

DEVELOPMENT OF ALGORITHMS FOR AUTOMATIC COMPUTATION OF ARTERIOLAR-VENULAR RATIO IN RETINAL FUNDUS IMAGES

Thesis submitted by

Sandip Sadhukhan

Doctor of Philosophy (Engineering)

Electrical Engineering Department
Faculty Council of Engineering & Technology
Jadavpur University
Kolkata, India

2024

1. Title of the Thesis:

DEVELOPMENT OF ALGORITHMS FOR AUTOMATIC
COMPUTATION OF ARTERIOLAR-VENULAR RATIO IN
RETINAL FUNDUS IMAGES

2. Name, Designation and Institution of the Supervisors:

(a) **Dr. Gautam Sarkar**

Professor

Department of Electrical Engineering

Jadavpur University, Kolkata– 700032, India

(b) **Dr. Ashis Kumar Dhara**

Assistant Professor

Department of Electrical Engineering

National Institute of Technology, Durgapur– 713209, India

3. List of Publications:

Journal Published

- **Sandip Sadhukhan**, Arpita Sarkar, Debprasad Sinha, Goutam Kumar Ghorai, Gautam Sarkar and Ashis Kumar Dhara “Attention Based Fully Convolutional Neural Network for Simultaneous Detection and Segmentation of Optic Disc in Retinal Fundus Images” International Journal of Medical and Health Science Vol. 14 No: 8, pp. 200-204, 2020.
- **Sandip Sadhukhan**, Goutam Kumar Ghorai, Debprasad Sinha, Souvik Maiti, Gautam Sarkar and Ashis Kumar Dhara “Segmentation of optic disc in retinal images using fully convolutional network” Journal of Current Indian Eye Research Vol. 6, Issue 2, December 2019, pp. 40-47.

Journal Communicated

- **Sandip Sadhukhan**, Debprasad Sinha, Ashis Kumar Dhara, Gautam Sarkar “Attention AV-Net for Artery and vein discrimination in Retinal Fundus Images”, Signal, Image and Video Processing Journal, Springer, 2023.

National/International Conference

- **Sandip Sadhukhan**, Goutam Kumar Ghorai, Souvik Maiti, Gautam Sarkar and Ashis Kumar Dhara, “Optic Disc Localization in Retinal Fundus Images using Faster R-CNN” Proceedings of Fifth International Conference on Emerging Applications of Information Technology (EAIT 2018), January 12-13, Kolkata, India.
- **Sandip Sadhukhan**, Goutam Kumar Ghorai, Souvik Maiti, Vikrant Anil-rao Karale, Gautam Sarkar and Ashis Kumar Dhara, “Optic Disc Segmentation in Retinal Fundus Images Using Fully Convolutional Network and Removal of False-Positives Based on Shape Features” Proceedings of 4th International Workshop, DLMIA 2018 and 8th International Workshop, ML-CDS 2018, September 20, 2018, Granada, Spain.
- **Sandip Sadhukhan**, G K Ghorai, S Maiti, D Sinha, G Sarkar and A K Dhara “Fully Convolutional Network for Segmentation of Optic Disc In Retinal Fundus Images” Proceedings of IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019) April 8-11, 2019, Venice, Italy.

Conference Communicated

- **Sandip Sadhukhan**, Ashis Kumar Dhara, and Gautam Sarkar, “Automatic Measurement of Arteriolar-to-Venular width Ratio (AVR) in Retinal Fundus Images” Proceedings of IEEE Calcutta Conference (CALCON), 2024, Kolkata, India.

4. List of Patents: Nil

Statement of Originality

I, Sandip Sadhukhan, registered on 24th January, 2017, do hereby declare that this thesis entitled "DEVELOPMENT OF ALGORITHMS FOR AUTOMATIC COMPUTATION OF ARTERIOLAR-VENULAR RATIO IN RETINAL FUNDUS IMAGES", contains literature survey and 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 8%.

Signature of Candidate

Sandip Sadhukhan

Sandip Sadhukhan

Date: 14.02.2024

Certified by the Supervisors: (Signature with date and seal)

Gautam Sarkar
14/02/2024

Dr. Gautam Sarkar
(Professor)
Electrical Engineering Department
Jadavpur University
Kolkata-700032, WB, India

Prof. Gautam Sarkar
Electrical Engineering Department
Jadavpur University
Kolkata, INDIA

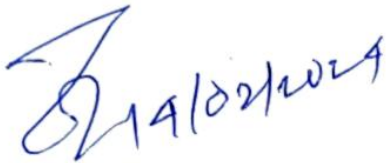
Prof. Gautam Sarkar
Electrical Engineering Department
Jadavpur University
Kolkata, INDIA

Ashis Kumar Dhara
14/02/2024

Dr. Ashis Kumar Dhara
(Assistant Professor)
Electrical Engineering Department
National Institute of Technology
Durgapur-713209, WB, India

CERTIFICATE OF APPROVAL

This is to certify that the thesis entitled **Development of Algorithms for Automatic Computation of Arteriolar-Venular Ratio in Retinal Fundus Images**, submitted by **Sandip Sadhukhan**, who got his name registered on 24.01.2017, for the award of PhD (Engineering) degree of Jadavpur University is absolutely based upon his own work under the supervision of Dr. Gautam Sarkar and Dr. Ashish Kumar Dhara and that neither his thesis nor any part of the thesis has been submitted for any degree or diploma or any other academic award anywhere before.



Dr. Gautam Sarkar
(Professor)
Electrical Engineering Department
Jadavpur University
Kolkata-700032, WB, India

Prof. Gautam Sarkar
Electrical Engineering Department
Jadavpur University
Kolkata, INDIA



Dr. Ashis Kumar Dhara
(Assistant Professor)
Electrical Engineering Department
National Institute of Technology
DURGAPUR - 713209
Durgapur-713209, WB, India

Dedicated to my parents, wife and daughter

Acknowledgment

I extend my deepest gratitude to the numerous individuals and institutions that have supported and guided me throughout my doctoral journey. Completing this PhD thesis would not have been possible without the unwavering support of my family, friends, and mentors.

I am profoundly thankful to my supervisors Professor Gautam Sarkar and Dr. Ashis Kumar Dhara whose guidance, encouragement, and expertise have been invaluable. Your informative input and constructive criticism have had a big impact on the course of this research, encouraging me to pursue new ideas and refine my thesis.

My appreciation also goes to the Electrical Engineering Department, Jadavpur University for providing a conducive research environment and resources necessary for the successful completion of this project.

I am indebted to my colleagues and fellow researchers Debprasad Sinha, Souvik Maiti, and Debasis Maji who have shared their knowledge and experiences contributing to the development of my ideas. I am thankful to Dr. Biswapati Jana of Computer Science Department, Vidyasagar University for his valuable guidance. I also like extend my thanks to Professor Amitava Chatterjee of Electrical Engineering Department, Jadavpur University and Professor Jitendra Nath Bera for their valuable suggestion and comments during different review sessions.

Lastly, heartfelt thanks to my friends and family for their constant support, understanding

Thanking You,

Sandip Sadhukhan

Sandip Sadhukhan 14.02.2024

Abstract

This thesis aims to develop the algorithm for Automatic Computation of Arteriolar-venular Ratio in Retinal Fundus Images. Arteriolar-venular ratio (AVR) is correlated with many diseases like coronary heart disease, stroke, atherosclerosis etc. A high AVR has been associated with higher cholesterol levels, high blood pressure diseases of the pancreas etc. A low AVR is caused by abnormal widening of the veins which results chronic high blood sugar levels. Globally around 314 million individuals suffer from glaucoma, hypertensive retinopathy, and diabetic retinopathy. These conditions gradually contribute to vision loss, which is a significant concern in developing nations. Timely detection of these vascular changes may allow preventive treatments to reduce the risk of vision loss. These data can be used in telemedicine and remote diagnostic areas. Manually classifying arteries and veins requires expertise in retinal image interpretation and is a very labour-intensive process. Different tasks associated with automatic computation of AVR in a fundus image. They are - a) Automated detection of Region of Interest (ROI) from where the vessels of interest will be taken into consideration, b) Artery-vein patch extraction based on vessel central-line, c) Classification of vessels into artery and vein and finally d) Computation of AVR. Below section gives a brief description of each task and approach we have followed:

Automated detection of Region of Interest (ROI) - For this accurate detection and localization of Optic Disc (OD) in the fundus image is the major task. We utilized a fully convolutional network based on the U-Net architecture to segment the optic disc (OD). The U-Net design proves highly effective in image segmentation, especially in scenarios with limited availability of input images. After successful segmentation of OD and removal of false positives, based on the diameter of the OD, the region of interest is identified. It is an annular region that falls under two concentric circles of radius 1 and 1.5 times the diameter of the optic disc.

Artery-vein patch extraction based on vessel central-line - The next step is extraction of artery vein mask from the fundus image dataset. In this case, we have used INSPIRE-AVR dataset for which artery-vein vessels mask ground truth is available. The centre-line coordinates were then determined using these masks. Subsequently, separate patches for arteries and veins were generated, each measuring 32×32 , based on the received artery and vein centre-line coordinates.

Classification of vessels into artery and vein and finally - A robust automatic method for classifying arteries and veins (A/V) is necessary to automate large-scale retinal image processing. We have addressed this classification task by employing residual attention-based deep neural networks. The attention mechanism reduces the impact of irrelevant features while selectively amplifying important ones, allowing the network to improve features at different levels. The Deep Convolutional Neural Network's overall performance has been enhanced through the use of both residual and attention learning processes.

Computation of AVR - The AVR value is calculated from the calibers of the vessels inside the region of interest (RoI). For this purpose, six largest arteries and six largest veins are considered. The arteriolar to venular ratio is defined as the ratio of CRAE and CRVE where CRAE is the Central Retinal Artery Equivalent and CRVE is the Central Retinal Vein Equivalent. We have applied our method for AVR computation on the publicly available INSPIRE-AVR dataset. We have reported the mean error and the correlation coefficient which is comparable to the reference AVR values.

Keywords: Retinal fundus image, Optic disc detection and segmentation, Fully convolutional neural network, Convolution neural network, Faster R-CNN, Computer-aided diagnosis of diabetic retinopathy, Ocular diseases, Arteries and veins discrimination, Attention mechanism, cardiovascular disorders.

Contents

Abstract	x
List of Abbreviations	xvi
List of Figures	xix
List of tables	xix
1 Introduction	1
1.1 Deep learning for diagnosis of hypertensive retinopathy	3
1.1.1 Detection	4
1.1.2 Segmentation	5
1.1.3 Classification	5
1.2 Eye Anatomy	6
1.3 Diagnosis of Hypertensive retinopathy	7
1.4 Literature review	8
1.4.1 Review on Optic Disc localization and segmentation	9
1.4.2 Review on Artery-vein classification	10
1.4.3 Review on Computation of Arteriolar-to-venular (AVR) width ratio	12
1.4.4 Overall Review	12
1.5 Objective	14
1.6 Outline of the Thesis	15
1.6.1 Chapter 1: Introduction	15
1.6.2 Chapter 2: Localization of optic disc	16
1.6.3 Chapter 3: Segmentation of optic disc	16
1.6.4 Chapter 4: Classification of vessels into artery and vein	16

CONTENTS

1.6.5	Chapter 5: Computation of AVR	17
1.6.6	Chapter 6: Conclusion	17
2	Optic Disc localization in fundus image using Faster R-CNN	19
2.1	Introduction	20
2.2	Reported Works on Optic Disc localization in fundus image	21
2.3	Methodology	22
2.3.1	Using Convolution neural networks for object localization	22
2.3.2	Proposed localization framework	23
2.4	Results and Discussion	24
2.5	Conclusion	26
3	Optic disc segmentation in fundus image Using Attention Based Fully Convolutional Network	27
3.1	Introduction	28
3.2	Methodology	29
3.2.1	Preprocessing	29
3.2.2	Segmentation using fully convolutional network	30
3.3	Results and Discussions	31
3.4	Conclusion	34
4	Artery-vein classification in fundus image using Attention based CNN	35
4.1	Introduction	36
4.2	Reported Works on Artery-vein classification in fundus image	37
4.3	Materials and method	39
4.3.1	Materials	39
4.3.2	Method	41
4.3.2.1	Determination of ROI	41
4.3.2.2	A/V Patch Extraction	42
4.3.3	A/V Discrimination	43
4.3.3.1	Attention Mechanism	43
4.3.3.2	Network Architecture	44
4.3.3.3	Training of Network	45
4.4	Results and discussion	46
4.5	Conclusion	48

CONTENTS

5	Computation of Arteriolar-to-venular (AVR) width ratio	51
5.1	Introduction	52
5.2	Reported Works on Arteriolar-to-venular (AVR) width ratio computation	53
5.3	Materials and Method	54
5.3.1	Materials	54
5.3.2	Method	54
5.3.2.1	Optic disc localization in fundus image	54
5.3.2.2	Optic disc segmentation in fundus image	55
5.3.2.3	Determination of RoI and AV Patch creation	56
5.3.2.4	Artery-Vein classification	56
5.3.2.5	AVR computaion	57
5.4	Results and Discussion	57
5.5	Conclusion	58
6	Conclusion	59
6.1	Summary of Studies	59
6.2	Contributions of the Thesis	61
6.3	Future Scope	62
A	Appendix	63
A.1	Deep Learning Fundamentals	63
A.1.1	Hardware	63
A.1.2	Datasets and benchmarks	64
A.1.3	Algorithmic advances	64
A.2	Building blocks of Deep learning network	65
A.2.1	Neural network	65
A.2.1.1	Weights and Bias	65
A.2.1.2	Activation Functions	66
A.2.2	Loss Function	67
A.2.2.1	Optimizer	69
A.2.2.2	Backpropagation	71
A.2.2.3	Batch Normalization	71
A.2.2.4	Regularization	72
A.2.2.5	Dropout	73
A.2.2.6	Pooling Layers	74
A.2.2.7	Convolutional Layers	75

CONTENTS

A.2.2.8 Fully-Connected Layer	76
A.2.3 Deep learning for computer vision	76
References	79

List of Abbreviations

ACM	Active contour model
2D	Two dimensional
ADAM	Adaptive Moment Estimation
AMD	Age-related macular degeneration
AUC	Areas under receiver operating characteristics curve
AV	Artery-Vein
AVR	Arteriolar Venular ratio
CFI	Color fundus imaging
CIFAR	Canadian Institute For Advanced Research
CNN	Convolutional neural netowrk
CRAE	Central Retinal Artery Equivalent
CRVE	Central Retinal Vein Equivalent
DR	Diabetic retinopathy
DRIVE	Digital Retinal Images for Vessel Extraction
DSC	Dice similarity coefficient
FC	Fully connected
FN	False negative
FP	False positive
GD	Gradient descent
GPU	Graphics processing unit
HR	Hypertension retinopathy
HRF	High-Resolution fundus
INSPIRE	Infrastructure for Spatial Information in the Euro- pean Community
MAD	Mean Absolute Distance
OD	Optic disc

CONTENTS

OM	Overlap Measure
RCNN	Region based CNN
ReLU	Rectified linear unit
RNN	Recurrent neural network
ROC	Receiver operating characteristics
ROI	Region of interest
RPN	Region proposal network
SGD	Stochastic Gradient Descent
SR	Success Rate
SVM	Support vector machine
TN	True positive
TP	True negative

List of Figures

1.1	Eye Anatomy. (original image is taken from - https://www.thoughtco.com/how-the-human-eye-works-4155646)	8
2.1	CIFAR-10 Network Architecture	23
2.2	Qualitative results of optic disc localization in retinal fundus images	26
3.1	Block diagram of Attention based fully convolutional neural network	31
4.1	Framework for automated AV classification	40
4.2	Determination of RoI (Region of Interest) for AVR computation. The annular regions around the optic disc pointed out by the arrows are the RoI.	42
4.3	Artery-vein Patches	43
4.4	Block Diagram of AV-Net	44
4.5	ROC Curve for AV-DRIVE Dataset	48
4.6	ROC Curve for AV-INSPIRE Dataset	49
4.7	ROC Curve for AV-WIDE Dataset	49
5.1	Framework for automated AV classification and AVR calculation . .	53
5.2	Original image in the left column and corresponding Region of in- terest for AVR (delimited by the two blue circles) in the right column	55
5.3	Network diagram for Artery-vein classification	56
A.1	Neural Network (Source: https://www.ibm.com/topics/neural-networks)	66
A.2	Different activation functions	68
A.3	Underfitting and overfitting	72
A.4	Dropout in a network (Source: https://www.dremio.com/)	74
A.5	Pooling operation	75
A.6	CNN Architecture	77

List of Tables

2.1	Description of the proposed network	25
2.2	Accuracy of Optic Disc Localization using CIFAR-10 Network Architecture over 1200 images of MESSIDOR	25
2.3	OD Localization method comparison on MESSIDOR Dataset	26
3.1	Comparative result of optic disc segmentation	33
4.1	Comparative study of the proposed and competing methods	47
5.1	Comparison of AVR values for all images of the INSPIRE-AVR dataset	58

Introduction

Arteriolar-venular diameter ratio (AVR) in fundus images is correlated with many diseases like coronary heart disease, stroke, diabetic retinopathy [1] etc. AVR of fundus images could be used for the prediction of these diseases. To measure the AVR, the first top-six wide arteries and top-six wide veins within 1D to 1.5D from the Optic Disc (OD) where D is the diameter of the optic disc are considered. AV ratio is calculated using the iterative method proposed by Knudtson et al. [2] using the formula: $AVR = CRAE / CRVE$, CRAE being Central Retinal Artery Equivalent and CRVE as Central Retinal Vein Equivalent. Higher blood pressure, high cholesterol, and pancreatic illnesses have all been linked to a higher AVR [3] [4] [5] etc. A low AVR is caused by abnormal widening of the veins which results in chronic high blood sugar levels. Timely detection of these vascular changes may allow preventive treatments to reduce the risk of vision loss. These data can be used in telemedicine and remote diagnostic areas. Manually classifying arteries and veins requires expertise in retinal image interpretation and is a very labour-intensive process (An expert ophthalmologist needs 10 minutes to compute the AVR of a fundus image).

Various diseases impact the flow of blood in the vessels, which causes either a thickening or a narrowing of arteries and veins [6], [7]. Specifically, the arteriolar-

Introduction

venular ratio (AVR) [8], which measures asymmetric changes in retinal arteriolar and venular diameter, has been linked to several disorders, including atherosclerosis [9], coronary heart disease, and stroke. Furthermore, a high AVR has been associated with increased levels of cholesterol levels [9] and inflammatory indicators, such as high-sensitivity C-reactive protein, among other factors. Other conditions that can cause an abnormal AVR include high blood pressure and diseases of the pancreas [10]. Moreover, a low AVR is a clear indicator of diabetic retinopathy (DR), which stands as the primary reason for vision impairment among working-age individuals in developed nations [10]. When the veins widen, it results in low AVR and it happens because of retinal hypoxia, which arises secondary to microvascular injury in the setting of chronic high blood sugar levels. If these vascular changes are identified early on, preventive treatment can be done, lowering the chance of any vision loss.

Different imaging techniques like fundus photography, fluorescein angiography, and optical coherence tomography are used to capture the vascular changes in the retina. Among these, fundus photography is less expensive than the others, as a result in the field of telemedicine this is mostly used. However, it is highly challenging to compute the AVR or any other measure of interest given a fundus image of a patient's arteries and veins, even after accurately segmenting the vasculature from the image. It takes a great deal of labor to manually classify arteries and veins. In a high-resolution image, we may frequently identify over a hundred vessels or vessel segments, many of which are ambiguous and need close examination to be identified. Due to this reason, traditional AVR measurement considers only top-six wide arteries and top-six wide veins close to the optic disc [2]. However, as smaller arteries are more susceptible to variations in blood pressure, computing a global AVR using the widths of smaller vessels may produce an even earlier biomarker of an underlying disease. However, because there are so many vessels that must be measured, manually calculating this global AVR is impractical. Hence, there is a requirement for automated or semi-automated approaches to address the time limitations associated with manual classification. Historically, the predominant focus of computer-aided vessel analysis has been on segmenting vessels within the image [11], [12], [13], [14], [15], [16], [17]. The Artery-Vein (AV) classification takes an additional step by aiming to categorize segmented vessels

1.1 Deep learning for diagnosis of hypertensive retinopathy

as either arteries or veins. But even with a perfect segmentation, automatically classifying arteries and veins is a very challenging computational operation due to the previously described ambiguity of small and mid-sized vessels.

The majority of heart attacks occur without apparent signs of previously detected high blood pressure. Hypertensive retinopathy (HR) serves as an indicator of high blood pressure, characterized by narrowed retinal arteries, retinal bleeding, and cotton wool spots. Thus, it's crucial to recognize early HR symptoms for effective prevention and treatment. Diagnosis typically involves fundus image analysis conducted by an ophthalmologist to identify stages of HR disease symptoms accurately.

1.1 Deep learning for diagnosis of hypertensive retinopathy

An essential requirement is the implementation of a computerized system to enhance the precision of retinal disease detection, diagnosis, and treatment for ophthalmologists. Deep learning has become a formidable tool in the analysis of medical images, particularly in the emerging method of computer-aided diagnosis from fundus images for detecting retinal diseases. Despite the rapid development of ophthalmic imaging in recent years, the application of deep learning algorithms to comprehend eye images has only recently gained traction.

The automated categorization of eye diseases is one of the numerous classification tasks for which machine learning has been used for many years. However, a significant portion of the efforts has been directed towards feature engineering, which entails the computation of explicit features as designated by experts [18], [19]. A class of machine learning approaches known as "deep learning" is characterized by the use of several computing layers, enabling an algorithm to acquire predictive features through examples, eliminating the need for manually engineered features. Lately, deep convolutional neural networks, a specialized form of deep learning tailored for images, have been utilized to create very accurate algorithms for diagnosing diseases like melanoma and diabetic retinopathy, as well

as hypertensive retinopathy [20], [21], based on medical images. These algorithms demonstrate accuracy levels comparable to those of human experts.

A deep neural network model is constructed through a sequence of mathematical operations executed on the input data, such as the pixel values found in an image. This mathematical function may have millions of parameters, also known as weights. [22]. Deep learning involves the use of multiple training cycles in an iterative manner to determine the right parameter values so that a function can carry out a certain task. For example, the model can generate a meaningful prediction by analyzing the values of the pixels within a fundus image. The training and tuning datasets make up the development dataset. The term tuning dataset is often used interchangeably with the validation dataset, although we aim to prevent confusion with a clinical validation dataset, representing data not utilized during the model's training. In the training phase, the neural network's parameters start with some random values. Subsequently, for every image, the model's prediction is evaluated against the known label in the training dataset. The model's parameters are then adjusted slightly to reduce the error for that particular image. The model undergoes this iterative process for each image in the training dataset until the model learns how to correctly determine the label based on the pixel intensities of all images in the training set. Through fine-tuning with ample data, the outcome is a model that possesses broad applicability, capable of predicting labels, such as cardiovascular risk factors, for novel images. Many studies utilize basic convolutional neural networks (CNNs) for examining color fundus imaging (CFI). They serve a diverse range of purposes, including the segmenting anatomical structures, the diagnosis of eye disorders, the segmentation and detection of retinal defects, and the evaluation of image quality, among others.

