcal{I}_g^m$ and $\mathcal{I}_g^n$, that is, $\mathcal{I}_g^{\{l|l=m,n\}}$. Let us assume that $\mathcal{I}_b^{\{l|l=m,n\}}$ and $\mathcal{I}_{\text{od}}^{\{l|l=m,n\}}$ are two sets representing the background region and the optic disc region, respectively. At the beginning, $\mathcal{I}_b^{\{l|l=m,n\}}$ and $\mathcal{I}_{\text{od}}^{\{l|l=m,n\}}$ are initialized as empty set, that is, $\mathcal{I}_b^{\{l|l=m,n\}} = \Phi$ and $\mathcal{I}_{\text{od}}^{\{l|l=m,n\}} = \Phi$. Now, at every iteration, the pixel having minimum intensity is computed from $\mathcal{I}_{\text{vr}}^{\{l|l=m,n\}}$ for further processing as

$$\hat{m} = \min\left(\left\{v | \forall p_{x,y,v} \in \left(\mathcal{I}_{\text{vr}}^{\{l|l=m,n\}} - \mathcal{I}_b^{\{l|l=m,n\}}\right)\right\}\right). \tag{3}$$

Subsequently, based on the value of $\hat{m}$, all pixels in $\mathcal{I}_{\text{vr}}^{\{l|l=m,n\}}$ are partitioned into two groups, namely the background pixel set $\mathcal{I}_b^{\{l|l=m,n\}}$ and the OD pixel set $\mathcal{I}_{\text{od}}^{\{l|l=m,n\}}$, using by the following equations in order, as

$$\mathcal{I}_b^{\{l|l=m,n\}} = \left\{p_{x,y,v} | \forall p_{x,y,v} \in \mathcal{I}_{\text{vr}}^{\{l|l=m,n\}}, v \leqslant \hat{m}\right\} \tag{4}$$

$$\mathcal{I}_{\text{od}}^{\{l|l=m,n\}} = \left\{p_{x,y,v} | \forall p_{x,y,v} \in \left\{\mathcal{I}_{\text{vr}}^{\{l|l=m,n\}} - \mathcal{I}_b^{\{l|l=m,n\}}\right\}\right\}. \tag{5}$$

After which, the average intensity of pixel elements in $\mathcal{I}_b^{\{l|l=m,n\}}$ is computed and compared with the average intensity of $\mathcal{I}_{\text{od}}^{\{l|l=m,n\}}$, which can be defined in the following equation:

$$\left(\frac{\sum_{x=1,y=1}^{h,w}\left\{v| \quad \forall p_{x,y,v} \in \mathcal{I}_{\text{od}}^{\{l|l=m,n\}}\right\}}{N_{\text{od}}} - \frac{\sum_{x=1,y=1}^{h,w}\left\{v| \quad \forall p_{x,y,v} \in \mathcal{I}_b^{\{l|l=m,n\}}\right\}}{N_b}\right) > t. \tag{6}$$

In the above equation, N_{od} and N_b signify the number of pixels in $\mathcal{I}_{\text{od}}^{\{l|l=m,n\}}$ and $\mathcal{I}_b^{\{l|l=m,n\}}$, respectively, and t is a threshold whose value is selected empirically as 150 in this work. If the difference is less than t, the process will stop, and the resulting set $\mathcal{I}_{\text{od}}^{\{l|l=m,n\}}$ will only contain the approximate OD region pixels. Otherwise, we repeat the earlier steps till the conclusion of the algorithm (as shown in Fig. 3). After the completion, the set $\mathcal{I}_{\text{od}}^{\{l|l=m,n\}}$ is treated as a binary mask set for the OD region, denoted by $q_{x,y,0}, q_{x,y,1}$ notations. After OD region identification, the coherence measure $\widehat{C}(\mathcal{I}_g^m, \mathcal{I}_g^n)$ can be defined formally as

$$\widehat{C}\left(\mathcal{I}_g^m, \mathcal{I}_g^n\right) = n\left(\left\{q_{x,y,1} \mid \forall(x,y), q_{x,y,1} \in \left(\mathcal{I}_{\text{od}}^m \cap \mathcal{I}_{\text{od}}^n\right)\right\}\right). \tag{7}$$

In the above equation, $n(.)$ is the cardinality of a set. It is noted that the coherence $\widehat{C}$ needs to be measured until the

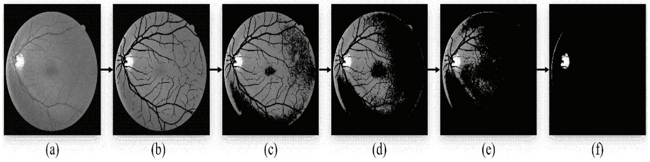

(a) (b) (c) (d) (e) (f)

Fig. 3. Computation of the OD region for $t = 150$: (a) green-channel image and (b) vessel-removed image using (3). (c)–(f) OD region based on the computation of (7) having output values 109.04, 83.15, 126.13, and 153.78, respectively.

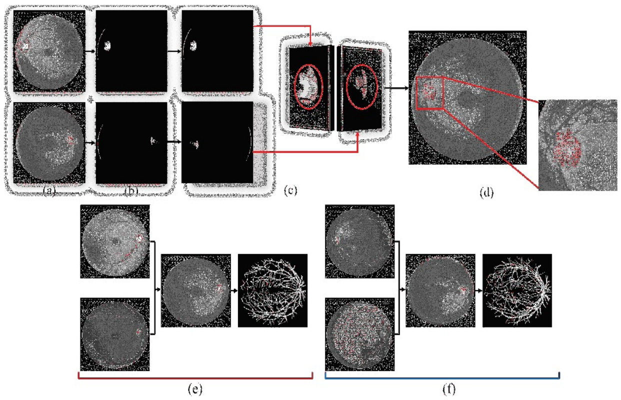


Fig. 4. Computation of the augmented image: (a) green-channel images, (b) isolation of the OD region, (c) OD alignment with $\widehat{C}$ equal to 98.43%, (d) λCMgC^2 image with enlarge optic disc region, and (e) and (f) image pairs and corresponding augmented image and respective ground truth.

OD region set $\mathcal{I}^n_{\text{od}}$ of $\mathcal{I}^n_g$ has maximal intersection with the OD region set $\mathcal{I}^m_{\text{od}}$ of $\mathcal{I}^m_g$. We thus perform rotation operation on the respective green-channel images I^m_g, I^n_g until $\widehat{C}$ is maximized. For simplicity, we have kept the image $\mathcal{I}^m_g$ unchanged and rotated $\mathcal{I}^n_g$ through an angle of 10^o at every iteration and subsequently measured the coherence.

Let $\mathcal{A}_{\widehat{C}} = \{(\widehat{C}(\mathcal{I}^m_g, \widehat{\mathcal{I}}^n_g), \theta) | \forall\theta, \theta \in [0^o, 360^o], \widehat{\mathcal{I}}^n_g = \text{ROTATE}(\mathcal{I}^n_g, \theta)\}$ be a set comprising of elements of the pair form $\langle\widehat{C}, \theta\rangle$. Here, $\widehat{C}$ denotes the coherence measure computed between images $\mathcal{I}^m_g$ and output of function ROTATE$(\mathcal{I}^n_g, \theta)$. ROTATE$(., \theta)$ denotes a function which rotates an input image by θ^o and returns the output image, respectively. Finally, we select the element pair $\langle\widehat{C}, \theta\rangle$ from $\mathcal{A}_{\widehat{C}}$ for which $\widehat{C}$ is the highest and rotate the image $\mathcal{I}^n_g$ by the corresponding pair value of θ, as shown in the following equation:

$$\theta = \arg_\theta\left(\max_{\widehat{C}}\left(\mathcal{A}_{\widehat{C}}\right)\right). \tag{8}$$

Let $\widehat{\mathcal{I}}^n_g$ denote the final θ^o rotated image of I^n_g. It is noted that the ground truth $\mathcal{I}^n_{\text{gt}}$ of $\mathcal{I}^n_g$ is also rotated with the same angle of θ^o and denoted as $\widehat{\mathcal{I}}^n_{\text{gt}}$. Finally, we define a λ-weighted linear injective function $g(.)$ to map the set $\{\mathcal{I}^m_g, \mathcal{I}^n_g\}$ to a green-channel image plane to generate the augmented image $\mathcal{I}^{\text{aug}}$ along with ground truth $\mathcal{I}^{\text{aug}}_{\text{gt}}$ as

$$\begin{aligned}\mathcal{I}^{\text{aug}} &= \lambda\mathcal{I}^m_g + (1-\lambda)\widehat{\mathcal{I}}^n_g \\ \mathcal{I}^{\text{aug}}_{\text{gt}} &= \mathcal{I}^m_{\text{gt}} \cup \widehat{\mathcal{I}}^n_{\text{gt}}.\end{aligned} \tag{9}$$

In the above equation, $\lambda \in (0, 1)$. Fig. 4 shows the image pairs and corresponding augmented image and respective ground truth. In this regard, to prove the structural integrity of the augmented image $\mathcal{I}^{\text{aug}}$, we state the following theorem.

Theorem 1: Given λ-weighted linear injective function $g : \mathbb{D}^2 \rightarrow \Re^2$, any vessel region perceptible in $\mathcal{I}_g^{\{l|l=m,n\}}$ remains perceptible in $\mathcal{I}^{\text{aug}}$.

The proof of the theorem is presented in the Appendix.

B. 2pCePd-Net Model

The proposed 2pCeP*d*-Net architecture is inspired by the design of an encoder–decoder structure as the backbone with three distinct sections, namely, 2pCd$^+$ encoder, CERg, and probabilistic bandpass fusion ($\hat{p}$BPf) module. Fig. 5 illustrates the structure of the proposed model. As can be observed in the figure, we denote each horizontal segment of the proposed network as a level, denoted by l. We use this notation extensively across our subsequent model discussion. Sections II-B1–II-B4 provide a detailed explanation of each individual section of the network.

1) Two-Path Cross-Dictionary Residual Encoder: Our proposed model starts from the encoder section as highlighted in Fig. 5. The encoder module is structured in alignment with an encoder–decoder module with skip connections, as is common in literature [29]. However, such a structure also introduces challenges in the vessel segmentation tasks in RFIs [30]. Notably, the continuous downsampling of features using a generic max-pooling layer does not preserve the subtle features necessary for decoding thin vessel regions [3]. To address this issue, we introduce a novel 2pCd$^+$ encoder that preserves the important features for segmentation by introducing a cross-dictionary-based attention module, as depicted in Fig. 6.

