BULGARIAN ACADEMY OF SCIENCES

INSTITUTE OF OPTICAL MATERIALS AND TECHNOLOGIES
"Acad. Jordan Malinowski"

SYNOPSIS

of the dissertation
for the education and science degree "doctor of philosophy"

Science field 4.1 "Physics", Science specialization "Physics of wave processes"

Deep Learning Models for Vascular Image Reconstruction and Analysis in Swept-Source Optical Coherence Tomography

Maryam Viqar

Supervisors:
Assoc. Prof. Violeta Madjarova, PhD (IOMT-BAS, Bulgaria)
Prof. DSc Elena Stoykova (IOMT-BAS, Bulgaria)
Dr. Erdem Sahin, PhD (Tampere University, Finland)
Prof. Atanas Gotchev, PhD (Tampere University, Finland)

Sofia, 2025

The dissertation was reviewed and approved for defence at a meeting of the Colloquium of the Institute of Optical Materials and Technologies Bulgarian Academy of Sciences (IOMT-BAS), held on ……..

The dissertation consists of a declaration of originality, a list of main symbols and abbreviations, five chapters including the introduction chapter, main contributions, a list of the author's publications and presentations, referenced in the dissertation, a list of citations of the author's publications and a bibliography of sources in total. The document comprises 128 pages, 33 figures, and 12 table.

The defence of the dissertation will take place on ……. at ……. in Room ….. of Block 109 of IOMT-BAS in an open session of the scientific jury, composed of:

List of key terms and main abbreviations

ABBREVIATION	MEANING
ADD	Addition layer
AMD	Age-related macular degeneration
AO-OCT	Adaptive optics OCT
ASPP	Atrous Spatial Pyramid Pooling
BN	Batch Normalization
CNN	Convolutional Neural Networks
CVD	Chronic venous diseases
DL	Deep-learning
DSP	Digital Signal Processing
ESDM	Efficient Segmentation-Denoising Model
FD	Fourier domain
FDML	Fourier Domain mode locked
FFT	Fast Fourier Transform
$\boldsymbol{FOV_{axial}}$	Axial field of view
$\boldsymbol{FOV_{lateral}}$	Lateral field of view
FWHM	Full Width at Half Maximum
GSPS	Giga Samples per Second
GVD	Group velocity dispersion
MCME	Multi-scale convolution mixture of experts
NA	Numerical aperture
NAION	Anterior ischemic optic neuropathy
NDFT	Non-uniform Discrete Fourier transform
OCT	Optical Coherence Tomography
PPM	Pyramid Pooling Module
PSF	Point spread function
PSNR	Peak Signal to Noise Ratio
RNFL	Retinal Nerve Fibre Layer
SD-OCT	Spectral Domain Optical Coherence Tomography
SNR	Signal-to-noise ratio
SSIM	Structural Similarity Index Metric
SS-OCT	Swept Source Optical Coherence Tomography
TD	Time-domain
TNode	Tomographic non-local-means despeckling
Up	Upsampling layers
WNNM	Weighted nuclear norm minimization

Chapter 1 Introduction

Optical Coherence Tomography (OCT) is a non-invasive imaging modality that acquires images of biological samples to generate volumetric data with micrometre resolution and a penetration depth of few millimetres. The full capabilities of OCT imaging remain underexplored in addressing challenges across many bio-medical fields. With the swift evolution of deep-learning (DL) techniques, automation of imaging tasks has witnessed significant progress, leading to more efficient workflows. The DL advanced bio-medical imaging is being progressively refined to bring forward automated methods that can extract meaningful information by rapid analysis on large data volumes even in real-time. The manual interpretation of such data is highly time-consuming and labour intensive especially in high-throughput clinical scenarios. In general, the automated methods provide high level of standardization and reproducibility, whereas human generated outcomes are usually prone to variability. Moreover, some subtle details and features can be overlooked by the human eyes.

Despite significant advancements in biomedical research through the deployment of state-of-the-art imaging modalities, numerous challenges remain unresolved across various domains. The domain of vascular studies is one such area where highly precise non-destructive imaging modality is required for precise analysis. Thorough evaluation of the veins is critical for understanding their functionality, to design effective veinous substitutes and to gain deeper insights into the biomechanical aspects of venous diseases. The segmentation maps are crucial for understand the shape and to extract meaningful information from this data. Existing state-of-the-art DL-based segmentation models suffer with high computational complexity, driven by the large number of parameters involved. This limits their applicability in the clinical settings. Another challenge is the development of an integrated model that is capable of segmentation and physical characteristic estimation (vein wall thickness). So that it can simultaneously achieve high-level semantic segmentation of varicose veins and estimate the thickness of the venous layers precisely. Current state-of-the-art approaches either handle the two tasks individually or compromise on the efficiency.

Additionally, there is considerable scope for enhancing the OCT processing pipeline to improve image reconstruction and quality, reduce computational complexity, and accelerate processing times. To obtain depth-resolved information, typical Swept Source OCT (SS-OCT) systems perform resampling in the wavenumber domain as a fundamental processing step. This process either demands additional hardware resources or increases the overall computational burden. To tackle the high computational demands while simultaneously generating high-fidelity images, this thesis analyses directly the raw OCT data that are non-linear in the wavenumber domain without required resampling in wavenumber domain. Along these directions, multidisciplinary approaches - combining signal processing, machine learning, and wave optics - are explored in this thesis to tackle key challenges and propose innovative computational methods for SS-OCT system.

The aim of the thesis is to address to the following research questions:

RQ1. How accurately can a DL model estimate segmentation maps of venous layers while maintaining low complexity through minimal number of model parameters, using the volumetric data acquired via SS-OCT?
RQ2. How can an integrated model be developed that simultaneously provides highly accurate segmentation maps - reliably identifying true positives and negatives, minimizing false detections - and provides accurate and consistent thickness estimations across diverse SS-OCT volumetric datasets of varicose veins?
RQ3. How can a DL network be developed to reconstruct SS-OCT images directly from data linear in wavelength domain, bypassing the k-linearization step, thereby reducing hardware cost and computational demands, while maintaining image quality comparable to standard SS-OCT processing pipeline?

RQ4. How can an enhanced reconstruction DL network be developed to optimize data in both spatial and Fourier domains, producing high-fidelity, speckle-reduced images with improved Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Metric (SSIM) compared to conventional processing of SS-OCT raw data?
RQ5. Is it feasible to develop a low-complexity DL-based alternative to the standard reconstruction pipeline that achieves high-fidelity image reconstructions with reduced complexity, enabling real-time imaging?
In summary, this thesis explores SS-OCT as a promising imaging modality for evaluating soft vascular structures, such as veins, and analysing their mechanical characteristics. The synergy between the SS-OCT pipeline and DL demonstrates that a strong potential as a robust reconstruction framework for addressing the above research questions. The methods developed throughout this thesis have shown high efficiency and performance, thereby advancing the state-of-the-art in the field of OCT.
The obtained results are presented in Chapter 3 (RQ1 and RQ2), Chapter 4 (RQ4-RQ5) and Chapter 5 provides the conclusions. Chapter 2 is dedicated to a literature survey on the historical background, the main areas of applications and the well-established algorithms.

Chapter 2 Background

OCT is a 3D imaging technique that employs low coherence interferometry to generate depth resolved images with micrometre scale resolution. Since its inception around 1990's [1], OCT has advanced into an indispensable medical imaging technique. Although it is primarily used in ophthalmology field, its applications are rapidly expanding in clinical domains ranging from domain of dermatology [2] to oncology [3], and many other medical fields [4]-[10]. OCT images contain the depth profiles that provide the information about the reflectivity from points within the scanned sample [11]. They are used to construct depth resolved cross-sectional 3D volume. As the refractive index, absorption and scattering properties vary across the layers of the sample, this is manifested in the form of varying intensity peaks across the interference profile [12].

The OCT systems can be broadly categorized into Time-domain (TD) and Fourier domain (FD) systems [11]. The depth profile of reflectivity in TD-OCT is obtained by mechanical scanning of the reference mirror. In FD-OCTs this profile is obtained by recording the interference pattern as a function of the light source's wavelength, and consecutively Fourier transform it [13], [14]. While the TD-OCT systems come with the potential advantages of cost and deeper penetration [15], [16], the FD-OCT systems provide much higher acquisition speeds [17], better SNR and higher axial resolution [18]. The FD-OCT systems can be further divided into SS-OCT and SD-OCT (Spectral domain OCT_ systems. The FD-OCT systems have tremendously contributed in terms of imaging speeds [19] and have evolved from the initial TD-OCT systems operating at nearly 2 A-scans per second to several million A-scans per second of the state-of-the-art SS-OCT systems [13]. Hence, the system of interest for this thesis is SS-OCT. In this thesis, for the acquisition of the OCT data, we used a commercial system from Optores GmBH with a FDML laser source operating at central wavelength of 1309 nm, bandwidth of 100 nm and sweeping frequency of 1.6 MHz [20], [21]. In the SS-OCT systems, the signal is recorded sequentially over time using a single detector forming the spectral interferogram. By applying a Fourier transform to this signal, the complete depth profile (axial structure) is reconstructed simultaneously, utilizing the full captured spectrum for image formation.

The advancements in the field of digital signal processing (DSP) allow faster implementations, improving the computational costs, real-time processing, high degree of reliability and scalability in realm of OCT processing. Here, we list the steps involved in data processing [22] of the SS-OCT systems: raw interferogram data acquisition, DC Removal, K-linearization, Dispersion Compensation, Windowing, 1D-FFT, Log Scaling, Denoising and Rendering. The signal in the SS-OCT systems is sampled uniformly in the wavelength domain, and hence uniformly sampled in the wavenumber domain. To retrieve the depth information from the spectral interferogram, Fourier transformation is required. For this transformation, the data should be linear in wavenumber domain.

Therefore, resampling and k-linearization is done in the k-space, to ensure high fidelity reconstruction, usually based on the different interpolation techniques to estimate the resampled points [23]-[25]. However, these techniques may introduce problems, for instance increased noise, imaging artifacts, and greater computational complexity, potentially impacting real-time imaging performance. Moreover, such methods often depend on a remapping function derived through an independent calibration step and may require additional hardware support. Hardware based techniques like k-clocks [26] often encounter issues especially when operating at rates exceeding 1.3 Giga Samples per Second (GSPS). The use of dual-channel acquisition in Micro-Electro-Mechanical Systems is highly effective [27], but it also introduces added system complexity. Laser based method [28] deploying akinetic lasers for wavenumber linearization is effective. However, the high cost associated with such lasers stand as a hurdle in the way to commercial deployment.

Recent progress made in the domain of machine vision driven by the advancement in deep-learning methodologies, has significantly transformed the several dimensions of image processing such as feature detection, enhancement, denoising, classification, reconstruction across a broad spectrum of imaging applications from healthcare to research-based biomedical applications. These models are not only capable of producing high-quality, visually appealing reconstructions but also of providing diagnostically relevant and interpretable scans. The state-of-the-art DL models are deployed for various image processing applications with U-Net [29], ResNet [30], SegNet [31], PSPNet (Pyramid Scene Parsing Network) [32], Deeplab3+ [33], attention [34], as their backbone architectures. While these architectures have significantly advanced the field of image processing, each has limitations when applied to the OCT data either in terms of computational complexity or low accuracy posed by challenges in capturing long-range dependencies and many more. Nevertheless, future advances in DL designs aim to meet the demands of specific application and the imaging sources.

Chapter 3 Analysis of veins using SS-OCT and Deep-learning

The studies in this chapter refer to author's papers P1-P2. They focus on the OCT based study of vascular structures found in a human body, the physical attributes and related issues such as the varicose veins. Methods are proposed for the shape and thickness estimation of varicose veins using DL. Several models exist in the literature that can perform segmentation for physical analysis of veins, but it's highly desirable to have an optimized model with improved computational complexity by minimizing the number of parameters, along with highly accurate segmentation results as presented in publication I. This chapter elaborates on how the direct calculation of the thickness from the semantic segmentation maps can be erroneous. The DL model proposed in the publications II, overcomes this problem using a transfer-learning based method in a unique manner, which enables to estimate the thickness of veins very precisely and accurately.

3.1 Introduction to veins and venous diseases

3.1.1 Veins and their importance

The human circulatory system consists of heart and the vessels that carry blood, oxygen and essential nutrients to various parts of the body. The arteries carry oxygenated blood and further divide into finer vessels called capillaries. Veins are responsible for carrying nearly 70% of the total blood volume and the cardiac activity is highly dependent on veins for its functionality. They can be affected by several diseases such as Chronic venous diseases commonly seen as the varicose veins. Hence, vein mechanics must be studied and analysed to better understand the factors that alter the properties of the veins in the diseased patients. Understanding the changes such as wall composition alterations, stiffness, etc. are crucial for gaining insights into disease progression and developing vein-specific innovations for treatment.

For all the commonly used mechanical tests, analysis of the parameters is based on geometry of the samples under test. These parameters can include estimation of the 3D shape, thickness,

circumference, length, cross-sectional area, etc. The primary variable parameters for analysis are shape and size. This requirement makes it's crucial to have a visual 3D shape model for the vascular structures under test, and to estimate the physical properties with high precision and accuracy. With the advancements in the field of imaging, the image-based 3D models can be easily built for the vascular structures to capture their morphology. This allows us to process the volumes and visualise across various planes at different angles. Accurate segmentation opens room to estimate dimensions such as diameter, thickness and cross-sectional area. The aim of this thesis was to acquire ex-vivo veinous volumes using an OCT system, to analyse, study and measure several physical characteristics of the veins in diseased patients suffering from varicose vein disease. DL based methods are further deployed to extract meaning full information in form of segmentation maps and thickness values from these volumetric scans described in sections below.

3.1.2 Veinous dataset: design, acquisition and processing

The 3D imaging models of veins are established using the SS-OCT modality to provide the spatially-organised structural information and to supplement the DL models with morphological details. To establish them, collaboration was done with Sofia Med University Hospital, Sofia, Bulgaria where surgical intervention was performed under supervision of Prof. Dr. Dimitar Nikolov, Head of Department of Vascular surgery. Ethical approval for this investigation was obtained from the Ethical Committee of the Bulgarian Academy of Sciences (permission no. 1-44/11.06.2021). All experiments were conducted in compliance with the principles set forth in the Declaration of Helsinki (1975), as amended in 2013.

In this study, the samples were obtained with the procedure of stripping the varicose vein from the diseased patients. Each patient was prepared as per the standard surgical protocol. The extracted veins were preserved using the saline solution at the discretion of the doctor. Within the 24 hours of the surgical procedure, the veins were used for volumetric data acquisition. The sample was placed in a sterile petri dish for precise imaging avoiding contaminations. During the acquisition procedure, the sample was moistened using the solution to prevent surface dryness. By carefully maintaining the control conditions, the vein's physical and optical characteristics were preserved, to perform precise OCT image acquisition.

The venous data were collected using the SS-OCT system from Optores GmbH [20],[21] as described in Chapter 2. The venous dataset comprises five different volumes, containing 1300 B-scans acquired from four patients. The details regarding the sex and age of the patients and the number of volumes acquired from each patient for the venous dataset can be found in Table 3.1. Each B-scan, encompassing 1024 A-scans, was subsequently cropped, and resized to 512 × 256 pixels to focus on the region of interest and address computational constraints while processing. The images are in the form of B-scans and they are affected by noise, mainly speckle noise. Hence

Table 3.1 Patient description for the venous dataset

Patient	Sex	Age	Volumes	Vein	B-scans (each volume)
P1	Male	57	1	Saphenous Vein	(200)
P2	Male	68	2	Saphenous Vein	(650,200)
P3	Male	33	1	Saphenous Vein	(100)
P4	Female	38	1	Saphenous Vein	(150)

denoising is performed using the weighted nuclear norm minimization (WNNM) [35].

3.2. Segmentation and Thickness measurement in Veins

This section demonstrates the problem formulation for segmentation of veins followed by the method presented for performing segmentation of veinous layers by optimization of the number of parameters. The segmentation approach aims at segregation of the inner and outer boundaries.