1.1.1 Detection

In any anatomical object segmentation task including organs or landmarks, localization of the same as the preprocessing step is very crucial. The predominant and widely adopted approach for identifying organs, regions, and landmarks appears to be localization using Convolutional Neural Networks (CNNs) in the context of

1.1 Deep learning for diagnosis of hypertensive retinopathy

2D image classification, giving satisfying results. Identifying objects of interest or abnormalities in images is a crucial aspect of diagnosis and represents one of the most labor-intensive tasks for clinicians. Typically, the tasks consist of the localization and identification of small lesions in the full image space. There exists a longstanding research tradition in the development of computer-aided detection systems aimed at automatically identifying lesions, intending to enhance detection accuracy or reduce the time required for human experts to interpret medical images. The majority of deep learning object detection systems currently rely on Convolutional Neural Networks (CNNs) for pixel classification. followed by the post-processing step that is typically employed to acquire potential object candidates.

1.1.2 Segmentation

Segmenting organs and other substructures in medical images enables quantitative analysis of clinical parameters associated with volume and shape. Segmentation is commonly described as the process of identifying the group of pixels that constitute either the contour or the interior of the object(s) of interest. This technique is commonly used in medical imaging analysis to identify and isolate specific structures, organs, or abnormalities within the images. In medical image processing, U-net [23] is the most widely recognized novel CNN architecture. RNNs have recently become more popular for segmentation tasks. When applying deep learning methods, the segmentation of lesions combines the difficulties of object recognition and organ and its substructure segmentation. Accurate segmentation often requires both global and local contexts. This forces employing multistream networks that incorporate various scales or patches sampled non-uniformly.

1.1.3 Classification

The technique of classifying and labeling groups of pixels or vectors within an image according to predetermined criteria is known as image classification. Classification in medical image analysis contributes to medical research by identifying patterns and correlations that may not be immediately apparent to human observers. There are two primary approaches to classification: supervised and unsu-

Introduction

pervised methods. The unsupervised classification approach is entirely automated and operates independently of training data. The supervised classification method of image classification involves using labeled training data to teach a computer algorithm to categorize pixels or regions within an image into predefined classes or categories. This process typically requires human intervention to provide examples of each class, allowing the algorithm to learn patterns and characteristics associated with each category. Once trained, the algorithm can then classify unseen or new images based on the patterns it has learned from the training data. The key area of medical image analysis where deep neural networks are mostly used is image classification. Classification models can be utilized to track the progression of diseases over time. By analyzing sequential medical images, these models can identify changes in the size, shape, or characteristics of abnormalities, providing valuable information for treatment adjustments and prognosis. Classification algorithms help automate the interpretation of images, reducing the workload on healthcare professionals and speeding up the diagnostic process. In summary, classification in medical image analysis is instrumental in enhancing diagnostic accuracy, treatment planning, and overall patient care. It brings automation, efficiency, and objectivity to the interpretation of medical images, contributing significantly to advancements in healthcare.

1.2 Eye Anatomy

The structure of the eye's anatomy is shown in Figure 1.1. Since human eyes are similar to camera lenses in that they concentrate light onto film, they function similarly to cameras. The way our eyes work is analogous to camera functionality where the retina of the eye is similar to the film, while the cornea and lens are comparable to the lens of a camera. The main parts of the human eye are the cornea, iris, pupil, aqueous fluid, lens, vitreous humor, retina, and optic nerve. The functionality of different components are as follows:

Cornea: It is the transparent outer layer situated in front of the eye. Light is refracted by the cornea, which produces around two-thirds of the optical power in the eye. The majority of light that enters the eye is focused by the cornea. It is devoid of blood arteries, in contrast to the majority of human tissues.

1.3 Diagnosis of Hypertensive retinopathy

Aqueous Humor: The clear fluid in the front segment of the eye is called aqueous humour. Its role involves nourishing the eye, maintaining intraocular pressure and keeping it inflated. This also helps to improve defence against various infections, wind, dust, and pollen grains.

Vitreous Humor: It is also known as a vitreous fluid which is a gel-like, colorless, and transparent substance that occupies the space between the lens and the retina in the eye. It mainly contains water along with small proportions of collagen, glycosaminoglycans (sugars), electrolytes (salts), and proteins.

Iris and Pupil: The pupil, an aperture in the cornea, allows light to enter the aqueous humor and its size controls the amount of light entering the eye. The iris, a contractile ring connected to eye color, determines the size of the pupil.

Lens: The primary function of the lens is to concentrate light onto the retina through transmission. About 80% of total refraction is contributed by the cornea, while the lens fine-tunes light focus onto the retina.

Retina and the Optic Nerve: The retina is the layer that covers the inner back part of the eye. Rods and cones are the two cell types in the retina that become active when light enters the area. Light is transformed by rods and cones into an electric signal that travels from the optic nerve to the brain. To create an image, the brain translates nerve impulses. Through comparison of the variations in the images created by each eye, three-dimensional information can be obtained.

1.3 Diagnosis of Hypertensive retinopathy

Hypertensive retinopathy is a disease that is generally asymptomatic in nature that damages the retina of the eye and results in loss of vision and is closely associated with high blood pressure. Severe hypertensive retinopathy cases result in systemic illnesses that have the potential to cause mortality, heart and kidney

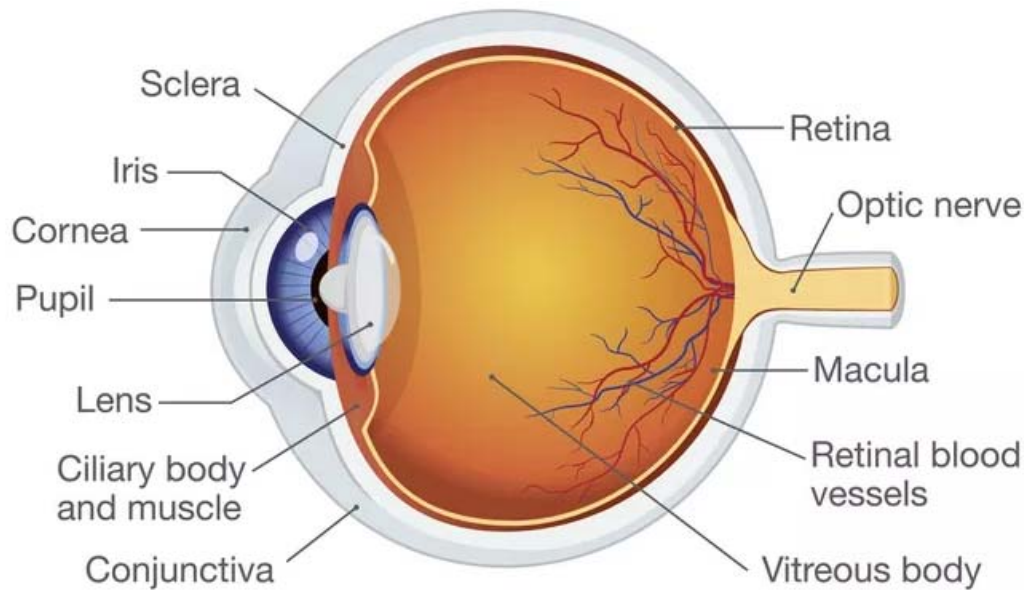


Figure 1.1: Eye Anatomy. (original image is taken from - <https://www.thoughtco.com/how-the-human-eye-works-4155646>)

failure, vision loss, and cardiovascular diseases. As a result, early diagnosis of the disease and therapy are essential. Only qualified professionals like ophthalmologists can provide a proper diagnosis of the disease.

Imaging techniques such as fundus photography may be used to capture detailed images of the retina and its blood vessels. Physicians can examine blood vessels non-invasively solely through the retina [24]. The blood vessels have historically been investigated due to their importance in clinical evaluation, as vessel diameter has been one of the most important elements in the diagnosis of hypertension. As a result, the arteriolar-venular ratio (AVR) gives a very important indication to medical experts for the possible presence of high blood pressure as arteries show a decrease in width for hypertensive patients.

1.4 Literature review

Our work has been divided into two parts. The first part concentrates on accurate and fast optic disc detection and segmentation methods using an attention-based

1.4 Literature review

fully convolutional network. Accurate segmentation of the Optic disc (OD) is very important for computer-aided diagnosis of several ocular diseases such as glaucoma, diabetic retinopathy, and hypertensive retinopathy. The subsequent section discusses an automated approach for assessing the Arteriolar-Venular Ratio (AVR) in retinal color images. This method involves identifying the optic disc's location, defining a suitable region of interest (ROI), categorizing vessels into arteries or veins, gauging vessel diameters, and ultimately computing the AVR.

1.4.1 Review on Optic Disc localization and segmentation

Several works have been reported on the automatic localization of the OD in the retinal fundus image based on the shape and texture feature of the retina. However, these methods fail in case of the presence of exudates due to similarity in appearance. Barrett et al. [25], Blanco et al. [26] localized OD assuming that would be of circular shape. The assumption is good in most of the cases but it fails if the object is of some other shape or if some other objects are of circular shape. Hoover et al. [27] address this issue by considering the anatomical feature that retinal vessels originate from OD. In the same concept, Mendonca et al. [28] used entropy of vascular directions as the anatomical feature. Soares et al. [29] combines the local appearance before determining the location. The disadvantage of using the local feature is that some pathologies may impact the feature. To overcome this, Foracchia et al. [30] used a model of the geometrical directional pattern of the retinal vascular system. This method works well for images having pathologies but can fail in irregular or incomplete vessel network scenarios. Niemeijer et al. [31] used a mixed approach of considering both local and global features and designed a point distribution model. Xiangqian et al. [32] localize the OD by integrating the information of both OD appearance and vessel networks. So it is seen that all these local and global features are sometimes impacted by pathologies and other abnormal factors.

Optic disc (OD) identification and segmentation are done using a variety of techniques. Lowell et al. [33] applied a direction sensitive gradient-based technique to remove the vessel obstructions and deformable ACM for finding the OD boundary in low resolution images. Welfer et al. [34] used morphological proce-

dures and the watershed transformation technique to determine the OD boundary. Aquino et al. [35] utilized morphological operations, an edge detection method, and a circular Hough transformation technique to segment the OD. Inpainting was used as a preprocessing technique by Morales et al. [36] to eliminate blood vessels and employed stochastic watershed transformation to delineate the boundary of the optic disc (OD). Xu et al. [37] applied the active contour model (ACM) and proved better segmentation. by addressing the fluctuations within the OD region, Joshi et al. [38] enhanced the robustness of ACM proposed by Chan and Vese [39]. Principal component analysis was used by Morales et al. [36] to determine the optic disc's boundaries.

1.4.2 Review on Artery-vein classification

Most of the automatic classical A/V classification methods [40] [41] [8] [42] [43] follow few common steps. It starts with preprocessing of the image and then assigns artery-vein probability for each pixel, next from topological structure of the vessel determines the label of the artery-vein. In the prepossessing step, issues like uneven illumination and background in-homogeneity are corrected using techniques like luminosity normalization and histogram equalization. A vessel binary map is generated once the retinal vessels have been segmented. Next, to assign the pixel-wise A/V probability, at first, all vessel centre-line pixels are analyzed to extract intensity-based features using which a probability is assigned to each centre-line pixel. Pixel-wise classification can be further improved by creating a topological structure of the vascular network. This structure basically establishes how each individual segment is connected. At last, using local and contextual information every section of the vessel is categorized as either an artery or a vein and eventually the entire vessel tree is labelled. Fan Huang et al. [44] proposed the categorization of arteries and veins based on features like colours, topological, geometrical and morphological properties of the vascular structure. Memari et al. [45] presented an artery-vein segmentation method by enhancing fuzzy c-means clustering and level sets. More the accuracy is achieved in segmentation, the more precise will be the A/V classification. Jebaseeli et al. [46] enhanced the Pulse Coupled Neural Network model using some parameter optimization to

1.4 Literature review

segment the vessel structure. Chen et al. [47] described topological connectivity for classifying the artery and vein, based on which a topology discriminator they proposed to enhance the accuracy of the classification.

For the classification task, of late many deep learning-based methods are being employed. Welikala et al. [48] proposed an artery-vein classification convolution neural network (CNN) model from retinal fundus images. The network presented by them is able to learn the existing complex features automatically instead of using handcrafted features made up of three convolution layers with three fully connected layers. Before classifying, the vessel must be segmented and the centre line extracted using the prescribed method. Lepetit-Aimon et al. [49] employed fully convolutional network a large receptive field for segmentation of the blood vessels for high-resolution images with minimum computation cost and comparatively low memory. The method's inability to accurately categorize blood vessels can be attributed to its ignorance of the vessels' true topological structure. Hemelings et al. [50] introduced a segmentation and classification approach based on the classical U-Net model. Girard et al. [51] employed deep learning techniques along with graph propagation for vessel segmentation and classification of arteries and veins. However for small diameter vessels and arteriovenous intersections, classification results were not up to the mark. For similar work, Ma et al. [52] has shown 94.5% pixel-wise accuracy using a multi-task neural network on the AV-Drive dataset. Wang et al. [53] presented a multi-task siamese network, that simultaneously performs segmentation and classification tasks to provide more robust deep features from the fundus images. Using the attention mechanism Chowdhury et al. [54] developed a multiscale encoder-decoder based attention network that can both segment and classify arteries and veins. IK Gupta et al. [55] applies the U-Net framework for segmentation and employs the mayfly optimization kernel extreme learning [56] method for classification. Sathananthavathi et al. [57] suggested a deep semantic segmentation architecture for AV classification. Hu et al. [58] tested a model for AV classification on DRIVE and HRF dataset that employs an encoder-decoder segmentation network along with a point set classification maps of the AV skeleton. Toptas et al. [59] created artery vein patches separately after preprocessing the fundus image and fed them as input to their

proposed deep learning network architecture to finally classify them whether the patches are of arteries or veins.

1.4.3 Review on Computation of Arteriolar-to-venular (AVR) width ratio

Several approaches to automatically calculate AVR have been previously introduced [40] [60] [61]. The main difference in those studies is in the area of artery-vein classification. In 2003, the automatic classification of AV was first published in a work by Grisan et al. [62]. Ruggeri et al. [60] presented a method where they first located the OD, identified RoI around the OD and finally classified the vessel pixels as an artery or vein inside the RoI. An extended version of this work has been presented by Tramontan et al. [61] where they changed both vessel tracking and the features for discriminating between structural arteries and veins. Niemeijer et al. [40] used a linear classifier for vessel classification and calculated AVR using Knudtson's revised formula [2]. Khanal et al. [63] in their recently devised a fully automated method to calculate AVR in fundus image after classifying artery-vein on a complete vascular network.

1.4.4 Overall Review

Diabetic retinopathy, hypertensive retinopathy, and glaucoma are common causes of visual impairment and blindness [64]. Early diagnosis and appropriate referral for treatment of these diseases can prevent visual loss. Research is going on in the development of a computer-aided diagnosis system for the accurate identification of different parts and pathologies in retinal fundus images to assist ophthalmologists. The optic disc is the entry point of the major blood vessels in the retina [33] and considered as landmark in retinal fundus image. Disc size and cup area are used for diagnosis of glaucoma [38], [65]. The centre of the optic disc is an important reference for detecting the macula and grading macular pathologies, such as diabetic maculopathy, macular edema, and macular ischemia [66]. Disc size is also an important parameter for the determination of the region of interest, where the

1.4 Literature review

width of the artery and vein needs to be computed for diagnosis of hypertensive retinopathy [67]. Along with the position of the optic disc, the vessel origin is another important feature for vasculature analysis [35]. Automated detection and segmentation of the optic disc is a challenging problem due to the variation in size, shape, color, and the variation introduced by the field of view, inhomogeneous illumination, and pathological abnormalities. Shape and brightness [25], [68], [26] convergence of blood vessels [27], [30] and orientation of blood vessels [30], [69] have been investigated for detection of optic disc. The assumption of the circular shape of the optic disc does not hold good, where the Optic Disc is partly present in the retinal image. Hoover *et al.* resolved this issue of poor contrast of the optic disc by considering the convergence of blood vessels in the optic disc. Orientation of blood vessels has been used by Foracchia *et al.* [30] and Youssif *et al.* [69] to improve the result of Optic Disc localization. Vessel templates were also investigated by Osareh *et al.* [70] and Lowell *et al.* [33]. The active shape model is used to extract the main blood vessels for localization of OD [71]. Brightness characteristics of OD and vessel density in the OD region are utilized by Giachetti *et al.* [72]. Soares *et al.* [29] focused on the local appearance of the OD region and orientation of main blood vessels to determine the centre of OD. Vessel directional and distribution of blood vessels is used by Zhang *et al.* [73] to improve the accuracy of OD localization. Roychowdhury *et al.* [74] used region-based features to classify the bright areas as OD and non-OD regions. The region with maximum vessel density and solidity is considered as the OD candidate.

The reported works on optic disc segmentation can be classified based on three categories - morphology, active contour model and active shape model. The OD boundary was segmented by Aquino *et al.* [35] by morphological operations, edge detection method and circular Hough transformation technique. Walter *et al.* [75], Reza *et al.* [76], and Welfer *et al.* [34] determined the OD boundary using morphological operations and watershed transformation technique. Morales *et al.* [36] applied inpainting as preprocessing for removing blood vessels and stochastic watershed transformation for determining the OD boundary. Caselles *et al.* [77] and Xu *et al.* [37] proved that segmentation can be performed better using active contour model (ACM). Snake-based ACM was used by Osareh *et al.* [70] to extract the OD boundary. In order to improve the accuracy of segmentation for low-resolution

images, Lowell *et al.* [33] applied a direction sensitive gradient-based technique to remove the vessel obstructions and deformable ACM for finding the OD boundary. Chrastek *et al.* [78] applied a distance map algorithm to remove the blood vessels and then segmented the OD by using a sequence of methods like morphological operation, Hough transformation, and ACM. The method presented by Joshi *et al.* [38] improved the robustness of ACM proposed by Chan and Vese [39] by taking care of the variations in the OD region. Li and Chutatape [71] presented an active shape model-based technique for OD segmentation. Hybrid level set algorithm [79] was implemented by combining the local and regional gradient information of the fundus images. Morales *et al.* [36] detected the boundary of the optic disc by the principal component analysis. Random forest [80], [81] and boosting [82], [83] have been used to predict the class probabilities of the pixels. These methods are prone to significant errors when certain geometric assumptions do not hold if precise identification fails.

Recently the ratio of people suffering from various kinds of eye diseases to that of available ophthalmologists, particularly in developing countries is increasing at an alarming rate. An automated computer based system for analysis of fundus images would reduce the burden on the health care system of these kinds of countries. The advancement of deep learning-based algorithms is playing an important role in designing such software systems. The success of convolutional neural network in object segmentation [84], [85], [86], [87] has motivated us to investigate the performance of fully convolutional network for optic disc detection and segmentation.

1.5 Objective

Cardiovascular diseases are very common in India, presenting a significant challenge to the healthcare sector in terms of early diagnosis and risk assessment. The current screening techniques for cardiovascular disease are extremely intricate and depend on a wide range of factors obtained from blood samples and the patient's medical history. The absence of certain parameters during a specific time period can result in inaccurate predictions regarding the risk of cardiovascular diseases.

The nature of retinal vessels is crucial for the early detection of various systemic

1.6 Outline of the Thesis

diseases like diabetes, hypertension, and cardiovascular disorders. Any change in the retinal vasculature can be observed through fundus photography, the preferred imaging method for telemedicine and remote diagnostics due to its cost-effectiveness and user-friendly nature.

Currently, there is a need to develop an indigenous imaging system suitable for deployment not just in major hospitals but also at grassroots-level primary healthcare units. It will enable extensive screening of the general population, particularly those unaware of the impact of cardiovascular diseases. Using this system eye specialists may distinguish between "normal" and "high-risk" patients, advising quick treatment for the latter. This initiative can fulfill the requirement of designing a rapid, accurate, dependable and cost-effective cardiovascular risk assessment tool based on the deep-learning-based analysis of retinal fundus images. So in summary, our aim of the whole work is :

- **Development of a robust classification technique for segregation of arteries from veins**
- **Investigation of diagnostically relevant properties of arteries and veins to compute AVR**
- **Improvement of algorithm for handling noisy images**

1.6 Outline of the Thesis

In this thesis, novel methods are developed for the detection and segmentation of Optic disc in the images affected by hypertensive retinopathy using robust and effective attention-based fully convolutional network and then vessel segmentation, Artery/Vein classification, vessel caliber measurement, and arteriovenous ratio calculation. The overview of the thesis is provided below:

1.6.1 Chapter 1: Introduction

Deep learning is an active research area in the fields of computer vision and pattern recognition. In the field of medical image analysis, deep learning techniques

achieved better performance in object detection, segmentation, and classification. This chapter provides a comprehensive review of how deep learning is applied in fundus image analysis for the treatment of Hypertensive retinopathy.

1.6.2 Chapter 2: Localization of optic disc

For the computer-aided diagnosis of several eye diseases e.g. hypertensive retinopathy, diabetic retinopathy, and glaucoma, accurate optic disc localization and subsequently segmentation is a very crucial step. The optic disc serves as the point of entry for the major blood vessels into the retina and is considered a significant landmark in retinal fundus images.

1.6.3 Chapter 3: Segmentation of optic disc

Accurate segmentation of the optic disc is very important for computer-aided diagnosis of several ocular diseases such as glaucoma, diabetic retinopathy, and hypertensive retinopathy. Chapter 3, comprises the methodology of an accurate and fast optic disc detection and segmentation method using an attention-based fully convolutional network. The network is trained from scratch using the fundus images of the extended MESSIDOR database and the trained model is used for the segmentation of the optic disc. The false positives are removed based on morphological operation and shape features. The attention-based fully convolutional network is robust and effective for the detection and segmentation of optic disc in the images affected by diabetic retinopathy.

1.6.4 Chapter 4: Classification of vessels into artery and vein

The significance of classifying arteries and veins lies in its pivotal role in the diagnosis and understanding of numerous cardiovascular disorders. This classification is essential for accurate medical assessments and effective treatment strategies. In this Chapter, we have focused on identifying the region of interest, extracting vessel central lines within that region, and classifying arteries and veins using an

1.6 Outline of the Thesis

attention-based deep neural network.

1.6.5 Chapter 5: Computation of AVR

In this Chapter, AVR measurement challenges and techniques have been discussed. The accurate estimation of AVR is challenging and requires optic disc detection, vessel segmentation, accurate vessel width measurement, vessel network analysis, and artery vein classification. Optic disc detection is necessary to determine the location of the region of interest (ROI) where the measurements are obtained. Vessel segmentation must be used to find the vessels themselves and the width of the vessels. Any AVR measurement system must identify which vessels are arteries and which are veins with high accuracy since small classification errors can have a large influence on the final AVR.

1.6.6 Chapter 6: Conclusion

Automatic Computation of Arteriolar-Venular Ratio in Retinal Fundus Images is a very critical area in clinical use. Automated detection of anatomical parts such as the optic disc, retinal vasculature, and macula is crucial for developing a screening tool. We have used the Faster R-CNN method for the optic disc detection work. In the next step, we used an attention-based fully convolutional network for the segmentation of the optic disc. Finally, we have presented an attention-based AV net to classify arteries and veins in the region around the optic disc known as RoI using which the Arterilar-venular ratio has been calculated. We intend to expand on this research by looking at unsupervised attention learning and classification.