First, we have deployed a chain of strided convolutions (stride = 2), which runs parallel to the encoder stream and acts as an auxiliary stream of features. It is noted that the max-pooling operation in the encoder stream enhances the translation invariance capabilities in CNN models, whereas the strided convolution in the auxiliary stream preserves better feature information [31]. To combine the advantages of both modules, a cross-dictionary-guided encoder block, named Cd$^+$, is designed in this work. Let us denote the auxiliary stream input for a given **Enc** block at level l as $f^{\text{aux}}_{\text{in}}$ and the encoder stream input from the **Enc** block at level *l-1* as $f^{\text{enc}}_{\text{in}}$, respectively. To enhance the feature representation in the downsampled encoder output, we take inspiration from the cross-attention [32] mechanism and compute the dictionary value matrices D^{enc}_v and D^{aux}_v, along with dictionary key matrix D^{enc}_k and dictionary query matrix D^{aux}_q, respectively, defined as follows:

$$\begin{aligned}D^{\text{enc}}_v &= \hat{f}^{\text{enc}}_{\text{in}} * \Phi^{\text{enc}}_v \\ D^{\text{enc}}_k &= \hat{f}^{\text{enc}}_{\text{in}} * \Phi^{\text{enc}}_k \\ D^{\text{aux}}_v &= f^{\text{aux}}_{\text{in}} * \Phi^{\text{aux}}_v \\ D^{\text{aux}}_q &= f^{\text{aux}}_{\text{in}} * \Phi^{\text{aux}}_q.\end{aligned} \tag{10}$$

In the above equation, $\hat{f}^{\text{enc}}_{\text{in}}$ represents the downsampled feature vector of $f^{\text{enc}}_{\text{in}}$ and $\Phi^{\text{enc}}_v, \Phi^{\text{enc}}_k, \Phi^{\text{aux}}_v, \Phi^{\text{aux}}_q$ represent the convolutional weight matrices, respectively. Subsequently, we merge the salient features from the two dictionaries and compute cross-attention on the downsampled feature space $\hat{f}^{\text{enc}}_{\text{in}}$ as follows:

$$f^{\text{temp}} = \text{softmax}\left(\frac{(D^{\text{enc}}_k)^T \odot D^{\text{aux}}_q}{\sqrt{d}}\right) \odot D^{\text{enc}}_v. \tag{11}$$

In the above equation, f^{temp} represents the cross-attention feature vector, d represents the channel dimension, $(D^{\text{enc}}_k)^T$ represents the transposed matrix, and $\odot$ represents the Hadamard product. Finally, the output of the Cd$^+$ encoder block ($f^{\text{Cd}^+}_{\text{out}}$) is computed as follows:

$$f^{\text{Cd}^+}_{\text{out}} = f^{\text{temp}} + \hat{f}^{\text{enc}}_{\text{in}} + \sigma\left(D^{\text{aux}}_v\right). \tag{12}$$

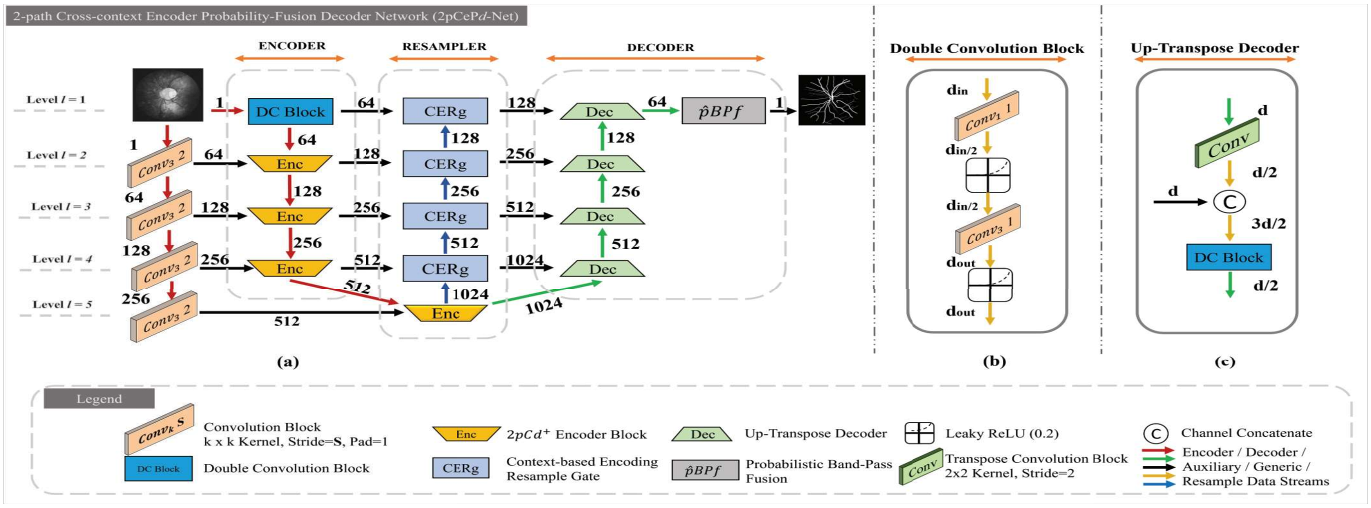


Fig. 5. Illustration of the proposed 2pCePd-Net model. (a) Symbolic block-based depiction of the proposed 2pCePd-Net. The symbolic interpretations of the network blocks are highlighted in the *legend*. (b) Detailed symbolic design of the double convolution block [marked as *dc block* in (a)]. (c) Detailed symbolic design of the up-transpose decoder block [marked as *Dec* in (a)].

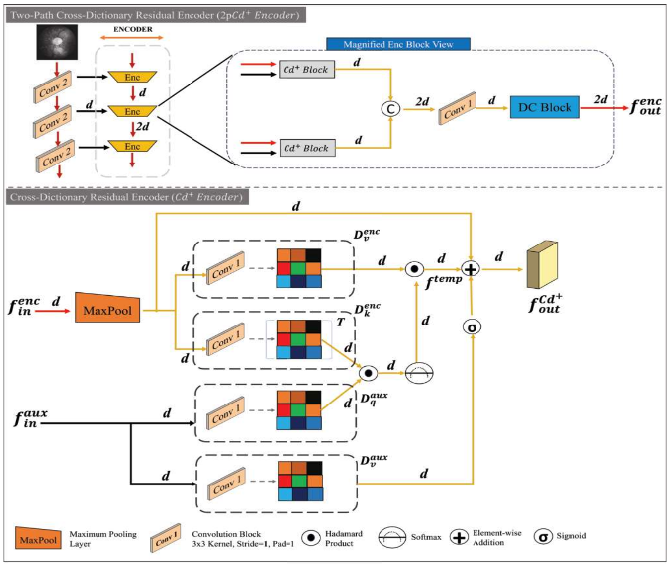


Fig. 6. Depiction of the proposed 2pCd^{+} encoder block.

In the above equation, $\sigma(.)$ denotes the Sigmoid function. The addendum of $\sigma(D_v^{\text{aux}})$ in (12) incorporates an additional pathway for f^{temp} to recover any depreciated feature of importance due to the initial max-pooling operation. Furthermore, the resultant feature $(f^{\text{temp}} + \sigma(D_v^{\text{aux}}))$ is added to the encoder stream input $f_{\text{in}}^{\text{enc}}$, similar to a residual block [33]. This enforces the network to learn the difference transformation [33] function between the input and output instead of the complete transformation, which leads to better encoding. However, it is noted that although our proposed Cd^{+} encoder block is capable of capturing relevant features, it might get biased toward certain patterns and dependencies, a common phenomenon in single-head attention mechanisms as discussed in [34]. To mitigate such issues, we employ a multiblock $n-$Cd^{+} encoder module to avoid local fitting and preferential dependencies, where n represents the number of Cd^{+} blocks. In this work, we empirically set the value of $n = 2$ for maintaining a balance between the quality of output with computational resources and thus design the proposed $2-$Cd^{+} encoder block as shown in Fig. 6. We combine the outputs obtained from each encoder as a channel concatenation followed by subsequent convolution operations, as shown in the figure, to obtain the final combined encoder output $f_{\text{out}}^{\text{enc}}$. The combined encoder architecture with multiple Cd^{+} blocks and combinatorial layers is denoted as 2pCd^{+} block in this work.

2) *Context-Aware Encoder Resampler Gate:* Similar to standard CNN architectures, in this work, we have incorporated a skip-connection bridge between the encoder section and the decoder section, respectively, for better segmentation. However, the usage of skip connections presents an alternate pathway to the network to bypass the deeper layers, leading to the dependence of the network on shallower features [13]. Additionally, traditional skip connections duplicate the less-informed features of the model from shallower layers which often lead to incorrect target boundary detection in medical image segmentation [35]. To simultaneously mitigate both the issues in traditional skip connections, we propose a novel CERg in this work. First, we present three axioms relevant to all learned encoded feature spaces in a CNN model for RV segmentation. **Axiom 1:** a well-learned feature space $\mathcal{F}$ would represent all vessel regions separate from background regions irrespective of local vessel properties. **Axiom 2:** a well-learned feature space $\mathcal{F}$ would be devoid of any influence from noise. **Axiom 3:** a feature space $\hat{F}$ derived from $\mathcal{F}$ would preserve all information present in $\mathcal{F}$. To incorporate the axioms in our network, we take inspiration from past context-sensitive networks [36]. A context-aware skip connection bridge is designed in this work to incorporate information from deeper layers. This introduces a framework to modulate an over-dependence of the network on less-informed features on skip connections, as shown in Fig. 7. As can be observed in the figure, the CERg block comprises three distinct submodules,

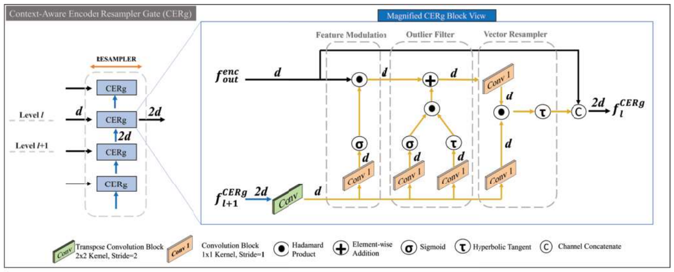


Fig. 7. Depiction of the proposed CERg resampler block.

namely feature modulation (f^{FM}), outlier filter (f^{OF}), and vector resampler (f^{VR}), respectively.