Further, the model is leveraged to accurately measure the thickness as a key physical parameter essential for understanding the functionality of the veins or their artificial substitutes. The doctors can delineate the layers and structures by segmentation of each slice from the OCT volume manually. However, this is highly time-consuming and demands a trained expert. There is an additional challenge in real-time applications. With the presence of AI-driven segmentation techniques for biomedical imaging, this task can be efficiently performed using a suitable model. Various state-of-the-art approaches based on DL are also investigated to perform pixel-wise segmentation and to obtain the thickness value using regression analysis.
The U-Net [29] is widely recognized for effective segmentation performance in the case of a small sized dataset, but its highly complex architecture incorporating millions of parameters imposes a significant burden on the hardware resources. Reducing the number parameters allows faster inference, smaller model size and use of less memory when handling big batches or higher resolution images during training. Another key challenge affecting the segmentation in encoder-decoder based architectures is direct upsampling of regions derived from the max-pooling performed on the encoder side, which possess a limited receptive field.

To establish a highly optimized DL model, we introduce a refined encoder-decoder based architecture as presented in Publication I. It is designed to reduce the complexity by minimizing the number of the parameters while ensuring accurate pixel-wise segmentation of veins. This is achieved by introducing a parallel-bridge block embedded with the atrous and depth-wise separable convolution layers. The model is also supplemented with residual connections [30] to mitigate the problem of vanishing gradients commonly encountered in the deeper architectures. Such connections between different network layers facilitate seamless information flow to successive layers. This improved flow of information is accomplished by propagating feature mappings from one layer to subsequent layers through identity mappings. Mathematically, the regular convolution can be represented using the input $X \in Z^{H \times W \times D}$, where H, W, and D denote the dimensions of the feature map and W_g is a filter of size (k^f, l), as follows:

$$\sigma_C(X, W_g)_{(i,j)} = \sum_{k^f,l,m}^{K^f,L,M} X(k^f, l, m) * W_g(i + k^f, j + l, m), \tag{3.1}$$

In Eq. (3.1), m represents the channels of the input, and the position on the input feature map is marked by indices (i,j). The atrous or dilated convolution with a dilation factor of r, can be expressed as:

$$\alpha(X, W_g)_{(i,j)} = \sum_{k^f,l}^{K^f,L} X(i + rk^f, j + rl) * W_g(k^f, l), \tag{3.2}$$

Systematic dilation using atrous convolution enables exponential enlargement of the receptive field while preserving both resolution and coverage. This facilitates the integration of contextual information across multiple scales. It is a sampling process wherein the sampling rate is expressed as the dilation value. The use of dilation enables the extraction of high-level relationships, helping to mitigate accuracy loss caused by pooling and upsampling processes. In Fig.3.1, the dilated

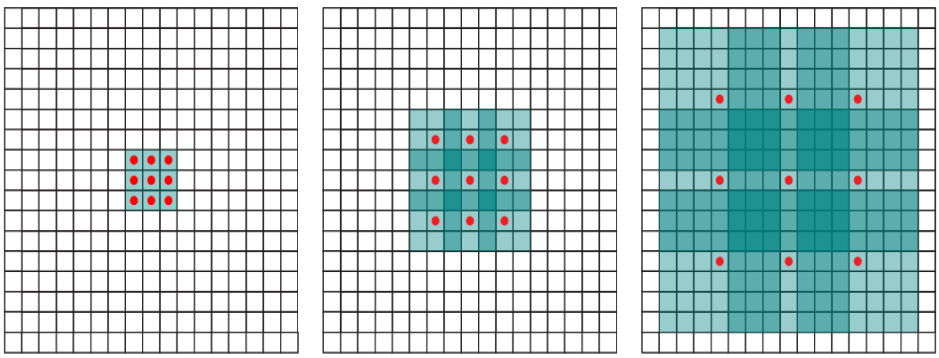

Figure 3.1 Dilated or atrous convolution (a) 1-dilated convolution with a receptive field of 3 × 3 (b) 2-dilated convolution with a receptive field of 7 × 7 (c) 4-dilated convolution with receptive field of 15 × 15

convolution is shown having different receptive fields and dilation rates.

The key distinction between the standard convolution and the depthwise convolution is how the convolution kernel interacts with input channels. The depthwise separable convolution as shown in Fig.3.2 follows a 2-step operation where the first step is to independently apply the spatial convolution to each channel of the input, followed by second step of 1×1 pointwise convolution, which maps the output channels from the depthwise convolution into a new channel space.
The two-step process involving pointwise and depthwise convolution is described in Eqs. (3.3) and (3.4) respectively as follows:

$$P(X, W_g)_{(i,j)} = \sum_m^M X(m) * W_g(i, j, m), \tag{3.3}$$

$$\partial(X, W_g)_{(i,j)} = \sum_{k^f,l}^{K^f,L} X(k, l) * W_g(i + k^f, j + l). \tag{3.4}$$

The depthwise separable convolution can be written as :

$$\delta(X_p, X_{(d)}, W_g)_{(i,j)} = P_{(i,j)}\left(X_p, \partial_{(i,j)}(X_{(d)}, W_g)\right). \tag{3.5}$$

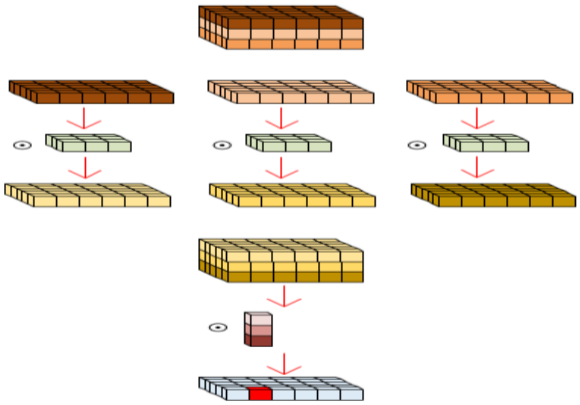

Figure 3.2 Depthwise separable convolution

3.2.1 Optimization for Segmentation of vein images

To design an efficient segmentation model, optimization is crucial by reducing the complexity, one of the aims in this thesis. A model is proposed here that is an encoder-decoder built upon the U-Net [29] architecture. The model is referred to as Opto-UNet [PI] and is proposed in Publication I. It incorporates a parallel bridge block that utilizes atrous and separable convolutions. This integration enhances the extraction of spatially extensive and distinct feature maps, resulting in improved performance compared to the state-of-art models. Additionally, the network is optimized with the use of depthwise separable convolution. This minimizes the number of parameters and

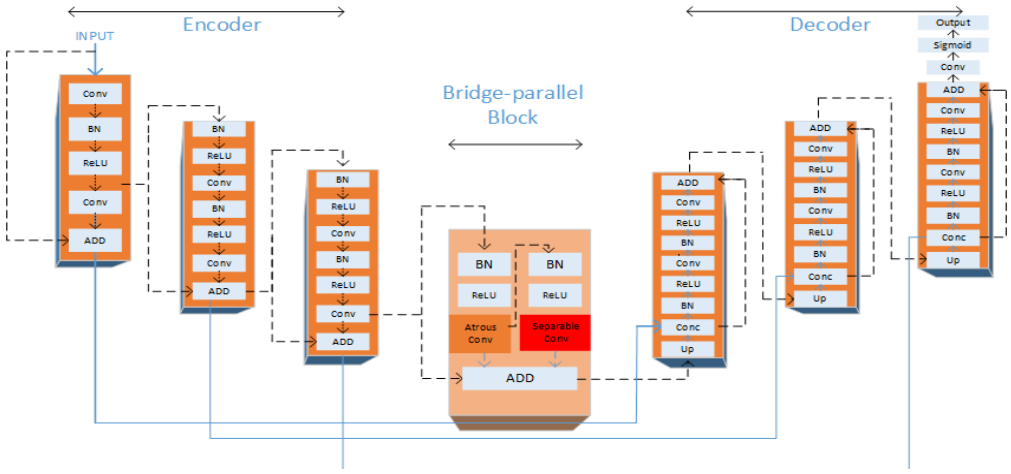

Figure 3.3 Block diagram of the model Opto-UNet for varicose vein segmentation (P.I)

significantly reduces the model's complexity.

The block diagram of the Opto-UNet is shown in Fig.3.3. The three main parts as highlighted here are the encoder, the decoder and the bridge parallel block. Such deeper networks offer the benefit of enhanced accuracy, however during training, they may also experience performance degradation primarily due to the prevalent issue of vanishing gradients. To address the issue, we incorporate residual units in the proposed network, enabling additional transfer of information between the subsequent layers in deep networks as shown in Fig.3.3 using black dashed lines. Each of the three blocks on the encoder side has different layers, namely Convolution (Conv), Batch Normalization (BN), Rectified Linear Unit (ReLU), and Addition layer (ADD). In addition to these layers, the three blocks on the decoder side also have Upsampling layers (Up) to perform the expansion of the feature map using upsampling operation and Concatenate layer (Conc) to skip connections between the encoder and decoder blocks marked with the blue lines. The crucial part of the proposed model is the bridge-parallel block placed at the centre of the network as shown in Fig. 3.3. Here the feature maps are in most compact form retaining the most crucial contextual information. It includes the atrous (orange block) and depthwise separable convolution (red block) in addition to BN and ReLU functions. The feature maps at the output of the bridge-parallel block are derived from two sources: one capturing spatial information and the other incorporating both spatial and depth details. This enables the model to learn diverse characteristics simultaneously. The block diagram illustrates that the outputs of the two newly introduced layers (atrous and separable convolution), along with the identity mapping operation, are combined in the final layer through addition. The output from the bridge-parallel block goes towards the decoder blocks. The decoder blocks perform the expansion of these feature maps and uses the information from the encoder blocks via skip connections to achieve precise localization of structural details. All the convolutions (Conv) within the encoder and decoder and parallel bridge blocks employ a 2D kernel of size 3×3, with a stride 1×1, padding is set to "same". The dilation rate is 2 for the atrous convolution.

The dataset for training the network consists of input and ground truth images from the FDML SS-OCT system by Optores GmbH, described in Chapter 2. The input images were cropped and resized B-scans containing 512×256 pixels obtained from the original OCT B-scans of size 1024×1152 due to hardware limitations. Intensities were normalized for all the input images in the range from 0 to 1. The total number of images used for this network was 800 B-scans, that were further divided into training, validation and testing images as 600, 100 and 100 respectively. The ground truth images were obtained by manually annotating the B-scans for the region of interest under the supervision of Prof. Dr. Dimitar Nikolov from the Department of Vascular surgery, also involved in vein extraction described in section 2.1. Certain areas of the veinous layer were excluded from the segmentation masks annotation, due to challenges in accurately distinguishing the border pixels. These ground truth images can be represented using $y_i \in \{0,1\}$ for each pixel value having total number of N^p pixels. The loss function used here is the Dice loss derived from Sorensen-Dice coefficient. This is employed to minimize the loss between the predicted output and the desired output (the ground truth). This loss is well known for handling the issue of class imbalance problem during segmentation. As in this work, the foreground occupies substantially less area compared to background, hence we employ the Dice-coefficient loss. The Dice coefficient can be expressed as follows:

$$DCC = \frac{\sum_{i=1}^{N^p} p_i y_i + \epsilon}{\sum_{i=1}^{N^p} p_i + y_i + \epsilon} + \frac{\sum_{i=1}^{N^p} (1-p_i)(1-y_i) + \epsilon}{\sum_{i=1}^{N^p} 2 - p_i - y_i + \epsilon}, \tag{3.6}$$

Here, in Eq. (3.6) for DCC, the $p_i \in [0,1]$ represents the predicted image pixel value, ϵ=1 is the smoothing factor to avoid zero denominator division. The DCC contains two terms - one for the foreground and the other for the object, but we take into consideration only the loss accounting for the object. The Dice-coefficient loss ($Loss^1$) can be written as:

$$Loss^1 = 1 - DC' \tag{3.7}$$

Here DC' in Eq.(3.7) includes only the first term of DC from the Eq.(3.6) considering only the foreground region loss. The proposed Opto-UNet model was executed on an NVIDIA GeForce RTX 3060 GPU featuring 12 GB of graphics memory and 64 GB of RAM. The training and testing

processes were carried out using the TensorFlow 2.0 framework, employing the RMSprop optimizer with a learning rate of 0.0001 using the Dice-coefficient loss ($Loss^1$) with a batch size of 16.

3.2.2 Experimental results and discussion

To analyse the performance of the proposed model Opto-UNet, we conducted experiments on the acquired venous datasets. The outcomes derived from the proposed model are presented upon systematic evaluations and compared with the prominent architectures namely U-Net [29], SegNet [31], and PSPNet [32]. These models were trained under the same conditions and environment as the proposed model Opto-UNet to ensure a fair comparison. The comparison with other segmentation models is based on key evaluation metrics including accuracy (Acc), intersection over union (IoU), Dice similarity coefficient (DSC), sensitivity (SEN), and specificity (SPE), which are mathematically given as:

$$\text{Acc}=\frac{TP+TN}{TP+TN+FP+FN}, \tag{3.8}$$

$$\text{IoU} =\frac{TP}{TP+FP+FN}, \tag{3.9}$$

$$\text{SEN}=\frac{TP}{TP+FN}, \tag{3.10}$$

$$\text{SPE}=\frac{TN}{TN+FP}, \tag{3.11}$$

$$\text{DSC}=\frac{2*SEN*SPE}{SEN+SPE}. \tag{3.12}$$

In Eqs. (3.8)-(3.12), TP, TN, FP, FN represent the number of pixels that are True Positive, True Negative, False Positive and False Negative respectively, for the output obtained from the model. Table 3.2 provides a summary of the different models and their performance for the varicose vein dataset, facilitating a comparison between Opto-UNet and state-of-the-art models. Overall, the model Opto-UNet achieves the best segmentation performance for all the metrics highlighted in bold in Table 3.2. It achieves a segmentation accuracy of 0.9830, highly above SegNet and comparable with the U-Net model. The number of the parameters used is significantly low compared to all other models. From this table, it can clearly be inferred that the performance of SegNet [31] is significantly worse particularly the values of SEN and IoU metrics, owing to high rate of false detections. While the proposed model exhibits slight improvement for all the evaluation metrics compared to the U-Net [29], expectation for a notable improvement in efficiency by using nearly 3.5 times fewer parameters is achieved. To visualise the results of the Opto-UNet model, we show in Fig. 3.4 (a) the

Table 3.2 Comparison of segmentation results for varicose veins (P.I)

Model	Acc	SEN	SPE	DSC	IoU	PARAMETERS
U-Net	0.9791	0.8346	0.9968	0.9085	0.8210	31.04M
SegNet	0.8868	0.1340	0.9031	0.2333	0.0245	29.45M
PSPNet	0.9619	0.8142	0.9813	0.8899	0.7127	48.70M
Opto-UNet	0.9830	0.8425	0.9980	0.9136	0.8395	8.54M

original image (resized B-scan), generated from the OCT system, and images in (b),(c) and (d)

contain three different scans, with upper and lower boundary delineated using blue and orange colours.
The arrows, shown in Fig.3.4(c) and (d), are used to demonstrate the accuracy of the segmentation model when the B-scans are afflicted with artifacts. It is also to be noted that the background region is afflicted with speckle noise even after smoothing the images by applying WNNM denoising [35]. The Opto-UNet model overall demonstrates highly precise performance, effectively handling the noise. To highlight the impact of the key functions performed within the parallel-bridge block by incorporating (i) atrous convolution, (ii) separable convolution, and (iii) addition, the output of individual channels for each layer is depicted in Figs.3.5(b), (c), and (d), respectively. These visualizations provide insight into the Opto-UNet model illustrating how the input image propagates through the various layers within the parallel-bridge block. The advantage of integrating these

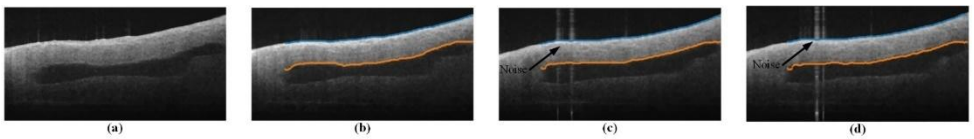

Figure 3.4 (a) Varicose vein B-scan using SS-OCT (b),(c),(d) Segmentation using Opto-UNet (P.I)

convolution layers for extracting structural details at spatial and depth-levels is evident in the feature map in Fig.3.5(d) derived after the addition operation. Further, in Fig.3.5(e), we illustrate the variation in loss and accuracy during model training and validation corresponding to the epochs. As the number of the epochs increases, the training loss (referred as loss in Fig.3.5(e)) gradually diminishes and around 15 epochs it is stabilized. A similar trend is observed in the accuracy curve, indicating a rapid convergence of the Opto-UNet model.