Optic Disc localization in fundus image using Faster R-CNN

Optic Disc localization in fundus image using Faster R-CNN

Now a days lot of people are suffering from Diabetic Retinopathy throughout the world. This is one kind of eye disease which affects people who have diabetes for a long time. If this is undiagnosed and not treated for a long time, it can lead to blindness. Several studies have shown that early detection and timely treatment are the only way to reduce the suffering from diabetic retinopathy. The development of an automated screening system is the right approach for screening diabetic retinopathy. Automated detection of several anatomical regions such as the optic disc, retinal vasculature and macula is important to design a tool for the screening purpose. In our work, we have presented a novel and fast optic disc detection method using Faster R-CNN. The proposed method is validated on 1200 fundus images from the MESSIDOR database [88] which is a widely accepted publicly available dataset for research purposes. We propose a supervised detection technique that uses a deep learning network trained on 6,992 retinal fundus images augmented using geometrical transformations from the MESSIDOR-II database. The proposed method shows satisfactory robustness on both normal and images affected by diabetic retinopathy. It outperforms many previous methods in terms of speed with satisfactory accuracy of optic disc localization.

2.1 Introduction

It has been observed that one of the main threats all over the world to public health is Diabetes mellitus which leads to Diabetic Retinopathy (DR) in the long term [89]. DR is a diabetic complication that causes changes in the retina. In the early stages DR shows no symptoms. The disease starts showing symptoms that become symptomatic only when it reaches the advanced stages, which is dangerous. The resulting blindness is a truly preventable complication via early detection of DR and timely treatment. The number of ophthalmologists is very low in India as compared to the number of patients suffering from various kinds of eye diseases. So it is important to develop a screening tool that can be the right option to detect diabetic retinopathy, its severity, and to decide which patients require referral for further investigation and possible treatment. An automated system for this kind of screening tool can be developed whose main purpose will be to filter out the retinal fundus images having the symptoms of DR. This can also be designed in such a manner that it can determine the stage of the DR which will be very helpful input to any ophthalmologist to start the right medication. There are a lot of advantages of these kinds of tools e.g. reduction in screening

2.2 Reported Works on Optic Disc localization in fundus image

time which will help to cater to more patients in less time, lesser cost of treatment due to early detection of the disease etc.

Generally the first pathological sign that appears in fundus image is microaneurysms when someone is attacked by DR. The other pathological signs that appear gradually are hemorrhages, exudates, macular edema etc. So the various information like size, and location regarding these pathological patterns help to distinguish DR infected images from normal fundus images [90]. Optic disc (OD) detection is very important for the early detection of Glaucoma, detection of the junction point of the blood vessel and tracking of the blood vessel, analysis of retinopathy of prematurity and to determine the distribution of different pathology present in retinal fundus image.

2.2 Reported Works on Optic Disc localization in fundus image

The reported work based on the shape and texture features of the retina has been discussed in Chapter 1. Automatic localization of OD based on these features faces challenges when exudates are present, as their similarity in appearance can lead to inaccuracies. To overcome this issue, some researchers [25], [26] have relied on the assumption that the OD is circular, This process fails in a few scenarios where other objects exhibit circular shapes. This problem has been addressed in the literature [27], [28] by using the anatomical feature of retinal vessels, and in [29] by combining the local appearance. All these methods perform well in images with pathologies but may falter in scenarios with irregular or incomplete vessel networks. A mixed approach, considering both local and global features, integrating information from both OD appearance and vessel networks have been presented in [31], [32]. Despite these efforts, it is evident that both local and global features can be influenced by pathologies and other abnormal factors in certain cases.

We have proposed Deep neural Networks using Faster-RCNN to overcome all these problems as it allows the network to learn features dynamically during the training process. Features of varying complexity will be extracted at the network's convolutional and pooling layers, and those will be used during testing. Our approach demonstrates robustness through training and testing on diverse datasets.

2.3 Methodology

2.3.1 Using Convolution neural networks for object localization

Convolutional Neural Networks (ConvNets) are a specific type of neural network that has shown excellent performance in many computer vision and machine learning problems. Nowadays there are various areas where ConvNets are used. Some well-known areas are self-driving cars, automatic face recognition, identification of objects, number plates of cars, giving vision to robots etc. The success of two new methods known as Region proposal and Region-based convolution neural networks (R-CNN) has dramatically helped the object detection task to be very successful in various types of images. R-CNN [91] was developed in 2013, then Fast R-CNN [92] came in 2015, in the same year Faster R-CNN [93] was developed to overcome the shortcomings of Fast R-CNN [92].

R-CNN technique [91] is Region proposals + CNN. Here Selective Search is used for localization which generates 2000 different regions where the probability of object existence is highest. These set of region proposals are then fed into a trained deep-learning CNN which then works on each region and extracts a feature vector. In the next step the vector is fed into a set of linear SVMs for the classification task. All these SVMs are already trained to classify each class. Also there is a bounding box regressor where this vector is given as input to get the most accurate coordinate.

Some notable drawbacks were observed in R-CNN, mainly in three areas - Multiple stages were required for the training like ConvNets to SVMs to bounding box regressor, it was extremely slow (on an average 53 seconds were taken by the RCNN to process one image) because computationally it was expensive. To improve the speed Fast R-CNN [92] incorporated two things - first shared computation of the conv layers between different proposals, secondly swapped the order of region proposal generation and running of the CNN. In this model, two inputs are given to the fully convolutional network - entire input image and multiple RoI (region of interest). After this each region of interest is pooled into a fixed size feature map. The fixed length feature vector that are extracted from the feature map are then fed into a series of fully connected (FC) layers. These FC layers finally give two outputs for each RoI - softmax probability and bounding box for each of the object classes.

The latest technique Faster R-CNN [93] has been developed which consists

2.3 Methodology

of two modules. The function of the first module which is a deep convolutional network is there to propose regions. The second module is Fast R-CNN detector described above that uses those proposed regions. When these two modules are combined it works as an object detection model. With respect to the earlier models, the main change is the designers inserted a region proposal network (RPN) after the last convolution network before the Fast R-CNN. This new layer informs the Fast R-CNN module where to look into the image. The major breakthrough in the object detection model has been achieved through this design because before this it was only possible to determine the existence of any specific object in an image but this model helps to determine the exact location of the object.

Refer to the CIFAR-10 Network architecture Fig. 2.1 used in our proposed work, in the next section it has been discussed in detail. The layer details of the network are given in Table 2.1.

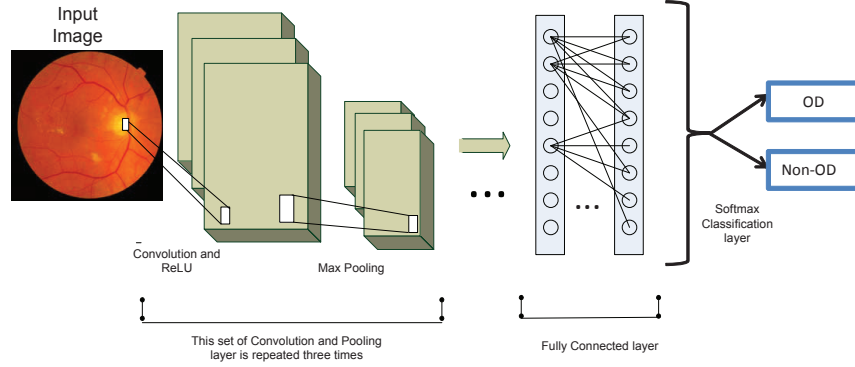


Figure 2.1: CIFAR-10 Network Architecture

2.3.2 Proposed localization framework

At first the fundus images are resized and enhanced by gamma correction. The resizing step is necessary because of the variety of sizes of the images in the MESSIDOR database [88]. All the images are resized to 224 X 224 pixels. This resizing helps improve the computation time without affecting the performance. Then 1000 nos. of positive and 1000 nos. of negative samples are created by randomly cropping images from MESSIDOR-II dataset. After this end-to-end training of CIFAR-10 network is being done with those positive and negative samples. This trained network is then fine-tuned with the help of images from MESSIDOR-II dataset. To introduce the robustness of this process, geometry-based augmenta-

tion is performed on MESSIDOR-II dataset and then fed those images into the network. We created a total of 6,992 training images by augmentation. We have trained our network using the images from MESSIDOR-II dataset and then tested on MESSIDOR dataset [88]. Faster R-CNN [93] method has been used to fine-tune this pretrained CIFAR-10 NET using 6,992 images using different epoch values. The optic disc was localized using the bounding box by the regression head of Faster R-CNN. Each bounding box on an input image indicates the probable location of OD. The actual OD location is indicated by a bounding box with the highest probability values. The centre of that bounding box is considered the centre of the optic disc.

2.4 Results and Discussion

The proposed optic disc localization framework is evaluated on 1200 images from Messidor database [88]. These images are publicly available for researchers in the area of retinal image processing, diabetic retinopathy etc. and these are maintained by French Techno-vision program. Out of 1200 eye fundus color numerical images of the posterior pole for the Messidor database [88], 800 images were acquired with pupil dilation and 400 without dilation. Qualitative results of optic disc localization are shown in Fig. 2.2. The method works fine for low-contrast as well as high-contrast images. The method achieves 98.75% accuracy of optic disc localization on Messidor database [88] as shown in Table 2.2. Our proposed method has achieved an average 0.15 milliseconds OD segmentation time per image which is far better than the reported performance on the MESSIDOR dataset [88]. The comparison of results has been given in Table 2.3.

A failure case is also given in Fig. 2.2(d), which is due to poor contrast between OD and background.

Euclidean distance is computed between the centre of the bounding box and the centre marked by experienced ophthalmologists to measure the perfectness of localization of the optic disc. Euclidean distance between two point is defined as $\mathbf{A} (a_1, a_2)$ and $\mathbf{B} (b_1, b_2)$

$$\text{Euclidean distance} = \sqrt{\sum_{i=1}^n (a_i - b_i)^2}, \quad (2.1)$$

The lower value of Euclidean distance is an indication of the perfectness of

2.4 Results and Discussion

Table 2.1: Description of the proposed network

Layer #	Layer Type	Description
Layer 1	Input Image	Input Image
Layer 2	Convolution	Filter Size=5x5 No. of Channels=3 No. of Filters=32 Stride=1, Padding=2
Layer 3,6,9,12	ReLU	Introduces non-linearity
Layer 4,7,10	Max pooling	Pool Size=3x3
Layer 5	Convolution	Filter Size=5x5 No. of Channels=32 No. of Filters=32 Stride=1, Padding=2
Layer 8	Convolution	Filter Size=5x5 No. of Channels=32 No. of Filters=64 Stride=1, Padding=2
Layer 11	Fully Connected	64 Nodes
Layer 13	Fully Connected	2 Nodes
Layer 14	Softmax	2 Classes
Layer 15	Classification Output	Bounding box around the object

Table 2.2: Accuracy of Optic Disc Localization using CIFAR-10 Network Architecture over 1200 images of MESSIDOR

Epoch Value	No. of Failed cases	Accuracy of OD localization (%)
1	24	98.00
5	15	98.75
10	15	98.75
15	15	98.75
20	15	98.75

Optic Disc localization in fundus image using Faster R-CNN

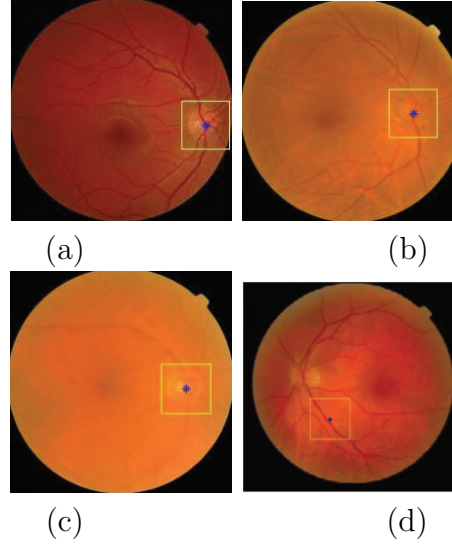


Figure 2.2: Qualitative results of optic disc localization in retinal fundus images

Table 2.3: OD Localization method comparison on MESSIDOR Dataset

	Time	No. of Failed Cases	Accuracy (%)
Yu [94]	4.7s	11	99.08
Lu [95]	5s	3	99.75
Aquino [96]	1.67s	14	98.83
Soares [29]	18.64s	9	99.25
Proposed	0.15ms	15	98.75

localization of the optic disc. The average Euclidean distance is 14 pixels for the proposed method.

2.5 Conclusion

We have investigated the performance of Faster R-CNN in optic disc localization. The proposed deep learning framework shows satisfactory robustness on normal and diabetic retinopathy fundus images. It outperforms previous methods in terms of speed. However, the accuracy of optic disc localization needs to be improved by introducing robust preprocessing technique and suitable deep-learning architecture.

Optic disc segmentation in fundus image Using Attention Based Fully Convolutional Network

Optic disc segmentation in fundus image Using Attention Based Fully Convolutional Network

Accurate segmentation of the optic disc is very important for computer-aided diagnosis of several ocular diseases such as glaucoma, diabetic retinopathy, and hypertensive retinopathy. The paper presents an accurate and fast optic disc detection and segmentation method using an attention based fully convolutional network. The network is trained from scratch using the fundus images of the extended MESSIDOR database and the trained model is used for the segmentation of the optic disc. The false positives are removed based on morphological operation and shape features. The result is evaluated using three-fold cross-validation on six public fundus image databases such as DIARETDB0, DIARETDB1, DRIVE, AV-INSPIRE, CHASE DB1 and MESSIDOR. The attention based fully convolutional network is robust and effective for the detection and segmentation of optic disc in the images affected by diabetic retinopathy and it outperforms existing techniques.

3.1 Introduction

Visual impairment and blindness are a major problem in developing countries [97]. Diabetic retinopathy, hypertensive retinopathy, glaucoma are common causes of visual impairment and blindness [64]. Early diagnosis and appropriate referral for treatment of these diseases can prevent visual loss. Research is going on in the development of a computer-aided diagnosis system for accurate identification of different parts and pathologies in retinal fundus images to assist ophthalmologists.

The Optic disc is the entry point of the major blood vessels in the retina [33] and is considered a landmark in the retinal fundus image. Disc size and cup area are used for diagnosis of glaucoma [38], [65]. The centre of the optic disc is an important reference for detecting the macula and grading macular pathologies, such as diabetic maculopathy, macular edema, and macular ischemia [66]. Disc size is also an important parameter for determination of the region of interest, where the width of artery and vein needs to be computed for diagnosis of hypertensive retinopathy [67]. Along with the position of the optic disc, the vessel origin is another important feature for vasculature analysis [35]. Automated detection and segmentation of the optic disc is a challenging problem due to the variation in size, shape, colour, and the variation introduced by the field of view, inhomogeneous illumination and pathological abnormalities. Shape and brightness [25], [68], [26], convergence of blood vessels [27], [30] and orientation of blood vessels [30], [69] have been investigated for detection of optic disc. The assumption of the circular shape of the optic disc does not hold good, where the optic disc is partly present

3.2 Methodology

in the retinal image. Hoover *et al.* resolved this issue of poor contrast of the optic disc by considering the convergence of blood vessels in the optic disc. Orientation of blood vessels has been used by Foracchia *et al.* [30] and Youssif *et al.* [69] to improve the result of optic disc localization. Vessel templates were also investigated by Osareh *et al.* [70] and Lowell *et al.* [33]. The Active shape model is used to extract the main blood vessels for localization of optic disc [71]. Brightness characteristics of the optic disc and vessel density in the optic disc region are utilized by Giachetti *et al.* [72]. Soares *et al.* [29] focused on the local appearance of the optic disc region and the orientation of main blood vessels to determine the centre of the optic disc. Vessel directional and distribution of blood vessels are used by Zhang *et al.* [73] to improve the accuracy of optic disc localization. Roychowdhury *et al.* [74] used region-based features to classify the bright areas as optic disc and non-optic disc regions. The region with maximum vessel density and solidity is considered as the optic disc candidate.

The success of convolutional neural network in object segmentation [84], [85], [86], [87] has motivated us to investigate the performance of attention based fully convolutional network for optic disc detection and segmentation. The attention modules help to increase the performance of segmentation for natural images. To suppress false-positive regions and to highlight informative regions, we have developed attention based fully convolutional network. The spatial attention provides the location of the features, while the channel attention utilises the features that can be found in the available channels. Attention module helps to focus on relevant features to improve the segmentation results. Channel and spatial attention combine both the spatial context as well as the semantic information of the optic disc. The purpose of the proposed framework is the development of attention based deep network for determining the informative region of the fundus image similar to the optic disc and the selection of informative images for fast and efficient transfer learning.

3.2 Methodology

3.2.1 Preprocessing

The images of different databases have different sizes. Therefore, the images are resized to 512×512 pixels for all databases. This process of resizing not only reduces the storage space of the database but also decreases the computational time

Optic disc segmentation in fundus image Using Attention Based Fully Convolutional Network

without hampering the performance of the algorithm. The remaining operations are carried out on these resized images. Red channel image is thresholded. Morphological opening, closing and erosion operations with square structuring element are used to create a mask of circular retinal fundus region-of-interest, which allows focusing only on the foreground of retinal images. The fundus image is cropped based on the bounding box of this mask. The segmentation algorithm is applied to the cropped image to reduce the processing time.

3.2.2 Segmentation using fully convolutional network

In the case of convolution neural networks (CNN), a fully connected layer is added at the end of the network, whereas an FCN uses a convolutional layer only without adding the fully connected layers. In classification tasks, CNN works well because the output required is a label to which the class belongs. In contrast to this, semantic segmentation outputs an image where each pixel has been labeled. Thus it is required that the convolutions are performed at pixel-level and also that both spatial and semantic information is preserved. By removing the fully connected layer and replacing it with convolutional layers an upsampling procedure can be performed once the pixels in the image have been labeled.

Attention modules provide a way to improve features that are relevant for classification tasks. Different approaches have been explored in recent years. The attention modules can either work in the direction of hard attention or soft attention. Hard attention is non-differentiable and stochastic. Soft attention is probability-based [98] and deterministic, which makes it easier to use during training. In this work, soft attention is discussed as it is better suited for optimisation as it is compatible with back-propagation [99]. Early works on attention modules, where both local features and global features are extracted and the classification decision is based on the weighted local features. By using the weighted local features the network is guided towards learning only relevant features. Fan et al. [100] worked on re-designing CNN architectures to improve semantic segmentation tasks. They used channel attention and spatial attention modules [100].

The channels containing high-level features could be improved by channel attention whereas the spatial attention could enhance the spatial connectivity of features. The proposed fully convolutional neural network with an attention module is shown in Fig. 3.1. The network consists of four blocks and three attention blocks are added to refine the feature map of the last three blocks. In the convolu-

3.3 Results and Discussions

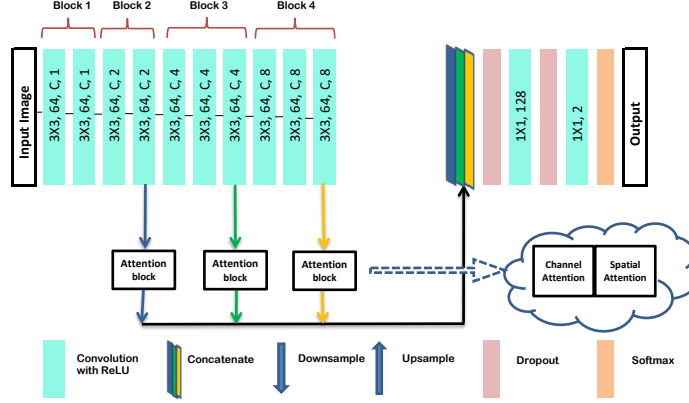


Figure 3.1: Block diagram of Attention based fully convolutional neural network

tion block attention module (CBAM) [101], a channel attention module is applied on a feature map. The result is further refined by a spatial attention module. The architecture focuses on both channel and spatial attention to learn about what and where about features. The features extracted from the last three blocks are refined by CBAM and the final feature map is obtained by depthwise concatenation of refined features.

A stochastic gradient-based optimization *ADAM* [102] is applied to minimize the cross-entropy based cost function. *ADAM* utilizes the first and second moments of gradients for updating and correcting the moving average of the current gradients. The learning rate for *ADAM* optimizer is set to 0.0001, weights of background and foreground are initialized as 1:10, and training was performed up to 30 epochs. Over-fitting is reduced by using dropout [103].

3.3 Results and Discussions

Training has been performed in a Linux environment using an 8 GB GPU on a system with Core-i7 processor and 32 GB RAM. The network architecture is implemented in Python using the PyTorch library. The segmentation results are evaluated by considering 3-fold cross-validation.

The performance of optic disc detection is evaluated using Success Rate (SR) and it represents the percentage of retinal images in a dataset where the centroid

Optic disc segmentation in fundus image Using Attention Based Fully Convolutional Network

of the optic disc is successfully localized within the boundary of the ground truth mask of the optic disc. The performance of optic disc segmentation is evaluated in terms of a region-based metric Overlap Measure (OM) and a contour-based metric Mean Absolute Distance (MAD) [74]. The OM represents the ratio of the intersecting area between the actual optic disc and the segmented optic disc and it is defined as

$$OM = \frac{N_{tp}}{N_{tp} + N_{fp} + N_{fn}}. \quad (3.1)$$

where N_{tp} , N_{fp} and N_{fn} are the number of true positive, false positive and false negative pixels, respectively. MAD represents the mean of the shortest distances from the boundary of the actual optic disc to the boundary of the segmented optic disc. Let $P_{i'}$, where $i' = 1, \dots, m_C$ and $Q_{j'}$ where, $j' = 1, \dots, n_{\hat{C}}$ are the boundary points obtained on the segmented and actual optic disc boundaries respectively. The shortest distance from each boundary point of the segmented optic disc to the boundary points on the boundary points of ground truth mask of the optic disc is calculated using Equation-3.2 and the shortest distance from each boundary point of the ground truth mask optic disc to the boundary points on the boundary points of the segmented optic disc is calculated using Equation-3.3. The mean of the obtained shortest distances is calculated using Equation-3.4.

$$m_{C,\hat{C}}(i') = \min_{j'} ||P_{i'} - Q_{j'}||_2 \quad (3.2)$$

$$m_{\hat{C},C}(j') = \min_{i'} ||Q_{j'} - P_{i'}||_2 \quad (3.3)$$

$$MAD = \frac{1}{2} \left\{ \frac{1}{n_C} \sum_{i'}^{m_C} m_{C,\hat{C}}(i) + \frac{1}{n_{\hat{C}}} \sum_{j'}^{n_{\hat{C}}} m_{\hat{C},C}(j) \right\} \quad (3.4)$$

The proposed framework consists of an off-line training step for learning global features and a transfer learning step for learning database-specific features. The network is trained from scratch using the fundus images of the extended MESSIDOR database and the trained model is fine-tuned by a few images of a particular database to learn database-specific features. Multi-scale features extracted in off-line training are concatenated with database-specific features for improved segmentation. The false positives are removed based on the geometrical features of salient regions in initial segmentation. The result is evaluated on six public fundus image databases such as DIARETDB0, DIARETDB1, DRIVE, AV-INSPIRE,

3.3 Results and Discussions

CHASE DB1 and MESSIDOR. The comparison of the proposed and competing techniques is provided in Table 3.1. The attention based fully convolutional network is robust and effective for the detection and segmentation of optic disc in the images affected by diabetic retinopathy and it outperforms existing techniques. The method is successful in optic disc localization and segmentation when tested on both dilated and non-dilated types of fundus images acquired from different medical centres. The performance of this algorithm does not degrade while handling images containing strong distractors like yellowish exudates which prove the effectiveness and robustness of the proposed process.