Let us consider a CERg module on skip connection at level l. It is well documented in the literature that encoder outputs at level $\hat{l} < l$ are more susceptible to noisy inputs [29] than deeper layers. A simple skip connection at shallower layers would transmit a noise-infused feature onto the decoder, inducing an additional difficulty for the decoder to adhere to Axiom 1. In such scenarios, an emphasis on Axiom 2 might aid in mitigating an additional workload on decoders. Furthermore, it is also important to preserve the relevant features on the skip connection for improving segmentation. These observations lead to the inspiration of the CERg skip connection module. First, a feature modulation unit takes an encoder output $f_{\mathrm{out}}^{\mathrm{enc}}$ at level l as input as well as the CERg block output at level $l+1$ and computes f^{FM}. In this context, it is noted that encoder outputs at deeper layers are less prone to noise [17] while preserving the relevant features necessary for segmentation (Axiom 3). Thus, f^{FM} is equated mathematically as

$$f^{\mathrm{FM}} = f_{\mathrm{out}}^{\mathrm{enc}} \odot \sigma\left(\hat{f}_{l+1}^{\mathrm{CERg}} * \Phi_{1,l+1}^{\mathrm{CERg}}\right). \tag{13}$$

In the above equation, $\hat{f}_{l+1}^{\mathrm{CERg}}$ represents de-convoluted matrix and $\Phi_{1,l+1}^{\mathrm{CERg}}$ represents the convolutional weights, respectively. Equation (13) highlights a modulation where less represented features from deeper layers are suppressed in f^{FM} (due to $\sigma(.)$), thus emphasizing useful information. An outlier filter unit is appended to the feature modulation unit to further boost the informative signal in the output f^{FM} while depreciating the less-represented features. Mathematically, it is represented as

$$\begin{aligned} f^{\mathrm{OF}} = f^{\mathrm{FM}} &+ \left(\sigma\left(\hat{f}_{l+1}^{\mathrm{CERg}} * \Phi_{2,l+1}^{\mathrm{CERg}}\right)\right) \\ &\odot \left(\tanh\left(\hat{f}_{l+1}^{\mathrm{CERg}} * \Phi_{3,l+1}^{\mathrm{CERg}}\right)\right). \end{aligned} \tag{14}$$

In the above equation, $\Phi_{2,l+1}^{\mathrm{CERg}}, \Phi_{3,l+1}^{\mathrm{CERg}}$ represents the convolutional weights. Equation (14) highlights that a less represented feature in $\hat{f}_{l+1}^{\mathrm{CERg}}$ would generate a tangent value closer to 0 or even negative, whereas a well-represented feature would generate a value closer to 1. This property of hyperbolic tangent is used as an additive input to boost or depreciate the features, respectively. Due to the additive nature of the design, a sigmoid function is multiplied with the tangent output to potentially limit the range of feature fluctuation for stabilized learning. Finally, to limit the feature values within range $[-1, 1]$ (a zero-centered distribution [31]) and prevent gradient abnormality [37], a vector resampler unit is introduced. It is desirable that informative features have a higher representation in the feature vector in comparison to less informative features in the final output. Thus, we resample the feature space in the vector resampler unit and compute the final output as

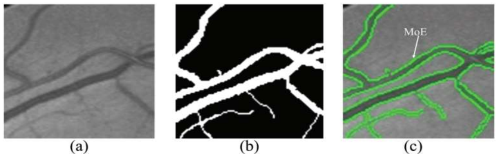


Fig. 8. Illustrates the MoE. (a) Fundus image patch. (b) Corresponding ground-truth patch. (c) MoE (green line).

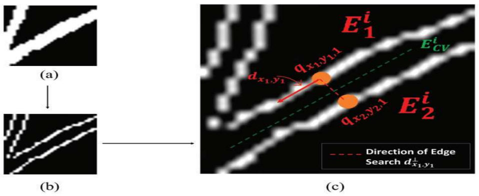


Fig. 9. (a) Vessel segment from a ground truth. (b) Vessel boundary after applying Canny-edge detector. (c) Edge pair searching in context of point $p_{x_1,y_1,1}$ of edge E_1^i.

$$f^{\mathrm{VR}} = \tanh\left\{\left(f^{\mathrm{OF}} * \Phi^{\mathrm{OF}}\right) \odot \left(\hat{f}_{l+1}^{\mathrm{CERg}} * \Phi_{4,l+1}^{\mathrm{CERg}}\right)\right\}. \tag{15}$$

In the above equation, $\Phi_{4,l+1}^{\mathrm{CERg}}, \Phi^{\mathrm{OF}}$ represents the convolutional weights. The obtained feature map f^{VR} is channel-concatenated with the input $f_{\mathrm{out}}^{\mathrm{enc}}$ and sent as an input to the decoder module as represented in Fig. 5.

3) Probabilistic Bandpass Fusion ($\hat{p}$BPf): We incorporate up-transpose decoder modules in the decoder section of our proposed network, as shown in Fig. 5. An up-transpose decoder at level l takes inputs from the CERg block at level l and an up-transpose decoder block from level $l+1$, respectively. It is noted that a model performance not only depends on model architecture but also on the training methodology which directs the gradients in the network to adjust the weights optimally. In this work, we emphasize and direct the model to optimally segment a challenging region, defined herewith as an MoE region (shown in Fig. 8). The MoE region is defined as a vessel region having poor contrast with the adjacent background region. This enhances the difficulty in identifying the boundary of the vessels as well as the proper delineation of thin vessels in the RV segmentation task. It is trivial to state that all thin vessel regions are subsets of MoE regions as well. To better augment the learning process, especially in the MoE region, we propose the use of a probability map in this work to direct the model learning as discussed next.

Let us consider $\mathcal{I}_{\mathrm{gt}}^{\mathrm{aug}}$ be the ground-truth image of the fundus image $\mathcal{I}^{\mathrm{aug}}$. Initially, we apply Canny edge detector [38] on the image $\mathcal{I}_{\mathrm{gt}}^{\mathrm{aug}}$ to obtain a binary vessel edge map $\mathcal{I}_{\mathrm{gt}}^{C}$, which highlights the vessel edges in the ground truth, as shown in Fig. 9. We calculate the local pixel direction $d_{x,y}$ for $q_{x,y,1} \in \mathcal{I}_{\mathrm{gt}}^{C}$ as highlighted in [38]. Additionally, we denote $d_{x,y}^{\perp}$ as the orthogonal direction to $d_{x,y}$ which is directed toward the vessel. To illustrate our formulation, we consider a fragment of a

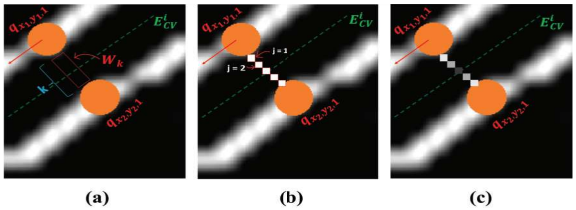


Fig. 10. Computation of the probability map for vessel i at point $p_{x_1,y_1,1}$. (a) Window W_k of length $k \times 1$ at point $p_{x_1,y_1,1}$. (b) Computation of probability at point q_j (13). (c) Probability map generated for vessel i at point $p_{x_1,y_1,1}$.

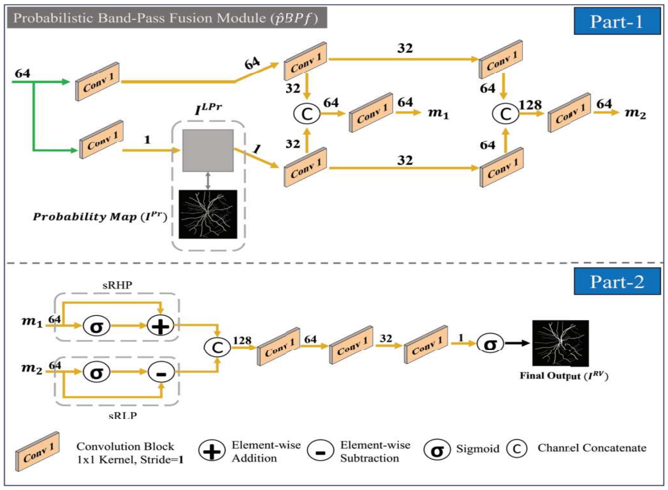


Fig. 11. Depiction of the proposed $\hat{p}$BPf fusion block.

vessel i comprising of two vessel edge fragments E^i_1 and E^i_2 which form the boundary of the vessel i, as shown in Fig. 9. For each pixel $q_{x_1,y_1,1} \in E^i_1$, we traverse along $d^{\perp}_{x_1,y_1}$ till we reach pixel $q_{x_2,y_2,1} \in E^i_2$. We define a hypothetical edge E^i_{CV}, parallel to d_{x_1,y_1}, which passes through the midpoint between pixels $q_{x_1,y_1,1}$ and $q_{x_2,y_2,1}$, as shown in Fig. 9. We define a window W_k of size $k \times 1$, positioned on E^i_{CV} and oriented along $d^{\perp}_{x_1,y_1}$, intersecting $q_{x_1,y_1,1}$ and $q_{x_2,y_2,1}$ (shown in Fig. 10). Here, k is equal to the Euclidean distance between pixels $q_{x_1,y_1,1}$ and $q_{x_2,y_2,1}$. We denote all pixels, which lie within window W_k, as u_j with $j = 0, \ldots, k-1$, where $u_0 = q_{x_1,y_1,1}$ and $u_{k-1} = q_{x_2,y_2,1}$. The probability of a pixel u_j to be the edge pixel is thus computed as

$$\mathcal{P}(q_j) = \delta^j (1-\delta)^{|(k/2)-j|}. \tag{16}$$

In the above equation, $\delta = 0.25$ is empirically set in this work. Equation (16) assigns the highest probability to the edge pixels $q_{x_1,y_1,1}$, $q_{x_2,y_2,1}$ and gradually decreases as j approaches the edge E^i_{CV}, as shown in Fig. 10. The process is repeated across all pixels $q_{x,y,1}$ in $\mathcal{I}^C_{\mathrm{gt}}$ to generate the final vessel boundary map I^{Pr}. Finally, the generated probability map $\mathcal{I}^{\mathrm{Pr}}$ is used as an auxiliary ground truth to train our proposed model.