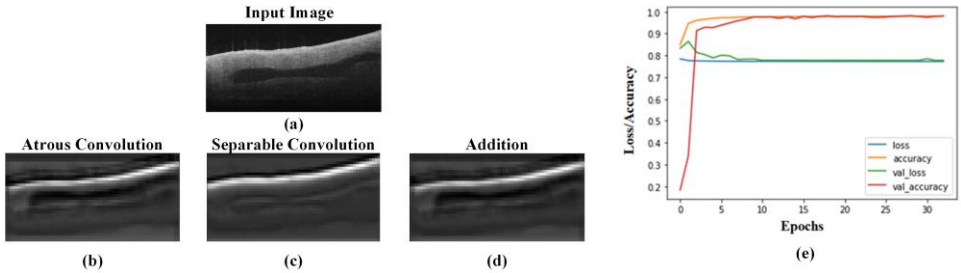

Figure 3.5 (a) Original Varicose vein B-scan using SS-OCT; (b),(c),(d) Segmentation using Opto-UNet, (e) loss and accuracy curves during training and validation of the Opto-UNet (P.I)

To summarize, we were able to achieve better results by incorporating layers with the parallel-bridge block, along with residual connection and implementing the Dice-coefficient loss in the proposed Opto-UNet model. The reduction in the number of parameters used, compared to state-of-the-art models is quite significant. Overall, the proposed model demonstrates fast convergence and high accuracy, making it a promising solution for varicose vein segmentation. However, as one of the initial experimental studies in this thesis, the main limitation of the work done in Publication I lies in the dataset's scope, which includes images from only a single patient diagnosed with varicose vein disease. This approach effectively preserves the separation of information across depth channels at various spatial positions, resulting in a more informative and deeper network. We investigated on thickness estimation for the veins. The thickness can be directly estimated using the segmentation map obtained from the different segmentation models [29]-[33],[36]. However, this can cause error accumulation due to potential inaccuracies in these maps. We put forward a dual-prediction model based on self-transfer learning to predict the thickness along with the segmentation map. The transfer learning approach was used to refine the features learned during the initial optimization phase for segmentation. This ensemble model is capable of precisely predicting the thickness value with high accuracy in addition to segmentation maps trained sequentially employing

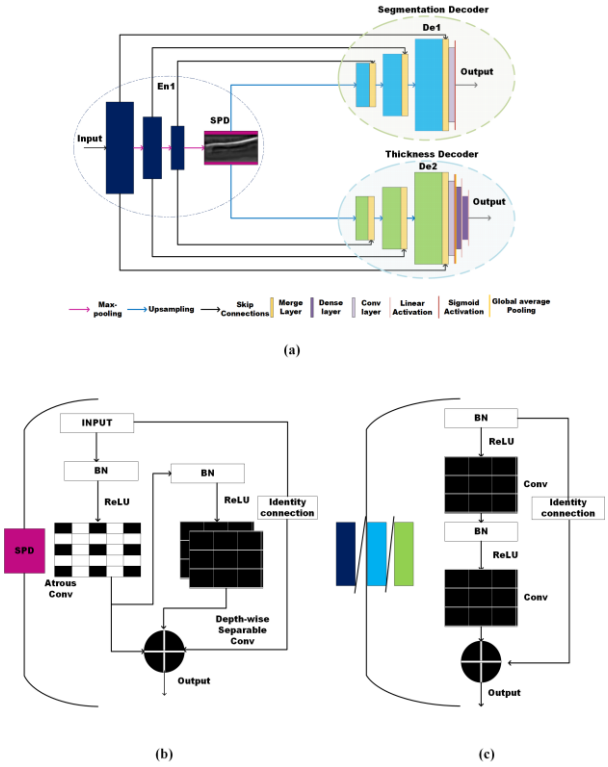

Figure 3.6 (a) Block diagram of the proposed model showing the overall architecture of the encoder and the two decoders, (b) Sensitive Pattern Detector module, (c) the encoder or the decoder blocks with the layers and flow of information. (P.II)

a single encoder and dual decoder architecture. Section 3.3 describes the proposed model for segmentation and simultaneously thickness estimation on a larger dataset.

3.3. Segmentation and Thickness measurement using Transfer Learning

To enhance the understanding of the venous characteristics, we have developed an automated model designed for simultaneous segmentation and thickness estimation that can estimate the shape and thickness for the region of interest. This ensemble model is named as TREE (TransfeR learning-based approach for thicknEss Estimation). It follows a modified encoder-decoder architecture where a single encoder serves as a backbone, branching into two decoders. These decoders respectively specialize in learning to predict the segmentation map and thickness values.

The network architecture and the experimental details are discussed below. Fig.3.6 presents the model -TREE, which follows a single-encoder and dual-decoder design. The main architecture of the encoder-decoder here, is inspired by the U-Net [29]. The bottleneck module between the encoder and the decoder in the U-Net [29] contains high-level structural details and plays a crucial role in generating accurate predictions. The proposed TREE model uniquely utilizes these features to facilitate transfer of the details from the segmentation task to thickness estimation. It consists of three blocks: a single encoder and two decoders. The decoder named as De1 (shown in Fig.3.6 (a)) produces the segmentation map using semantic segmentation of pixels, while the other decoder named as De2 (shown in Fig.3.6 (a)) estimates vein thickness. The encoder named as En (shown in Fig.3.6 (a)) is linked to both these decoders, creating two distinct branches: En-De1 and En-De2. The encoder's output is passed through a bottleneck module (shown in pink in Fig. 3.6 (a)), that serves both decoders. This bottleneck module enhances model performance by focusing on the region of interest within each B-scan.

As seen in Fig.3.6(a), the encoder (En) and the two decoders (De1, De2) contain modules highlighted using dark-blue, light blue and light green rectangular boxes respectively. Further, these modules are elaborated in Fig. 3.6 (c), to demonstrate the stack of layers within them. All these modules follow the process as: (i) Normalization using the Binary Normalization (BN) layer, followed by ReLU activation, (ii) two consecutive convolution (Conv) operations with

intermediatory operations using BN and ReLU layers, (iii) addition of the output feature map to the input map, using identity mapping connections known as residual connections [30], as shown in Fig.3.6 (c). These connections help the network achieve consistent training as the count of layers increases when the model becomes deeper. They mainly address the vanishing or exploding gradient problem.

A block named as Sensitive Pattern Detector (SPD) is placed in this model as the junction of the encoder and decoder branches shown as a pink box in Fig.3.6 (a) and (b). On the encoder (En) side, each module is followed by a max pooling layer, to contract the size of feature maps by using the maximum value for a given region. After passing through the SPD block, the feature map is enriched with global information and progresses towards the decoder side for expansion of feature maps using upsampling. The max-pooling and the up-sampling layers are designed to extract details at multiple resolutions. These global and local details from the encoder and decoder sides respectively are fused together to promote the flow of information at the same resolution of maps.

Following from the mathematical definition of regular convolution given by Eq. (3.1), we can write the over-all processing of the encoder block / decoder block as follows:

$$Y_{en} = \sigma_C(ReLU(BN(\sigma_C(ReLU(BN(X)), W_g\ ), W_g\ )(ReLU(BN(X)), W_g\ ). \quad (3.13)$$

To obtain the global structural information defining the boundaries or layers, the output from each encoder block, is down sampled by employing the max-pooling layer. After passing through the encoder blocks, the output passes through the bottle-neck junction that consists of a special combination of atrous (dilation rate=2) and depthwise separable convolution as described above in (Publication I). Standard convolution operation puts constrains on the feature learning mechanism due to its limited receptive field. This leads to restricting the information exchange to a small region defined by the kernel size (typically 3×3). This poses challenges when segmentation masks involve long-range spatial relationships. The veins exhibit elongated structures that require capturing extended dependencies. To address this issue, depthwise separable and atrous convolutions are utilized to extract deeper and long-ranged feature maps, enabling the model to learn correlations more effectively. As shown in Fig.3.6, the SPD, contains BN and ReLU activation functions along with the depthwise separable and atrous layers in parallel. In contrast to the bottleneck module in Publication I, the depthwise separable convolution is also supplemented with the same dilation rate (equal to 2) as in atrous convolution to preserve the opaqueness and transparency of these two layers during their integration upon addition operation. After getting the integrated output from the convolution operations at the last layer, the final feature map is obtained by combining them with a residual connection to enhance feature representation as shown in Fig.3.6. Following from Eq. (3.5), the overall expression for the SPD block can be formulated as:

$$Y_{bb} = \alpha(ReLU(BN(X)), W) + \delta'(ReLU(BN(X)), W) + \varphi(BN(X), W). \quad (3.14)$$

In Eq. (3.14), we use α, δ', φ to represent the atrous, depthwise separable (the same as δ from Eq. (3.5), along with dilation rate equal to 2 and residual operation respectively. The global level features extracted with the help of the SPD block facilitate the learning process for both the segmentation and thickness estimation of varicose veins. To achieve pixel-level accuracy, these feature maps are subsequently upsampled, ensuring precise segmentation and thickness estimation. The decoder De1 and De2 blocks are shown in Figs.3.6 (a) and (c) using blue and green colours respectively. In De1, the final layers after the three main decoder blocks perform convolution, and Sigmoid activation aims at generation of segmentation maps. On the other hand, as shown in Fig. 3.6 (a), in De2, the blocks are followed by a different combination of convolution, Sigmoid activation, Global average pooling and dense layers, designed for obtaining a thickness estimate. It is to be noted that all convolutions within the encoder and decoder blocks, including the SPD module, employ a 2D kernel of 3×3, with a stride of 1×1 and padding set to maintain the same spatial dimensions.

To capture the salient feature, required for segmentation and thickness estimation, the learning of the proposed model TREE is carried out in a step-by-step process as depicted in Fig.3.7. Firstly, the model is trained for the segmentation map using the branch 1 that incorporates the En and De1. The estimation of the thickness values is subsequently carried out by training the second branch involving En and De2. The sequential training takes the pixel-level segmentation map coming from SPD and En to further advance through the layers of the second decoder (De2) and effectively learn the details required for precise thickness estimation. In the second training phase, only decoder-De2 is trained to capture new features for accurate thickness estimation. At this stage, the weights for the encoder (En) and SPD are inherited from the initial learning meant for segmentation.

In addition to the distinctive architecture of the proposed model, we also introduce an innovative method of training the two decoders in a sequential manner. Initially, the first encoder and decoder combination learn to predict segmentation and afterwards the second decoder is exclusively trained to estimate the thickness. In this manner, we deploy a transfer learning approach to support the thickness estimation using the features already learnt during segmentation training.

3.3.1 Experimental results and discussion

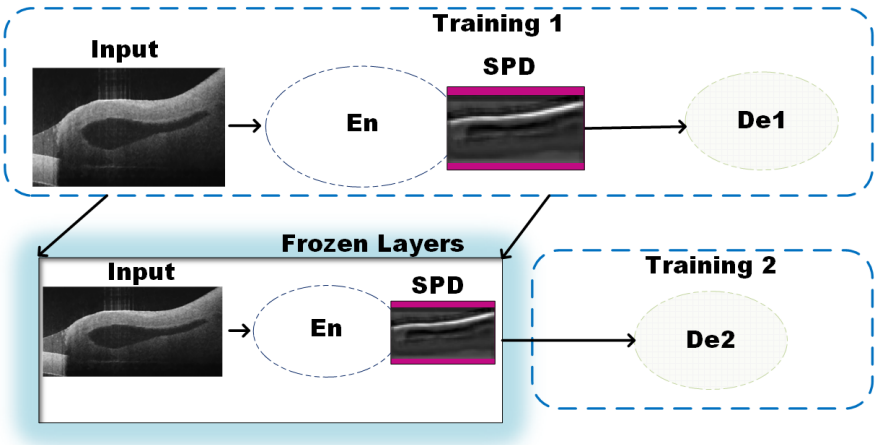

Figure 3.7 (a) Block diagram demonstrating the training phase of the proposed model incorporating the Encoder En and the two decoders De1, De2 respectively. (P.II)

In this section, we illustrate the performance of the proposed model TREE and compare it with that of the state-of-the-art models, mainly U-Net [29], PSPNet [32], RESUNET [36], DeepLab3+[33], and SegNet [31] for varicose vein dataset segmentation and thickness estimation. The dataset used for this study includes 5 volumes containing 1300 images (B-scans) obtained from 4 different patients as described in Section 3.1. The size of these images after cropping and resizing was 512 × 256, to accommodate the region of interest and to consider the computational constraints. They were processed using the denoising method [35] and then the images were normalised in range [0,1]. The main dataset, containing four volumes, was randomly partitioned into training, validation, and testing sets in proportions of 70%, 20%, 10% respectively. The fifth volume (referred to as the independent-test set) containing 200 images was randomly chosen to test for generalization capability of the model. This volume was not exposed to the model during training or validation and was only tested to represent the independent-test results. Each B-scan of the dataset represented different veins, demonstrating various shapes and sizes. To enhance dataset diversity, geometric transformations such as translation and flipping were applied during the training phase. The images were translated by 20% of their height and width, while flipping was performed in both horizontal and vertical orientations. The segmentation maps, serving as ground truth and shown in Fig.3.8, were manually outlined under the supervision of a vascular surgery expert Prof. Dr. Dimitar Nikolov from Department of Vascular surgery, also referred in section 3.1. Due to limitations associated with the depth penetration in SS-OCT system used for data acquisition, the lower surface remains obscured. Hence, in this work only the top 2D layer of the vein is considered. Subsequently, the ground truth thickness for each row along the axial direction is computed, forming what is referred to as the thickness map.

The system used for the experiment was equipped with an AMD Ryzen 7 processor, an NVIDIA GeForce RTX 3060 GPU, 32 GB RAM, and a 1TB SSD, utilizing the Keras API within the TensorFlow framework. The optimizer employed was Adam, having a learning rate of 0.0001. The batch size was 16, and a maximum of 200 epochs was used for training. To ensure a fair comparison, these training parameters were consistently applied across all state-of-the-art models. To guide the optimization process of the proposed model TREE, we incorporate two different loss functions to address two different problems. Initially, the ensemble of encoder (En) and decoder (De1) are trained by minimization of the Binary cross-entropy loss (Loss1) function to accurately perform the segmentation of venous layers, mathematically represented as:

$$Loss1 = -\frac{1}{N_b}\sum_{i=1}^{N_b}(Y_i^g.\log\left(\mathrm{p}(\hat{Y}_i)\right) + \left(1 - Y_i^g\right).\log\left(1 - p(\hat{Y}_i)\right) \tag{3.15}$$

Here, in Eq.(3.15), N_b is the total number of pixels, $\hat{Y}_i$ is the predicted probability for ith pixel, Y_i^g ground truth label for ith pixel. This loss (Loss1) is optimized between the output of the ensemble (En-De1) and the manually annotated mask (ground truth) whose pixels are classified as 0 or 1, depicting the foreground and object regions, respectively.

$$Loss2 = \frac{1}{N^C}\sum_{c=1}^{N^C}(Y_c - X_c)^2. \tag{3.16}$$

In the next phase of training as described in section 3.3 and using Fig.3.7, the ensemble of (En-De2) is trained using the Mean Squared Error (MSE) referred to as *Loss2* as given in Eq.(3.16), to estimate

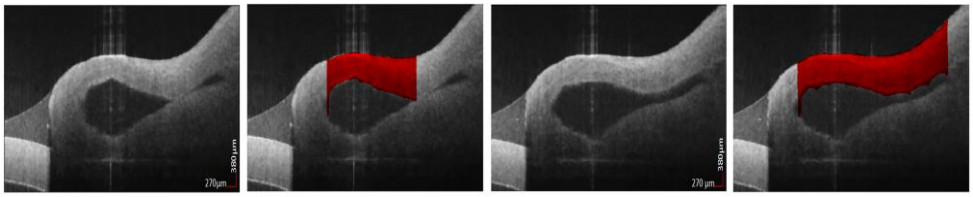

Figure 3.8 (a) original images and their segmentation map overlayed on the original image in red (P.II)

the thickness of the vein where N^C, Y_c and X_c represent total number of columns in the image, the ground truth and the predicted thickness (pixels) for the c^{th} column respectively. It is to be noted

that the weights are frozen for the En block along with SPD module during this second phase of training. Hence, the output of the SPD block based on the weights learnt during segmentation optimization serves as the input to De2 during its optimization stage.