Table 3.1: Comparative result of optic disc segmentation

Method	Author	OM	MAD	SR
DIARETDB1	Proposed	0.92	1.98	100
	Roychowdhury <i>et al.</i> [74]	0.80	4.82	100
	Morales <i>et al.</i> [36]	0.82	2.88	100
	Salazar <i>et al.</i> [104]	0.76	6.38	96.7
	Welfer <i>et al.</i> [34]	0.43	8.31	97.7
DIARETDB0	Proposed	0.83	4.26	97.75
	Roychowdhury <i>et al.</i> [74]	0.78	4.91	-
DRIVE	Proposed	0.86	2.58	97.5
	Roychowdhury <i>et al.</i> [74]	0.81	5.01	100
	Morales <i>et al.</i> [36]	0.72	5.85	100
	Salazar <i>et al.</i> [104]	0.71	6.68	97.5
	Welfer <i>et al.</i> [34]	0.42	5.74	100
MESSIDOR	Proposed	0.92	1.95	99.92
	Roychowdhury <i>et al.</i> [74]	0.84	3.9	100
	Marin <i>et al.</i> [105]	0.87	6.17	99.75
	Giachetti <i>et al.</i> [72]	0.88	-	99.83
	Aquino <i>et al.</i> [35]	0.86	-	98.83
	Yu <i>et al.</i> [94]	0.83	7.7	99.08
CHASE_DB1	Proposed	0.81	7.89	100
	Roychowdhury <i>et al.</i> [74]	0.81	5.19	-
AV_INSPIRE	Proposed	0.83	4.63	100

3.4 Conclusion

Attention based fully convolutional network is developed for the segmentation of optic disc. Attention modules are a reasonably new approach used in object detection and classification. Most of the reported architectures have been designed to work well for object detection. The placement of attention modules in a fully convolutional network could enhance the features of different layers. The method is successful in optic disc detection and segmentation when tested on both dilated and non-dilated types of fundus images acquired from different medical centers. The performance of this algorithm does not degrade while handling images containing strong distractors like yellowish exudates which prove the effectiveness and robustness of the proposed process.

Artery-vein classification in
fundus image using Attention
based CNN

Artery-vein classification in fundus image using Attention based CNN

Automatic classification of the vessels in retinal fundus images into arteries and veins could help in the diagnosis of a wide range of cardiovascular disorders. Manual discrimination of arteries and veins is very difficult because they are very similar in many respects like appearance, contrast, and geometry. A novel framework for detecting the region of interest, extraction of vessel central-lines within the region of interest, and classifying the arteries and veins using an attention-based deep neural network named Attention AV-Net is presented in this paper. The proposed framework with an attention mechanism achieves high classification accuracy by focusing on informative features and suppressing irrelevant features. The suggested method has been evaluated using three public databases AV-DRIVE, AV-INSPIRE and AV-WIDE where it achieves accuracy of 96.7%, 97.2% and 92.4% respectively. The suggested method outperforms the other established advanced methods and could be useful for the development of computer-aided diagnosis of cardiovascular disorders.

4.1 Introduction

The fundus image which gives precise details of the human retina shows the optic disc, vessels, macula, fovea, exudate, microaneurysms etc. The method of capturing fundus images is relatively easy and inexpensive, as a result, there is a lot of potential for using the images in extensive screening programs and comparative statistical analysis. In diseases like diabetes mellitus, sometimes irregular blood vessel development gives early indications of retinal complications [106]. Hypertension [107] and other cardiovascular illnesses [108] are also linked to changes in retinal blood vessels. Factors like aging, problems in the cardiovascular system, diabetes, habits of smoking etc. can also lead to changes in blood vessels. Unusual changes are seen in the diameter of the vessels in retinal fundus images. Investigations by Rotterdam and Wisconsin studies [9], [109], have quantified the relationships between systemic factors and vascular parameters like diameters and AVR. Diseases like hypertension, high blood sugar, high cholesterol level, retinopathy [9] etc. are related to lower AVR values whereas higher AVR is associated with diseases like atherosclerosis, proliferative and diabetic retinopathy, Aortic valve calcification etc. A lower value of AVR can occur either due to narrowing of artery diameter or venular widening and the reverse happens in case of higher AVR. In clinical practice, manual estimating of these changes requires lots of patience and expertise and needs automation. The clinicians often refer to the ratio of the average arteriolar and average venous diameter of the retinal vessels

4.2 Reported Works on Artery-vein classification in fundus image

known as AVR within a specific annular region around the optic disc to determine the gravity of the diameters changes [2]. High-resolution fundus imaging is typically quick and inexpensive, as a result, large amounts of data are produced by any retinal screening programs. In order to automate large-scale retinal image processing, a comprehensive automatic system for classifying arteries and veins (A/V) is essential. The prior step for AVR computation is the classification of artery and vein and this paper deals with this classification using attention based deep neural networks.

4.2 Reported Works on Artery-vein classification in fundus image

Most of the automatic classical A/V classification methods [40] [41] [8] [42] [43] follow few common steps. It starts with preprocessing of the image then assigns artery-vein probability for each pixel, next forms the topological structure of the vessel and finally determines the label of the artery-vein. In the preprocessing step, issues like uneven illumination and background in-homogeneity are corrected using techniques like luminosity normalization and histogram equalization. A vessel binary map is generated once the retinal vessels have been segmented. Next, to assign the pixel-wise A/V probability, at first, all vessel centre-line pixels are analyzed to extract intensity-based features using which a probability is assigned to each centre-line pixel. Pixel-wise classification can be further improved by creating a topological structure of the vascular network. This structure basically establishes how each individual segment is connected. At last, using local and contextual information every section of the vessel is categorized as either an artery or a vein and eventually the entire vessel tree is labelled. Fan Huang et al. [44] proposed the categorization of artery and vein based on features like colours, and topological, geometrical and morphological properties of the vascular structure. Memari et al. [45] presented an artery-vein segmentation method by enhancing fuzzy c-means clustering and level sets. More the accuracy is achieved in segmentation, the more precise will be the A/V classification. Jebaseeli et al. [46] enhanced the Pulse Coupled Neural Network model using some parameter optimization to segment the vessel structure. Chen et al. [47] described topological connectivity for classifying the artery and vein, based on which a topology discriminator they proposed to enhance the accuracy of the classification.

Artery-vein classification in fundus image using Attention based CNN

For the classification task, of late many deep learning-based methods are being employed. Welikala et al. [48] proposed an artery-vein classification convolution neural network (CNN) model from retinal fundus images. The network presented by them is able to learn the existing complex features automatically instead of using handcrafted features made up of three convolution layers with three fully connected layers. Before classifying, the vessel must be segmented and the centre line extracted using the prescribed method. Lepetit-Aimon et al. [49] employed a fully convolutional network with a large receptive field for segmentation of the blood vessels for high-resolution images with minimum computation cost and comparatively low memory. The method's inability to accurately categorize blood vessels can be attributed to its ignorance of the vessels' true topological structure. Hemelings et al. [50] introduced a segmentation and classification approach based on the classical U-Net model. Girard et al. [51] employed deep learning techniques along with graph propagation for vessel segmentation and classification of arteries and veins. But for small diameter vessels and arteriovenous intersections, classification results were not up to the mark. For similar work, Ma et al. [52] has shown 94.5% pixel-wise accuracy using a multi-task neural network on the AV-Drive dataset. Wang et al. [53] presented a multi-task siamese network, that simultaneously performs segmentation and classification tasks to provide more robust deep features from the fundus images. Using the attention mechanism Chowdhury et al. [54] developed a multiscale encoder-decoder based attention network that can both segment and classify arteries and veins. IK Gupta et al. [55] applies the U-Net framework for segmentation and employs the mayfly optimization kernel extreme learning [56] method for classification. Sathananthavathi et al. [57] suggested a deep semantic segmentation architecture for AV classification. Hu et al. [58] tested a model for AV classification on the DRIVE and HRF dataset that employs an encoder-decoder segmentation network along with a point set classification maps of the AV skeleton. Toptas et al. [59] created artery vein patches separately after preprocessing the fundus image and fed them as input to their proposed deep learning network architecture to finally classify them whether the patches are of arteries or veins.

Distinguishing between arteries and veins in fundus images at the pixel level remains a difficult endeavour due to the significant resemblance between the two, poor contrast, and artery-vein intersection close to the optic disk. In the proposed framework, our contributions are as follows:

- The novelty of this work lies in the automated detection of ROI, extraction

4.3 Materials and method

of artery and vein patch based on vessel central, and development of AV-Net to classify artery and vein with acceptable accuracy.

- Multiple attention blocks have been added to generate different types of attention-aware features needed for attention residual learning of AV-Net for the classification of arteries and veins.
- The utilization of both residual learning and attention learning mechanisms has helped to improve the overall capability of the Deep Convolutional Neural Network.
- The proposed method is evaluated on three public databases such as AV-DRIVE, AV-INSPIRE and AV-WIDE. The attention mechanism helps to improve features at various levels by selectively highlighting important features and reducing the irrelevant features which leads to improved performance.

This paper contains the following sections: Section 4.3 contains a description of the publicly available datasets and a description of the proposed method, Section 4.4 describes the results and Section 4.5 contains the conclusion.

4.3 Materials and method

4.3.1 Materials

Three public datasets such as AV-DRIVE, AV-INSPIRE and AV-WIDE have been used for evaluating the proposed and competing techniques as described.

The AV-DRIVE is a set of forty high-quality fundus images of size [565 x 584] pixels which are used for research and development in the field of ophthalmology and artificial intelligence (AI). For AV-DRIVE, three individual human graders conducted manual labelling of vessel pixels, and subsequently, they cross-referenced their findings to determine the ultimate A/V labels. The progress of AI algorithms for the automated analysis and interpretation of fundus images is supported by the AV-DRIVE Fundus Image Dataset. It contains a diverse range of fundus images captured from different sources, including hospitals, clinics, and research institutions. These were obtained during a diabetic retinopathy screening program on an age group of 25-90 years in The Netherlands. The dataset is carefully curated to include images with various pathologies, anatomical variations,

Artery-vein classification in fundus image using Attention based CNN

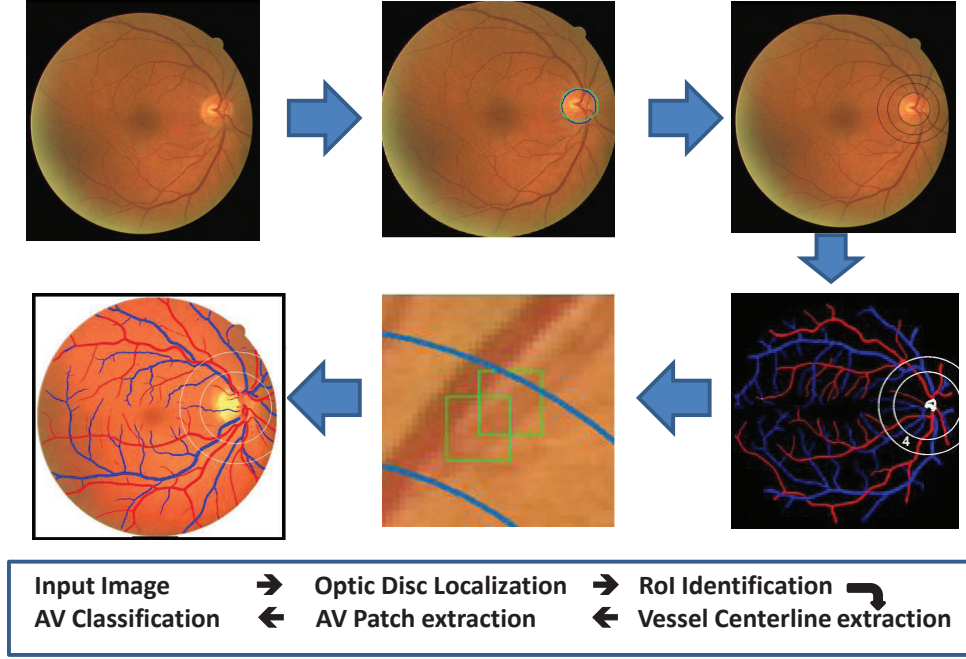


Figure 4.1: Framework for automated AV classification

and image quality characteristics to ensure its suitability for a wide range of research purposes. Each image in the AV-DRIVE dataset is accompanied by relevant metadata, such as patient demographics, clinical annotations, and imaging parameters, which provide valuable context for analysis and algorithm development. The dataset also includes ground truth annotations, which are expert-verified labels that indicate the presence or severity of specific eye conditions in the images. These notations function as benchmark criteria for assessing the effectiveness of AI algorithms.

The AV-WIDE dataset [110] contains 30 high-resolution, wide-field fundus [3900 x 3072] pixels images. These images were taken by the Duke University Medical Center, USA and these consist of some healthy eyes and some eyes infected with neovascular AMD and age-related macular degeneration (AMD). As part of post-processing, these images were cropped to half of their original size which helps to wipe out non-retinal areas, eyelashes etc.

The AV-INSPIRE Fundus dataset is a comprehensive collection of retinal fundus images curated for medical research and development in the field of ophthal-

4.3 Materials and method

mology. Fundus images capture the posterior segment of the eye which provides valuable insights into various ocular diseases and conditions. This database comes from the INSPIRE-AVR dataset, which has 40 images [2392 X 2048] pixels. The dataset contains a wide range of fundus images, including those obtained from both healthy individuals and patients with eyes infected with diseases e.g. diabetic and hypertensive retinopathy, macular degeneration due to age, and glaucoma among others. These images are captured using specialized equipment, such as fundus cameras, which enable high-resolution imaging of the retinal structures.

4.3.2 Method

The general approach an ophthalmologist utilizes for AVR computation are detection of ROI, separation of arteries and veins in ROI and computation of AVR. The current study focuses on the classification of artery-vein in the ROI. Figure 4.1 displays the proposed method's block diagram. Inside the retina, the optic disc is a bright round-shaped area. Major vessels can be detected surrounding the optic disc because all of the vessels originate from this area. These vessels become thinner as their distance increases from the optic disc. Major vessels are mainly the area of interest for most of the research work.

4.3.2.1 Determination of ROI

Faster R-CNN network has been employed for the identification of the optic disc within the retinal fundus image. The said network tries to localize the object in those proposed regions with the help of a region proposal network and detector. The fundus images are subjected to resizing and gamma correction to improve their quality. The bounding box generated by the regression head of Faster R-CNN is employed to locate the optic disc. Each bounding box on an input image shows the potential position of the optic disc, whereas the actual location is indicated by the bounding box with the highest probability value. It is considered that the optic disc's centre is in the middle of the bounding box. The region of interest lies between the annular region that falls under two concentric circles of radius 1 and 1.5 times the diameter of the optic disc. The analyses and assessments conducted in this study rely on measurements taken within the annular region as shown in Figure 4.2.

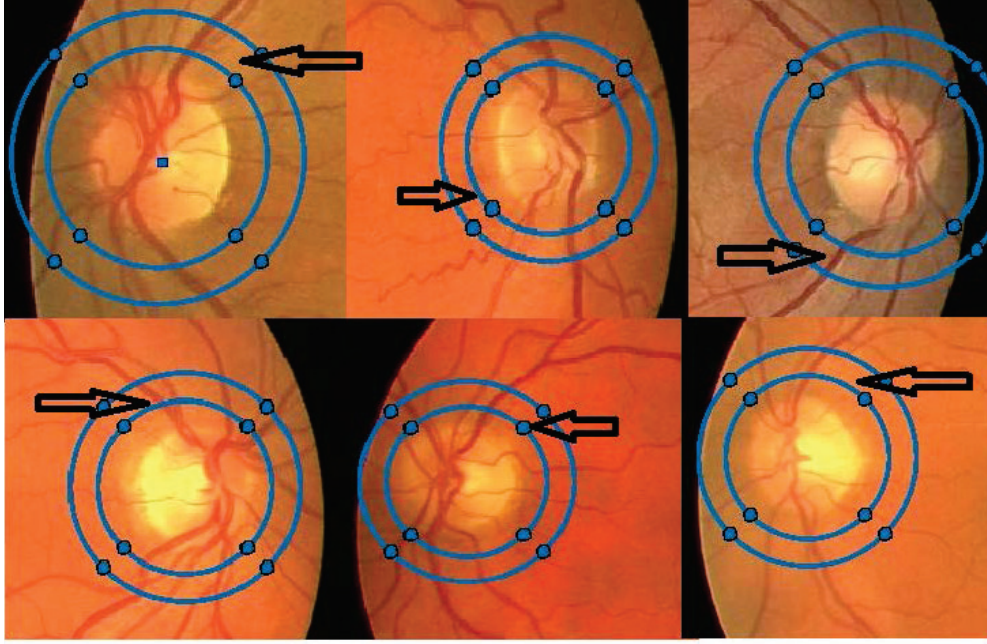


Figure 4.2: Determination of RoI (Region of Interest) for AVR computation. The annular regions around the optic disc pointed out by the arrows are the RoI.

4.3.2.2 A/V Patch Extraction

Arteries carry oxygenated blood and veins carry deoxygenated blood. As a result, compared with veins the arteries appear brighter and slimmer, and the interior region of vessels exhibits the central reflex. In addition, the width of the vessels as well as its colour also helps to trace them. For thinner or smaller vessels, it depends on structural characteristics at intersections where crossing of two different types of vessels happens and bifurcation points where they branch out in the vascular network. Bifurcation points are where a vessel splits into two smaller ones. A few other important features ophthalmologists use for artery vein classification are: (i) arteries do not intersect with other arteries, it is true for veins too. Therefore, of the two vessels at any cross-over location, one is an artery and the other is a vein (ii) The blood vasculature system has a binary tree-like structure, thus all three vessel segments at any T-junction are of the same type. For each dataset such as AV-DRIVE, AV-INSPIRE, and AV-WIDE, the following steps have been followed.

- Then artery and vein mask is extracted for all the images based on the ground truth information provided by ophthalmologists Figure 4.3.

4.3 Materials and method

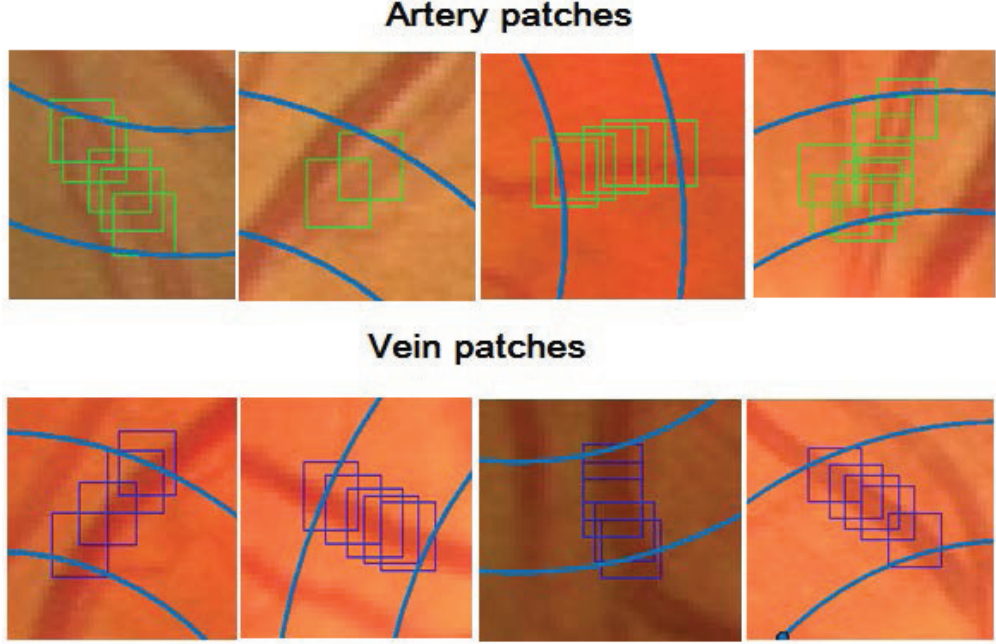


Figure 4.3: Artery-vein Patches

- Using the artery and vein mask, centre-line coordinates are determined.
- Based on the artery and vein centre-line coordinates that are received, separate artery and vein patches are created of size 32x32.

4.3.3 A/V Discrimination

4.3.3.1 Attention Mechanism

This mechanism operates much like the human perception process, employing higher-level information to steer the forward flow of lower-level data. At first, the attention mechanism in natural language processing [111] [112]. Attention mechanisms have also utilized for crafting medical reports [113] [114] and combined image-text classification [115]. Attention blocks are utilized to highlight the relevant information and suppress the irrelevant information and refine the feature map and used for image categorization [99] [116], [117], segmentation [118], and captioning [119] [120]. The channel attention looks for semantic information (what), with each channel assigned with specific weights. The spatial attention concentrates on specific locations (where) and the feature map is updated with corresponding soft weights. According to Jetley et al. [99], attention blocks learned

Artery-vein classification in fundus image using Attention based CNN

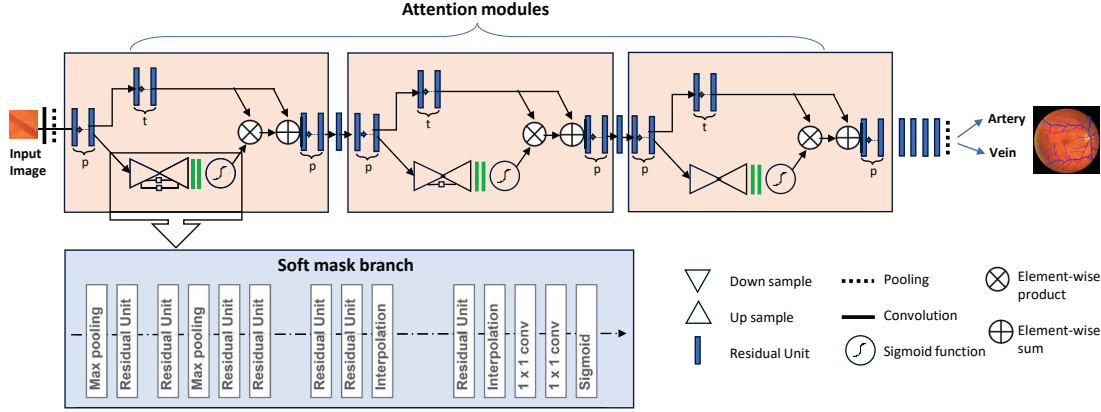


Figure 4.4: Block Diagram of AV-Net

during the full network’s end-to-end training, improve performance.

Hu et al. [121] introduced the Squeeze-and-excitation (SE) module, where channel-wise features are corrected based on finding non-linear correlations between channels. The modules for spatial and channel attention can be integrated either adding or concatenating. Long-range contextual dependencies are captured in dual attention network [122] by fusion of spatial and channel attention. The channel and spatial attention modules are cascaded in convolutional block attention module [101].