We propose a novel probabilistic bandpass fusion module to incorporate the probability map in the training process of the network, as shown in Fig. 11 (Part-1). As depicted in the figure, the probability map is convoluted and concatenated, at lower as well as higher-feature levels to generate two feature maps, namely, m_1 and m_2, respectively. The model is thus further tasked to predict the probability map along with the vessel segmentation, emphasizing the MoE region specifically. However, this emphasis on the MoE region introduces a susceptibility in the model to noise [39] which, in turn, gets encoded into the m_1 feature vector. To address the issue, we incorporate a novel self-rectified high pass gate (sRHP) which enables the feature map m_1 to rectify the encoded space and reduce the influence of low-frequency background noise (low-contrast regions). In addition, we also propose a novel self-rectified low-pass gate (sRLP) which enables the higher-order m_2 feature map to remove high-frequency spikes in intensities which manifest as speckle noise in segmented outputs of standard CNN models [40]. The sRHP and sRLP gates can be represented mathematically as

$$\begin{aligned} \mathrm{sRHP} &= m_1 + \sigma(m_1) \\ \mathrm{sRLP} &= m_2 - \sigma(m_2). \end{aligned} \tag{17}$$

In the above equation, it is noted that the sRHP module effectively enhances the prominence of dominant features due to the additive operation. Additionally, it enhances the bandwidth between low-frequency noise and information in the feature space, thus making noise more noticeable. In contrast, sRLP depreciates dominant features which leads to better representation of less-prominent features. Simultaneously, a higher frequency noise in the feature space gets better suppressed as well due to the subtractive operation. The combined *sRHP + sRLP gates* (coined as bandpass gates) thus possess the ability to counteract the effect of noise induced due to MoE detection. The final segmented output $\mathcal{I}^{\mathrm{RV}}$ is thus obtained from the $\hat{p}$BPf fusion module along with the predicted probability map $\mathcal{I}_{\mathrm{LPr}}$, as shown in Fig. 11 (Part-2). In the subsequent section, we discuss the training paradigm used in this work to train the proposed model.

4) Training and Implementation Details: The proposed 2*pCePd*-Net is trained using a generalized Dice loss function [41], denoted as $L_{\mathcal{I}^{\mathrm{RV}}}$, for a better overall segmentation. In addition, the network is trained with an auxiliary objective to optimize the predicted probability map, which is trained using a mean squared error (MSE) function [29], termed as $L_{\mathcal{I}^{\mathrm{LPr}}}$.

We combine the loss functions and present our final loss function as

$$L = \alpha * L_{\mathcal{I}^{\mathrm{RV}}} + \beta * L_{\mathcal{I}^{\mathrm{LPr}}}. \tag{18}$$

In the above equation, α and β are loss coefficient weights, empirically set to 0.6 and 0.4, respectively.

III. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we systematically assess the segmentation performance of the proposed method on four public benchmark datasets. First, the fundus images dataset used for experimentation is described. Second, the efficacy of the proposed data augmentation method λCMgC^2 is evaluated. Finally, the efficacy of the presented segmentation results with respect to individual blocks of the proposed 2pCeP*d*-Net model is evaluated. The proposed model is coded in the PyTorch environment having 64 GB of RAM with 2xRTX 3060 GPUs for faster processing. Each individual module of the platform is trained using the Adam optimizer with a maximum iteration of 1000 and a learning rate of 3e^{-}4, constrained on an early stopping criterion based on the loss per iteration. To prevent overfitting, an L2 regularization is used in the training as well. All modules are trained with a batch size of 3 to limit the need for excess memory.

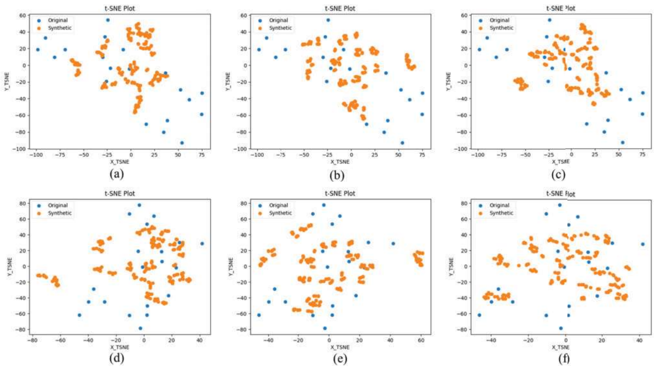
A. Database of Fundus Images

The fundus images used in this work are collected from four widely used databases: DRIVE [42], STARE [43], CHASEDB1 [44], and HRF [45]. The DRIVE dataset contains 40 color fundus images (20 for training and 20 for testing), whereas the STARE and CHASEDB1 datasets are randomly separated in this work into four sets each. Three sets are used to train the model, while one is used for testing. The procedures are performed iteratively to ensure that all sets are tested. Finally, the average of the four sets of test results is reported in this article. The images in the HRF dataset are separated into three sets, each containing 15 randomly selected images, and the network is trained iteratively on two of them while testing the third. The cumulative average of the results is reported in this work. It is noted that we have rescaled all images across all datasets to 512 × 512 pixels to train and evaluate the proposed model. Additionally, for all training sets, we have applied the proposed λCMgC^2 model for data-augmented model training and green-channel image for testing.

B. Performance Evaluation of the Proposed λCMgC^2 Method

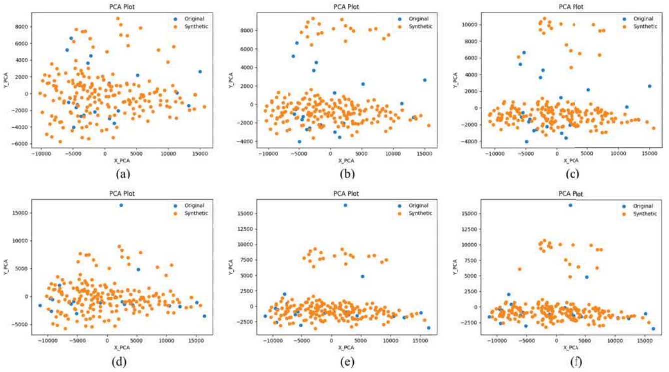
In this section, the ability of the proposed λCMgC^2 method to generate structurally coherent fundus images is evaluated. First, we evaluate the λCMgC^2 model using t-distributed stochastic neighbor embedding (t-SNE) and principal component analysis (PCA) plots. A robust visualization of real images and λCMgC^2 images is plotted for the DRIVE and STARE datasets (Figs. 12 and 13). Three sets of 200 unique synthetic images have been selected from the λCMgC^2-generated synthetic image dataset for visual comparison with real images. All synthetic images have been selected from the λCMgC^2 dataset having $\lambda \in 0.4, 0.5, 0.6$ to demonstrate better visualization between real images and synthetic images which are most dissimilar to original datasets. From Figs. 12 and 13, it can be seen that the generated images closely match the distribution of original images. Most of the synthetic images are positioned close to real images; however, Fig. 12 depicts that the synthetic image distribution is not centered in one region. This demonstrates the proposed method's ability to generate realistic data even with limited training data.

Furthermore, two commonly used metrics: Frechet inception distance (FID) and multiscale structural similarity index measure (MS-SSIM) [46] are computed to assess the discriminative features of the λCMgC^2-generated images in comparison to real images. FID is highly correlated with image visual quality. A low FID score is usually indicative of the visual authenticity of synthetic data. In addition, MS-SSIM scores of 600 randomly selected pairs of synthetic images are computed to assess the diversity of the images in the real and synthetic datasets. The higher mean MS-SSIM score correlates with higher similarity between the real and synthetic datasets. We have obtained an overall FID score of 1.5418 compared to a score of 3.5583 for traditional augmentation (TA) [13]. Similarly, we have obtained an MS-SSIM score of 0.6948 compared to 0.6119 for TA approaches. The lower FID scores of λCMgC^2 indicate that the images it generates are of higher quality when compared to the TA method. In addition, the λCMgC^2 method produced more coherent images with respect to the real dataset than the TA method due to the higher average MS-SSIM value. In medical imaging, it is important to not only generate new images but also medically sound and coherent images. Thus, a balance in similarity with existing data is more important along with diversity in the medical domain which we have achieved in this work.



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

Fig. 12. t-SNE plots to illustrate the distribution of real and λCMgC^2-generated images (a)–(c) for the DRIVE dataset and (d)–(f) for the STARE dataset.



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

Fig. 13. PCA plots to illustrate the distribution of real and λCMgC^2-generated images (a)–(c) for the DRIVE dataset and (d)–(f) for the STARE dataset.

C. Segmentation Performance Evaluation

This section presents an in-depth analysis of the segmentation results generated by the proposed method, both qualitatively and quantitatively, using the ground truth generated by the expert.

1) Qualitative Evaluation of the Segmentation Result: Fig. 14 illustrates examples of fundus images from the DRIVE, STARE, CHASEDB1, and HRF databases and the corresponding segmentation results obtained using the proposed model. The first row highlights the original fundus images from DRIVE, STARE, CHASEDB1, and HRF. The second and third

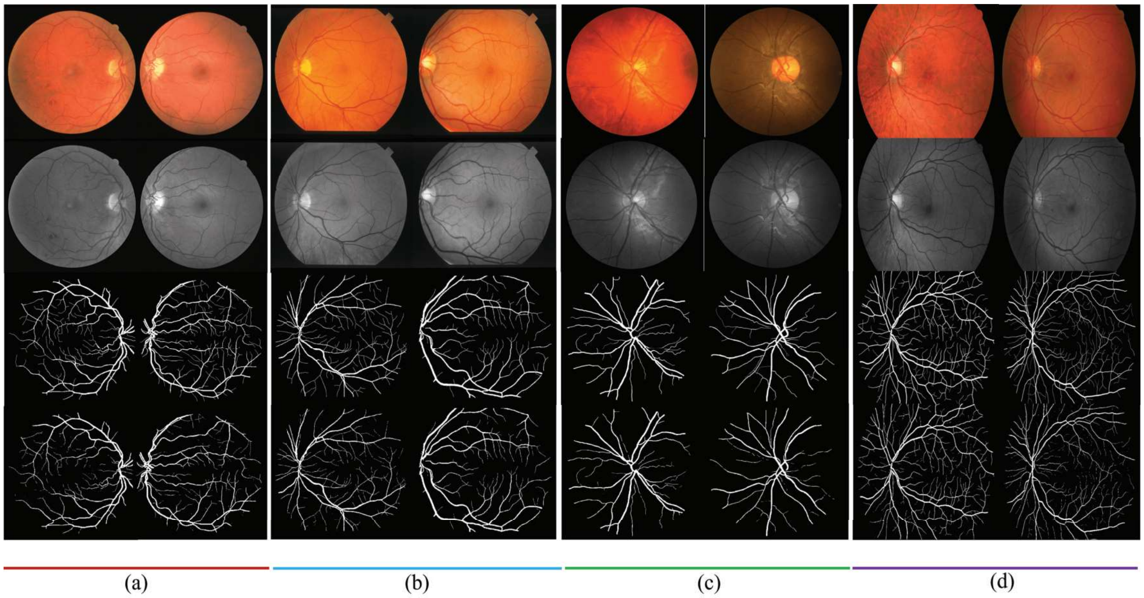


Fig. 14. Comparison of the proposed method's segmentation results and ground truths. (a) DRIVE, (b) STARE, (c) CHASEDB, and (d) HRF.