To assess the performance and to compare with the state-of-the-art works we use the following performance indicators: Accuracy, Area Under receiver Operator (AUC), Intersection over Union (IOU), TNR, Recall, Precision and F1 score. We incorporate these metrics to compare the results of the proposed model TREE to the state-of-the-art works namely: U-Net [29], SegNet [31], RESUNET [36], PSPNet [32], and DeepLab3+[33]. All models were trained under identical conditions as the proposed model for fair evaluation. To evaluate the segmentation results, the metrics are illustrated using line plots in Fig.3.9 for an objective comparison. The TREE model outperforms state-of-the-art methods across all indicators, except for the precision metric, where only the SegNet [31] model shows a slight advantage.

The thickness was calculated for the ground truth images and the state-of-the-art methods as follows. Firstly, to train the model TREE for thickness estimation, we use the ground truth segmentation maps marked under supervision of the expert to estimate the thickness value for each column falling under the region of interest. For the state-of-the-art works, U-Net [29], SegNet [31], RESUNET [36], PSPNet [32], and DeepLab3+[33], the thickness is calculated from their output segmentation masks. The pixels demarcating the two layers of region of interest i.e. the top and bottom layers (white pixels with pixel value y=1) were segregated from the segmentation map. The $\boldsymbol{M_{I\times J}}$ matrix is used here to represent the segmentation map, where each entity of the map is represented using $\boldsymbol{m_{ij}}$ where i and j represent the row and the column respectively. Let C_j represent a column of the image matrix, and let E^{kl} be 1-D matrix to represent the boundary of the region of interest expressed as follows:

$$C_j = [m_{1j}, m_{1j}, \ldots\ldots\ldots\ldots m_{I\times j}],^T \tag{3.17}$$

$$E^{kl} = [e_1, e_2, \ldots\ldots\ldots\ldots e_n]. \tag{3.18}$$

In Eq.(3.18), $kl \in$ {Layer1, Layer 2} and the coordinates of both Layers 1 and 2 (L1 and L2) in column C_j can be written as:

$$B^{L1} = \arg\min \text{i}\, C_j, \tag{3.19}$$

$$B^{L2} = \arg\max \text{i}\, C_j. \tag{3.20}$$

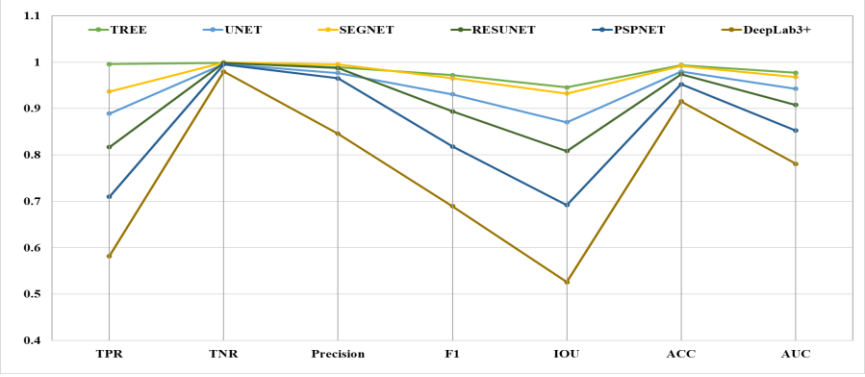

Figure 3.9 Plots of segmentation metrics. (P.II)

The thickness value (pixels) from the co-ordinates of columns estimated in Eqs. (3.19) and (3.20), we can calculate as follows:

$$N = |B^{L1} - B^{L2}|. \tag{3.21}$$

The vein thickness is represented as a row vector of dimensions 1×n, where n varies based on the spatial coverage of the region of interest within the vein. To evaluate the relationship between the predicted and actual thickness values determined using Eq. (3.21), error metrics were computed on the test dataset. These metrics include Mean Square Error (MSE), Mean Absolute Error (MAE),

Bias, Standard Deviation (SD), and Root Mean Square Error (RMSE). By analysing these standard thickness error measures, the accuracy and reliability of the proposed TREE model's predictions are assessed and compared to state-of-the-art works U-Net [29], SegNet [31], RESUNET [36], PSPNet [32], and DeepLab3+ [33] with respect to the ground truth (mask). These error results are tabulated in Table 3.3. It can be observed that the proposed model TREE exhibits superior accuracy in thickness estimation, showing minimal errors across various metrics. While its SD (1.541) is comparable to SegNet [31] (1.485), at the same time, it outperforms SegNet [31] over other error

Table 3.4 Thickness (mm) for segmentation outputs of DeepLab3+ [33], U-Net [29], PSPNet [32], RESUNET [36], SegNet [31], ground truth (mask) and thickness predicted using TREE models. (P.II)

Method	DeepLab3+	U-Net	PSPNet	RESUNET	SegNet	MASK	TREE
Thickness(mm)	0.7486	0.7076	0.7744	0.6887	0.7502	**0.7168**	**0.7160**

calculation metrics. It highly outperforms PSPNet [32] and RESUNET [36], especially when MSE, BIAS and MAE values are compared. To determine the measurable thickness from pixel values, thickness is computed in millimetres (mm) using the following formula:

$$\boldsymbol{T} = \frac{\boldsymbol{\mu}}{n_{st}} \times N_{pt} \times r_f \tag{3.22}$$

Table 3.5 Independent test set accuracy comparisons (P.II)

Method	DeepLab3+	U-Net	PSPNet	RESUNET	SegNet	TREE
Accuracy	0.8325	0.8796	0.8471	0.8668	0.8295	**0.9242**

Here, in Eq. (3.22), μ denotes the mean value of the pixel resolution (9.61 μm) and the refractive index of tissue is taken to be n_{st}= 1.38. The number of pixels involved in thickness estimation is

Table 3.3 Errors (pixels) in veins thickness estimation are represented using five metrics, namely MSE, MAE, Bias, SD and RMSE on test set. (P.II)

Method	MSE	MAE	BIAS	SD	RMSE
DeepLab3+	29.21	4.459	-2.155	4.957	5.405
U-Net	14.36	3.115	0.794	3.706	3.790
PSPNet	56.900	6.4701	-4.010	6.388	7.543
RESUNET	34.224	4.7930	2.154	5.439	5.850
SegNet	7.388	2.3684	-2.276	1.485	2.718
TREE	2.409	1.109	0.184	1.541	1.552

N_{pt}, and the rescaling factor is r_f=2 as resizing (for vertical direction) of the image was done prior to using it as the input for the model. The thickness for all models, including TREE, is computed using Eq. (3.22) and compared with the ground truth thickness (Mask). The averaged thickness values over the test dataset are presented in Table 3.4. This ground truth thickness is derived axially from expert-annotated masks, independent of any segmentation model, ensuring high accuracy. The results indicate that TREE provides highly precise thickness estimations. In addition to the thickness values, we also provide the number of parameters utilized by each model. It can be inferred easily that TREE incorporates the least number of parameters, due to the reduction in the number of blocks from four to three (on both encoder and decoder sides), compared with original U-Net [29] and other similar architectures like RESUNET [36]. Hence, owing to contributions made by the SPD module, the capability of the module enhances even with lesser number of layers.

To assess the generalizability of the proposed model, its performance is evaluated further on 200 B-scans obtained from the fifth data volume. This volume contains a new vein sample that was not

included in the training or validation phases. The segmentation results of TREE are compared with other models on this independent test dataset, as summarized in Table 3.5 demonstrate superior performance compared to other methods advocating its generalization capability.
To evaluate the impact of transfer learning by deploying the dual-decoder architecture and training the model in two phases, an ablation study was conducted. The models were trained on the venous volume dataset following the same experimental procedure described above. Five different models were examined:

1. **En-De1 Model** – The segmentation model (En-De1) is trained first, followed by additional three dense and activation layers (as in De2) to determine the vein thickness.
2. **En-De1-De2 Model** – In this approach, both decoders (De1 and De2) are trained simultaneously alongside the encoder, but without utilizing transfer learning.
3. **En-De2 Model** – This model predicts only the vein thickness without segmentation by training the encoder and De2 exclusively.
4. **En-De2-De1 Model** –the En-De2 model from the previous setup is incorporated, where encoder weights are fixed. The segmentation decoder (De1) is subsequently trained, allowing the model to first learn vein thickness estimation before segmentation.
5. **NO-SPD Model** – A variation of TREE where the SPD module is replaced with conventional bottleneck bridge layers, similar to the U-Net [29] architecture. Atrous and depthwise convolutions are substituted with standard convolutions while maintaining the same kernel size, stride, and layer structure as in TREE.

For these models, the performance is compared using accuracy and MSE metrics for segmentation and thickness prediction, respectively, on the test dataset. The calculation of mean thickness here is done by calculating the mean of the row vector containing the thickness of each column for every B-scan. This mean thickness is further used for calculation of MSE between the ground truth and predicted values as in Table 3.6. By computing the mean error across the columns, one can get sight about the estimates of thickness. This facilitates a deeper understanding of effects of the transfer learning, dual decoders, and sequential training proposed here. The first model En-De1 is based on the same segmentation framework as in TREE, leading to similar results. However, the major difference lies in the value of MSE for the thickness as no transfer learning is applied.

For the model En-De1-De2, the simultaneous training hinders the model's performance in predicting both segmentation and thickness values. Meanwhile, the En-De2 model exhibits significant errors in thickness estimation and fails to fulfil the objective of obtaining both the segmentation map and the thickness from a single network. The model trained sequentially (En-De2-De1), where the thickness was predicted first followed by segmentation, also exhibited inferior performance. Initiating the training for thickness estimation did not effectively translate to accurate pixel-wise segmentation, as this approach failed to capture the global spatial context necessary for segmentation, making transfer learning inefficient. Lastly, testing the model by replacing the SPD module, as mentioned above, revealed that the NO-SPD variant also gave poor performance when compared to TREE. The superior performance of TREE, as demonstrated above, is largely attributed to its transfer learning strategy, while the incorporation of the SPD module further enhances accuracy.

Table 3.6 Ablation study for segmentation and thickness. (P.II)

Method	En-De1	En-De1-De2	En-De2	En-De2-De1	NO-SPD	TREE
Accuracy	0.9910	0.9282	-	0.6636	0.9514	**0.9937**
MSE (column)	0.1140	0.3554	58	58	0.2651	**0.0537**

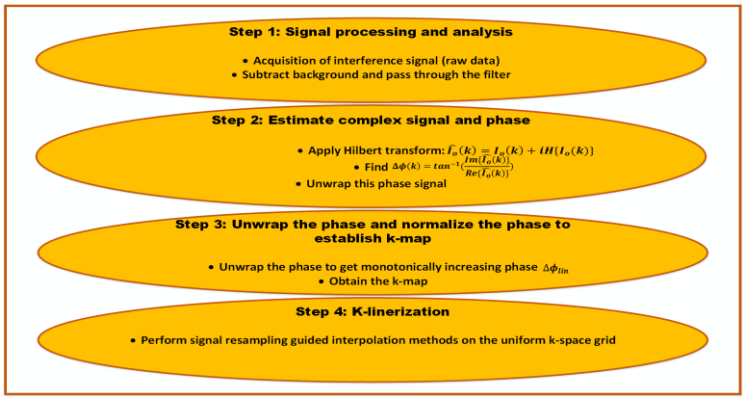

Figure 4.1 Flow diagram for the k-linearization procedure stepwise

Chapter 4 A Deep-learning framework for Image Reconstruction in SS-OCT

The studies in the Fourth Chapter refer to author's publication P III and P IV. Thus far in the thesis, the focus has been on bio-medical data acquisition for varicose vein and corresponding DL-based methods for automation of segmentation and thickness measurements. Meanwhile, it was observed that OCT systems suffer with image quality owing to speckle noise and complexity related issues either due to expensive hardware or software-based processing techniques. In this chapter we overview the image reconstruction pipeline in SS-OCT using the wavelength domain data. The focus of the chapter is set on putting forward highly optimized and efficient DL models that can reconstruct high-quality images in SS-OCT systems along with reduced time-complexity.

4.1 Wavelength domain reconstruction

The SS -OCT type of systems captures the data resolved in the spectral domain by employing the swept-source laser capable of rapidly changing its emission wavelength in time, i.e., sweeping across certain wavelengths in time. This technique encodes the depth-related reflectivity information of the test sample as a spectrum in Fourier space. It is commonly referred to as a ‘depth profile’ as it represents structural details corresponding to varying depths. To extract the information from the Fourier domain into A-scans (spatial domain), the wavelength encoded interference signal is subjected to Inverse Fourier Transform using IDFT. A typical limitation in most such systems, is the requirement for the interreference signal to be linear in the wavenumber domain (k-space) prior to applying the Fourier transform. Therefore, before the transformation, the wavelength domain interference signal is converted into the k-space signal through a process known as k-mapping.

In Fig.4.1, we represent the calibration procedure [22] for better understanding the concepts of calibration and k-linearization that are crucial in image reconstruction for achieving optimal axial resolution. The method described here begins with the acquisition of the raw data which incorporate an interferometric signal that is linear in wavelength domain. Then, to remove the fixed pattern noise, the background subtraction and filtering are performed enabling proper phase retrieval during Hilbert transform. Then, applying the Hilbert Transform to the real acquired signal, the complex signal is derived as shown in Step 2 in Fig.4.1. Then phase is determined and unwrapped to obtain a monotonically increasing phase, used in obtaining the k-map which reflects the linear in k-space indices. Interpolation methods can be used to resample the signal on a uniformly spaced k-grid. This completes the k-linearization process. Upon obtaining this linearized in k-space signal, the IDFT can be applied to get the reflective profile along axial direction known as an A-scan in the OCT system. To address the challenges associated with the calibration and k-linearization, a common strategy involves resampling the acquired data through different interpolation methods. Several recent studies have proposed hardware-based solutions, serving as alternatives to conventional software-based processing techniques for calibration and linearization in OCT systems. Such methods include optical k-clocks, dual-channel acquisition, akinetic laser to name a few. These

hardware-based methods have disadvantages such as high costs, complexity, and inefficient processing causing inaccuracies. These techniques have been largely driven by hardware-based solutions.

In cases where the sample exhibits dynamic behaviour, internal motion can introduce artifacts when speckle reduction is attempted via averaging. This temporal variability entails meticulous selection of the processing technique and associated time constraints, as both are critical for achieving high accuracy and reliability of an OCT-based imaging system. The DL based techniques presented in this thesis offer promising solutions to overcome the issue of speckle noise, avoid averaging as a solution to speckle reduction as it is not possible in case of dynamic samples, avoid high costs due to expensive hardware required for linearization along with reduced time-complexity for processing the data linear in wavelength domain.