4.3.3.2 Network Architecture

The architecture presented in Figure 4.4 has been designed based on residual attention network [116]. The network is constructed with multiple attention blocks to generate different types of attention-aware features. The attention mechanism enables the network to improve features at various levels by selectively amplifying important features while reducing the influence of irrelevant ones. The attention mask serves as a feature selector in the forward pass. Residual trunk and attention mask branches are integral components in training the network through attention-based residual learning. The purpose of the trunk branch (Figure 4.4) is to extract features and the mask branch is used to generate corresponding weights to refine the features by element-wise multiplication and added with trunk branch features. The trunk branch is constructed with two residual blocks in each stage. The mask branch is constructed with a bottom-up and top-down block, where the bottom-up block consists of residual blocks, max-pooling layers and the top-down

4.3 Materials and method

block consists of residual blocks and up-sampling layers. The skip connections are used between bottom-up and top-down blocks for efficient merging of high-resolution features with low-resolution features. The input image is operated with five residual blocks and the feature map is refined by three attention stages and finally classified as artery or vein.

4.3.3.3 Training of Network

The proposed network and competing methods are implemented in PyTorch and trained using a workstation with 24 GB GPU for up to 150 epochs. During the experiment, the learning rate that has been considered is 0.00001. The training of a deep neural network relies on the crucial contribution of optimizers. There are many popular optimizers used during training like Adam, RMSProp, Stochastic gradient descent etc. We opted for the Adam optimizer due to its ease of implementation, computational efficiency, minimal memory requirements, resilience to diagonal rescaling of gradients, and suitability for handling large-scale problems involving extensive data and/or parameters. Adam optimizer has been used during the network training by reducing cross-entropy loss to quantify the difference between the targets and predictions. The objective is to minimize this loss function, which is measured by $l(f(x_i), y_i)$ where (x_i, y_i) represents the sample input-output pair. For example, in the case of Binary Classification, cross-entropy is defined in Equation A.1:

$$-\frac{1}{N} \sum_{i=1}^N (y_i \log(p_i) + (1 - y_i) \log(1 - p_i)) \quad (4.1)$$

where p_i and y_i (0 or 1 in the case of binary classification) are the predicted probability and indicator of the i^{th} case respectively and N is the number of samples. A simple extension to a Multi-class Classification (say C classes) problem exists as follows Equation A.2:

$$-\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C (y_{ij} \log(p_{ij})) \quad (4.2)$$

where p_{ij} is the predicted output for the i^{th} sample and j^{th} class and y_{ij} is the true output for the i^{th} sample and j^{th} class.

4.4 Results and discussion

This paper evaluates the artery-vein classification performance of the proposed and competing methods using three publicly available datasets: AV-DRIVE, AV-WIDE and AV-INSPIRE using sensitivity Equation 4.3, specificity Equation 4.4, precision, accuracy Equation 4.5, and YI-score Equation 4.6 as mentioned below.

$$Sensitivity(Se) = \frac{TP}{TP + FN} \quad (4.3)$$

$$Specificity(Sp) = \frac{TN}{TN + FP} \quad (4.4)$$

$$Accuracy(Acc) = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.5)$$

$$YI - Score = 1 - (Sensitivity + Specificity) \quad (4.6)$$

here TP, TN, FP and FN means true positives, true negatives, false positives, and false negatives respectively. Three evaluation metrics namely sensitivity, specificity and accuracy are used to assess the effectiveness of the proposed method in artery-vein patch classification. YI(Youden's index) is a widely recognised diagnostic assessment that evaluates the effectiveness of a dichotomous test. In the case of YI, both sensitivity and specificity are considered to compute the trade-off between them. The value of YI ranges from 0 to 1, when both sensitivity and specificity are maximum it is 1, when they are both least it is 0.

Table 5.1 presents a comparative analysis of the proposed and other established methods for AV-DRIVE, AV-INSPIRE and AV-WIDE. We used a three-fold cross-validation approach to assess performance. It has shown better accuracy of 96.7%, 97.2% and 92.4% for AV-DRIVE, AV-INSPIRE and AV-WIDE respectively when compared with other approaches. The specificity was recorded as 98.0% for AV-DRIVE, 97.3% for AV-INSPIRE and 91.1% for AV-WIDE which is better than competing methods. The proposed model obtained sensitivity of 95.3% for AV-DRIVE, 97.5% for AV-INSPIRE and 93.7% for AV-WIDE. The YI-scores are 84.8% for AV-WIDE, 93.3% for AV-DRIVE, and 94.8% for AV-INSPIRE, respectively. It is observed that the sensitivities and YI-scores of the proposed network outperform other established methods.

The test performance of our method on the AV-DRIVE, AV-INSPIRE and AV-WIDE datasets has been shown using the ROC curves in Figure 4.5, Figure 4.6

4.4 Results and discussion

Table 4.1: Comparative study of the proposed and competing methods

Paper	AV-DRIVE				AV-INSPIRE				AV-WIDE			
	Se(%)	Sp(%)	Acc(%)	YI(%)	Se(%)	Sp(%)	Acc(%)	YI(%)	Se(%)	Sp(%)	Acc(%)	YI(%)
Estrada et al. [123]	93.0	94.1	-	87.1	91.5	90.2	-	81.7	91.0	90.9	-	81.9
Zhao et al. [124]	94.2	92.7	91.1	86.9	96.8	95.7	95.1	92.5	96.2	94.2	91.0	90.4
Srinidhi et al. [125]	95.0	91.5	93.2	86.5	96.9	96.6	96.8	93.5	92.3	88.2	90.2	80.5
Our method	95.3	98.0	96.7	93.3	97.5	97.3	97.2	94.8	93.7	91.1	92.4	84.8

Artery-vein classification in fundus image using Attention based CNN

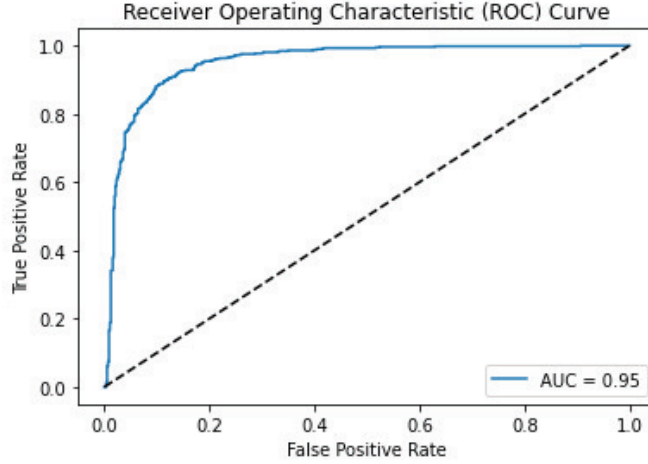


Figure 4.5: ROC Curve for AV-DRIVE Dataset

and Figure 4.7 respectively. These curves represent the effectiveness of the network for several operating points. In ROC, sensitivities are plotted with respect to 1-specificity. The single numerical value provided by the area under the curve (AUC) quantifies the classifier's overall performance. An ideal classifier achieves an AUC of 1, whereas a random classifier attains an AUC of 0.5. The closer the AUC is to 1, the better the classifier. In our case, it achieves the AUC values of 0.95, 0.94 and 0.93 for AV-DRIVE, AV-INSPIRE and AV-WIDE data sets respectively.

4.5 Conclusion

In our research, we have proposed a novel Convolutional Neural Network (CNN) utilizing the Resnet architecture, coupled with an Artery-vein discriminator, to classify retinal vessels. This model consists of multiple Attention Residual Learning Block to discriminate Artery from Vein. Here the approach uses the benefit of both the residual learning as well as attention learning mechanisms to enhance the nature of the Deep convolution neural network's capacity. We evaluated this network model on the AV-INSPIRE, AV-DRIVE, and WIDE datasets. Our findings demonstrate that our model can achieve superior results in artery-vein classification by adaptively focusing on the distinguishing aspects of artery-vein characteristics within the fundus image. We intend to expand on this research by looking at unsupervised attention learning and classification.

4.5 Conclusion

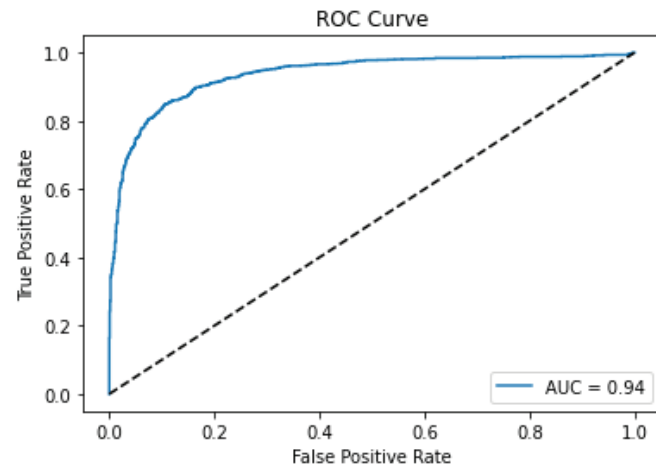


Figure 4.6: ROC Curve for AV-INSPIRE Dataset

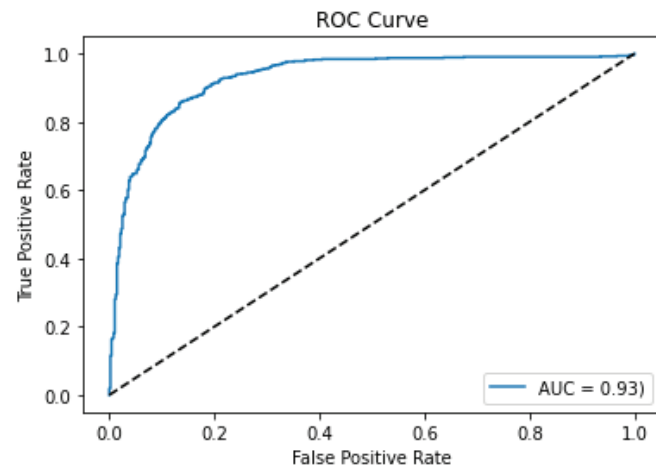


Figure 4.7: ROC Curve for AV-WIDE Dataset

CHAPTER 5

Computation of Arteriolar-to-venular (AVR) width ratio

Computation of Arteriolar-to-venular (AVR) width ratio

Accurate estimation of the Artery-Vein width ratio (AVR) in a fundus image is a very important step in the diagnosis of many diseases like diabetic retinopathy, glaucoma, hypertension or cardiovascular pathologies. The accuracy of AVR is highly dependent on Artery-vein (AV) classification around the close region of the Optic Disc known as Region-of-interest. This paper presents an automated approach for measuring the AVR in retinal fundus images. In our suggested process the steps are identifying the optic disc's location, segmenting it, determining the region of interest, classifying arteries/veins using an Attention-based AV net, and ultimately calculating the Arteriole-to-Venule Ratio. Evaluation of this method was conducted using a publicly available INSPIRE-AVR image dataset. We have reported the mean error and the correlation coefficient which are comparable to the AVR values mentioned as reference.

Ocular diseases, AVR, Retinal fundus image, Arteries and veins classification, Convolutional neural network, Attention mechanism, cardiovascular disorders

5.1 Introduction

Analysing the health status of microvasculature can provide an early alert to the person not only for significant cardiovascular diseases as well as for some systemic diseases like diabetic retinopathy[126], glaucoma[127], retinopathy of prematurity etc. which can lead to blindness[128],[129]. All these diseases affect the vascular width of segments or the overall length and even sometimes the structure. A 2D retinal fundus image which captures the back of the eye, is a cost-effective method of allowing a doctor to directly view a patient's blood vessels. It is captured using a non-invasive device called funduscope. Though eye physicians appreciate this fact but manually measuring the artery-vein width using these images is a very cumbersome and laborious process. Estimating AVR values manually poses a challenging endeavour, and nowadays most medical applications employ semi-automatic approaches to compute these index values. Detecting these diseases early can save a substantial economic burden and prevent a person from early blindness.

So the development of an automated technique for calculating the AVR might significantly influence clinical practice and result in a better evaluation of individuals having brain and cardiovascular illnesses. Any robust automatic AVR estimation system requires optic disc (OD) detection for region of interest (RoI) delineation and artery-vein classification with high accuracy as even minor classification errors can significantly impact the final AVR outcome.

5.2 Reported Works on Arteriolar-to-venular (AVR) width ratio computation

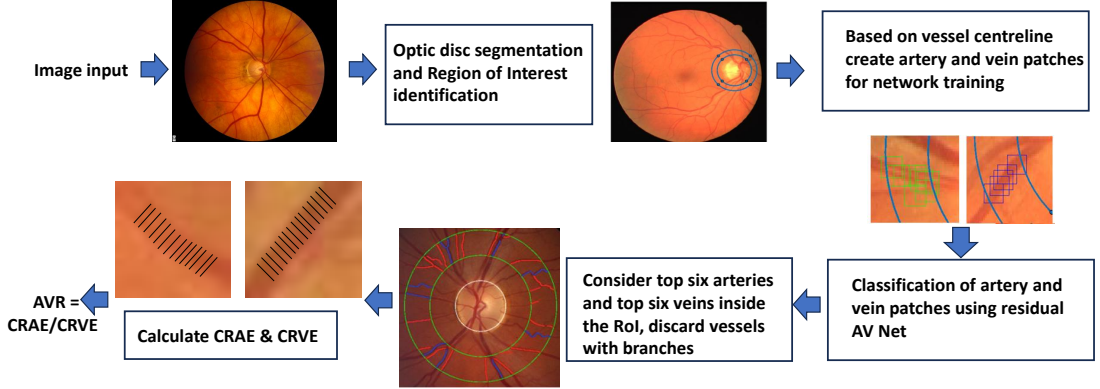


Figure 5.1: Framework for automated AV classification and AVR calculation

5.2 Reported Works on Arteriolar-to-venular (AVR) width ratio computation

Several approaches to automatically calculate AVR have been previously introduced [40] [60] [61]. The main difference in those studies is the different approaches to artery-vein classification. In 2003, the automatic classification of AV was first published in a work by Grisan et al. [62]. Ruggeri et al. [60] introduced a technique where they first located the OD, identified RoI around the OD and finally classified the vessel pixels as an artery or vein inside the RoI. Tramontan et al. [61] published an extended version of this work where they changed both the vessel tracking mechanism and the features for discriminating arteries and veins structure. Niemeijer et al. [40] used a linear classifier for vessel classification and calculated AVR using Knudtson’s revised formula [2]. Khanal et al. [63] in their recently devised a fully automatic method to calculate AVR in fundus image after classifying artery-vein on a complete vascular network.

In this paper, we have presented a fully automatic method for AVR calculation that includes the steps - OD segmentation, RoI detection, Vessel centre-line detection, AV classification using Attention-based AV net and finally AVR calculation.

The structure of the paper is outlined as follows: Section II introduces the proposed approach for estimating AVR. Section III details the outcomes of tests conducted on images from the INSPIRE-AVR database. Lastly, in Section IV, the conclusions drawn from this work are summarized.

5.3 Materials and Method

5.3.1 Materials

Our proposed method has been evaluated on the publicly available INSPIRE-AVR fundus image dataset. This dataset is a comprehensive collection of retinal fundus images curated for medical research and development in the field of ophthalmology. This dataset has 40 images $[2392 \times 2048]$ pixels. The dataset contains a wide variety of fundus images, including both individuals without health issues and those with eyes affected by diseases. infected with diseases e.g. diabetic and hypertensive retinopathy, macular degeneration due to age, and glaucoma among others. Fundus cameras, a specialised equipment that allows for high-resolution imaging of the retinal structures, are used to collect these images.

5.3.2 Method

We have followed the following steps for AVR computation - identification and segmentation of the optic disc, RoI detection, vessel centre-line extraction, classification of arteries and veins patches in the RoI and computation of AVR. Classification of artery-vein in the RoI is one of the main areas of this study. Vessel calibers present in the RoI are measured for which detection of OD is required. Figure 5.1 displays the proposed method's block diagram. Inside the retina, the optic disk is a bright round-shaped area. Major vessels can be detected surrounding the optic disc because all of the vessels originate from this area. As the vessels move away from the optic disc, it become thinner. Mainly, major vessels are the area of interest for most of the research work.

5.3.2.1 Optic disc localization in fundus image

The error-free measurement of AVR is a very challenging task and the first step of the whole workflow is optic disc detection. The identification of the optic disc is essential for determining the specific location of the region of interest (RoI). In our work, we have used Faster R-CNN [93] based Deep neural networks that can speed up the optic disc detection. The network is made up of two modules - the first module is a deep convolutional network is there to propose regions and the second module is a Fast RCNN detector that uses those proposed regions. When these two modules are combined it works as an object detection model. This design has led to a significant advancement in the object detection model because

5.3 Materials and Method

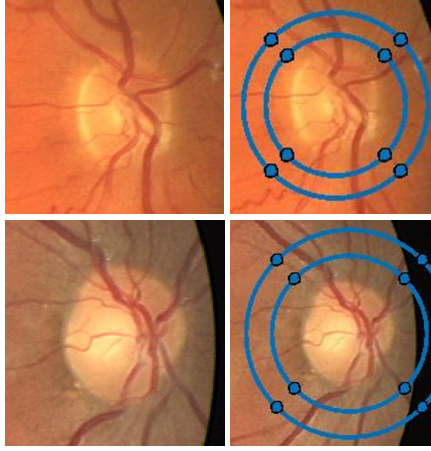


Figure 5.2: Original image in the left column and corresponding Region of interest for AVR (delimited by the two blue circles) in the right column

before this it was only possible to determine the presence of any specific object in an image but this model helps to determine the exact location of the object. The bounding box generated by the regression head of Faster R-CNN is employed to locate the optic disc. Each bounding box on an input image shows the potential position of the optic disc, whereas the bounding box with the highest probability value indicates the actual location. It is considered that the optic disc's centre is in the middle of the bounding box.

5.3.2.2 Optic disc segmentation in fundus image

Attention Based Fully Convolutional Neural Network has been utilized in our work for optic disc segmentation. The attention modules help to increase the performance of segmentation for natural images. It also helps to suppress false-positive regions and to highlight informative regions. The spatial attention provides the location of the features, while channel attention utilises the features that can be found in the available channels. The attention module helps to focus on relevant features to improve the segmentation results. The morphological operation and shape features are used to eliminate the false positives. This framework aims to create an attention-based deep network designed for identifying informative regions within fundus images, such as the optic disc and selecting informative images to facilitate rapid and efficient transfer learning. The channels containing high-level features could be improved by channel attention whereas the spatial attention could enhance the spatial connectivity of features.

Computation of Arteriolar-to-venular (AVR) width ratio

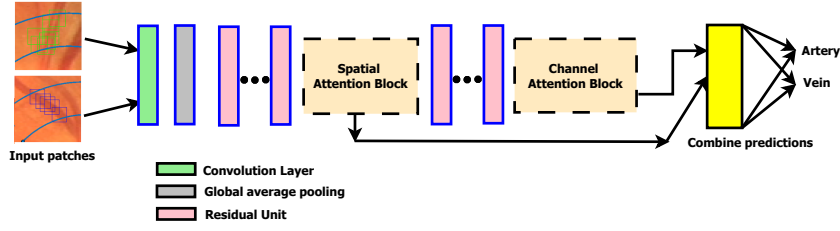


Figure 5.3: Network diagram for Artery-vein classification

5.3.2.3 Determination of RoI and AV Patch creation

Once the optic disc segmentation is complete, the region of interest lies between the annular region that falls under two concentric circles within 0.5 and 1.0 times the diameter of the optic disc, refer to Figure 5.2. The analyses and assessments conducted in this study rely on measurements taken within the annular region. For every image in the dataset, the artery and vein mask is extracted based on the ground truth information provided by ophthalmologists. Using the artery and vein mask, centre-line coordinates are determined. Based on the artery and vein centre-line coordinates that are received, separate artery and vein patches are created of size 32×32 .

5.3.2.4 Artery-Vein classification

We have used a residual attention network, refer to Figure 5.3, for the classification task where multiple attention blocks are used to generate different types of attention-aware features. The attention mechanism helps to improve features at various levels by selectively highlighting important features and reducing irrelevant features. The channel attention looks for semantic information (what), with each channel assigned with specific weights. The spatial attention concentrates on specific locations (where) and the feature map is updated with corresponding soft weights. The utilization of both residual learning and attention learning mechanisms has helped to enhance the overall capability of the Deep Convolutional Neural Network. The input image is operated with five residual blocks and the feature map is refined by three attention stages and finally classified as artery or vein. The proposed network and competing methods are implemented in PyTorch and trained using a workstation with 24 GB GPU for up to 150 epochs. During the experiment, the learning rate is considered 0.00001 because in this setting we have received the best performance. Adam optimizer has been used during the

5.4 Results and Discussion

network training by reducing cross-entropy loss to quantify the difference between the targets and predictions. A 3-fold cross-validation approach was used during training and testing.

5.3.2.5 AVR computaion

Once the vessel patches have been identified as arteries or veins, the distance between each vessel pixel and the nearest boundary point, d , is determined by applying a distance transform to the vessel segments inside the patches. Subsequently, the vessel caliber value is estimated for every pixel along the vessel centerline using the formula $2d-1$. The AVR value is computed based on the calibers of the vessels within the defined region of interest (RoI).

The AVR is determined by the ratio of CRAE (Central Retinal Artery Equivalent) and CRVE (Central Retinal Vein Equivalent). First, a set of the six largest arteries is chosen to calculate CRAE. The largest and smallest vessels in this group are then paired using an algorithm known as Knudtson’s formula [2] for the parent trunk width determination. In the event of insufficient measurement points, fewer widths overall may be employed. The parent trunk is then used to generate a new set of arteries replacing the largest and smallest arteries, and the process is continued until only one vessel is left, the width of which corresponds to the CRAE value. The CRVE is computed following the same method considering the width of the six largest veins.

5.4 Results and Discussion

This paper evaluates the artery-vein ratio calculation of our proposed and competing methods on the INSPIRE-AVR dataset. We need to perform the steps in the methods described above. The assessment of optic disc detection performance involves measuring the Success Rate (SR) that represents the percentage of images in which the optic disc’s centroid is successfully localized inside the boundary of the ground truth mask of the same. We have received 100% SR for this dataset. We have used two metrics to assess how well the optic disc segmentation worked. First is the Overlap Measure (OM) that shows the proportion of the overlapping area between the real and segmented optic disc. Second is the Mean Absolute Distance (MAD) which denotes the average of the minimum distances measured from the actual optic disc’s boundary to the segmented optic disc’s boundary. In our case, we have received 0.83 as OM value and 4.63 as MAD value. Using the resid-

Computation of Arteriolar-to-venular (AVR) width ratio

ual attention network, we have received 97.2% accuracy for the INSPIRE-AVR dataset images. The specificity and sensitivity have come to 97.3% and 97.5% respectively. Finally, for AVR values the ground-truth data is available for this dataset where two human graders have measured the AVR. We can achieve a mean error of 0.05 ± 0.05 with respect to the Observer-1 reference value and 0.05 ± 0.04 with respect to the 2nd grader. The comparison result with both the sets has been shown in Table 5.1. Our result is similar to the one of Observer-2 and is smaller than the error of the recent approach presented by Niemeijer et al. [130].