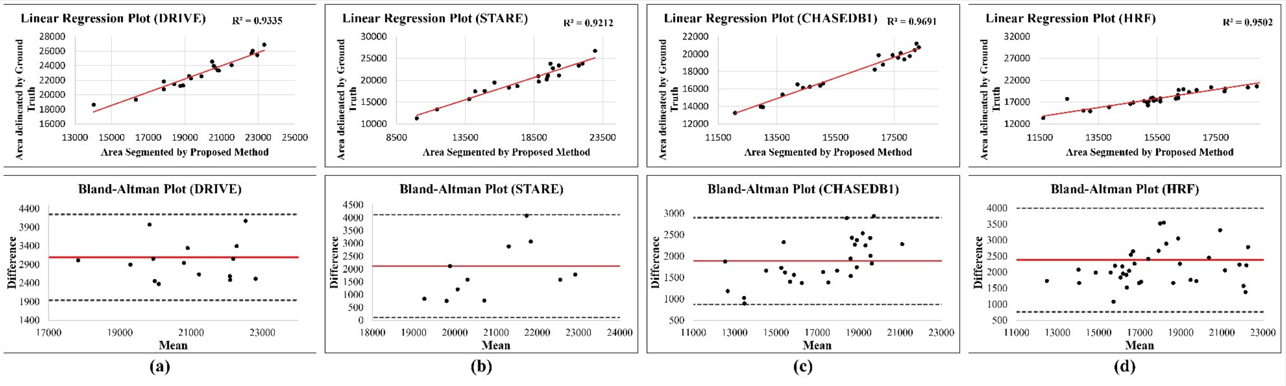
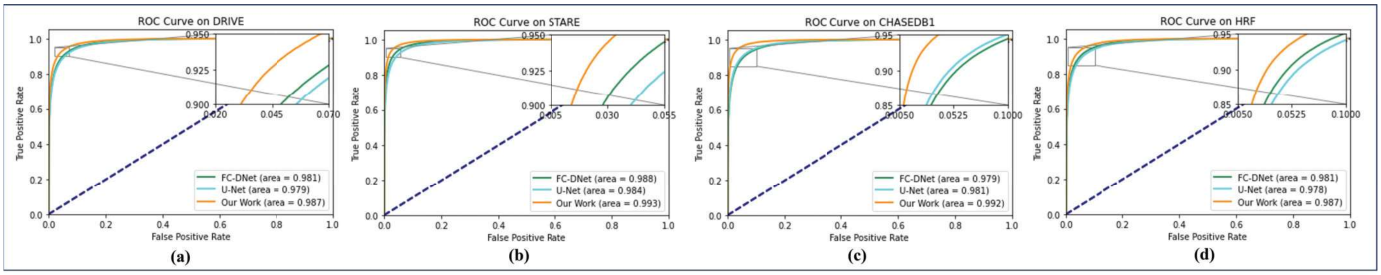


Fig. 15. Linear regression and Bland-Altman plots of 2pCePd-Net against the ground truths. (a) DRIVE, (b) STARE, (c) CHASEDB1, and (d) HRF.

rows depict corresponding green-channel images and ground truths. The fourth row highlights the segmentation results from the proposed model. Comparing the segmentation results of the proposed model with the respective ground truths, it can be observed that the proposed model holds an excellent agreement with the ground truths.

2) Quantitative Evaluation of the Segmentation Result: Quantitative analysis is performed in two stages. First, the consistency of the 2pCePd-Net segmentation results is evaluated with the ground truths annotated by experts. In this context, the linear regression plots and the Bland-Altman plots are constructed, as shown in Fig. 15. It can be observed that there is a high correlation (R^2) between the proposed model's segmentation results and the ground truths for all four databases, with values of 0.9335 (DRIVE), 0.9212 (STARE), 0.9691 (CHASEDB1), and 0.9502 (HRF), respectively. It is also seen that there are very few vessel regions outside of the 95% confidence interval for all four databases (see lower row of Fig. 15). Furthermore, we have calculated Cohen's kappa coefficient (κ) [13] to evaluate the agreement of the proposed method's segmentation results with the ground truths. In general, the value of k varies between 0 and 1, and $\kappa \geq 0.8$ signifies high consistency between the proposed model's results and that of the ground truths. The obtained kappa (κ) values for the four datasets are: 0.86 (DRIVE), 0.91 (STARE), 0.89 (CHASEDB1), and 0.85 (HRF), respectively. These results ensure that the proposed model is highly consistent with the ground truths.

Subsequently, we have compared our proposed model to two baseline networks having similar architectures, viz. the U-Net [29] and the FC-Densenet [41]. Four performance measures have been considered for evaluation, namely, Se, specificity (Sp), and accuracy (Ac) [47]. For a better comparative analysis, the baseline model performances have been evaluated on both TA [13] as well as λCMgC^2 augmentation (Table I). It is noted that in all tables, a darker shade of a color reflects a higher score and bold scores highlight the best performances. It can be seen that the proposed model achieves better overall Se, Sp, and Ac measures on test data. Additionally, a significant improvement in baseline model performance is



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

Fig. 16. ROC plot and the AUC scores of Table I. (a) DRIVE, (b) STARE, (c) CHASEDB1, and (d) HRF.

TABLE I
COMPARISON OF THE PROPOSED METHOD WITH THE BASELINE MODELS

		Traditional			$\lambda CMgC^2$		
DB	**Methods**	**Se**	**Sp**	**Ac**	**Se**	**Sp**	**Ac**
DRIVE [42]	FC-DNet [41]	0.801	0.974	0.960	0.826	0.976	0.965
	U-Net [29]	0.790	0.972	0.959	0.812	0.976	0.962
	SOTA	**0.836**	0.986	0.970	**0.836**	0.986	0.970
	Our Work	0.825	**0.986**	**0.972**	0.832	**0.989**	**0.976**
STARE [43]	FC-DNet [41]	0.799	0.980	0.970	0.830	0.981	0.973
	U-Net [29]	0.774	0.978	0.965	0.839	0.979	0.970
	SOTA	**0.865**	0.987	0.977	0.865	0.987	0.977
	Our Work	0.859	**0.987**	**0.978**	**0.867**	**0.989**	**0.981**
CHASE-DB1 [44]	FC-DNet [41]	0.810	0.979	0.962	0.835	0.982	0.966
	U-Net [29]	0.799	0.982	0.965	0.841	0.985	0.972
	SOTA	0.848	0.989	0.977	0.848	0.989	0.977
	Our Work	**0.851**	**0.989**	**0.979**	**0.884**	**0.990**	**0.982**
HRF [45]	FC-DNet [41]	0.799	0.981	0.965	0.811	0.981	0.969
	U-Net [29]	0.784	0.979	0.964	0.815	0.980	0.968
	SOTA	**0.823**	0.987	0.972	0.823	0.987	0.972
	Our Work	0.822	**0.988**	**0.972**	**0.826**	**0.989**	**0.977**

also observed when trained upon λCMgC^2 augmentation. This empirically validates the importance of the λCMgC^2-based training paradigm. To assess overall performance, metric-wise best-recorded scores across the literature are selected separately. State-of-the-art (SOTA) in Table I highlights the same. It is noted that 2pCeP*d*-Net achieves an overall better or at par performance, cumulatively across literature, even when trained on TA. This highlights the standalone contribution of the 2pCeP*d*-Net model in this work. The ROC plot and the AUC score are presented in Fig. 16. To further investigate the performance of our proposed approach, a systematic comparative analysis with contemporary works is presented in Table II. To ensure a fair comparison, the results of contemporary works have been directly quoted from the literature. Tables II and III illustrate the achieved Se, Sp, Ac, $F1$, and AUC scores for the proposed model and establishes a comparative benchmark with respect to existing literature.

3) Comparative Analysis With SOTA: Table II highlights that the proposed model achieves a significantly higher Se (0.832) and at par IVUC (0.987) score compared to the other methods listed on the DRIVE dataset. Furthermore, the overall Ac (0.976), Sp (0.989), and $F1$ (0.857) scores achieved by 2pCeP*d*-Net is highest compared to those of the other methods. Furthermore, the proposed model outperforms the existing methods in terms of Ac (0.981), Sp (0.989), Se (0.867), AUC (0.993), and $F1$ (0.869) scores on the STARE dataset. It is also noted that the achieved Se (0.867) is far superior in comparison to other models. For the CHASEDB1 dataset, 2pCeP*d*-Net outperforms the other competing methods in terms of Sp (0.990), Se (0.884), Ac (0.982), $F1$ (0.871), and AUC (0.992) scores. Table III highlights the achieved performance on the HRF dataset. It is observed that the proposed model outperforms existing literature with an overall score of Se (0.8259), Sp (0.9896), Ac (0.9770), $F1$ (0.8468), and AUC (0.9869). To further investigate the performances, a series of assessments have been conducted on the proposed work, and the observations are reported next.

D. Ablation Study

To evaluate the utility of the modules designed for our proposed 2pCeP*d*-Net model, a series of comparative assessments have been conducted with respect to the DRIVE and STARE datasets and the obtained results are as follows.