4.2 Problem Formulation

To model the reflectivity profile in a SS-OCT system based on a Michelson interferometer, the intensity I_o can be expressed in a practical scenario as follows:

$$I_o = I_{DC} + I_{CC} + I_{AC} + N_G(t). \tag{4.1}$$

In Eq. (4.1), $N_G(t)$ represents the additive white Gaussian noise, and I_{DC}, I_{CC}, I_{AC} denote the DC, cross-correlation, and autocorrelation components respectively. The depth-resolved reflectivity information from the sample points is encoded within the cross-correlation term that can be expanded as:

$$I_{CC} = S\big(k(t)\big)\left[\frac{\rho}{2}\sum_{n=1}^{N}\sqrt{R_{rm}R_{Sn}}\left(\cos\big(k(t)(z_R - z_{Sn})\big)\right)\right]. \tag{4.2}$$

In Eq.(4.2), $S(k(t))$ is the light source's spectral shape, R_{rm} and R_{sn} represent reflectivity from the reference mirror (at depth z_R) and nth sample particle at depth z_{Sn} respectively, N is the number of particles and ρ is the responsivity of the detector. The reflectivity profile of the sample can be retrieved by applying the IDFT as $\boldsymbol{\mathcal{F}^{-1}}$ in the k-space, expressed as in Eq. (4.3):

$$i_z = \boldsymbol{\mathcal{F}^{-1}}\{I_{CC}\}, \tag{4.3}$$

$$i_z = \frac{1}{2}\sum_{i\neq 1}^{\infty}\sqrt{R_{rm}R_{Sn}}(\alpha(z_R - z_{Sn}) + \alpha(-(z_R - z_{Sn}))). \tag{4.4}$$

Here, in Eq.(4.4), $\alpha(z)$ is for DFT dual of $S(k)$. Since the IDFT assumes equidistant sampling points, it is essential that the spectral interferometric signal be acquired with uniform spacing in k-space to ensure accurate reconstruction. This stands in contrast to the acquired spectral data, which are sampled uniformly in the wavelength domain over the temporal interval -Δt/2 to Δt/2, exhibiting a linear dependency on time described as:

$$\boldsymbol{\lambda} = \beta t + \boldsymbol{\lambda}_0. \tag{4.5}$$

In Eq.(4.5), β is the sweeping speed of the source, $\boldsymbol{\lambda}$ corresponds to the wavelength ranging from $\boldsymbol{\lambda}_{\min}$ to $\boldsymbol{\lambda}_{max}$ and $\boldsymbol{\lambda}_0$ represents the central wavelength. Further, if we deploy $\boldsymbol{\lambda} = 2\pi/\boldsymbol{k}$, then we can express as follows:

$$\mathrm{t} = \frac{1}{\beta}(\boldsymbol{\lambda} - \boldsymbol{\lambda}_0) = \frac{2\pi}{\beta}\left(\frac{1}{\boldsymbol{k}} - \frac{1}{k_0}\right). \tag{4.6}$$

By applying a power series expansion on Eq.(4.6), the expression can be reformulated as Eq. (4.7) below:

$$t = \frac{2\pi}{\beta}\left[-\frac{1}{k_0}\left(\frac{\boldsymbol{k}}{k_0} - 1\right) + \frac{1}{k_0}\left(\frac{\boldsymbol{k}}{k_0} - 1\right)^2 - \frac{1}{k_0}\left(\frac{\boldsymbol{k}}{k_0} - 1\right)^3 + \cdots\right]. \tag{4.7}$$

Furthermore, for advanced light sources, such as FDML lasers [20], the relationship between the wavelength and time exhibits a sinusoidal dependence, which can be mathematically expressed as:

$$\boldsymbol{\lambda}(t) = \boldsymbol{\lambda}_0 + \frac{\Delta\boldsymbol{\lambda}}{2}(\sin(2\pi f_t t)). \tag{4.8}$$

Here, in Eq. (4.8) the bandwidth is $\Delta\lambda$ of the light source, and the tuning frequency is f_t of the Fabry-Perot (FP) filter used in the light source.
Considering the above formulation, it can be observed that the fundamental aspect of the OCT systems lies in the depth-encoded spectral data. As indicated by Eq. (4.7) and Eq. (4.8), the spectral signal exhibits a non-linear dependency on the wavenumber. The presence of significant higher-order terms in the expansion series necessitates the application of calibration and wavenumber linearization techniques to enable accurate B-scan generation.
Further, as discussed in the background chapter of this thesis the OCT imaging system uses a coherent light source due to which it experiences the speckle phenomenon. Speckle noise degrades the image contrast, and obscures fine morphological features. Therefore, it is imperative to apply noise suppression techniques.
The publication III and IV aims at high fidelity image reconstruction by effectively suppressing this noise without compromising spatial resolution or structural integrity by using raw data. First, a method is presented in publication III to reconstruct high quality images directly from the raw wavelength domain OCT data surpassing the k-linearisation process by incorporating the CNN-based model as discussed in section 4.3. Further, to enhance the capability of such models, in publication IV a network is proposed that explores the data in spatial and frequency domains. This proposal is motivated by the fact that the FD-OCT systems are based on the fundamental Weiner-Khinchin theorem that relates the Fourier and the spatial domains. According to this theorem, the power spectral density of the detected interferometric signal is mathematically related to its auto-correlation function through a Fourier transform relationship. This Fourier-domain relationship is leveraged by the FD-OCT systems that acquire spectrally resolved interferometric signals. These signals are then processed using the IDFT to reconstruct depth-resolved reflectivity profiles in the spatial domain. This foundational principle inspired us to develop a dual-network framework, that integrates spatial and Fourier domain processing for enhanced OCT image reconstruction. The sub-section 4.4 of this chapter throws light on the method proposed that first processes raw data in spatial domain followed by processing in Fourier domain using CNN based network.

4.3 Reconstruction using the raw data for SS-OCT images

In this section, we summarize the work presented in Publication III wherein a DL model is advanced to generate OCT B-scans by using interferometry signal that is linear in λ-domain. As discussed above, if the interference signals that are non-linear in k-space are transformed using IDFT in spatial domain the resultant images lack sharpness and structural details. The proposed model is capable of directly processing the linear in wavelength-domain (λ-space) signal, surpassing the need of k-linearization process. It uses DL-based architecture deploying a modified U-Net [29] architecture. It comprises of attention mechanisms [34] and residual connections [30]. The framework also focuses on image quality degradation due to the presence of speckle noise.

4.3.1 Methodology for the reconstruction framework

The proposed framework named as WAVE-UNET is demonstrated in Fig.4.2 along with the standard processing involving k-linearisation. This framework initially processes raw data interferometric signal acquired in the non-linear k-space that corresponds to a single B-scan. This is done to reconstruct depth-resolved profiles and corresponding images, which are blurred due to non-linearity in the signal.

This pre-processing of raw λ-space signal is carried out in several stages:

1. The background spectrum is obtained without placing any sample. This spectrum is then subtracted to eliminate the source-induced spectral bias.
2. Then the signal goes through Hann window in order to suppress spectral leakage.

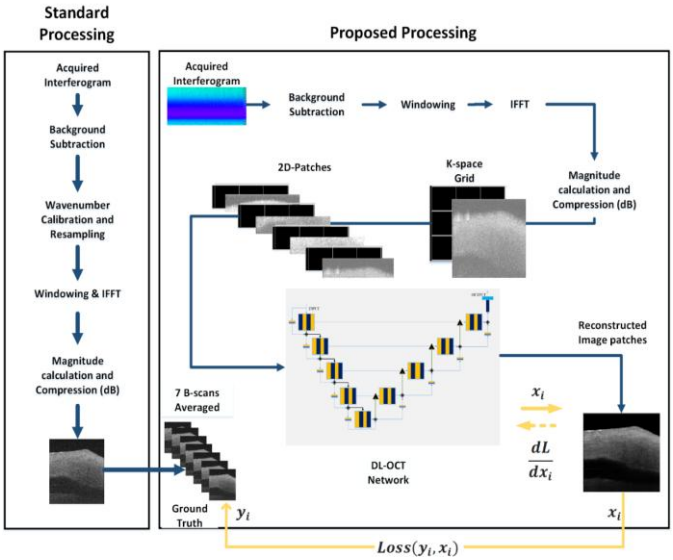

Figure 4.2 Stepwise processing in OCT (Standard Processing), the proposed framework WAVE-UNET for OCT image reconstruction (Proposed Processing). (P.III)

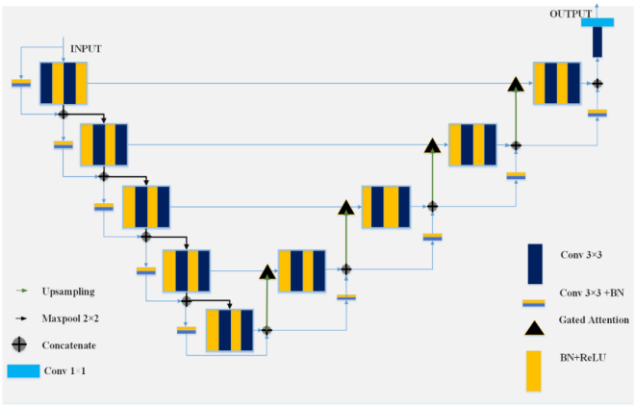

Figure 4.3 DL-based model used in WAVE-UNET; having modified U-Net with residual connections and gated attention blocks. (P.III)

3. The one-dimensional IDFT is executed along the vertical axis of each B-scan.
4. The magnitude of the resulting complex signal is computed, transformed into a logarithmic (dB) scale, and due to conjugate symmetry, the redundant part of the spectrum is discarded.

The outcome of this process is a depth-encoded OCT image derived directly from linear λ-space measurements. This resultant image is referred to here as the λ-space image. The images produced through this reconstruction, i.e. from non-uniformly sampled wavenumber (k-space) data, are inherently of low quality due to blurring artifact as discussed earlier. These low-quality images as illustrated in Fig.4.2 are subsequently fed as input to the proposed WAVE-UNET framework. The

WAVE-UNET is based on a DL-model inspired by the U-Net architecture [29] and enhanced with attention mechanisms [34] and residual connections [30] to improve feature representation. Attention modules are placed at each skip connection that connects the encoder and decoder blocks at respective levels, comprehensively presented in Fig.4.3. To incorporate prior knowledge of the wavenumbers into the learning model, a layer named as Wavenumber Sharing (WS) layer is introduced. This layer encodes a predefined grid of wavenumbers spanning from 4.62×10^6 rad/m to 4.99×10^6 rad/m, mapped across 1152 spatial positions corresponding to points on the A-line scan. Placed as a secondary channel, the WS layer is overlaid onto the input λ-space image (low-resolution B-scan), forming a composite two-channel input matrix, where the first channel contains the intensity values from the λ-space image and the second channel encodes the wavenumber grid. The wavelength spans the interval $(\boldsymbol{\lambda}_{min}, \boldsymbol{\lambda}_{max})$ which is divided into P^p points, s is the sampling point, where each point in λ-space can be represented in form of Eq. (4.9) and Eq. (4.10) below:

$$\lambda = \boldsymbol{\lambda}_{min} + \frac{s}{P^p - 1}(\boldsymbol{\lambda}_{max} - \boldsymbol{\lambda}_{min}) \tag{4.9}$$

$$k = \frac{2\pi}{\lambda} = \frac{2\pi}{\boldsymbol{\lambda}_{min} + s(\boldsymbol{\lambda}_{max} - \boldsymbol{\lambda}_{min})/(P^p - 1)} \tag{4.10}$$

This integration is visually represented as k-space grid in Fig.4.2. They are partitioned into four two-dimensional patches, each patch comprising both layers, i.e., the λ-space image and the wavenumber information. These patches serve as input to the proposed encoder-decoder architecture, where each side (encoder/decoder) of the network consists of four residual blocks. During training, the model processes batches of these 2D patches, resulting in the corresponding output patches x_i as the input propagates through the model. This dual-layered approach using the patch-based input enables the network to learn a more informed representation by leveraging both the spatial intensities and the associated spectral range. The loss between the generated output x_i, and the expected output, i.e. the ground truth y_i, is minimized to guide the learning process of the proposed model. For optimization, the Adam optimizer is deployed that iteratively updates its parameters. As illustrated in Fig.4.2, the solid yellow arrows represent the forward propagation, whereas the dashed arrows represent the backward propagation. All the residual units are composed of Binary Normalization (BN), convolutional operations, and ReLU activation functions, integrated with residual connections. The encoder module progressively reduces spatial dimensions by utilizing max pooling operations, while the decoder reconstructs back the same resolution by deploying upsampling layers. The attention mechanism between the encoder and decoder blocks along the skip connections helps to enhance the fidelity of image reconstruction by supressing irrelevant attributes [34].

4.3.2 Experiments and analysis for WAVE-UNET

In this section, we present the experiments conducted to analyse the performance of the WAVE-UNET framework based on a dataset obtained from the MHz SS-OCT with FDML source from Optores GmbH described in Chapter 2. The dataset overall consists of 3000 images. For each acquisition, 5 different samples were used, each of these volumes contains 600 images, and the corresponding raw data linear in wavelength domain. The samples used were lemon, cherry, human hand-finger, human-tooth, and varicose vein. The raw input consists of a λ-space data represented as a matrix having dimensions 2304 × 1024. This matrix is subjected to image pre-processing operations as described in preceding section. This processing involves Fourier domain transformation, hence, due to presence of conjugate symmetry causing redundancy in data, we utilized 1152 × 1024 pixels to extract the λ-space image. Further, due to memory constraints, region-of-interest for each image was restricted to a size of 1152× 512. Subsequently, image patches of dimensions (288 × 512 × 2) are obtained by interleaving the wavenumber matrix with the λ-space image and segmenting the B-scan (of size 1152 × 512) into four parts in vertical direction. Each patch is then standardized by performing mean subtraction followed by division using the standard deviation. The dataset comprising a total of 3000 images was partitioned into training, validation, and testing subsets to facilitate effective learning and hyperparameter tuning. Specifically, training contained 70%, validation contained 20% and testing the remaining 10% of the dataset. For the

ground truth images employed for training the network, the B-scans produced by the Optores OCT system were used after averaging seven consecutive scans. This averaging was performed to reduce the speckle noise with careful consideration to maintain the anatomical integrity and prevent the degradation of fine details that usually result from excessive averaging. These images were subjected to the OCT system's (Optores GmbH) built-in processing procedures, including k-linearization and calibration. The model during the training phase underwent 150 epochs. The batch size comprised 12 image patches, each having the dimensions 288 × 512 × 2. The model parameters were optimized deploying the ADAM optimizer with a learning rate - 0.0001. The entire framework was executed on a computational system equipped with 64 GB of RAM and NVIDIA RTX 3090 GPU.

The Mean Squared Error (MSE) loss function is employed here to estimate the quantitative difference between the predicted image patches generated by the WAVE-UNET model and the corresponding ground truth patches.

Fig.4.4 shows the output image patches having dimensions of 288 × 512 pixels corresponding to (a) vein, (b) cherry, (c) hand-finger and (d) tooth. Here, the columns represent as follows: column 1 -

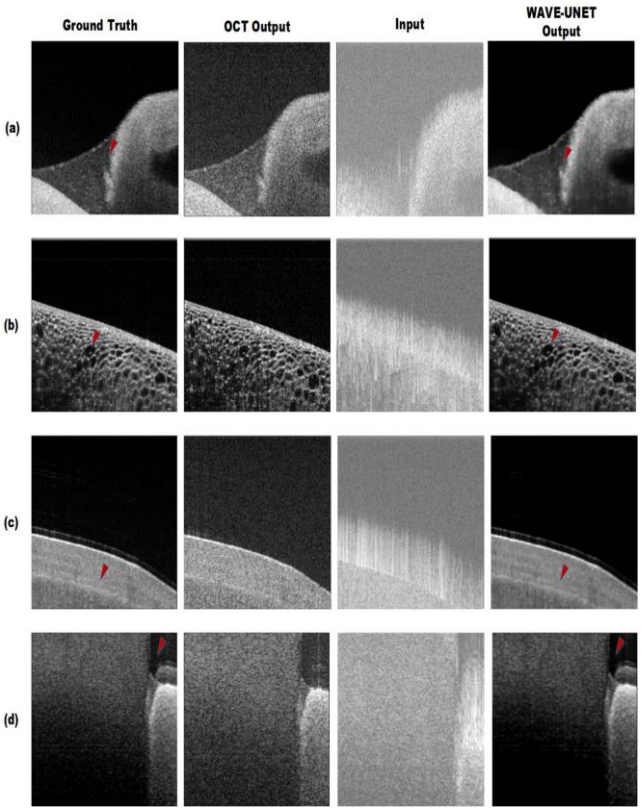

Figure 4.4 OCT B-scan patches from (a) vein, (b) cherry, (c) hand and (d) tooth, Columns 1 to 4 show: (1) Ground Truth (left), (2) OCT output, (3) Input Image and (4) WAVE-UNET reconstructed output image respectively. (P.III)

the ground truth, column 2 - OCT Output, column 3 – input and column 4- output of the WAVE-UNET. From the results depicted in Fig.4.4, it can be clearly observed that the WAVE-UNET model exhibits high-fidelity reconstructions when compared to the original blurred λ-space images (the input) and the corresponding OCT output produced by the Optores system. The regions highlighted by red arrows further advocate for the WAVE-UNET's ability to recover intricate structural attributes. These observations collectively substantiate the capability of the proposed DL based model in reconstructing high-quality, speckle-suppressed OCT images, while streamlining the reconstruction process in comparison to traditional OCT pipeline that relies on extensive averaging. Further, to estimate the performance quantitively, this work uses the following metrics: Structural Similarity Index Metric (SSIM), Peak Signal to Noise Ratio (PSNR) in dB, and MSE to better analyse the performance on various samples. Here, the ground truth serves as the reference standard