Table 5.1: Comparison of AVR values for all images of the INSPIRE-AVR dataset

Method Name	Mean AVR value (Stdev)	Mean error (Stdev)
Observer-1 (Ref)	0.67 (0.08)	-
Observer-2	0.66 (0.08)	0.05 (0.04)
Niemeijer et al. [130]	0.67 (0.07)	0.06 (0.04)
Khanal et al. [63]	0.64 (0.09)	0.07 (0.06)
Proposed method	0.67 (0.07)	0.05 (0.05)

5.5 Conclusion

In our research, an attention network for the artery-vein classifier has been utilized and we have calculated AVR on the INSPIRE-AVR dataset which shows promising results. Thus we can say this process has the potential to be used in clinical applications. In future, we shall emphasize the development of a better attention module for improved accuracy for artery-vein classification and subsequent improvement in AVR computation.

Conclusion

6.1 Summary of Studies

The Arteriolar-to-Venular Ratio (AVR) serves as an indicator utilized in the early detection of conditions like diabetes, hypertension, or cardiovascular disorders. In our work, we have presented fully automated methods for assessing the Arteriolar-Venular Ratio (AVR) measurement in retinal images. Our whole work comprises of few methods optic disc detection or segmentation, determination of the region of interest, classification of vessels into arteries and veins, and ultimately, AVR calculation. We have provided an overview of our research in this chapter and laid the groundwork for future uses of deep learning techniques in AVR calculation.

In **Chapter 2**, we have discussed optic disc localization in retinal fundus images. We have employed Faster R-CNN method for this work. We know R-CNN method introduced in 2013 comprises of region proposal plus CNN. There selective search is used for localization where the probability of object existence is highest. These set of region proposals are then fed into a trained deep learning CNN to extracts a feature vector. The vector is fed into a set of linear SVMs for the classification task. There is a bounding box regressor to get the most accurate coordinate. Then in 2015, Fast R-CNN came to overcome drawbacks of R-CNN like

Conclusion

Multiple stages were required for the training like ConvNets to SVMs to bounding box regressor. Further improvement happens in the Faster R-CNN architecture which consists of two module. The first module is a deep convolutional network to propose regions and second module is Fast R-CNN detector. Our proposed method shows satisfactory robustness on both normal and images affected by diabetic retinopathy. It outperforms many previous methods in terms of speed with satisfactory accuracy of optic disc localization. We tested accuracy of optic disc localization on AV-DRIVE, AV-INSPIRE and AV-WIDE Datasets where we have achieved 100% accuracy. We also measured OD localization accuracy in MES-SIDOR dataset where we have achieved 98.75% accuracy outperforming existing methods.

In **Chapter 3**, we presented simultaneous optic disc detection and segmentation in fundus image using attention based fully convolutional network. The network consists of four blocks of convolution layers and three attention blocks are added to refine the feature map of last three blocks. In the convolution block attention module (CBAM), a channel attention module is applied on a feature map. The result is further refined by a spatial attention module. The architecture focuses on both channel and spatial attention to learn about what and where about features. The features extracted from the last three blocks are refined by CBAM and final feature map is obtained by depth wise concatenation of refined features. We reported Overlap measure (OM), Mean absolute distance (MAD) and Success rate (SR) on DIARETDB0, DIARETDB1, DRIVE, AV-INSPIRE, CHASE DB1 and MESSIDOR datasets.

In **Chapter 4**, we discussed artery-vein classification in fundus image using attention based CNN. In this method, first artery and vein mask is extracted for all the images based on the ground truth information provided by ophthalmologists. Then Using the artery and vein mask, centre-line coordinates are determined. Based on the artery and vein centre-line coordinates that are received, separate artery and vein patches are created of size 32x32. Attention mechanism operates much like the human perception process, employing higher-level information to steer the forward flow of lower-level data. Attention blocks are utilized to highlight the relevant information and suppress the irrelevant information and refine the feature map and used for image categorization. The channel attention looks

6.2 Contributions of the Thesis

for semantic information (what), with each channel assigned with specific weights. The spatial attention concentrates on specific locations (where). The modules for spatial and channel attention can be integrated either adding or concatenating. The architecture has been designed based on residual attention network. Our network is constructed with multiple attention blocks to generate different types of attention aware features. The attention mechanism enables the network to improve features at various levels by selectively amplifying important features while reducing the influence of irrelevant ones. The input image is operated with five residual blocks and the feature map is refined by three attention stages and finally classified as artery or vein. Adam optimizer has been used during the network training by reducing cross-entropy loss. We reported sensitivity, specificity, accuracy and Y1 score of classification on Av-DRIVE, AV-INSPIRE and AV-WIDE datasets.

In **Chapter 5**, we calculated state-of-the art A-V ratio on the INSPIRE-AVR dataset. Top six wide arteries and six wide veins within 1D to 1.5D from the disc center has been extracted. AV ratio has been calculated using the iterative method proposed by Knudtson et al. using the formula : $AVR = CRAE(A)/CRVE(V)$ where A and V are list of width of arteries and veins in descending order, CRAE being Central Retinal Artery Equivalent, and CRVE as Central Retinal Vein Equivalent. We have compared this result with the work of Niemeijer et al.[40] and Khanal et al.[63].

6.2 Contributions of the Thesis

The contributions of the thesis towards the development of a self-learning tool are summarized as follows.

- Implementation of Faster R-CNN method for optic disc localization.
- The success of convolutional neural network in object segmentation has motivated us to investigate the performance of fully convolutional network for optic disc detection and segmentation. Attention module helps to focus on relevant features to improve the segmentation results. Channel and spatial

Conclusion

attention combine both the spatial context as well as the semantic information of optic disc. We have presented attention based fully convolutional network for optic disc detection and segmentation and suppress false-positive regions and to highlight informative regions, we have developed attention based fully convolutional network. The method is successful in optic disc segmentation when tested on both dilated and non-dilated types of fundus images acquired from different medical centres. The performance of this algorithm does not degrade while handling images containing strong distractors like yellowish exudates which prove the effectiveness and robustness of the proposed process.

- In our Artery-vein classification in fundus image work we have used a novel CNN (Convolutional Neural Network) framework utilizing the Resnet architecture for the detection of the region of interest (ROI), extraction of vessel central-lines within ROI, and development of AV-Net to classify artery and vein with acceptable accuracy. Multiple attention blocks are used to generate different types of attention-aware features. The attention mechanism helps to improve features at various levels by selectively highlighting important features and reducing irrelevant features. Here the approach uses the benefit of both the residual learning as well as attention learning mechanisms to enhance the nature of the Deep convolution neural network's capacity.

6.3 Future Scope

The current study opens up opportunities to further explore the ideas proposed in this thesis, paving the way for the creation of a diagnostic tool for AVR measurement. Potential future lines of investigation related to the thesis are listed below:

- We intend to expand on this research by looking at unsupervised attention learning and classification.
- Improvement of algorithm for handling noisy images

Appendix

A.1 Deep Learning Fundamentals

The deep learning-based applications started gaining the attention of the universe from 2012 onwards though the main ideas of deep learning for computer vision such as convolutional neural networks and backpropagation were known to researchers at the end of 1980s. There are three areas which have helped deep learning to flourish are - Hardware, Datasets and benchmarks, Algorithmic advances

A.1.1 Hardware

Hardware advancements have played a crucial role in the flourishing of deep learning, enabling the development and deployment of increasingly complex neural network models. Typical deep learning models for speech recognition and computer vision demand much more parallel processing power than any commonly available server. Graphics Processing Units (GPUs) are well-suited for this parallel processing, as they consist of thousands of cores that can handle multiple tasks simultaneously. This allows for a significant acceleration of training times compared to traditional central processing units (CPUs). After that tensor processing unit (TPU) came into the field, which is a brand-new chip architecture created to run deep neural networks and it is ten times faster and significantly more energy-efficient than the best GPUs. Next, distributed computing architectures, which involve connecting multiple computing devices to work collaboratively, have become essential for training large deep learning models. This approach allows for

the distribution of computational tasks across multiple GPUs or even across different machines, facilitating faster training times and increased model complexity. In essence, hardware advancements have provided the computational muscle needed to tackle the complexities of deep learning models. These advancements have not only accelerated the training and deployment of existing models but have also paved the way for the exploration and development of increasingly sophisticated and capable neural networks.

A.1.2 Datasets and benchmarks

Datasets are essential to the growth of deep learning. Because they provide the information needed for model evaluation and training. The development of the internet has changed everything by making it possible to gather and share very big datasets for machine learning. Large organizations now operate with datasets that could not have been gathered without the internet, including image, video, and natural language information. The ImageNet dataset, which consists of 1.4 million images that have been manually annotated with 1,000 image categories (one category per image), is arguably the dataset that has contributed most to the development of deep learning. Kaggle is another popular dataset used by the online community of data scientists and machine learning practitioners. In essence, datasets serve as the foundation upon which deep learning models are built. The quality, diversity, and size of datasets significantly influence the success and generalization capabilities of deep learning models, ultimately contributing to the ongoing evolution and flourishing of the field.

A.1.3 Algorithmic advances

Algorithmic advances have played a crucial role in facilitating the flourishing of deep learning. At the beginning, shallow methods such as SVMs and random forests were mostly used. The main drawback of these methods was as the number of layers increased, the feedback signal that was utilized to train neural networks used to disappear. With the introduction of several improved algorithmic advancements gradient propagation started giving much better performance. Among them, activation functions, such as Rectified Linear Units (ReLU), have addressed issues like vanishing gradients and enabled the training of deeper networks more effectively; better optimization techniques like RMSProp and Adam, helps deep learning models learn from data and improve their performance over time. Transfer learning leverages pre-trained models on large datasets for a specific task and fine-tunes them for a new, related task with limited data. This algorithmic approach allows the transfer of knowledge gained from one domain to

A.2 Building blocks of Deep learning network

another, making deep learning more accessible in scenarios with limited labeled data. During the last few years, with the advancement of new techniques like batch normalization, residual connections, and depthwise separable convolutions, we can now train thousands of layer-deep models from scratch.

A.2 Building blocks of Deep learning network

Deep learning networks are composed of several fundamental building blocks, each serving a specific purpose in the modeling and learning process. Here are the basic building blocks of a typical deep learning network:

A.2.1 Neural network

The building blocks of deep learning are neural networks. Deep learning networks can be defined as neural networks with more than three layers, which would include both the inputs and the output. A simple neural network consists of simply two or three layers. The word "deep" basically refers to the depth of layers in a neural network. Layers are the fundamental unit of a neural network that consists of multiple neurons. It is made up of an input layer that is used to receive the initial input data, an output layer that produces the final output of the network, and a few hidden layers that perform a computation (usually a weighted sum). The number of hidden layers varies depending on the problem. Every node, that is connected to one another has a weight and threshold. When the output of a particular node exceeds the designated threshold value, the node activates, and transmits data to the next layer of the network. Otherwise, no data is forwarded to the following layer of the network. (Refer Fig. A.1). The efficacy of neural networks depends on their ability to learn and get fine-tuned through training data. Once these algorithms are fine-tuned they become very powerful. They can classify and cluster data with very high accuracy.

A.2.1.1 Weights and Bias

The weights and biases play a crucial role in advancing data through a neural network. Weights are parameters that determine the strength of connections between neurons. Bias terms are additional parameters that provide flexibility to the model. Each connection has an associated weight, which is adjusted during the training process, known as forward propagation, to optimize model performance. After the completion of forward propagation, the neural network proceeds to enhance connections based on the errors identified during this phase. Subsequently, the flow reverses, traversing through layers to identify nodes and connections in

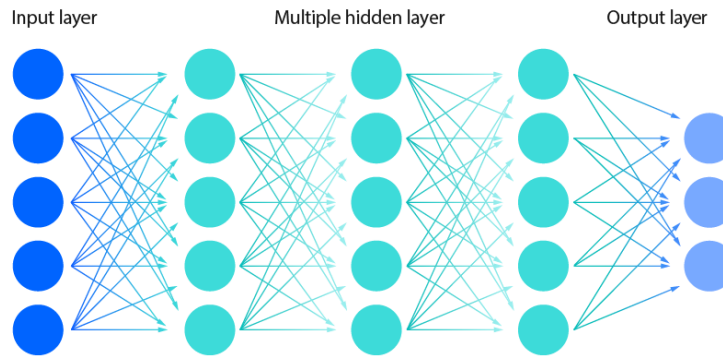


Figure A.1: Neural Network (Source: <https://www.ibm.com/topics/neural-networks>)

need of adjustment—a process referred to as backward propagation. Biases are additional parameters in neural networks that are associated with each neuron. The neural network can learn and express more complex functions using biases. They offer flexibility by enabling the network to take into account variations in input that the weights alone might not be able to fully capture. Like weights, biases are also adjusted during the training process to minimize the overall error. They contribute to the neuron's activation function, helping to introduce non-linearity into the network.

A.2.1.2 Activation Functions

Activation functions introduce non-linearity into the network, allowing it to learn complex patterns in data. Only linear layers would not improve a neural network's ability to learn and represent more complex relationships. Without activation functions, a neural network would only be a linear model. An activation function, used in neural networks, produces a low output when the input is below some set threshold and a higher output when inputs surpass the threshold. It "fires" if the inputs are sufficiently large; otherwise, it remains inactive. Here are some common activation functions used in deep learning, refer Fig. A.2 :

- **Sigmoid Function** It is frequently utilized in binary classification models' output layer since it reduces the output to a value that resembles probability. It transforms input values into a range between 0 and 1, like an S-shaped curve. Its formula is expressed as $f(x) = 1 / (1 + e(-x))$, where e is the base of the natural logarithm.
- **Hyperbolic Tangent (tanh) Function:** Similar to the sigmoid function

A.2 Building blocks of Deep learning network

but with a larger output range, often used in hidden layers of neural networks. The tanh function maps input values to the range $[-1, 1]$, sigmoidal (s-shaped). The benefit of this graph is that the zero inputs will be mapped close to zero and the negative inputs will be mapped considerably negatively. This function is mainly used in binary classification.

- **Rectified Linear Unit (ReLU):** It is an activation function frequently employed in artificial neural networks. It is a piecewise linear function that, in the case of a negative input, outputs zero and, otherwise, outputs the input directly. The function is defined as $f(x) = \max(0, x)$, where x is the input to the function. ReLU has gained popularity in deep learning due to its simplicity and effectiveness in mitigating the vanishing gradient problem, which can occur with other activation functions. This makes ReLU a common choice for hidden layers in neural networks.
- **Leaky-ReLU:** An alternative to the Rectified Linear Unit (ReLU) activation function used in artificial neural networks is the Leaky Rectified Linear Unit (Leaky ReLU). When the input is negative, it permits a small, non-zero gradient, resolving the problem of "dying" neurons that can arise with regular ReLU units. The formula for the Leaky ReLU function is $f(x) = \max(ax, x)$, where ' x ' is the function's input and ' a ' is a small positive constant (usually a very small number, like 0.01). Leaky ReLU, as opposed to ReLU, prevents complete inactivity and promotes information flow even in situations where it receives a negative input. ReLU outputs zero in such circumstances. Some of the drawbacks of conventional ReLU units are addressed by this tiny slope for negative inputs.

A.2.2 Loss Function

A loss function is a mathematical function that evaluates the difference between the predicted and target output values, showing the effectiveness of a neural network in representing the training data typically for a classification or regression task. The goal of training a neural network is to minimize this loss function. Several types of loss functions that are mostly used have been described here :

- **Cross-entropy loss function:** This is generally used in classification tasks for predicting probabilities whose output is a probability value between 0 and 1. The cross-entropy loss also known as log loss rises when the predicted probability deviates from the true label. The cross-entropy loss is minimized, with lower values indicating a superior model compared to higher values. A model achieving perfect probability predictions would have a cross entropy

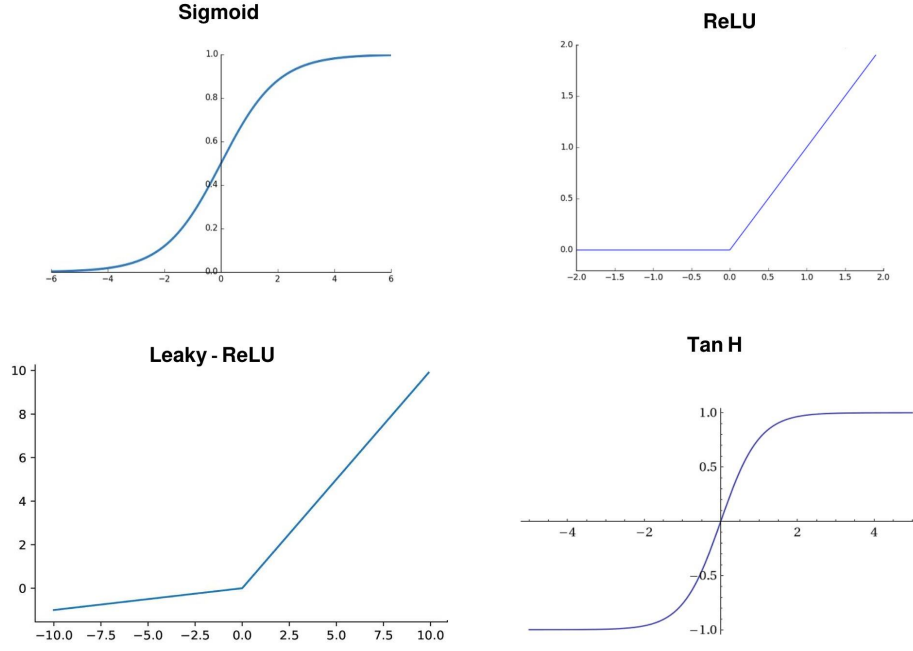


Figure A.2: Different activation functions

or log loss of 0.0. For binary classification, it is a two-class scenario, the computation of cross-entropy can be calculated as :

$$-\frac{1}{N} \sum_{i=1}^N (y_i \log(p_i) + (1 - y_i) \log(1 - p_i)) \quad (\text{A.1})$$

where p_i and y_i (0 or 1 in the case of binary classification) are the predicted probability and indicator of the i^{th} case respectively and N is the number of samples. A simple extension to a Multi-class Classification (say C classes) problem exists as follows Equation A.2:

$$-\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C (y_{ij} \log(p_{ij})) \quad (\text{A.2})$$

where p_{ij} is the predicted output for the i^{th} sample and j^{th} class and y_{ij} is the true output for the i^{th} sample and j^{th} class.

A.2 Building blocks of Deep learning network

- **Mean squared error loss function:** In regression tasks where the objective is to predict continuous numbers, this is used. Mean Squared Error (MSE) calculates the mean of the squared variances between the predicted outputs and the target values. It can be expressed as :

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (A.3)$$

where y_i is the predicted value and \hat{y}_i is the ground truth value. This function possesses various characteristics that makes it particularly suitable for loss calculation. It squares the difference, ensuring indifference to whether the predicted value is greater or lesser than the target value. Nevertheless, it penalizes values exhibiting substantial errors. Additionally, the Mean Squared Error (MSE) is a convex function with a well-defined global minimum, facilitating the application of gradient descent optimization for the adjustment of weight values in a more straightforward manner.

It's important to choose an appropriate loss function based on the nature of the problem. Different tasks may require different loss functions to effectively guide the training process and improve the model's performance.

A.2.2.1 Optimizer

It is the method by which the network adjusts its parameters in response to its loss function and the data it observes. The neural network must be trained using an optimization method, which looks for the optimal set of parameters to enable the model to give correct output on the testing data. Based on the gradients of the loss function with respect to those parameters, the optimizer modifies the neural network's parameters. Gradients show the extent and direction of the loss's steepest climb. Most optimizers are variants of gradient descent. To find a local minimum of the loss function, the main idea is to change the parameters in the opposite direction of the gradient. This process is repeated iteratively until convergence. One hyperparameter that regulates the step size in the parameter space at each optimization iteration is the learning rate. It affects how much the parameters of the model are changed in reply to the computed gradients. Selecting the right learning rate is essential to the convergence and stability of the training process. There are several optimizers commonly used in deep learning:

- **Stochastic Gradient Descent (SGD):** This is the basic form of gradient descent where the model parameters are updated on every iteration. It suggests that the model is updated and the loss function is checked following each training sample. The algorithm is based on a randomness that is implied

by the term "stochastic." For large amounts of data, this optimizer uses batches of the datasets for iteration, not the complete dataset at once to attain the local minima. In deep learning, an SGD optimizer is preferred when dealing with large amounts of data and significant computing time.

- **Adaptive Gradient Algorithm (Adagrad):** An optimizer that adapts the learning rate for each parameter based on the historical sum of squared gradients. In the case of sparse feature input with predominantly zero values, it is possible to use a higher learning rate to counteract the diminishing gradient associated with these sparse features. Conversely, when dealing with dense data, a slower learning rate may be more appropriate. Having an adaptable learning rate that can alter based on the input given is the answer to this problem. Adagrad optimizer aims to achieve this adaptability by reducing the learning rate proportionally to the evolving history of gradients.
- **Root Mean Square Propagation (RMSprop):** It is an improved version of the Adagrad optimizer. The full form of RMSprop is Root Mean Square Prop. It is adaptive in nature. For every weight, it maintains the moving average of the squared gradients, then divides that by the square root of the mean square. In this manner, it tries to overcome the diminishing learning rate problem of Adagrad. The learning rate in RMSprop is automatically modified, and a distinct learning rate is selected for every parameter. It indicates that parameters with small or stable gradients have larger effective learning rates than those with big and fluctuating gradients. This step also helps to scale down the learning rates for parameters that have large and frequent updates. This adaptive nature of this optimizer helps to overcome the vanishing gradient issue.
- **Adaptive Moment Estimation (ADAM):** This optimizer tries to assemble the advantage of both the Adagrad and RMSprop optimizer and make it one. Initially, the algorithm computes the gradient of the loss function concerning a batch of data. This algorithm retains a moving average of both the gradients (first moment) as well as the squares of the gradients (second moment). subsequently, Adam changes the learning rate of each parameter by taking the ratio of the square root of the second-moment estimate and the first-moment estimate. Adam includes a bias adjustment phase to take into account the initialization bias of the moving averages at the start of training. By doing this, the proper scaling of the moving averages is guaranteed. Considering all these behaviors, Adam is an effective choice for deep neural network training due to its momentum-like behavior and adjustable learning rates.

A.2 Building blocks of Deep learning network

It is required to understand the nature of the data and the model being trained to choose the appropriate optimizer for that scenario. Adam is a common option in reality because of its excellent performance in a wide range of instances. To determine which optimizer is most effective for a certain task, it is recommended to test out a variety of them.

A.2.2.2 Backpropagation

Backpropagation also known as reverse-mode differentiation is a supervised learning technique used in neural network learning. Backpropagation starts by taking the final loss value and proceeds to work in reverse, moving from the final layers to the initial layers. It employs the chain rule to calculate the impact that each parameter had on the loss value.