1) Computational Complexity: An assessment of the computational complexity of the proposed 2pCeP*d*-Net is conducted and relevant observations are reported in Table IV. The complexity analysis is evaluated across four major metrics, namely, structural parameters, floating-point operations per second (FLOPS), memory required, and inference time needed for each test sample. The respective Ac is also presented in Table IV for ease of readability. It is observed that the proposed model has a greater number of structural parameters in comparison to baseline models (such as U-Net [29]). However, the number of parameters is slightly lesser than the recent GT-DLA-dsHFF [3] model while achieving a higher overall Ac. Additionally, the achieved overall performance is significantly higher than the TransUNet model [48] (Table II) despite having nearly 24% fewer parameters. A similar trend is also observed with respect to the FLOPS metric. The 2pCeP*d*-Net model achieves a better Ac while requiring only 50% of FLOPS in comparison to models having a comparable number of parameters such as GT-DLA-dsHFF. Furthermore, the proposed work has a far lesser memory requirement of a modest 105.51 MB and an inferential time of 0.1083 seconds in contrast to recent models such as GT-DLA-dsHFF and DPL-GTF-EFA [47]. Noticeably, the proposed model has a lesser floating-point operation in comparison to the DPL-GTF-EFA model despite having a larger set of parameters. This can be attributed to the design of an end-to-end convolutional model in our work which reduces the activation memory footprint [37] and enhances the speed of execution. It can thus be inferred that the 2pCeP*d*-Net model comprises a moderate complexity and better memory utilization while

TABLE II

PERFORMANCE COMPARISON ON THE DRIVE, STARE, AND CHASEDB1 DATASETS

Databases/ Methods	**DRIVE**					**STARE**					**CHASEDB1**				
	Se	**Sp**	**Ac**	**F1**	**AUC**	**Se**	**Sp**	**Ac**	**F1**	**AUC**	**Se**	**Sp**	**Ac**	**F1**	**AUC**
DN-Net [13]	0.781	0.980	0.953	–	–	–	–	–	–	–	–	–	–	–	–
JSL-PWL [16]	0.765	0.981	0.954	–	0.975	0.758	0.984	0.961	–	0.980	0.763	0.980	0.961	–	0.978
DU-Net [20]	0.796	0.980	0.956	0.823	0.980	0.759	0.987	0.964	0.814	0.983	0.815	0.975	0.961	0.788	0.980
DD-Net [21]	0.813	0.978	0.960	–	–	–	–	–	–	–	–	–	–	–	–
CE-Net [22]	0.820	0.975	0.954	0.824	0.983	0.829	0.984	0.966	0.836	0.987	0.802	0.981	0.965	0.805	0.986
IterNet [23]	0.779	0.983	0.954	0.820	0.984	0.771	0.988	0.970	0.814	0.987	0.797	0.982	0.965	0.807	0.986
NFN+ [14]	0.799	0.981	0.958	0.829	0.983	–	–	–	–	–	–	–	–	–	–
HA-Net [15]	0.799	0.981	0.958	0.829	0.982	0.818	0.984	0.967	0.838	0.988	0.824	0.981	0.967	0.819	0.987
MPS-Net [17]	0.836	0.974	0.956	0.827	0.979	0.856	0.982	0.968	0.846	0.987	0.848	0.979	0.966	0.830	0.986
SCS-Net [18]	0.828	0.983	0.969	–	0.983	0.820	0.983	0.973	–	0.984	0.836	0.984	0.974	–	0.986
Genetic U-Net [19]	0.830	0.975	0.957	0.831	0.982	0.865	0.984	0.972	0.863	0.992	0.846	0.984	0.966	0.822	0.988
SkelCon [24]	0.832	0.986	0.946	–	–	–	–	–	–	–	0.781	0.979	0.956	–	–
TransUNet [48]	0.826	0.983	0.969	0.823	0.983	0.833	0.986	0.975	0.832	0.987	0.822	0.986	0.975	0.809	0.987
GT-DLA-dsHFF [3]	0.835	0.982	0.970	–	0.986	0.848	0.986	0.976	–	0.990	0.844	0.985	0.976	–	0.989
Wave-Net [49]	0.816	0.976	0.956	–	–	0.790	0.983	0.964	–	–	–	–	–	–	–
DPF-Net [30]	0.828	0.977	0.957	0.830	–	0.828	0.985	0.965	0.836	–	0.830	0.984	0.967	0.830	–
DPL-GTF-EFA [47]	**0.836**	0.983	0.970	0.832	0.987	0.853	0.987	0.977	0.847	0.992	0.796	0.989	0.977	0.814	0.990
Our Work	0.832	**0.989**	**0.976**	**0.857**	**0.987**	**0.867**	**0.989**	**0.981**	**0.869**	**0.993**	**0.884**	**0.990**	**0.982**	**0.871**	**0.992**

TABLE III

PERFORMANCE COMPARISON ON THE HRF DATASET

Methods	**Se**	**Sp**	**Ac**	**F1**	**AUC**
JSL-PWL [16]	0.8087	0.9417	0.9437	–	–
DU-Net [20]	0.7464	0.9874	0.9651	–	0.9831
HA-Net [15]	0.7803	0.9843	0.9654	0.8074	0.9837
SCS-Net [18]	0.8114	0.9823	0.9687	–	0.9842
Genetic U-Net [19]	0.8220	0.9818	0.9667	0.8179	**0.9872**
TransUNet [48]	0.8116	0.9834	0.9702	0.8068	0.9841
GT-DLA-dsHFF [3]	0.8178	0.9823	0.9698	–	0.9853
DPL-GTF-EFA [47]	0.8225	0.9840	0.9716	0.8165	0.9867
Our Work	**0.8259**	**0.9896**	**0.9770**	**0.8468**	0.9869

TABLE IV

COMPARISON OF COMPUTATIONAL COMPLEXITY

Methods	**Params. (1x10^6)**	**FLOPS (1x10^9)**	**Mem. (MB)**	**Time (s)**	**Acc**
U-Net [29]	8.03	51.04	30.67	0.071	0.962
CE-Net [22]	13.40	124.48	110.77	0.076	0.954
Iter-Net [23]	13.60	194.65	52.05	0.107	0.954
TransUNet [48]	**36.46**	225.40	142.14	0.092	0.969
DPF-Net [30]	3.63	56.99	40.44	0.124	0.957
GT-DLA-dsHFF [3]	26.08	**473.92**	125.68	**0.445**	0.970
DPL-GTF-EFA [47]	16.86	256.09	**194.73**	0.153	0.970
Our Work	25.14	198.32	105.51	0.108	**0.976**

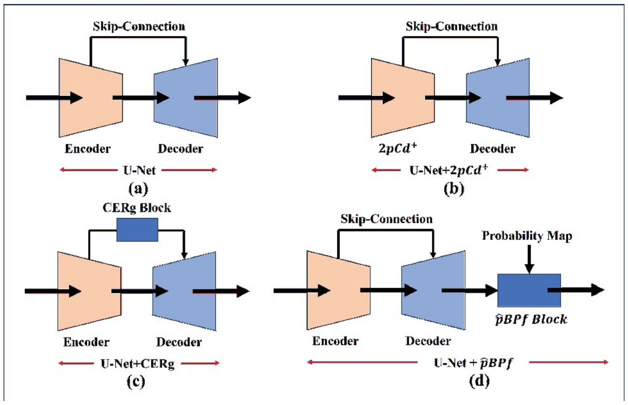


Fig. 17. Network block diagram used for the ablation study. (a) U-Net, (b) U-Net +2pCd$^+$, (c) U-Net + CERg, and (d) U-Net +$\hat{p}$BPf.

TABLE V

ABLATION STUDY OF THE NETWORK BLOCKS WITH U-NET. (↑ INDICATES THE INCREASE COMPARED TO THE BASELINE U-NET)

DB	**Methods**	**Se (↑x100)**	**Sp (↑x100)**	**Ac (↑x100)**
DRIVE	U-Net	0.812	0.976	0.962
	U-Net+$2pCd^+$	0.815 (↑ 0.3)	**0.981 (↑ 0.5)**	**0.967 (↑ 0.5)**
	U-Net+CERg	0.814 (↑ 0.2)	0.978 (↑ 0.2)	0.964 (↑ 0.2)
	U-Net+(w/o)	0.817 (↑ 0.5)	0.978 (↑ 0.2)	0.963 (↑ 0.1)
	U-Net+$\hat{p}BPf$	**0.821 (↑ 0.9)**	0.978 (↑ 0.2)	0.965 (↑ 0.3)
STARE	U-Net	0.839	0.979	0.970
	U-Net+$2pCd^+$	0.842 (↑ 0.3)	0.981 (↑ 0.2)	0.972 (↑ 0.2)
	U-Net+CERg	0.841 (↑ 0.2)	0.979 (– 0.0)	0.970 (– 0.0)
	U-Net+(w/o)	0.843 (↑ 0.4)	0.981 (↑ 0.2)	0.972 (↑ 0.2)
	U-Net+$\hat{p}BPf$	**0.849 (↑ 1.0)**	**0.983 (↑ 0.4)**	**0.974 (↑ 0.4)**

achieving an improved performance compared to other SOTA approaches. The moderate quantity of parameters coupled with low memory and inferential time establishes the practicality of the approach.

2) Structural Block-Wise Performance: We analyze the individual performances of the proposed 2pCd$^+$ encoder, CERg block, and $\hat{p}$BPf fusion block, separately, by considering a U-Net architecture [29] with level $l = 5$ as a baseline for the ablation study [as shown in Fig. 17(a)].

a) Efficacy with 2pCd$^+$ Block: To test the capabilities of the **Enc** block, we consider a simple U-Net model and replace the encoder arm with our proposed 2pCd$^+$ encoder [as shown in Fig. 17(b)]. Table V highlights the statistical results obtained for both DRIVE and STARE datasets, with and without the proposed block. The obtained results for the U-Net +2pCd$^+$ structure demonstrates an overall increase in the performance across different metrics, viz. a mean increase of 0.35% in Ac, 0.35% in Sp, and 0.3% in Se in the presence of the 2pCd$^+$ block. To gauge the feature representation, the 2pCd$^+$ block output at $l = 5$ (denoted by f^{enc}) is extracted. A U-Net-inspired autoencoder is subsequently trained on the annotated vessel images and the bottleneck feature output (f^{auto}) is extracted. f^{auto} is selected for this comparison as it comprises a compact feature representation on vascular structure [50]. Finally, the correlation measure of mutual information ($\Im(.)$) is selected for feature comparison in this work. A correlation measure of $\Im(f^{\text{auto}}, f^{\text{enc}}) = 0.121$ is achieved for the 2pCd$^+$ block. In contrast, the bottleneck layer of the baseline U-Net (f^{unet}) achieved a score of $\Im(f^{\text{auto}}, f^{\text{unet}}) = 0.061$, respectively. This highlights that f^{enc} encodes better vascular structure information in encoded space, in alignment with the annotated vessel map.