Table 4.1 Comparison of images generated by SS-OCT system and WAVE-UNET (P.III)

Metric		VEIN	LEMON	CHERRY	TOOTH	HAND
PSNR	I/P	17.06	15.41	15.56	14.57	14.24
	OCT	21.94	19.18	14.11	17.56	19.65
	WAVEUNET	25.50	27.34	19.21	19.75	22.81
SSIM	I/P	0.05	0.008	0.05	0.09	0.05
	OCT	0.21	0.04	0.03	0.08	0.03
	WAVEUNET	0.59	0.51	0.41	0.29	0.46
MSE	I/P	1.31	1.30	1.33	0.72	1.12
	OCT	0.42	0.54	1.86	0.36	0.32
	WAVEUNET	0.18	0.08	0.57	0.21	0.15

for computing these evaluation metrics. It is also used to understand quantitively the difference between the input (raw λ-space image), the OCT system Output (OCT) and the output of the proposed model (WAVE-UNET) as tabulated in Table 4.1 for 10 % of image patches chosen randomly from each test set sample, namely Vein, Lemon, Cherry, Tooth and Hand. It is clear that the proposed approach surpasses the original SS-OCT imaging system by Optores in terms of image quality across all the measured quality parameters. A substantial enhancement around 4 dB in PSNR can be observed from the output by the OCT system compared to the reconstructed output for the lemon, vein, and cherry sets. Likewise, the SSIM demonstrates improvement for all samples. It is also important to highlight that for the samples from the cherry, tooth, and hand the measured SSIM of the raw λ-space images exceeds the corresponding OCT-processed outputs. This is due to the fact that OCT images are significantly degraded by speckle noise. Thus, the SSIM values alone may not function as a fully reliable indicator of image quality, especially in the case of speckle inflicted images. A similar trend is observed in the MSE values, where the lemon dataset exhibits the smallest error, and the cherry dataset presents the highest. Hence, this comparative observation justifies that the proposed method demonstrates significantly superior performance compared to standard OCT processing by Optores system. Moreover, the proposed approach is also suitable for applications demanding high-speed reconstructions. It can achieve an inference time of 90 seconds for a volume comprising 600 B-scans, making it markedly faster than the 792 seconds traditionally required by the OCT Optores system (using its inbuilt processing pipeline) tested on the same GPU-accelerated system described above in this section.

4.4 Reconstruction of OCT images by optimizing the data in Fourier and spatial domains

This section describes the work presented in Publication IV, which is based on a method to perform reconstruction of OCT images by taking into consideration the Fourier-spatial domain relationship described by the Wiener–Khinchin theorem as elaborated in Section 4.2. This theoretical ground is integrated as a physics-based prior to enhance the learning process of the proposed CNN based framework. The proposed framework is implemented by incorporating two networks along with processing at different levels, that are leveraged sequentially. The first network, named as SD-CNN (Spatial Domain Convolution Neural Network), as indicated by the name, optimizes in the spatial domain. The second network, named as FD-CNN (Fourier Domain Convolution Neural Network), optimizes in the Fourier domain. The SD-CNN initiates by utilizing the severely deteriorated images generated by applying a Fourier transform to the λ-domain interference signal as input. The aim of this network is to restore intricate morphological structures along with suppression of the unwanted noise. The FD-CNN aims to serve as a higher-level refinement network in the Fourier domain for high quality image reconstruction. This network mainly emphasizes on restoring fine structural details that were inadequately captured by the SD-CNN network. This dual network-based optimization strategy enables the overall framework to more efficiently generate B-scans demonstrating enhanced visual quality.

4.4.1 Methodology for reconstruction OCT images using SD-CNN and FD-CNN based framework

The framework composing SD-CNN and FD-CNN for OCT image reconstruction, proposed in Publication IV, is illustrated in Fig.4.5 presented in Publication IV. The framework utilizes the spatial domain raw-data in the form of λ-space images obtained after processing. The processing is performed in the same sequence as described in section 4.3.1 involving background subtraction from the acquired interferogram, application of the Hann windowing, followed by applying Fourier transform using IDFT and magnitude scaling. The DL network used is shown in Fig.4.6 (a) and elaborated in detail below. The DL-model used in SD-CNN and FD-CNN is shown in Fig.4.6 (a) being nearly the same in architecture and having only minor variations. It is built upon the U-Net as the base architecture, comprising the bottleneck along four encoder and decoder blocks respectively. It includes the attention gating mechanism and residual connections to enhance its operational efficiency. The attention mechanism as stated earlier, works by suppression of irrelevant features

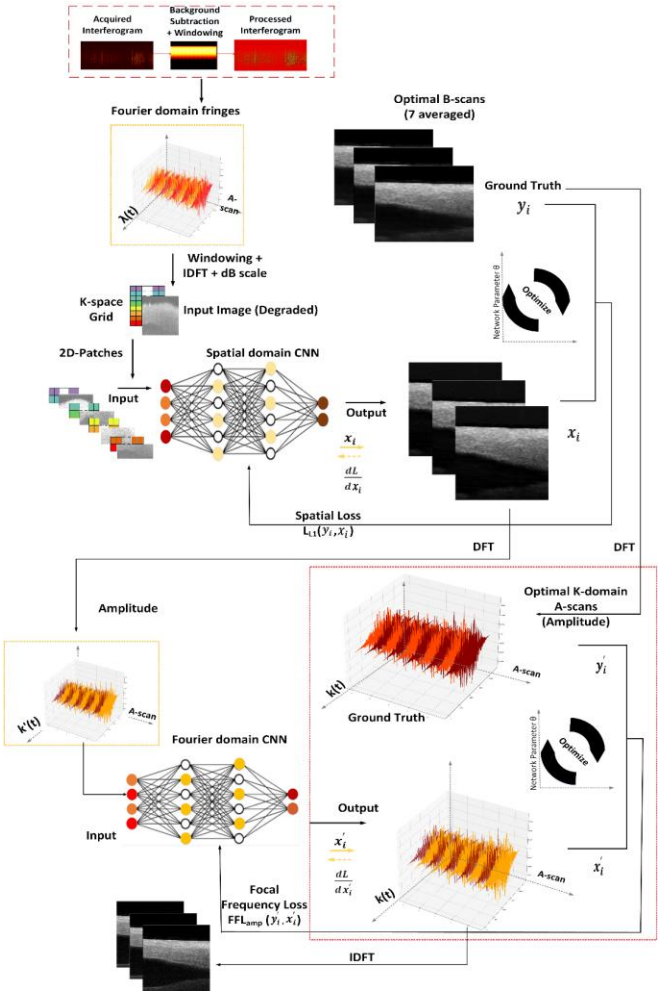

Figure 4.5 Schematics of the workflow of proposed framework including SD-CNN and FD-CNN (P.IV)

for superior performance. The encoder block comprises following layers: BN, the activation function Parametric ReLU (PReLU), and convolution layers, arranged in the stack depicted in Fig.4.6 (a). In contrast, the decoder blocks use the Upsampling layers to upscale the resolution, along with Conv Transpose, activation as PReLU, and normalization using BN layers. The intermediate skip connections between the encoder and the decoder assist the network by transfer of information at the corresponding levels (resolution). However, these connections can transfer low-level redundant features, which are filtered by incorporating the attention mechanism. This mechanism reduces activations in regions having redundant information.

The ground truth images (y_i) used by this network are generated from the Optores OCT system. They are obtained by averaging 7 consecutive images from the acquired volume. This is feasible as there exists high level of similarity across these adjacent images. To visually understand how the OCT images look after averaging adjacent scans, we present in Fig.4.6(b) the OCT B-scan obtained from the Optores system, followed by averaging 5, 7, and 9 adjacent scans (in columns 2, 3, and 4 respectively) consisting of vein, lemon and cherry samples shown in each row. The single B-scan images across different types of samples display significant amount of the speckle noise.
Visually comparing the various samples, we can observe and interpret that for the vein sample, a noticeable enhancement can be seen for structural details indicated by the red arrow, for the cherry

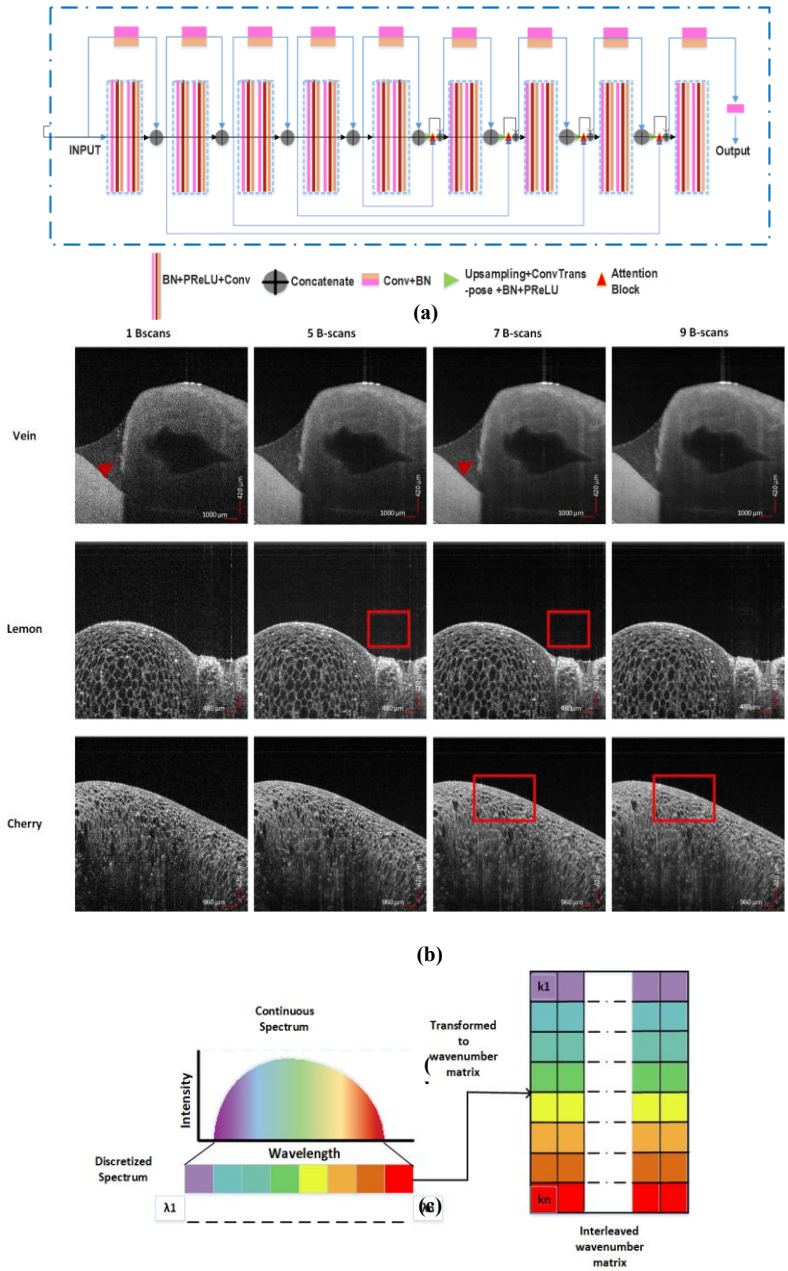

Figure 4.6 (a) DL network used as the FD-CNN and SD-CNN (b) 1 B-scan and images upon averaging 5, 7, 9 B-scans for vein, lemon and cherry, *(c)* Wavenumber layer interleaved with the input of SD-CNN (P.IV)

sample, a gradual increase in over-smoothing is experienced as a greater number of B-scans are averaged and lastly, for the lemon sample, a prominent decrease in background noise, particularly for the fifth and seventh B-scans highlighted using the red box, can be seen. Considering the balance between speckle noise suppression, the effects of over-smoothing, and the preservation of lateral resolution to prevent excessive blurring challenges we adopt the average of 7 B-scans as the ground truth (y_i) for this study.

The framework SD-CNN used the degraded image as the input, after pre-processing the raw interferometric data as described above. The information regarding the wavenumber range corresponding to the acquired data was integrated into the network as an auxiliary input layer during the training of the SD-CNN. This was named as the K-space grid as shown in Fig.4.5 and elaborated in Fig.4.6(c). The resolution of these input λ-space images is compromised since non-uniformly spaced wavelength domain fringes are directly subjected to IDFT. Initially, this parallel layered input is utilized by SD-CNN network to learn the reconstruction of OCT images. These points are calculated using the Eq. (4.9) and Eq. (4.10), and they supplement the network with information about the non-linearity in the k-space. Since identical wavenumber values exist for each A-scan, the column vector is replicated across all rows of the raw input image. Initially, this secondary layer is of size (1152 × 256), but then it is divided into patches of size 288 × 256 for improved convergence and hardware limitations. This step is important, as these patches correspond to the axial (depth) pixel positions in each A-scan through a Fourier transform relationship.

The output of SD-CNN (x_i) is further used by FD-CNN as its input upon processing. This input is generated by transforming the SD-CNN's output into the Fourier domain and applying the Discrete Fourier Transform (DFT) to obtain the amplitude component. The ground truth (y_i') for this network is the amplitude derived by applying DFT on ground truth images described above. Each of the 7 images used for generating the ground truth, undergoes the k-linearization prior to IDFT and, therefore, has approximately uniformly spaced k-space data. Given the one-dimensional structure of dependencies in A-scans, the FD-CNN employs 1-D convolutional kernels for extracting features. During the training phase, the FD-CNN processes the amplitude values (in Fourier domain) and learns to minimize the loss function to produce a uniformly spaced wavenumber-domain. For final prediction, the optimized amplitude output x_i' from the FD-CNN model undergoes the IDFT operation, while the corresponding phase information is extracted from the Fourier-transformed output of the SD-CNN.

4.4.2 Experiments and analysis for SD-CNN and FD-CNN based framework

The dataset used in Publication IV was acquired by the Optores system. The samples used for dataset acquisition were vein, finger, lemon, tooth, cherry, flounder egg, and seed (pea). All the B-scans from different volumes consisted of 1024 A-scans, with every A-scan containing 2304 depth points corresponding to the acquired raw spectrum. At the end of the processing pipeline, both the raw data and their associated B-scan images were cropped to dimensions of 2304 × 256 prior to applying the IDFT, considering the memory constraints. Following the IDFT step described in Section 4.4.1, only half of the signal owing to mirror symmetry is retained, resulting in a final size of 1152 × 256. The raw data are directly accessible through the software application 'MHz-OCT Processing' of the Optores system, along with individual OCT scans and 7-averaged OCT scans. The obtained raw data and the ground truth images were standardized respectively by subtraction of mean and normalization using standard deviation.