During the forward pass, a specified loss function, which calculates the difference between the expected and true values, is used to compare the predicted output to the actual target values. Neural networks compute gradient descent, an optimization process that leads the user to the maximum or minimum of a function, using backpropagation as a learning algorithm. The gradients are transmitted from the neurons of the output layer, traversing through the hidden layers towards the neuron of the input layer. This allows neurons to make adjustments along the way if they contributed to the generation of the error. The aim of this process is to minimize the loss function and improve the model's accuracy in predicting the target values. In a nutshell, backpropagation plays a crucial role in deep learning training by enabling neural networks to learn from data through iterative weight adjustments meant to reduce the discrepancy between expected and actual outputs.

A.2.2.3 Batch Normalization

Normalization is a broad category of methods that seek to make different samples seen by a machine-learning model more similar to each other, which helps the model learn and generalize well to new data. Normalization is a process that tries to put different input data into a neural network on a similar scale. The purpose of this is to enhance the ability of the learning process to generalize effectively to new data. In order to have zero mean and unit variance, it requires normalizing the inputs of each layer in a mini-batch. During the forward pass, this normalization is applied prior to the activation function. The popular method of data normalization involves centering the data around 0 through the subtraction of the mean and providing the data with a unit standard deviation by dividing it by its standard deviation.

When training, the mean and variance of the data can change, yet batch normalization is a kind of layer that can adaptively normalize the data. The way it functions is by internally preserving an exponential moving average of the training data's batch-wise mean and variance. Batch normalization primarily facilitates gradient propagation, enabling the training of deeper networks. In certain cases, the inclusion of multiple Batch Normalization layers becomes essential for effectively training very deep networks. Batch Normalization offers several benefits including learning stabilization, training acceleration, allowing higher learning rates, making networks less sensitive to weight initialization etc.

A.2.2.4 Regularization

It is a common phenomenon that in any machine learning model while validated on the validation set data, after a few epochs of iteration the performance begins to degrade after reaching its peak. The model started to overfit on the training data. The performance of the model depends a lot on how it balances between optimization and generalization. Optimization is the process of refining a model to attain the best performance possible on the training data while generalization assesses the model's performance on test data. While the ultimate objective is to achieve strong generalization, but it is only achievable by adjusting the model based on its training data. In general, at the beginning of the training process, the model remains in underfit status which indicates the model needs more training. At that stage, a decrease in the loss on the training data corresponds to a decrease in the loss on the test data. However, as training progresses and a certain number of iterations are completed, generalizations stop to improve. Validation metrics also gradually started deteriorating after reaching their peak. This indicates that the model is entering an overfitting stage (Refer Fig. A.3). In this phase, the model starts learning patterns specific to the training data, but these patterns may be misleading when applied to unseen data.

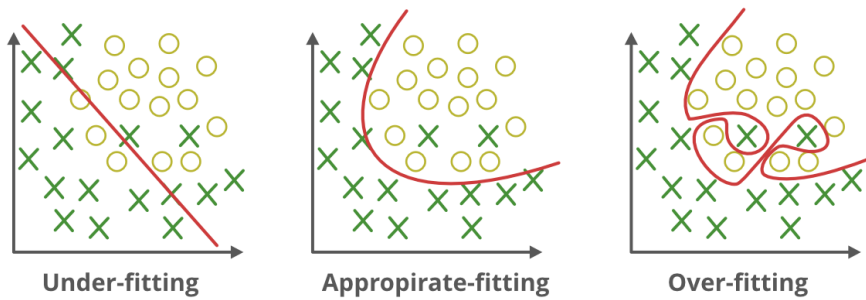


Figure A.3: Underfitting and overfitting

A.2 Building blocks of Deep learning network

One solution to the above problem is to train the model on more training data, but if that is not available then the best solution is to prevent the model from storing less information or add some rule to restrict it from storing what it is allowed to store. Regularization is the approach used to address the overfitting problem. There are a few techniques employed for regularization:

- **Weight Regularization:** A commonly employed strategy to address overfitting involves imposing constraints on a network's complexity by compelling the weights of the network to assume small values. There are two ways it is implemented - a) a penalty is added to the loss function proportional to the absolute values of the model weights known as L1 regularization and b) adds a penalty to the loss function proportional to the square of the model weights known as L2 regularization
- **Dropout:** Dropout is a technique where randomly selected neurons are ignored during training. This helps prevent co-adaptation of neurons and reduces overfitting. During each training iteration, a random subset of neurons is "dropped out" or set to zero, forcing the network to learn more robust features.

Regularization methods are applied during the training phase, and their effectiveness depends on the specific problem, dataset, and architecture. Choosing the right combination of regularization techniques often involves experimentation and tuning.

A.2.2.5 Dropout

Dropout is a regularization technique commonly used in neural networks to improve model performance by removing the effect of overfitting. When a model fits itself with the training data too well it does not perform well with new data, this is known as overfitting. To prevent this, during training time a certain percentage of input and hidden units are dropped from producing the output (Refer Fig. A.4). By doing this, the model becomes more resilient and less prone to overfitting the training data, improving generalization and reducing dependency on particular units. This concept was first developed by Geoffrey Hinton and his students at the University of Toronto.

Let's consider a scenario where a specific layer, under normal circumstances, would produce a vector $[0.2, 0.5, 1.3, 0.8, 1.1]$ for a given input sample during training. When dropout is applied, a few random entries in this vector become zero, resulting in, for instance, $[0, 0.5, 1.3, 0, 1.1]$. The dropout rate represents the proportion of features that are zeroed out, typically set between 0.2 and 0.5. To allow the fact that more units are active during test time than during training,

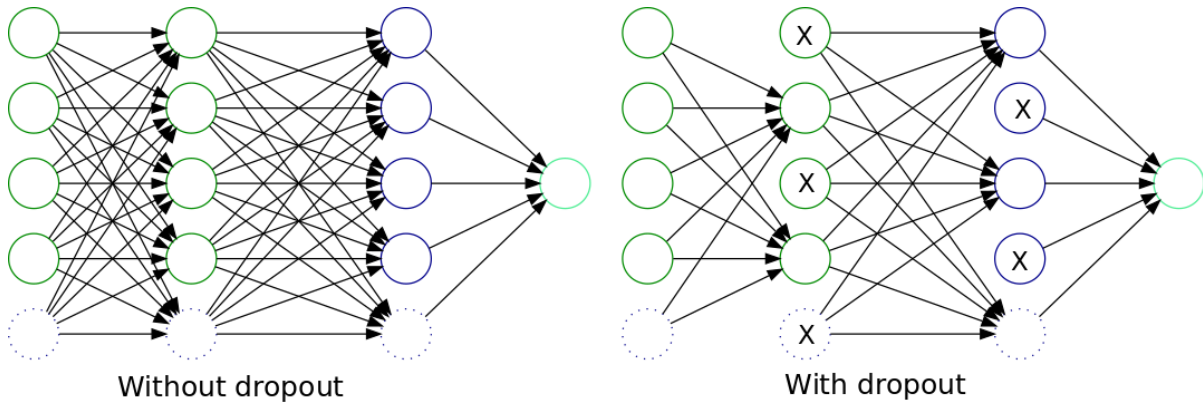


Figure A.4: Dropout in a network (Source: <https://www.dremio.com/>)

the layer's output values are scaled down by a factor equal to the dropout rate. Generally, no units are dropped during the testing period.

The efficiency of this strategy depends a lot on the architecture of the network and also on the dataset used for training. So even if this is a popular strategy, it is also required to understand the present working scenario and whether this fits or not.

A.2.2.6 Pooling Layers

Pooling layers are a key component of convolutional neural networks (CNNs) which are used in deep learning. They are used to minimize the width and height of the input data while preserving the most significant information. Sometimes this layer is used multiple times in the whole network to progressively reduce the dimension of the feature maps. The functionality of pooling layers consists of a few steps - a) first the input data is divided into non-overlapping regions known as windows b) then some aggregate function like maximum or average is applied on each window to get a single value for each window (Refer Fig. A.5) c) at last all those values generated from the second step are combined to generate a down-sampled representation of the input data.

Pooling layers are incorporated for several reasons. By reducing the number of parameters these layers help to reduce the computational complexity in the subsequent layers of the model. These layers extract the most significant information from several spatial places to offer a type of translation invariance. By combining lower-level characteristics to produce higher-level features, pooling layers help in the creation of a hierarchical representation of the input data.

A.2 Building blocks of Deep learning network

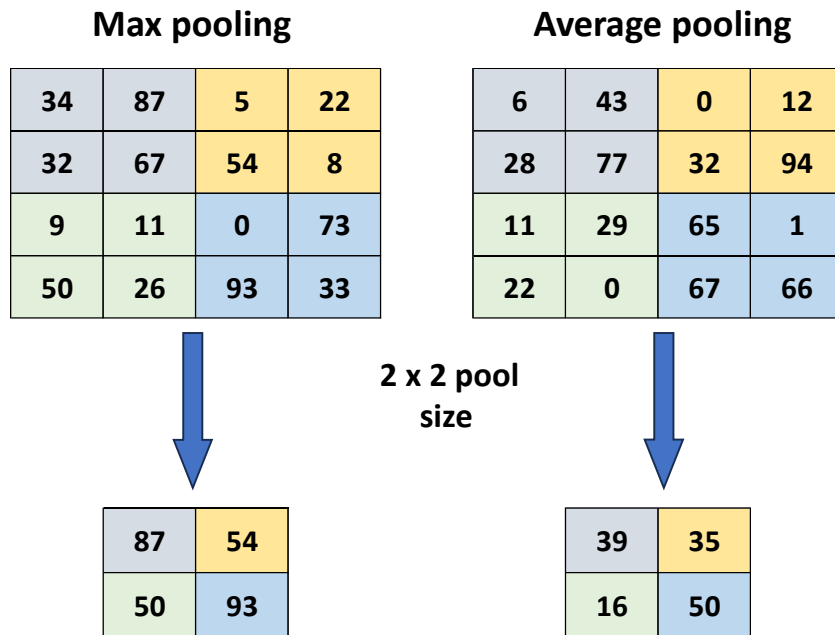


Figure A.5: Pooling operation

Pooling Layers are used in many different fields like image classification, object detection etc. It's important to note that while pooling layers offer several advantages, the choice of pooling strategy (e.g., max pooling, average pooling) and the degree of downsampling can impact the network's performance.

A.2.2.7 Convolutional Layers

Convolutional Layers are a fundamental component of neural networks known as Convolutional Neural Network (Convnet) especially when it comes to computer vision tasks like segmentation, detecting objects, and image categorization. Applying spatial filtering operations to the input data, convolutional layers enable the network to directly acquire significant features and patterns from the input raw data.

Convolutional Layers operate by employing a collection of learnable filters on the input data. Each filter conducts a convolution operation, moving across the input data to calculate a weighted sum of the values within its filter size. This mechanism helps the network to identify local patterns and spatial hierarchies in the data. Furthermore, convolutional Layers frequently integrate non-linear activation functions like ReLU to introduce non-linearity to the network.

The basic difference between convolution layers and other dense layers is that

the former learn local patterns whereas the latter learn global patterns from the input feature space. For this characteristic, the convnets possess two special features:

- They can learn translation invariant patterns, which means once they learn a specific pattern anywhere in an image, they can identify the same pattern elsewhere in the image. As a result, convnets can learn features more efficiently even with small no. of input datasets.
- They can learn spatial hierarchies of patterns, which means the initial convolution layer captures small local patterns like edges, while subsequent layers build upon these features to learn larger, more complex patterns. This hierarchical approach enables convnets to effectively acquire progressively intricate and abstract visual concepts.

To summarize, convolutional layers in neural networks play a crucial role in retaining spatial relationships, capturing hierarchical characteristics, and facilitating effective processing of structured data, especially in computer vision tasks.

A.2.2.8 Fully-Connected Layer

Fully-connected layer is a type of layer in a neural network where each neuron or node in the layer is connected to every neuron in the layer before it. It is sometimes known as densely connected layer. The fully-connected layer calculates the output as per the following logic - multiplies the input layer by the weight of the connections between the input and the neurons, bias vector is added to the weighted sum and finally, an activation function is applied element-wise to the weighted sum to add non-linearity to the model (Refer Fig. A.6).

A.2.3 Deep learning for computer vision

Deep learning network has shown remarkable performance in different computer vision-related tasks such as image classification, object detection, image segmentation etc. They are made up of layers such as convolutional, pooling, and fully linked layers. Apart from these some key concepts and techniques in deep learning for computer vision are - Transfer Learning, Attention Mechanisms, Data Augmentation etc.

Deep learning frameworks such as TensorFlow, PyTorch, and Keras provide tools and abstractions to build, train, and deploy deep learning models for computer vision tasks. The field continues to evolve rapidly, with ongoing research and development pushing the boundaries of what is possible in computer vision using deep learning.

A.2 Building blocks of Deep learning network

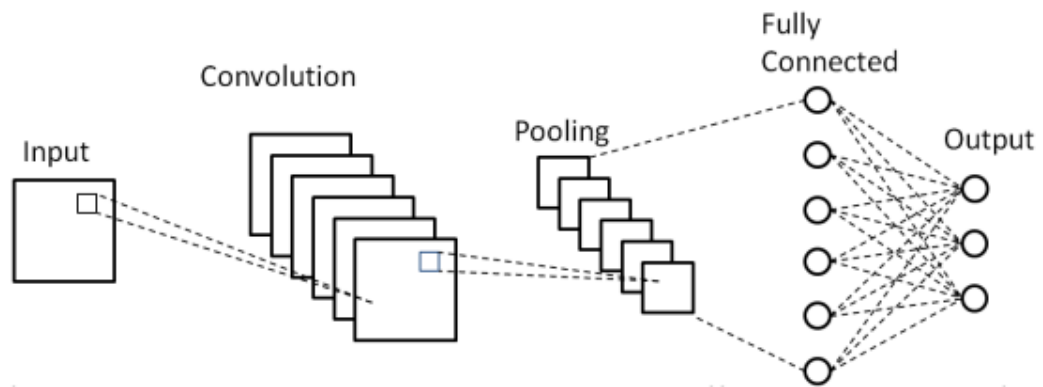


Figure A.6: CNN Architecture

References

- [1] C. Sun, J. J. Wang, D. A. Mackey, and T. Y. Wong, “Retinal vascular caliber: systemic, environmental, and genetic associations,” *Survey of ophthalmology*, vol. 54, no. 1, pp. 74–95, 2009.
- [2] M. D. Knudtson, K. E. Lee, L. D. Hubbard, T. Y. Wong, R. Klein, and B. E. Klein, “Revised formulas for summarizing retinal vessel diameters,” *Current eye research*, vol. 27, no. 3, pp. 143–149, 2003.
- [3] T. Y. Wong, A. Shankar, R. Klein, B. E. Klein, and L. D. Hubbard, “Prospective cohort study of retinal vessel diameters and risk of hypertension,” *bmj*, vol. 329, no. 7457, p. 79, 2004.
- [4] L. D. Hubbard, R. J. Brothers, W. N. King, L. X. Clegg, R. Klein, L. S. Cooper, A. R. Sharrett, M. D. Davis, J. Cai, A. R. in Communities Study Group *et al.*, “Methods for evaluation of retinal microvascular abnormalities associated with hypertension/sclerosis in the atherosclerosis risk in communities study,” *Ophthalmology*, vol. 106, no. 12, pp. 2269–2280, 1999.
- [5] S. Lesage, T. Mosley, T. Wong, M. Szklo, D. Knopman, D. J. Catellier, S. Cole, R. Klein, J. Coresh, L. Coker *et al.*, “Retinal microvascular abnormalities and cognitive decline: the aric 14-year follow-up study,” *Neurology*, vol. 73, no. 11, pp. 862–868, 2009.
- [6] 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, 2012.
- [7] M. K. Ikram, C. Y. Cheung, M. Lorenzi, R. Klein, T. L. Jones, T. Y. Wong *et al.*, “Retinal vascular caliber as a biomarker for diabetes microvascular complications,” *Diabetes care*, vol. 36, no. 3, pp. 750–759, 2013.
- [8] B. Dashtbozorg, A. M. Mendonça, and A. Campilho, “An automatic method for the estimation of arteriolar-to-venular ratio in retinal images,” in *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*. IEEE, 2013, pp. 512–513.
- [9] M. K. Ikram, F. J. de Jong, J. R. Vingerling, J. C. Witteman, A. Hofman, M. M. Breteler, and P. T. de Jong, “Are retinal arteriolar or venular diameters associated with markers for cardiovascular disorders? the rotterdam study,” *Investigative ophthalmology & visual science*, vol. 45, no. 7, pp. 2129–2134, 2004.
- [10] C. Kondermann, D. Kondermann, and M. Yan, “Blood vessel classification into arteries and veins in retinal images,” in *Medical Imaging 2007: Image Processing*, vol. 6512. International Society for Optics and Photonics, 2007, p. 651247.

-
- [11] F. Benmansour and L. D. Cohen, "Tubular structure segmentation based on minimal path method and anisotropic enhancement," *International Journal of Computer Vision*, vol. 92, no. 2, pp. 192–210, 2011.
 - [12] R. Estrada, C. Tomasi, M. T. Cabrera, D. K. Wallace, S. F. Freedman, and S. Farsiu, "Exploratory dijkstra forest based automatic vessel segmentation: applications in video indirect ophthalmoscopy (vio)," *Biomedical optics express*, vol. 3, no. 2, pp. 327–339, 2012.
 - [13] J. Jan, J. Odstrcilik, J. Gazarek, and R. Kolár, "Retinal image analysis aimed at blood vessel tree segmentation and early detection of neural-layer deterioration," *Computerized Medical Imaging and Graphics*, vol. 36, no. 6, pp. 431–441, 2012.
 - [14] H. Jiang, D. C. DeBuc, T. Rundek, B. L. Lam, C. B. Wright, M. Shen, A. Tao, and J. Wang, "Automated segmentation and fractal analysis of high-resolution non-invasive capillary perfusion maps of the human retina," *Microvascular research*, vol. 89, pp. 172–175, 2013.
 - [15] N. Lee, A. F. Laine, and R. T. Smith, "A hybrid segmentation approach for geographic atrophy in fundus auto-fluorescence images for diagnosis of age-related macular degeneration," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2007, pp. 4965–4968.
 - [16] J. V. Soares, J. J. Leandro, R. M. Cesar, H. F. Jelinek, and M. J. Cree, "Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification," *IEEE Transactions on medical Imaging*, vol. 25, no. 9, pp. 1214–1222, 2006.
 - [17] C. Xiao, M. Staring, Y. Wang, D. P. Shamonin, and B. C. Stoel, "Multiscale bi-gaussian filter for adjacent curvilinear structures detection with application to vasculature images," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 174–188, 2012.
 - [18] M. R. K. Mookiah, U. R. Acharya, C. K. Chua, C. M. Lim, E. Ng, and A. Laude, "Computer-aided diagnosis of diabetic retinopathy: A review," *Computers in biology and medicine*, vol. 43, no. 12, pp. 2136–2155, 2013.
 - [19] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi, "Dream: diabetic retinopathy analysis using machine learning," *IEEE journal of biomedical and health informatics*, vol. 18, no. 5, pp. 1717–1728, 2013.
 - [20] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *Jama*, vol. 316, no. 22, pp. 2402–2410, 2016.
 - [21] D. S. W. Ting, C. Y.-L. Cheung, G. Lim, G. S. W. Tan, N. D. Quang, A. Gan, H. Hamzah, R. Garcia-Franco, I. Y. San Yeo, S. Y. Lee *et al.*, "Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes," *Jama*, vol. 318, no. 22, pp. 2211–2223, 2017.
 - [22] C. Angermueller, T. Pärnamaa, L. Parts, and O. Stegle, "Deep learning for computational biology," *Molecular systems biology*, vol. 12, no. 7, 2016.
 - [23] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

-
- [24] T. T. Nguyen, J. J. Wang, and T. Y. Wong, "Retinal vascular changes in pre-diabetes and prehypertension: new findings and their research and clinical implications," *Diabetes care*, vol. 30, no. 10, pp. 2708–2715, 2007.
 - [25] S. Barrett, E. Naess, and T. Molvik, "Employing the hough transform to locate the optic disk," *Biomedical sciences instrumentation*, vol. 37, pp. 81–86, 2001.
 - [26] M. Blanco, M. G. Penedo, N. Barreira, M. Penas, and M. J. Carreira, "Localization and extraction of the optic disc using the fuzzy circular hough transform," in *International Conference on Artificial Intelligence and Soft Computing*. Springer, 2006, pp. 712–721.
 - [27] A. Hoover and M. Goldbaum, "Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels," *IEEE transactions on medical imaging*, vol. 22, no. 8, pp. 951–958, 2003.
 - [28] A. Mendonça, F. Cardoso, A. Sousa, and A. Campilho, "Automatic localization of the optic disc in retinal images based on the entropy of vascular directions," *Image analysis and recognition*, pp. 424–431, 2012.
 - [29] I. Soares, M. Castelo-Branco, and A. M. Pinheiro, "Optic disc localization in retinal images based on cumulative sum fields," *IEEE journal of biomedical and health informatics*, vol. 20, no. 2, pp. 574–585, 2016.
 - [30] M. Foracchia, E. Grisan, and A. Ruggeri, "Detection of optic disc in retinal images by means of a geometrical model of vessel structure," *IEEE transactions on medical imaging*, vol. 23, no. 10, pp. 1189–1195, 2004.
 - [31] M. Niemeijer, M. D. Abramoff, and B. van Ginneken, "Segmentation of the optic disk, macula and vascular arch in fundus photographs," *IEEE Trans Medical Imaging*, vol. 26(1), pp. 116–127, 2007.
 - [32] X. Wu, B. Dai, and W. Bu, "Optic disc localization using directional models," *IEEE Transactions on Image Processing*, vol. 25, no. 9, pp. 4433–4442, 2016.
 - [33] J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, E. Fletcher, and L. Kennedy, "Optic nerve head segmentation," *IEEE Transactions on medical Imaging*, vol. 23, no. 2, pp. 256–264, 2004.
 - [34] D. Welfer, J. Scharcanski, C. M. Kitamura, M. M. Dal Pizzol, L. W. Ludwig, and D. R. Marinho, "Segmentation of the optic disk in color eye fundus images using an adaptive morphological approach," *Computers in Biology and Medicine*, vol. 40, no. 2, pp. 124–137, 2010.
 - [35] A. Aquino, M. E. Gegúndez-Arias, and D. Marín, "Detecting the optic disc boundary in digital fundus images using morphological, edge detection, and feature extraction techniques," *IEEE transactions on medical imaging*, vol. 29, no. 11, pp. 1860–1869, 2010.
 - [36] S. Morales, V. Naranjo, J. Angulo, and M. Alcañiz, "Automatic detection of optic disc based on pca and mathematical morphology," *IEEE transactions on medical imaging*, vol. 32, no. 4, pp. 786–796, 2013.
 - [37] C. Xu and J. L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Transactions on image processing*, vol. 7, no. 3, pp. 359–369, 1998.
 - [38] G. D. Joshi, J. Sivaswamy, and S. Krishnadas, "Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment," *IEEE transactions on medical imaging*, vol. 30, no. 6, pp. 1192–1205, 2011.