b) Efficacy with the CERg Block: Similar to the above study, the capabilities of the CERg block are tested by designing a U-Net architecture with our proposed CERg skip-connection block [as shown in Fig. 17(c)]. Table V highlights the obtained results. Only a marginal improvement is noticeable in the model performance. To further investigate the strength of the CERg block in noise filtration, Gaussian noise [13] is added to the test dataset, and the respective performances are noted. It is observed that the baseline U-Net model achieves Se, Sp, and Ac mean scores of 0.681, 0.919, and 0.905, respectively, on the Gaussian noise-induced data. In contrast, *U-Net + CERg* model achieves a superior 0.741, 0.947, and 0.934 scores, respectively, in the same task. Thus, the introduction of the CERg block enhanced the noise tolerance efficacy in the model, as was the objective.

c) Efficacy with the $\hat{p}$BPf Fusion Block: To test the $\hat{p}$BPf fusion block, we have replaced the last output layer of a U-Net architecture [as shown in Fig. 17(d)] and recorded the impact of the proposed block. Two sets of tests are reported in Table V, viz., 1) without the bandpass gate (marked *U-Net + w/o*) and 2) with both modules (marked *U-Net* $+\hat{p}BPf$). As can be observed in Table V, our proposed fusion block enhances the capabilities of a standard U-Net model with an overall mean increase of 0.35% in Ac, 0.3% in Sp, and 0.95% in Se, respectively. It is noted that the $\hat{p}$BPf fusion block has the highest overall impact on STARE dataset performance. This can be attributed to the use of a secondary objective function (MSE loss) which leads to improvement. However, the presence of lower contrast regions in DRIVE enforces additional complexity on MoE detection [13]. A well-informed feature space is thus necessary to tackle this challenge. Hence, 2pCd$^+$ achieves a better performance on DRIVE. Furthermore, *w/o* set achieves a slightly diminished performance in comparison. This can be attributed to poor performance on noise filtration, introduced by the MoE objective [39], without the bandpass gate. The complete $\hat{p}$BPf layer, however, does not suffer from such issues and achieves a better performance. Thus, it is noted that all three modules have individual roles in the network and are co-dependent to achieve the overall performance.

3) Segmentation in Complex Regions: To evaluate the robustness of our model on diseased images, 2pCeP*d*-Net performance is evaluated on seven diseased images from the DRIVE dataset and ten diseased images from the STARE dataset. Table VI presents the comparative analysis of the achieved results. It is observed that the proposed model achieves a better overall performance on the diseased data compared to other contemporary models. In comparison to 2pCeP*d*-Net performance in Table II, the proposed model experiences a modest drop of 0.3% in Ac across both datasets on diseased images. Notably, a marginal drop of only 0.1% in comparison to the literature is noted in the Sp on DRIVE dataset. However, the Se attained is far superior in contrast to existing literature. This establishes the performance attained by our proposed work in the diseased dataset. Fig. 18 visually represents the model performance on diseased images.

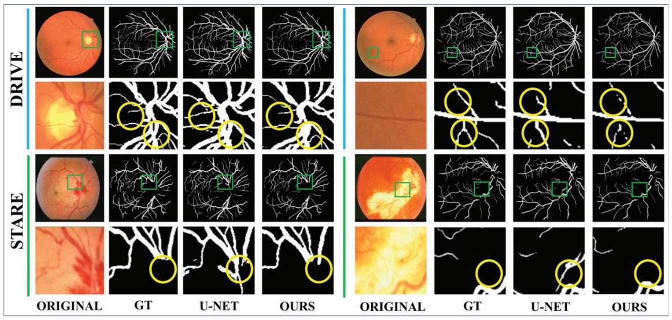


Fig. 18. Qualitative analysis of 2pCeP*d*-Net segmentation results in diseased images. Odd rows display the images with critical regions marked in green. Even rows display the magnified view of the marked green regions. Yellow markers highlight the areas of major challenges.

TABLE VI

COMPARATIVE ASSESSMENT ON DISEASED IMAGES

Database	Method	Se	Sp	Ac
DRIVE	U-Net [29]	0.7546	0.9790	0.9573
	CE-Net [22]	0.8149	0.9794	0.9650
	Iter-Net [23]	0.8234	0.9818	0.9679
	GT-DLA-dsHFF [3]	0.8259	**0.9821**	0.9684
	Our Work	**0.8309**	0.9811	**0.9692**
STARE	U-Net [29]	0.7783	0.9779	0.9631
	CE-Net [22]	0.7945	0.9829	0.9689
	Iter-Net [23]	0.8087	0.9791	0.9665
	GT-DLA-dsHFF [3]	0.8328	0.9835	0.9724
	Our Work	**0.8451**	**0.9847**	**0.9749**

TABLE VII

COMPARATIVE ASSESSMENT OF CROSS-VALIDATION RESULTS

Train	Test	Model	Se	Sp	Acc
DRIVE	STARE	DPF-Net [30]	0.749	0.983	0.953
		DPL-GTF-EFA [47]	**0.825**	0.984	0.972
		Our Work	0.822	**0.988**	**0.977**
STARE	DRIVE	DPF-Net [30]	0.746	0.984	0.948
		DPL-GTF-EFA [47]	0.810	0.982	0.967
		Our Work	**0.819**	**0.984**	**0.970**

4) Cross-Validation: To evaluate the generalization of our proposed 2pCeP*d*-Net, a cross-validation experiment was performed on the DRIVE and STARE datasets, respectively. Table VII highlights the comparative assessment of the statistical results. It is noted from Table VII that our proposed work achieves a better overall performance compared to the literature. The model trained on DRIVE and tested on the STARE dataset achieves an overall increase of 0.4% and 0.5% with respect to current benchmark approaches across Sp and Ac, respectively. A drop of 0.3% in Se is noted which can probably be attributed to a better-generalized vessel feature learned from the DRIVE dataset in the DPL-GTF-EFA model [47]. Nonetheless, in our proposed approach, better overall performance is achieved when trained on STARE and tested

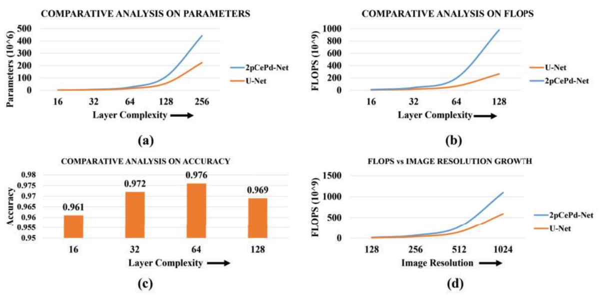


Fig. 19. Comparative assessment of limitations. (a) and (b) Scalability analysis of computational complexity. (c) and (d) Speed versus Ac analysis.

on the DRIVE dataset with an increase of 0.9%, 0.2%, and 0.3%, respectively, across the three parameters of evaluation. This indicates the robustness and generalization potential of our proposed work.

E. Limitations and Future Scope

Our proposed 2pCeP*d*-Net with λCMgC^2 augmentation is shown to improve upon existing works in literature. However, one point of the study is the scalability. As shown in Fig. 19(a), the number of parameters rises in cubic series if the number of kernels is increased in each layer. In Fig. 19(a), the x-axis corresponds to the number of kernels in layer $l = 1$ and denotes an equivalent increase across all model layers overall. A similar trend is, however, also observed in a baseline U-Net model [as shown in Fig. 19(a)]. It is thus noted that scalability is an architectural challenge inherent to encoder–decoder models. Simultaneously, the choice of speed against Ac is an additional challenge in the DL paradigm. In this work, the number of floating-point operations increases with increasing layer complexity, in alignment with the DL paradigm [shown in Fig. 19(b)]. However, in contrast, Ac is observed to increase with layer complexity and reach a peak at 64 [Fig. 19(c)]. The subsequent increase in complexity leads to overfitting which diminishes the achieved Ac. As diagnostic Ac is crucial in medical image segmentation, providence has been given to Ac over speed in this work and layer complexity of 64 is chosen. It is also noted, from Fig. 19(d), that an increase in image resolution leads to an exponential demand for FLOPS. This limits the current model to consider image resolutions up to 512 × 512 dimension for inference in reasonable computing time. To reduce the complexity, more sophisticated operations can be further investigated in the future to minimize the parameters gap noted in Fig. 19(a), without sacrificing Ac. Additionally, future studies in this work can further explore the λCMgC^2 data space by introducing more variations to achieve better performance on challenging regions in RFIs. Finally, the noise tolerance in the feature space can be studied further in future work for better MoE detection, especially in diseased images.

IV. CONCLUSION

In this work, a λCMgC^2 data augmentation model has been proposed to alleviate data scarcity in RV segmentation tasks from fundus images. The learning potentials of the augmented dataset are empirically established in this work with a comparative analysis of performance. Subsequently, a proposed 2pCeP*d*-Net model, comprising of a 2pCd$^+$ encoder, a CERg skip connection, and a $\hat{p}$BPf fusion block, is trained on the λCMgC^2 dataset to predict the probability-based vessel segmentation. The 2pCd$^+$ block uses a cross-dictionary-based attention mechanism to attain a rich encoded feature space as output. The encoded feature space is passed as an input to the CERg Skip Connection and Decoder arm to generate a low-noise feature, conducive to segmentation. Finally, a $\hat{p}$BPf fusion block is used to integrate the probability-map-based MoE region information to generate the segmented output of our proposed model. The individual advantages of the proposed modules are statistically evaluated and reported in this work. 2pCd$^+$ Encoder has been noted to achieve a better encoding in comparison to standard baseline models in the literature. Furthermore, the noise filtration capabilities of the CERg and $\hat{p}$BPf modules are also established in this work. Cumulatively, the proposed 2pCeP*d*-Net is empirically shown to achieve improved performance across literature on four public benchmark datasets. Finally, the limitations of the proposed work are also highlighted to provide scope for future improvements on the designed model.

APPENDIX
PROOF OF THE λCMGC^2 VESSEL PERCEPTION

Proof: We define $\mathcal{I}_{\text{vr}}^{\text{aug}}$ as a vessel removed augmented image. Let us define neighborhood intensity difference set $N_8^-(p_{x,y,v}) = \{\tilde{v} - v \mid \forall p_{\tilde{x},\tilde{y},\tilde{v}} \in N_8(p_{x,y,v})\}$ for any pixel $p_{x,y,v}$, where N_8 is the 8-neighborhood of a pixel. Thus, from this formulation, it is trivial to state that for any pixel $p_{x,y,v} \in (\mathcal{I}^{\text{aug}} - \mathcal{I}_{\text{vr}}^{\text{aug}})$.