To develop a highly generalizable framework, we designed the dataset in a manner so that it comprises samples acquired under varying conditions. This includes variations in the field of view, calibration patterns, and material characteristics. The lateral (x-y) field of view across the various samples is as follows: lemon (4 mm × 4 mm), vein (8 mm × 8 mm), cherry (8 mm × 8 mm), tooth (4 mm × 4 mm), finger (8 mm × 8 mm), seed (10 mm × 10 mm), and flounder egg (3 mm × 3 mm). Further, in Fig.4.7, distinct k-linearization fringes are shown for the samples, highlighting the variation between the calibration pattern applied for each volume, at the start of data acquisition process employed by the Optores OCT system. Also, the refractive indices for the chosen samples vary as follows: human vein tissue (1.3–1.4), lemon and cherry (1.47), human finger skin (1.42), human tooth (2.6–3.1), seed (1.5–1.7), and flounder egg (1.3–1.4), enabling creation of diverse

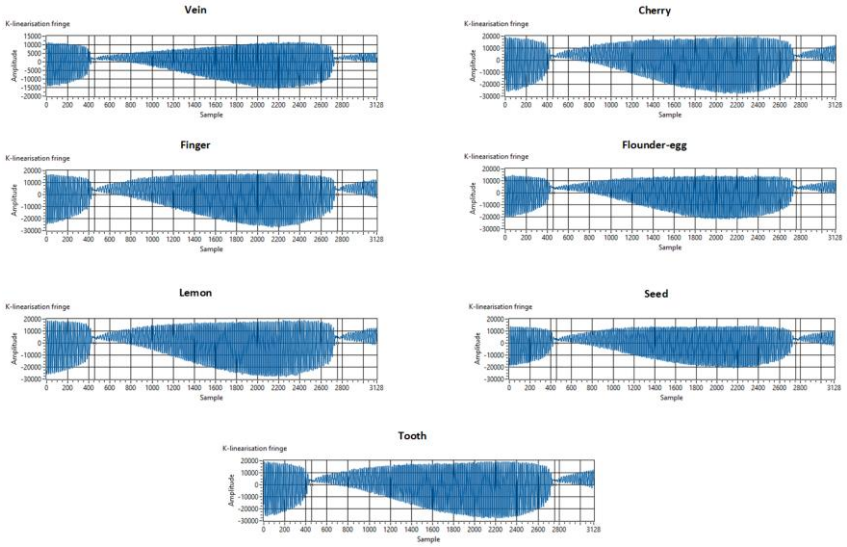

Figure 4.7 Analysis of k-linearization fringes across various sample volumes, including vein, finger, lemon, tooth, cherry, flounder egg, and pea seed. (P.IV)

spectral patterns tailored to samples with varying material characteristics. Overall, such variations support the training and evaluation of the model with the objective of higher adaptability with the diversity in the dataset incorporated using different samples and acquisition parameters and more robustness by incorporating samples afflicted with varying speckle noise. Amongst the seven data volumes described earlier, five - namely lemon, vein, cherry, tooth, and finger, comprising a total of 3000 B-scans (600 from each volume) - were randomly divided into subsets for the purpose of training, validation, and testing in the ratio 70:20:10 for the proposed framework. The remaining two volumes, seed (pea) and flounder egg, each containing 200 B-scans having dimensions of 1152 × 256 pixels, were exclusively used to estimate the model's generalization capability, as outlined in the independent test procedure below. These two volumes were excluded completely during the training or validation phases.

The training phase begins with the independent training of the SD-CNN. Upon completion of this training, the weights of SD-CNN are frozen, and the FD-CNN is subsequently trained using the processed outputs from the SD-CNN, as detailed in Section 4.4.1. During the inference phase, the SD-CNN first processes a degraded input image derived from the uneven wavenumber domain, along with the k-space grid, following the approach illustrated in Fig.4.5. DFT is then applied to the SD-CNN's output to extract the amplitude, which is input to the FD-CNN. The predicted output is then passed through an IDFT to generate the final OCT image. This integrated framework operates by sequentially deploying both models to infer the reconstructed output image. The optimizer deployed for both models is Adam, and the learning rate was set to 10^{-4}. The SD-CNN attains convergence after 200 training epochs, whereas the FD-CNN requires additional fine-tuning and takes approximately 400 epochs to achieve convergence.

Two different loss functions were used, namely L_{L1} and FFL, by SD-CNN and FD-CNN respectively. Here, L_{L1} represents the mean absolute error and is calculated between the output generated by the SD-CNN (LR) and the 7-averaged ground truth image (HR) from Optores system. This loss L_{L1} can be written (in Eq. (4.11)) demarcating the spatial indices as i and j:

$$L_{L1} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}|HR - LR|, \tag{4.11}$$

The FD-CNN uses the Focal Frequency Loss (FFL) [150], estimated for the output of the model ($F_\lambda(u,v)$) and the corresponding ground truth ($F_k(u,v)$), where u and v represent the indices of the frequency coefficients. This loss FFL can be written as:

$$FFL(u,v) = \frac{1}{UV}\sum_{u=0}^{U-1}\sum_{v=0}^{V-1} w(u,v)\,|F_k(u,v) - F_\lambda(u,v)|^2 \tag{4.12}$$

$$w(u,v) = |F_k(u,v) - F_\lambda(u,v)|^\alpha \tag{4.13}$$

Here, in Eq. (4.12), the spectrum size is represented using U and V and in Eq. (4.13) w(u,v) is the weight matrix, and scaling factor $\alpha = 1$. The error minimization is applied only to the amplitude component denoted as FFL_{amp}, since B-scan reconstruction from the raw data relies solely on the spectral amplitude.

The framework was tested using the test dataset described above using the following metrics: MSE, SSIM, CNR_r (Contrast to Noise Ratio), β_s- smoothness parameter and PSNR for analysing the reconstruction results. The CNR_r can be mathematically calculated over the r[th] region as:

$$CNR_r = 10log_{10}\left(\frac{|\mu_f - \mu_b|}{\sqrt{{\sigma_f}^2 + {\sigma_b}^2}}\right), \tag{4.14}$$

In Eq. (4.14), the foreground mean is μ_f, background mean is μ_b, and the standard deviations are σ_f and σ_b respectively for the foreground and background regions. By denoting mean as μ, the 2D image pixel indices are i and j, and – 'out' and 'in' corresponds to the output and input respectively in Eq.(4.15) for the smoothness parameter as follows:

$$\beta_s = \frac{\Gamma(I_{out} - \mu_{out}, I_{in} - \mu_{in})}{\sqrt{\Gamma(I_{out} - \mu_{out}, I_{out} - \mu_{out}).\Gamma(I_{in} - \mu_{in}, I_{in} - \mu_{in})}}, \tag{4.15}$$

where $\Gamma(I_1, I_2) = \sum_{i,j}[I_1(i,j).I_2(i,j)]$.

The analysis is performed using the degraded input, OCT output by the Optores system, the proposed

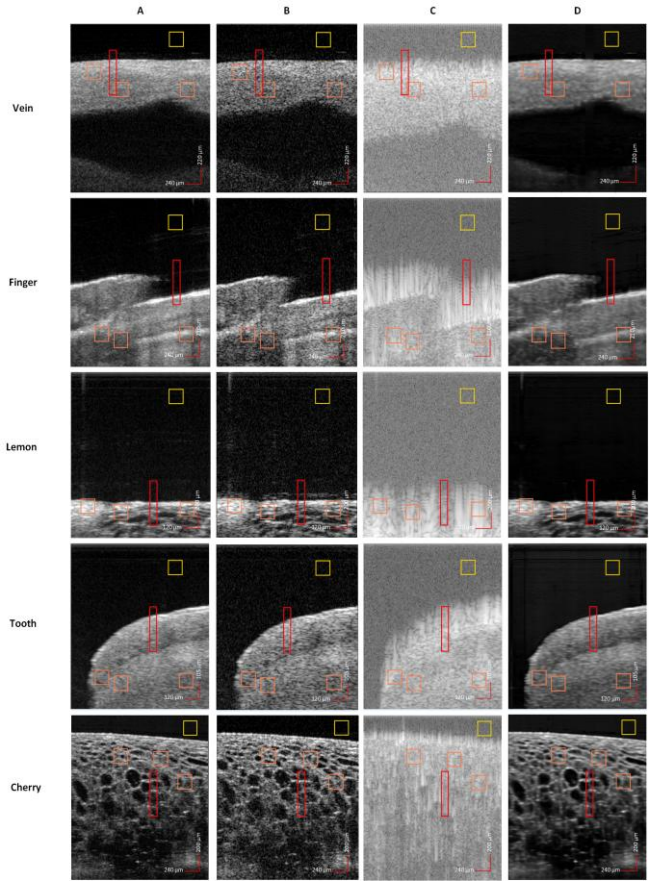

Figure 4.8 Comparison of B-scans from five different volumes. (A) Ground truth (seven B-scans of OCT system output averaged), (B) OCT system output, (C) OCT system raw data input, (D) output of the proposed framework. (P.IV)

framework's output and the high-quality ground truth images on the test dataset. Fig.4.8 illustrates the ground truth (A), commercial OCT output without averaging (B), degraded raw input to the SD-CNN (C), and the final output from the proposed DL framework (D) from all the 5 samples in the

test data set. The final reconstructions by the proposed framework demonstrate significant improvement over the degraded inputs, closely preserving fine details in alignment with the ground truth. The proposed method attains superior image quality and reduced noise when compared to the standard OCT output from the Optores system. Notably for samples like vein, tooth, and finger, each showing larger homogeneous regions, the framework effectively suppresses speckle noise, enhancing quality to a greater extent.

In Table 4.2 we compare the proposed framework using the metrics PSNR, SSIM, CNR and β_s on the test dataset. The PSNR reflects a gain of 2dB and 13 dB by the proposed framework with respect to what the OCT Output (by Optores) and Input (degraded input to SD-CNN) delivers. The lemon and cherry samples, which have more fine structural details compared with other samples, exhibit relatively high PSNR values strongly advocating high-fidelity image reconstruction capability of the framework. The comparison in terms of β_s parameter reveals an improvement in image quality, as the metric increases from 0.87 to 0.93, highlighting the proposed framework's

Table 4.2 Comparison of scores for PSNR, SSIM, CNR, β parameter (P.IV)

Method	Dataset	PSNR	SSIM	CNR	β
Input	Overall	8.94	0.08	-	0.71
	Vein	12.76	0.14	4.44	0.68
	Finger	7.92	0.06	4.62	0.75
	Lemon	7.14	0.05	5.69	0.64
	Tooth	10.11	0.12	4.31	0.77
	Cherry	8.79	0.07	5.04	0.73
OCT Output	Overall	19.95	0.35	-	0.87
	Vein	21.20	0.22	5.21	0.88
	Finger	19.93	**0.45**	3.04	0.92
	Lemon	20.40	0.26	6.73	0.78
	Tooth	**22.11**	**0.56**	0.75	**0.93**
	Cherry	17.55	0.30	3.96	0.85
Proposed	Overall	**22.30**	**0.46**	-	**0.93**
	Vein	**21.98**	**0.43**	**8.71**	**0.93**
	Finger	**21.75**	0.42	**4.86**	**0.94**
	Lemon	**25.74**	**0.54**	**7.98**	**0.91**
	Tooth	21.61	0.32	**6.48**	0.92
	Cherry	**21.67**	**0.62**	**4.86**	**0.95**

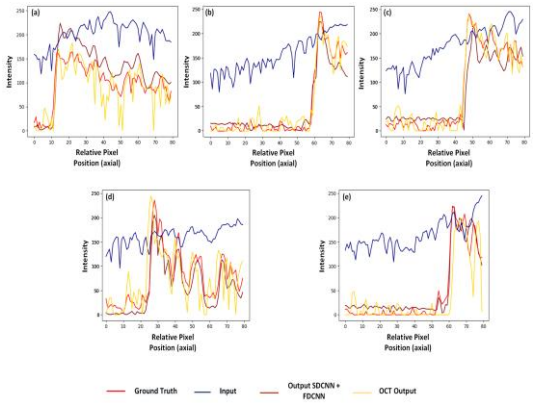

Figure 4.9 Comparison of line plots to show the variation of intensity (A.U.) for the central column of the red rectangle marked in Figure 4.8, here plots correspond to (a*) vein*, (b) f*inger* (c) lemon, (d) *tooth* and (e) *cherry samples*, for ground truth, OCT output and proposed framework outputs shown using different lines in each plot. (P.IV)

capability to suppress speckle noise effectively. Further, the CNR metric was calculated using Eq. (4.14) for the highlighted areas as depicted in Fig.4.8. Here, r = 3, the orange boxes are used for the foreground area and the yellow box for the background region. It is evident from the findings that the model effectively highlights critical image structures over noise across various sample types.
To assess image quality with respect to high-frequency components like edges, we examine a vertical pixel column located at the centre of the red rectangular region highlighted in Fig.4.8. This assessment is represented by the intensity curves displayed in Fig.4.9 for all five sample categories: the raw input, ground truth, standard OCT output, and the final output produced by the proposed method.

The x-axis for the plots in Fig.4.9 represents a relative pixel index, where the top row of the marked rectangle is labelled as position 1. These positions are relative and do not correspond to the original image's absolute pixel coordinates. The ground truth curve, shown in red, is closely tracked by the maroon curve depicting the proposed framework's output, indicating high fidelity in reconstruction. The OCT output by Optores system exhibits repeated intensity fluctuations, especially in uniform or smooth regions owing to persistent speckle noise. As anticipated, the input data show a significantly deviated profile dominantly influenced by blur and noise, because of non-linear mapping in the wavenumber-domain spectrum.

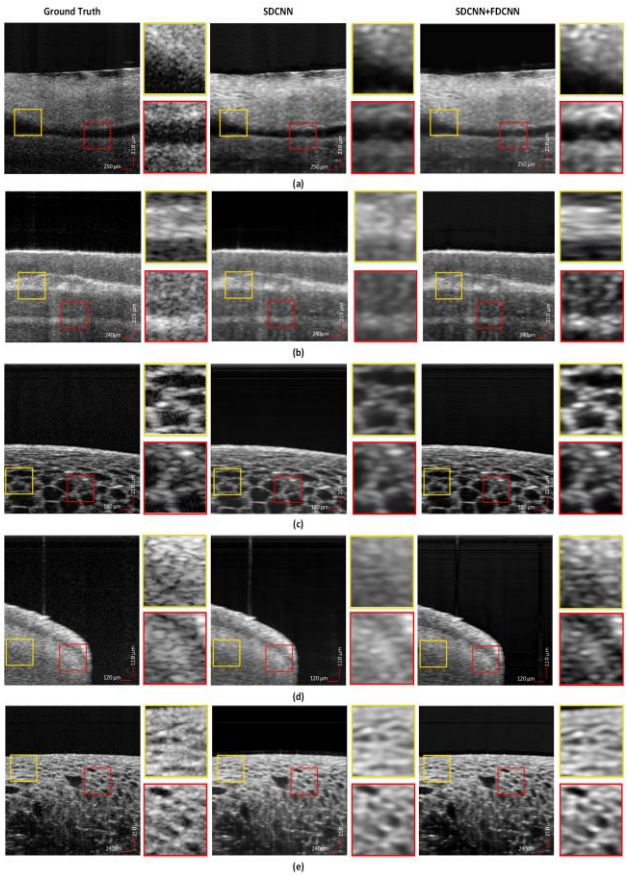

Figure 4.10 Comparison of output of SD-CNN and FD-CNN from the ground truth. The samples are (a) *vein*, (b) *finger*, (c) *lemon,* (d) *tooth* and (e) *cherry*. For all the images, the magnified regions are shown for better comparison. (P.IV)

The effectiveness of SD-CNN and FD-CNN was evaluated in Fig.4.10 through a visual comparison that includes images of the samples from the test data namely vein, finger, lemon, tooth and cherry (shown in rows 1-5) and their respective ground truth references. Here, each image highlights two boxes in yellow and red colours, they are enlarged nearly five times and placed beside the originals to facilitate better analysis. We observe that SD-CNN performs well when the image contains highly uniform regions whereas the FD-CNN strengthens the framework by its ability to reconstruct high frequency details like edges and boundaries. The overall framework provides

consistent high quality image reconstruction owing to the integration of a physics-based prior guiding the optimization across spatial and Fourier domains.

The performance is further analysed using the independent test dataset containing samples from a seed (pea) and a flounder egg consisting of 200 images in Table 4.3 using PSNR and SSIM values. In Fig.4.11, comparison is done between the ground truth, OCT system output and the proposed framework's output on the independent test data. These results confirm the framework's robustness and adaptability, showcasing its strong capacity to generalize even on unseen data.