-
- [39] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Transactions on image processing*, vol. 10, no. 2, pp. 266–277, 2001.
 - [40] M. Niemeijer, X. Xu, A. V. Dumitrescu, P. Gupta, B. Van Ginneken, J. C. Folk, and M. D. Abramoff, "Automated measurement of the arteriolar-to-venular width ratio in digital color fundus photographs," *IEEE Transactions on medical imaging*, vol. 30, no. 11, pp. 1941–1950, 2011.
 - [41] V. S. Joshi, J. M. Reinhardt, M. K. Garvin, and M. D. Abramoff, "Automated method for identification and artery-venous classification of vessel trees in retinal vessel networks," *PloS one*, vol. 9, no. 2, p. e88061, 2014.
 - [42] Q. Hu, M. D. Abramoff, and M. K. Garvin, "Automated construction of arterial and venous trees in retinal images," *Journal of Medical Imaging*, vol. 2, no. 4, p. 044001, 2015.
 - [43] S. Vázquez, B. Cancela, N. Barreira, M. G. Penedo, M. Rodríguez-Blanco, M. P. Seijo, G. C. de Tuero, M. A. Barceló, and M. Saez, "Improving retinal artery and vein classification by means of a minimal path approach," *Machine vision and applications*, vol. 24, no. 5, pp. 919–930, 2013.
 - [44] F. Huang, B. Dashtbozorg, T. Tan, and B. M. ter Haar Romeny, "Retinal artery/vein classification using genetic-search feature selection," *Computer methods and programs in biomedicine*, vol. 161, pp. 197–207, 2018.
 - [45] N. Memari, A. R. Ramli, M. I. B. Saripan, S. Mashohor, and M. Moghbel, "Retinal blood vessel segmentation by using matched filtering and fuzzy c-means clustering with integrated level set method for diabetic retinopathy assessment," *Journal of Medical and Biological Engineering*, vol. 39, pp. 713–731, 2019.
 - [46] T. Jemima Jebaseeli, C. Anand Deva Durai, and J. Dinesh Peter, "Retinal blood vessel segmentation from depigmented diabetic retinopathy images," *IETE Journal of Research*, vol. 67, no. 2, pp. 263–280, 2021.
 - [47] W. Chen, S. Yu, J. Wu, K. Ma, C. Bian, C. Chu, L. Shen, and Y. Zheng, "Tr-gan: Topology ranking gan with triplet loss for retinal artery/vein classification," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part V 23*. Springer, 2020, pp. 616–625.
 - [48] R. Welikala, P. Foster, P. Whincup, A. R. Rudnicka, C. G. Owen, D. Strachan, S. Barman *et al.*, "Automated arteriole and venule classification using deep learning for retinal images from the uk biobank cohort," *Computers in biology and medicine*, vol. 90, pp. 23–32, 2017.
 - [49] G. Lepetit-Aimon, R. Duval, and F. Cheriet, "Large receptive field fully convolutional network for semantic segmentation of retinal vasculature in fundus images," in *Computational Pathology and Ophthalmic Medical Image Analysis: First International Workshop, COMPAY 2018, and 5th International Workshop, OMIA 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16–20, 2018, Proceedings 5*. Springer, 2018, pp. 201–209.
 - [50] R. Hemelings, B. Elen, I. Stalmans, K. Van Keer, P. De Boever, and M. B. Blaschko, "Artery–vein segmentation in fundus images using a fully convolutional network," *Computerized Medical Imaging and Graphics*, vol. 76, p. 101636, 2019.

-
- [51] F. Girard, C. Kavalec, and F. Cheriet, "Joint segmentation and classification of retinal arteries/veins from fundus images," *Artificial intelligence in medicine*, vol. 94, pp. 96–109, 2019.
 - [52] W. Ma, S. Yu, K. Ma, J. Wang, X. Ding, and Y. Zheng, "Multi-task neural networks with spatial activation for retinal vessel segmentation and artery/vein classification," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22*. Springer, 2019, pp. 769–778.
 - [53] Z. Wang, X. Jiang, J. Liu, K.-T. Cheng, and X. Yang, "Multi-task siamese network for retinal artery/vein separation via deep convolution along vessel," *IEEE Transactions on Medical Imaging*, vol. 39, no. 9, pp. 2904–2919, 2020.
 - [54] A. E. Chowdhury, G. Mann, W. H. Morgan, A. Vukmirovic, A. Mehnert, and F. Sohel, "Msganet-rav: A multiscale guided attention network for artery-vein segmentation and classification from optic disc and retinal images," *Journal of Optometry*, vol. 15, pp. S58–S69, 2022.
 - [55] I. K. Gupta, A. Choubey, and S. Choubey, "Mayfly optimization with deep learning enabled retinal fundus image classification model," *Computers and Electrical Engineering*, vol. 102, p. 108176, 2022.
 - [56] K. Zervoudakis and S. Tsafarakis, "A mayfly optimization algorithm," *Computers & Industrial Engineering*, vol. 145, p. 106559, 2020.
 - [57] V. Sathananthavathi and G. Indumathi, "Atrous fully convolved depth concatenated neural network with enriched encoder for retinal artery–vein classification," *IETE Journal of Research*, pp. 1–10, 2022.
 - [58] J. Hu, L. Qiu, H. Wang, and J. Zhang, "Semi-supervised point consistency network for retinal artery/vein classification," *Computers in Biology and Medicine*, p. 107633, 2023.
 - [59] B. Toptaş and D. Hanbay, "Separation of arteries and veins in retinal fundus images with a new cnn architecture," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 11, no. 4, pp. 1512–1522, 2023.
 - [60] A. Ruggeri, E. Grisan, and M. De Luca, "An automatic system for the estimation of generalized arteriolar narrowing in retinal images," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2007, pp. 6463–6466.
 - [61] L. Tramontan, E. Grisan, and A. Ruggeri, "An improved system for the automatic estimation of the arteriolar-to-venular diameter ratio (avr) in retinal images," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2008, pp. 3550–3553.
 - [62] E. Grisan and A. Ruggeri, "A divide et impera strategy for automatic classification of retinal vessels into arteries and veins," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439)*, vol. 1. IEEE, 2003, pp. 890–893.
 - [63] A. Khanal and R. Estrada, "Fully automated artery-vein ratio and vascular tortuosity measurement in retinal fundus images," *arXiv preprint arXiv:2301.01791*, 2023.

-
- [64] D. Pascolini and S. P. Mariotti, "Global estimates of visual impairment: 2010," *British Journal of Ophthalmology*, vol. 96, no. 5, pp. 614–618, 2012.
 - [65] J. Cheng, J. Liu, Y. Xu, F. Yin, D. W. K. Wong, N.-M. Tan, D. Tao, C.-Y. Cheng, T. Aung, and T. Y. Wong, "Superpixel classification based optic disc and optic cup segmentation for glaucoma screening," *IEEE Transactions on Medical Imaging*, vol. 32, no. 6, pp. 1019–1032, 2013.
 - [66] R. J. Qureshi, L. Kovacs, B. Harangi, B. Nagy, T. Peto, and A. Hajdu, "Combining algorithms for automatic detection of optic disc and macula in fundus images," *Computer Vision and Image Understanding*, vol. 116, no. 1, pp. 138–145, 2012.
 - [67] R. Estrada, C. Tomasi, S. C. Schmidler, and S. Farsiu, "Tree topology estimation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 8, pp. 1688–1701, 2015.
 - [68] C. Sinthanayothin, J. F. Boyce, H. L. Cook, and T. H. Williamson, "Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images," *British Journal of Ophthalmology*, vol. 83, no. 8, pp. 902–910, 1999.
 - [69] A. A.-H. A.-R. Youssif, A. Z. Ghalwash, and A. A. S. A.-R. Ghoneim, "Optic disc detection from normalized digital fundus images by means of a vessels' direction matched filter," *IEEE Transactions on Medical imaging*, vol. 27, no. 1, pp. 11–18, 2008.
 - [70] A. Osareh, M. Mirmehdi, B. Thomas, and R. Markham, "Comparison of colour spaces for optic disc localisation in retinal images," in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, vol. 1. IEEE, 2002, pp. 743–746.
 - [71] H. Li and O. Chutatape, "Automated feature extraction in color retinal images by a model based approach," *IEEE Transactions on biomedical engineering*, vol. 51, no. 2, pp. 246–254, 2004.
 - [72] A. Giachetti, L. Ballerini, and E. Trucco, "Accurate and reliable segmentation of the optic disc in digital fundus images," *Journal of Medical Imaging*, vol. 1, no. 2, pp. 024001–024001, 2014.
 - [73] D. Zhang and Y. Zhao, "Novel accurate and fast optic disc detection in retinal images with vessel distribution and directional characteristics," *IEEE journal of biomedical and health informatics*, vol. 20, no. 1, pp. 333–342, 2016.
 - [74] S. Roychowdhury, D. D. Koozekanani, S. N. Kuchinka, and K. K. Parhi, "Optic disc boundary and vessel origin segmentation of fundus images," *IEEE journal of biomedical and health informatics*, vol. 20, no. 6, pp. 1562–1574, 2016.
 - [75] T. Walter and J.-C. Klein, "Segmentation of color fundus images of the human retina: Detection of the optic disc and the vascular tree using morphological techniques," *Medical data analysis*, pp. 282–287, 2001.
 - [76] A. W. Reza, C. Eswaran, and S. Hati, "Automatic tracing of optic disc and exudates from color fundus images using fixed and variable thresholds," *Journal of medical systems*, vol. 33, no. 1, p. 73, 2009.
 - [77] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *International journal of computer vision*, vol. 22, no. 1, pp. 61–79, 1997.

-
- [78] R. Chrstek, M. Wolf, K. Donath, H. Niemann, D. Paulus, T. Hothorn, B. Lausen, R. Lmmer, C. Y. Mardin, and G. Michelson, "Automated segmentation of the optic nerve head for diagnosis of glaucoma," *Medical Image Analysis*, vol. 9, no. 4, pp. 297–314, 2005.
 - [79] Y. Zhang, B. J. Matuszewski, L.-K. Shark, and C. J. Moore, "Medical image segmentation using new hybrid level-set method," in *BioMedical Visualization, 2008. MEDIVIS'08. Fifth International Conference*. IEEE, 2008, pp. 71–76.
 - [80] X. He, R. S. Zemel, and M. Á. Carreira-Perpiñán, "Multiscale conditional random fields for image labeling," in *Computer vision and pattern recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE computer society conference on*, vol. 2. IEEE, 2004, pp. II–II.
 - [81] C. Russell, P. Kohli, P. H. Torr *et al.*, "Associative hierarchical crfs for object class image segmentation," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 739–746.
 - [82] V. Lempitsky, A. Vedaldi, and A. Zisserman, "Pylon model for semantic segmentation," in *Advances in neural information processing systems*, 2011, pp. 1485–1493.
 - [83] A. Delong, A. Osokin, H. N. Isack, and Y. Boykov, "Fast approximate energy minimization with label costs," *International journal of computer vision*, vol. 96, no. 1, pp. 1–27, 2012.
 - [84] G. Papandreou, I. Kokkinos, and P.-A. Savalle, "Modeling local and global deformations in deep learning: Epitomic convolution, multiple instance learning, and sliding window detection," in *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*. IEEE, 2015, pp. 390–399.
 - [85] X. Chen, R. Mottaghi, X. Liu, S. Fidler, R. Urtasun, and A. Yuille, "Detect what you can: Detecting and representing objects using holistic models and body parts," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1971–1978.
 - [86] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3213–3223.
 - [87] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab:," *arXiv preprint arXiv:1606.00915*, 2016.
 - [88] E. Decencière, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay *et al.*, "Feedback on a publicly distributed image database: the messidor database," *Image Analysis & Stereology*, vol. 33, no. 3, pp. 231–234, 2014.
 - [89] Q. Mohamed, M. C. Gillies, and T. Y. Wong, "Management of diabetic retinopathy: a systematic review," *Jama*, vol. 298, no. 8, pp. 902–916, 2007.
 - [90] N. Patton, T. M. Aslam, T. MacGillivray, I. J. Deary, B. Dhillon, R. H. Eikelboom, K. Yogesan, and I. J. Constable, "Retinal image analysis: Concepts, applications and potential," *Prog Retin Eye Res*, vol. 25(1), pp. 99–127, 2006.
 - [91] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.
 - [92] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

-
- [93] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” in *Advances in neural information processing systems*, 2015, pp. 91–99.
 - [94] H. Yu, E. S. Barriga, C. Agurto, S. Echegaray, M. S. Pattichis, W. Bauman, and P. Soliz, “Fast localization and segmentation of optic disk in retinal images using directional matched filtering and level sets,” *IEEE Transactions on information technology in biomedicine*, vol. 16, no. 4, pp. 644–657, 2012.
 - [95] S. Lu and J. H. Lim, “Automatic optic disc detection from retinal images by a line operator,” *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 88–94, 2011.
 - [96] A. Aquino, M. E. Gegundez, and D. Marin, “Automated optic disc detection in retinal images of patients with diabetic retinopathy and risk of macular edema,” *International Journal of Biological and Life Sciences*, vol. 8, no. 2, pp. 87–92, 2012.
 - [97] R. R. Bourne, S. R. Flaxman, T. Braithwaite, M. V. Cicinelli, A. Das, J. B. Jonas, J. Keeffe, J. H. Kempen, J. Leasher, H. Limburg *et al.*, “Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis,” *The Lancet Global Health*, vol. 5, no. 9, pp. e888–e897, 2017.
 - [98] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, “Attention u-net: Learning where to look for the pancreas,” *arXiv preprint arXiv:1804.03999*, 2018.
 - [99] S. Jetley, N. A. Lord, N. Lee, and P. H. Torr, “Learn to pay attention,” *arXiv preprint arXiv:1804.02391*, 2018.
 - [100] L. Fan, W.-C. Wang, F. Zha, and J. Yan, “Exploring new backbone and attention module for semantic segmentation in street scenes,” *IEEE Access*, vol. 6, pp. 71 566–71 580, 2018.
 - [101] S. Woo, J. Park, J.-Y. Lee, and I. So Kweon, “Cbam: Convolutional block attention module,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19.
 - [102] D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
 - [103] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, “Improving neural networks by preventing co-adaptation of feature detectors,” *preprint arXiv:1207.0580*, 2012.
 - [104] A. G. Salazar-Gonzalez, Y. Li, and X. Liu, “Optic disc segmentation by incorporating blood vessel compensation,” in *Computational Intelligence In Medical Imaging (CIMI), 2011 IEEE Third International Workshop On*. IEEE, 2011, pp. 1–8.
 - [105] D. Marin, M. E. Gegundez-Arias, A. Suero, and J. M. Bravo, “Obtaining optic disc center and pixel region by automatic thresholding methods on morphologically processed fundus images,” *Computer methods and programs in biomedicine*, vol. 118, no. 2, pp. 173–185, 2015.
 - [106] T. Y. Wong, R. Klein, A. R. Sharrett, M. I. Schmidt, J. S. Pankow, D. J. Couper, B. E. Klein, L. D. Hubbard, B. B. Duncan, A. Investigators *et al.*, “Retinal arteriolar narrowing and risk of diabetes mellitus in middle-aged persons,” *Jama*, vol. 287, no. 19, pp. 2528–2533, 2002.

-
- [107] A. J. Houben, M. C. Canoy, H. A. Paling, P. J. Derhaag, and P. W. de Leeuw, "Quantitative analysis of retinal vascular changes in essential and renovascular hypertension," *Journal of hypertension*, vol. 13, no. 12, pp. 1729–1733, 1995.
 - [108] T. Y. Wong, R. Klein, A. R. Sharrett, B. B. Duncan, D. J. Couper, J. M. Tielsch, B. E. Klein, and L. D. Hubbard, "Retinal arteriolar narrowing and risk of coronary heart disease in men and women: the atherosclerosis risk in communities study," *Jama*, vol. 287, no. 9, pp. 1153–1159, 2002.
 - [109] R. Klein, B. E. Klein, S. E. Moss, T. Y. Wong, L. Hubbard, K. J. Cruickshanks, and M. Palta, "The relation of retinal vessel caliber to the incidence and progression of diabetic retinopathy: Xix: The wisconsin epidemiologic study of diabetic retinopathy," *Archives of ophthalmology*, vol. 122, no. 1, pp. 76–83, 2004.
 - [110] R. Estrada, C. Tomasi, S. C. Schmidler, and S. Farsiu, "Tree topology estimation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 8, pp. 1688–1701, 2015.
 - [111] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.
 - [112] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," *arXiv preprint arXiv:1508.04025*, 2015.
 - [113] Z. Zhang, P. Chen, M. Sapkota, and L. Yang, "Tandemnet: Distilling knowledge from medical images using diagnostic reports as optional semantic references," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 320–328.
 - [114] Z. Zhang, Y. Xie, F. Xing, M. McGough, and L. Yang, "Mdnet: A semantically and visually interpretable medical image diagnosis network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 6428–6436.
 - [115] X. Wang, Y. Peng, L. Lu, Z. Lu, and R. M. Summers, "Tienet: Text-image embedding network for common thorax disease classification and reporting in chest x-rays," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 9049–9058.
 - [116] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3156–3164.
 - [117] B. Zhao, J. Feng, X. Wu, and S. Yan, "A survey on deep learning-based fine-grained object classification and semantic segmentation," *International Journal of Automation and Computing*, vol. 14, no. 2, pp. 119–135, 2017.
 - [118] M. Ren and R. Zemel, "End-to-end instance segmentation and counting with recurrent attention. arxiv 2016," *arXiv preprint arXiv:1605.09410*, vol. 2.
 - [119] J. Lu, C. Xiong, D. Parikh, and R. Socher, "Knowing when to look: Adaptive attention via a visual sentinel for image captioning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 375–383.
 - [120] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," in *International conference on machine learning*. PMLR, 2015, pp. 2048–2057.

-
- [121] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.
- [122] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3146–3154.
- [123] R. Estrada, M. J. Allingham, P. S. Mettu, S. W. Cousins, C. Tomasi, and S. Farsiu, "Retinal artery-vein classification via topology estimation," *IEEE transactions on medical imaging*, vol. 34, no. 12, pp. 2518–2534, 2015.
- [124] Y. Zhao, J. Xie, H. Zhang, Y. Zheng, Y. Zhao, H. Qi, Y. Zhao, P. Su, J. Liu, and Y. Liu, "Retinal vascular network topology reconstruction and artery/vein classification via dominant set clustering," *IEEE transactions on medical imaging*, vol. 39, no. 2, pp. 341–356, 2019.
- [125] C. L. Srinidhi, P. Aparna, and J. Rajan, "Automated method for retinal artery/vein separation via graph search metaheuristic approach," *IEEE Transactions on Image Processing*, vol. 28, no. 6, pp. 2705–2718, 2019.
- [126] M. D. Abràmoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE reviews in biomedical engineering*, vol. 3, pp. 169–208, 2010.
- [127] A. K. Schuster, C. Erb, E. M. Hoffmann, T. Dietlein, and N. Pfeiffer, "The diagnosis and treatment of glaucoma," *Deutsches Ärzteblatt International*, vol. 117, no. 13, p. 225, 2020.
- [128] K. Guan, C. Hudson, T. Wong, M. Kisilevsky, R. K. Nrusimhadevara, W. C. Lam, M. Mandelcorn, R. G. Devenyi, and J. G. Flanagan, "Retinal hemodynamics in early diabetic macular edema," *Diabetes*, vol. 55, no. 3, pp. 813–818, 2006.
- [129] A. S. Neubauer, M. Luedtke, C. Haritoglou, S. Priglinger, and A. Kampik, "Retinal vessel analysis reproducibility in assessing cardiovascular disease," *Optometry and Vision Science*, vol. 85, no. 4, pp. E247–E254, 2008.
- [130] M. Niemeijer, X. Xu, A. Dumitrescu, P. Gupta, B. van Ginneken, J. Folk, and M. Abramoff, "Inspire-avr: Iowa normative set for processing images of the retina-artery vein ratio," 2011.

List of Publications:

Journals

- **Sandip Sadhukhan**, Arpita Sarkar, Debprasad Sinha, Goutam Kumar Ghorai, Gautam Sarkar and Ashis Kumar Dhara “Attention Based Fully Convolutional Neural Network for Simultaneous Detection and Segmentation of Optic Disc in Retinal Fundus Images” International Journal of Medical and Health Science Vol. 14 No: 8, pp. 200-204, 2020.
- **Sandip Sadhukhan**, Goutam Kumar Ghorai, Debprasad Sinha, Souvik Maiti, Gautam Sarkar and Ashis Kumar Dhara “Segmentation of optic disc in retinal images using fully convolutional network” Journal of Current Indian Eye Research Vol. 6, Issue 2, December 2019, pp. 40-47.

Journals communicated

- **Sandip Sadhukhan**, Debprasad Sinha, Ashis Kumar Dhara, Gautam Sarkar “Attention AV-Net for Artery and vein discrimination in Retinal Fundus Images”, Signal, Image and Video Processing Journal, Springer, 2023.

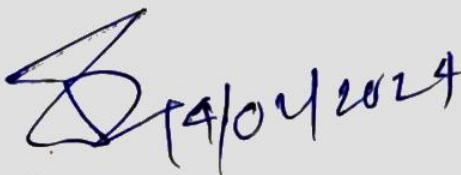
Conferences


- **Sandip Sadhukhan**, Goutam Kumar Ghorai, Souvik Maiti, Gautam Sarkar and Ashis Kumar Dhara, “Optic Disc Localization in Retinal Fundus Images using Faster R-CNN” Proceedings of Fifth International Conference on Emerging Applications of Information Technology (EAIT 2018), January 12-13, Kolkata, India.

- **Sandip Sadhukhan**, Goutam Kumar Ghorai, Souvik Maiti, Vikrant Anil-rao Karale, Gautam Sarkar and Ashis Kumar Dhara, "Optic Disc Segmentation in Retinal Fundus Images Using Fully Convolutional Network and Removal of False-Positives Based on Shape Features" Proceedings of 4th International Workshop, DLMIA 2018 and 8th International Workshop, ML-CDS 2018, September 20, 2018, Granada, Spain.
- **Sandip Sadhukhan**, G K Ghorai, S Maiti, D Sinha, G Sarkar and A K Dhara "Fully Convolutional Network for Segmentation of Optic Disc In Retinal Fundus Images" Proceedings of IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019) April 8-11, 2019, Venice, Italy.

Conference Communicated

- **Sandip Sadhukhan**, Ashis Kumar Dhara, and Gautam Sarkar, "Automatic Measurement of Arteriolar-to-Venular width Ratio (AVR) in Retinal Fundus Images" Proceedings of IEEE Calcutta Conference (CALCON), 2024, Kolkata, India.


Prof. Gautam Sarkar
 Electrical Engineering Department
 Jadavpur University
 Kolkata, INDIA


Dr. Ashis Kumar Dhara
 Assistant Professor
 Electrical Engineering Department
 National Institute of Technology
 DURGAPUR - 713209

Author's Biography

Sandip Sadhukhan is a PhD candidate from the Department of Electrical Engineering, at Jadavpur University, WB, India. He completed his B.Tech and M.Tech degrees in Computer Science & Engineering from the University of Calcutta. He is currently working at Dinabandhu Andrews Institute of Technology & Management, Kolkata as Assistant Professor since 2018. Prior to this, he worked in the IT industry for 18 years.

Contact Information

Email: sansad@gmail.com

Research Interests

Machine Learning, Medical Image Analysis.