$(N_8^-(p_{x,y,v}) \subset \mathcal{O}) \implies (\nexists q_{\tilde{x},\tilde{y},0} \in \mathcal{I}_{\text{gt}}^{\text{aug}})$, where $\tilde{x} = x \pm 1, \tilde{y} = y \pm 1$ and $\mathcal{O}$ represents set of infinite zero elements.

Let us consider a hypothetical pixel $p_{x,y,v} \in (\mathcal{I}^{\text{aug}} - \mathcal{I}_{\text{vr}}^{\text{aug}})$ such that $N_8^-(p_{x,y,v}) \subset \mathcal{O}$ and $\exists q_{\tilde{x},\tilde{y},0} \in \mathcal{I}_{\text{gt}}^{\text{aug}}$. That implies $\exists p_{\tilde{x},\tilde{y},\tilde{v}} \in N_8(p_{x,y,v})$ such that $\tilde{v} = v$. We consider $\tilde{v}^{\{l|l=m,n\}}, v^{\{l|l=m,n\}}$ as pixel values at $(\tilde{x}, \tilde{y})$ and (x, y) in $\mathcal{I}_g^{\{l|l=m,n\}}$, respectively. It follows thus from the formulation of the associated injective function $g(.)$, $[\lambda.\tilde{v}^m + (1 - \lambda).\tilde{v}^n] - [\lambda.v^m + (1 - \lambda).v^n] = 0$. Rearranging, we obtain $\lambda.[\tilde{v}^m - v^m] + (1 - \lambda).[\tilde{v}^n - v^n] = 0$. As the theorem is applicable for all $0 < \lambda < 1$, we may conclude that $\tilde{v}^m = v^m$ and $\tilde{v}^n = v^n$, respectively, thus implicating that the respective background pixel and vessel pixel in images $\mathcal{I}_g^m, \mathcal{I}_g^n$ are imperceptible. This is a fallacy, as otherwise, the original annotations would not have been possible due to imperceptibility. Hence, all vessel pixels perceptible in $\mathcal{I}_g^m, \mathcal{I}_g^n$ remain perceptible in $\mathcal{I}^{\text{aug}}$ and the theorem is proved by the law of contradiction.

ACKNOWLEDGMENT

The authors would like to thank the CMATER Lab at Jadavpur University for the requisite support provided in pursuit of this work.

REFERENCES

[1] M. M. Fraz et al., “Blood vessel segmentation methodologies in retinal imags—A survey,” *Comput. Methods Programs Biomed.*, vol. 108, no. 1, pp. 407–433, 2012.

[2] J. V. B. Soares, J. J. G. 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, Sep. 2006.

[3] Y. Li et al., "Global transformer and dual local attention network via deep-shallow hierarchical feature fusion for retinal vessel segmentation," *IEEE Trans. Cybern.*, vol. 53, no. 9, pp. 5826–5839, Sep. 2023.

[4] C. L. Srinidhi, P. Aparna, and J. Rajan, "Recent advancements in retinal vessel segmentation," *J. Med. Syst.*, vol. 41, no. 4, pp. 1–22, Apr. 2017.

[5] 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, Jul. 2001.

[6] 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, Jan. 2003.

[7] U. T. V. Nguyen, A. Bhuiyan, L. A. F. 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, Mar. 2013.

[8] B. Sheng et al., "Retinal vessel segmentation using minimum spanning superpixel tree detector," *IEEE Trans. Cybern.*, vol. 49, no. 7, pp. 2707–2719, Jul. 2019.

[9] C. Li, C. Xu, C. Gui, and M. D. Fox, "Distance regularized level set evolution and its application to image segmentation," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3243–3254, Dec. 2010.

[10] G. Akbarizadeh, "A new statistical-based kurtosis wavelet energy feature for texture recognition of SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4358–4368, Nov. 2012.

[11] G. Akbarizadeh, G. A. Rezai-Rad, and S. B. Shokouhi, "A new region-based active contour model with skewness wavelet energy for segmentation of SAR images," *IEICE Trans. Inf. Syst.*, vol. E93.D, no. 7, pp. 1690–1699, 2010.

[12] P. K. Karn, B. Biswal, and S. R. Samantaray, "Robust retinal blood vessel segmentation using hybrid active contour model," *IET Image Process.*, vol. 13, no. 3, pp. 440–450, Feb. 2019.

[13] P. Liskowski and K. Krawiec, "Segmenting retinal blood vessels with deep neural networks," *IEEE Trans. Med. Imag.*, vol. 35, no. 11, pp. 2369–2380, Nov. 2016.

[14] Y. Wu, Y. Xia, Y. Song, Y. Zhang, and W. Cai, "NFN+: A novel network followed network for retinal vessel segmentation," *Neural Netw.*, vol. 126, pp. 153–162, Jun. 2020.

[15] D. Wang, A. Haytham, J. Pottenburgh, O. Saeedi, and Y. Tao, "Hard attention net for automatic retinal vessel segmentation," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 12, pp. 3384–3396, Dec. 2020.

[16] Z. Yan, X. Yang, and K.-T. 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, Sep. 2018.

[17] Z. Lin et al., "A high resolution representation network with multi-path scale for retinal vessel segmentation," *Comput. Methods Programs Biomed.*, vol. 208, Sep. 2021, Art. no. 106206.

[18] H. Wu, W. Wang, J. Zhong, B. Lei, Z. Wen, and J. Qin, "SCS-Net: A scale and context sensitive network for retinal vessel segmentation," *Med. Image Anal.*, vol. 70, May 2021, Art. no. 102025.

[19] J. Wei et al., "Genetic U-Net: Automatically designed deep networks for retinal vessel segmentation using a genetic algorithm," *IEEE Trans. Med. Imag.*, vol. 41, no. 2, pp. 292–307, Feb. 2022.

[20] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "DUNet: A deformable network for retinal vessel segmentation," *Knowl.-Based Syst.*, vol. 178, pp. 149–162, Aug. 2019.

[21] L. Mou, L. Chen, J. Cheng, Z. Gu, Y. Zhao, and J. Liu, "Dense dilated network with probability regularized walk for vessel detection," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1392–1403, May 2020.

[22] Z. Gu et al., "CE-Net: Context encoder network for 2D medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019.

[23] L. Li, M. Verma, Y. Nakashima, H. Nagahara, and R. Kawasaki, "IterNet: Retinal image segmentation utilizing structural redundancy in vessel networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 3645–3654.

[24] Y. Tan, K.-F. Yang, S.-X. Zhao, and Y.-J. Li, "Retinal vessel segmentation with skeletal prior and contrastive loss," *IEEE Trans. Med. Imag.*, vol. 41, no. 9, pp. 2238–2251, Sep. 2022.

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

[26] H. Zhao, H. Li, S. Maurer-Stroh, Y. Guo, Q. Deng, and L. Cheng, "Supervised segmentation of un-annotated retinal fundus images by synthesis," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 46–56, Jul. 2018.

[27] X. Sun et al., "Robust retinal vessel segmentation from a data augmentation perspective," in *Proc. 8th Int. Workshop OMIA*, vol. 12970. Cham, Switzerland: Springer, Jan. 2021, pp. 189–198.

[28] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "Mixup: Beyond empirical risk minimization," 2017, *arXiv:1710.09412*.

[29] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234–241.

[30] J. Li, G. Gao, L. Yang, G. Bian, and Y. Liu, "DPF-Net: A dual-path progressive fusion network for retinal vessel segmentation," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–17, 2023.

[31] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," 2014, *arXiv:1412.6806*.

[32] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30. Red Hook, NY, USA: Curran Associates, Dec. 2017, pp. 6000–6010.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[34] J. Tang, B. Zou, C. Li, S. Feng, and H. Peng, "Plane-wave image reconstruction via generative adversarial network and attention mechanism," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–15, 2021.

[35] Y. Chen, C. Xu, W. Ding, S. Sun, X. Yue, and H. Fujita, "Target-aware U-Net with fuzzy skip connections for refined pancreas segmentation," *Appl. Soft Comput.*, vol. 131, Dec. 2022, Art. no. 109818.

[36] S. Das, D. Jyotishi, and S. Dandapat, "Automated detection of heart valve diseases using stationary wavelet transform and attention-based hierarchical LSTM network," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–10, 2023.

[37] N. C. Thompson, K. Greenewald, K. Lee, and G. F. Manso, "The computational limits of deep learning," 2020, *arXiv:2007.05558*.

[38] R. C. Gonzalez, *Digital Image Processing*. London, U.K.: Pearson, 2009.

[39] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1395–1403.

[40] S. Das, D. Jyotishi, and S. Dandapat, "Heart valve diseases detection based on feature-fusion and hierarchical LSTM network," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.

[41] S. Jegou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional DenseNets for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 11–19.

[42] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. 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, Apr. 2004.

[43] A. D. 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, Mar. 2000.

[44] C. G. Owen et al., "Measuring retinal vessel tortuosity in 10-year-old children: Validation of the computer-assisted image analysis of the retina (CAIAR) program," *Investigative Ophthalmol. Vis. Sci.*, vol. 50, no. 5, pp. 2004–2010, 2009.

[45] T. Köhler, A. Budai, M. F. Kraus, J. Odstr{č}ili, G. Michelson, and J. Hornegger, "o}hler, A. Budai, M. F. Kraus, J. Odstr{č}ili, G. Michelson, and J. Hornegger," in *Proc. 26th IEEE Int. Symp. Comput.-Based Med. Syst.*, Jun. 2013, pp. 95–100.

[46] A. Borji, "Pros and cons of GAN evaluation measures," *Comput. Vis. Image Understand.*, vol. 179, pp. 41–65, Feb. 2019.

[47] Y. Li et al., "Diffusion probabilistic learning with gate-fusion transformer and edge-frequency attention for retinal vessel segmentation," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–13, 2024.

[48] J. Chen et al., "TransUNet: Transformers make strong encoders for medical image segmentation," 2021, *arXiv:2102.04306*.

[49] Y. Liu, J. Shen, L. Yang, H. Yu, and G. Bian, "Wave-Net: A lightweight deep network for retinal vessel segmentation from fundus images," *Comput. Biol. Med.*, vol. 152, Jan. 2023, Art. no. 106341.

[50] A. Ashraf, N. M. Nawi, T. Shahzad, M. Aamir, M. A. Khan, and K. Ouahada, "Dimension reduction using dual-featured auto-encoder for the histological classification of human lungs tissues," *IEEE Access*, vol. 12, pp. 104165–104176, 2024.