Table 4.3 Independent test dataset results (P.IV)

Volume	PSNR	SSIM
Flounder egg	22.09	0.45
Seed (pea)	21.53	0.42

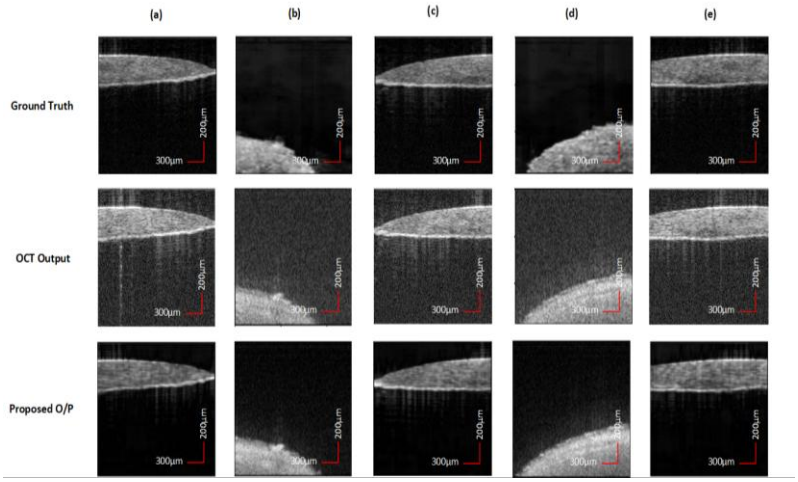

Figure 4.11 Comparison of independent test set results on (a),(c),(e) *flounder egg* and (b),(d) *seed (pea)* samples for ground truth, OCT output and reconstructions obtained by the proposed method. (P.IV)

To quantify the individual contributions of the SD-CNN and FD-CNN models to the over-all framework ablation study was performed. In Table 4.4 the ablation results for the individual models and the combined model (as SD-CNN+FD-CNN) are tabulated. It can be observed that the SD-CNN is responsible for reconstruction of images owing to higher PSNR and SSIM values. However, FD-CNN mainly compliments the performance by marginal improvement in reconstruction, as the design and the feature extraction mechanism of CNNs allow them to achieve higher accuracy in the spatial domain than when applied solely in the Fourier domain.

Table 4.4 Ablation analysis conducted to evaluate the individual contributions of FD-CNN and SD-CNN on image reconstruction. (P.IV)

SD-CNN	FD-CNN	SD-CNN+FD-CNN	PSNR		SSIM	
			Avg	Std	Avg	Std
✓	-	-	20.81	2.64	0.42	0.11
-	✓	-	10.97	0.80	0.03	0.01
-	-	✓	22.30	2.51	0.46	0.08

The time associated with various intermediate processes within the OCT system such as calibration, resampling and FFT for one B-scan and averaging to remove speckle (7 images) is presented in Table 4.5. This does not correspond to all the steps incorporated in the processing pipeline of the Optores OCT system. From the time-complexity measures in Table 4.5, it is evident that the resampling demands a significant processing time of 0.17 seconds. Moreover, in real-world scenarios, an individual B-scan is significantly impacted by speckle noise, necessitating techniques

such as averaging multiple scans or using registration-based approaches over multiple volumes to suppress it, introducing more time-complexity required for generating multiple B-scans. We also analyse the time complexities associated with volume reconstruction containing 600 images, each of dimensions 1152 × 1024. For these experiments, the raw data were acquired, and the reference images were produced along with minimization of the speckle (averaging over 7 B-scans) using the Optores OCT system.

Table 4.5 Time comparison (P.IV)

Operations		**Time (s)**	
Calibration		0.008	
Resampling in K-space		0.17	
FFT		0.05	
Averaging (Speckle reduction)		0.07	
Overall time Complexity -Volume (s)			
OCT system [30]	*Fourier Domain-CNN*	*Spatial Domain-CNN*	*Fourier Domain-CNN + Spatial Domain-CNN*
792	68.55	79.723	142.5

The proposed framework processes the entire volume in 142.5 seconds, significantly outperforming the commercial system [20],[21] which requires 792 seconds. When the models are executed individually, the FD-CNN and SD-CNN components take 68.55 and 79.723 seconds, respectively. The approach of averaging the adjacent scans used by the OCT system [20],[21] to suppress the speckle noise, resulting in smoother B-scan outputs, increases computational load. As opposed to this, the proposed framework achieves speckle noise suppression without utilizing multiple B-scans which requires generation (of B-scans) followed by averaging, thereby offers a significantly faster processing speed. This accelerated processing with the improved efficiency allows seamless integration of the OCT systems with the modern technologies enabling prompt and reliable functionalities.

Chapter 5 Conclusions

As a biomedical imaging modality, Optical Coherence Tomography (OCT) has achieved notable improvements in terms of automation with the integration of deep learning approaches. These approaches are capable of extracting complex as well multi-level abstract features. Such attributes are more robust and efficient compared to conventional hand-annotated features. As a result, they deliver highly accurate outcomes in an efficient manner.

OCT systems are highly effective for imaging soft tissues, enabling various measurements and facilitating the analysis of their underlying physical characteristics. With micrometre scale resolution and millimetric scale depth penetration, OCT proves particularly valuable in generating volumetric data of varicose veins, as investigated in this thesis. The integration of DL models has further enhanced the understanding of these veins by enabling accurate layer segmentation thickness measurements using neural networks specifically optimized for these tasks.

Chapter 3 presented the key components required for estimating thickness and automatically predicting segmentation maps from SS-OCT venous images, thus eliminating the need for manual expert input. The dataset used for this purpose was collected from patients with diverse physical characteristics. The segmentation models were evaluated in terms of their performance and efficiency. To address the challenge of model complexity due to a high number of parameters, the first approach referred to as Opto-UNet and built on the U-Net architecture [29], offered a highly optimized structure for accurate segmentation of venous images. The method was tested and compared to several state-of-the-art methods wherein the model outperformed the others in terms of its segmentation capabilities. The second model introduced in Chapter 3, TREE, aimed at predicting both segmentation maps and thickness values from SS-OCT images. The model employed a transfer

learning approach, allowing to first learn segmentation maps that subsequently facilitate more effective thickness estimation. TREE integrated a specialized SPD module, which combines atrous and depthwise separable convolutions in a tailored configuration. This integrated framework allows to extract comprehensive shape and thickness information within a single end-to-end network. Extensive experiments were conducted to evaluate the model and quantitatively compare it with the state of the art. The results demonstrated that TREE outperformed existing methods in terms of accuracy and precision for both segmentation and thickness measurements of veins.

The reconstruction of high-quality images with low computational complexity and supressed speckle noise is crucial for effective OCT imaging. Chapter 4 introduced DL based models developed to enhance image reconstruction in OCT systems. The model WAVE-UNET generates images directly from raw data that is linear in wavelength space – a format that typically yields poor-quality reconstructions when processed in Fourier domain. Instead of performing k-linearization and remapping, the DL models take the raw interference fringes as the input and process them further to generate the B-scans. It was demonstrated that blurred input OCT images can be used by the neural network for generating the B-scans of high contrast. Further, the reconstruction framework incorporating DL models in both spatial and Fourier domains successfully generated intricate morphological structures while significantly suppressing speckle noise – both crucial factors in biomedical imaging. This framework was also highly optimized for low inference time, ensuring efficient performance. The proposed methods were rigorously evaluated and visualised across a diverse set of samples, including vein, cherry, lemon, finger, tooth, flounder egg and pea seed each with distinct material properties. Quantitative and qualitative results consistently demonstrated the superior performance of the models compared to existing approaches.

Author contributions in the dissertation work

The primary contributions of the dissertation lie in the applied-theoretical domain and serve as the foundation for developing and optimizing DL-models as a solution for venous measurements, reconstruction of OCT images from the linear in wavelength domain data, contributing to the physics of wave processes. As a result of a systematic theoretical and experimental study, new solutions for advanced data processing were proposed, and novel implementations of DL-based methods were developed. Building on the aforementioned conclusions, the thesis aims to address five main research questions identified as research gaps:

RQ1. How accurately can a DL model estimate segmentation maps of venous layers while maintaining low complexity through minimal number of model parameters, using the volumetric data acquired via SS-OCT?

We incorporated the ability of atrous and depthwise separable convolutions and proposed an encoder-decoder architecture based on U-Net using residual connections. The main contributions were in proposing the new bridge-parallel block, which helped in minimizing the number of parameters associated with the model along with giving highly accurate segmentation results for the venous layers.

RQ2. How can an integrated model be developed that simultaneously provides highly accurate segmentation maps - reliably identifying true positives and negatives, minimizing false detections - and provides accurate and consistent thickness estimations across diverse SS-OCT volumetric datasets of varicose veins?

We achieved it by leveraging the transfer learning in a single encoder and dual decoder-based model, trained specifically to learn the features necessary segmentation, followed by thickness measurements. The bottleneck module contains compressed and highly abstract representations

learned during the segmentation process, encapsulating the most informative, distinctive, and semantically meaningful features. These representations are subsequently employed for thickness measurements via transfer learning, enabling the model to attain higher accuracy and precision.

RQ3. How can a DL network be developed to reconstruct SS-OCT images directly from data linear in wavelength domain, bypassing the k-linearization step, thereby reducing hardware cost and computational demands, while maintaining image quality comparable to standard SS-OCT processing pipeline?

The acquired interferograms (linear in wavelength domain) were subjected to background subtraction, windowing and Fourier transform to obtain the low-quality blurred images. These blurred images were used as the input for the reconstruction network allowing to surpass the k-linearization step. The proposed DL model was trained with high quality ground truth images generated through a 7-fold B-scan averaging approach to minimize speckle noise. The model was able to achieve high performance in terms of superior image quality compared to OCT images generated by the commercial Optores system without introducing any additional hardware to perform linearization.

RQ4. How can an enhanced reconstruction DL network be developed to optimize data in both spatial and Fourier domains, producing high-fidelity, speckle-reduced images with improved PSNR and SSIM compared to conventional processing of SS-OCT raw data?

To obtain high-quality outputs, the ground truth images were derived from the SS-OCT system. To enhance intricate details and structures the DL model was optimized not only in spatial domain but also in Fourier domain for higher-level refinement. This multi-domain optimization enabled the framework to preserve high-frequency features, such as edges, while simultaneously suppressing noise without over-smoothing.

RQ5. Is it feasible to develop a low-complexity DL-based alternative to the standard reconstruction pipeline that achieves high-fidelity image reconstructions with reduced complexity, enabling real-time imaging?

We proposed a DL-based model capable of surpassing multiple stages of the conventional SS-OCT pipeline from transforming acquired interferogram to final OCT B-scans, thereby significantly improving the time-complexity. The single unified DL network achieves low inference time due to its use of a pre-trained model with parallel computations that incorporate batch processing and high level of optimizations.

Considering the above answers provided in context of the research questions identified in Chapter 1, it is also important to highlight some of the associated shortcomings with the presented works. The segmentation and thickness model proposed in Publication II uses the dataset incorporating the data from only four patients. However, the dataset should also incorporate normal veins and more samples to analyse the robustness of the model. The Publications III and IV use only one type of SS-OCT for data acquisition – the commercial Optores GmbH system thus featuring only one laser source. This limits the assessment of the generalization capability of the model. In this work, the 7-averaged speckle suppression method was used. However more advanced noise suppression methods can be studied in future works.

In light of these contributions and the shortcomings, we further highlight here some perspectives for extending and deepening the proposed research work. In this dissertation work, for the segmentation and thickness studies only, vessels affected by the varicose vein disease were considered for analysis. The study could be extended further by including more veins from diseased patients along with healthy subjects. The focus could also be laid on other vein diseases such as

Chronic Venous Insufficiency for better generalization capability and enhanced robustness to variability of the DL-model.

For the reconstruction of SS-OCT images from the raw data, only one type of light source was considered. In future studies, data could be acquired from diverse OCT systems incorporating different light sources. Further work could be done by incorporating dispersion compensation into the processing pipeline to further improve the image quality. Additionally, time-complexity could further be reduced to enable real-time processing through advanced optimization techniques and the use of recently proposed state-of-the-art neural networks.

References

[1] D. Huang et al., “Optical Coherence Tomography,” Science (80-.)., vol. 20, pp. 1–26, 1991..

[2] Sattler, Elke, Raphaela Kästle, and Julia Welzel. "Optical coherence tomography in dermatology." Journal of biomedical optics 18, no. 6 (2013): 061224-061224.

[3] Boone, Marc Alm, Sarah Norrenberg, G. B. E. Jemec, and Véronique Del Marmol. "Imaging of basal cell carcinoma by high-definition optical coherence tomography: histomorphological correlation. A pilot study." British journal of dermatology 167, no. 4 (2012): 856-864.

[4] Kermany, Daniel S., Michael Goldbaum, Wenjia Cai, Carolina CS Valentim, Huiying Liang, Sally L. Baxter, Alex McKeown et al. "Identifying medical diagnoses and treatable diseases by image-based deep learning." cell 172, no. 5 (2018): 1122-1131.

[5] Ahmed, Hanya, Qianni Zhang, Ferranti Wong, Robert Donnan, and Akram Alomainy. "Lesion Detection in Optical Coherence Tomography with Transformer-Enhanced Detector." *Journal of Imaging* 9, no. 11 (2023): 244.

[6] Wan, B., Clarisse Ganier, X. Du-Harpur, N. Harun, F. M. Watt, Rakesh Patalay, and M. D. Lynch. "Applications and future directions for optical coherence tomography in dermatology." British Journal of Dermatology 184, no. 6 (2021): 1014-1022.

[7] Zhou, Kang, Shenghua Gao, Jun Cheng, Zaiwang Gu, Huazhu Fu, Zhi Tu, Jianlong Yang, Yitian Zhao, and Jiang Liu. "Sparse-gan: Sparsity-constrained generative adversarial network for anomaly detection in retinal oct image." In 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pp. 1227-1231. IEEE, 2020.

[8] Sigsgaard, V., L. Themstrup, P. Theut Riis, J. Olsen, and G. B. Jemec. "In vivo measurements of blood vessels’ distribution in non-melanoma skin cancer by dynamic optical coherence tomography—a new quantitative measure?." Skin Research and Technology 24, no. 1 (2018): 123-128.

[9] Yang, Chin-Yi, Ja-Hon Lin, and Chien-Ming Chen. "Optical Coherence Tomography as a Diagnosis-Assisted Tool for Guiding the Treatment of Melasma: A Case Series Study." Diagnostics 14, no. 18 (2024): 2083.

[10] Lindert, Judith, Kianusch Tafazzoli-Lari, Ludger Tüshaus, Beke Larsen, Anna Bacia, Marie Bouteleux, Tina Adler et al. "Optical coherence tomography provides an optical biopsy of burn wounds in children—a pilot study." Journal of Biomedical Optics 23, no. 10 (2018): 106005-106005.

[11] Bouma, Brett E., Johannes F. de Boer, David Huang, Ik-Kyung Jang, Taishi Yonetsu, Cadman L. Leggett, Rainer Leitgeb et al. "Optical coherence tomography." Nature Reviews Methods Primers 2, no. 1 79, (2022).

[12] Drexler, Wolfgang, and James G. Fujimoto, eds. Optical coherence tomography: technology and applications. Springer Science & Business Media, 2008.

[13] De Boer, Johannes F., Rainer Leitgeb, and Maciej Wojtkowski. "Twenty-five years of optical coherence tomography: the paradigm shift in sensitivity and speed provided by Fourier domain OCT." Biomedical optics express 8, no. 7 (2017): 3248-3280.

[14] P. H. Tomlins, and R. K. Wang. "Theory, developments and applications of optical coherence tomography." Journal of Physics D: Applied Physics, 38, no. 15, 2519 (2005).

[15] Zhou, Shi-you, Chun-xiao Wang, Xiao-yu Cai, David Huang, and Yi-zhi Liu. "Optical coherence tomography and ultrasound biomicroscopy imaging of opaque corneas." Cornea 32, no. 4 (2013): e25-e30.

[16] Ramos, Jose Luiz Branco, Yan Li, and David Huang. "Clinical and research applications of anterior segment optical coherence tomography–a review." Clinical & experimental ophthalmology 37, no. 1 (2009): 81-89.

[17] Chen, Teresa C., Barry Cense, Mark C. Pierce, Nader Nassif, B. Hyle Park, Seok H. Yun, Brian R. White, Brett E. Bouma, Guillermo J. Tearney, and Johannes F. De Boer. "Spectral domain optical coherence tomography: ultra-high speed, ultra-high resolution ophthalmic imaging." Archives of ophthalmology 123, no. 12 (2005): 1715-1720.
[18] Leitgeb, R. A., W. Drexler, A. Unterhuber, B. Hermann, T. Bajraszewski, T. Le, A. Stingl, and A. F. Fercher. "Ultrahigh resolution Fourier domain optical coherence tomography." Optics express 12, no. 10 (2004): 2156-2165.
[19] Potsaid, Benjamin, Bernhard Baumann, David Huang, Scott Barry, Alex E. Cable, Joel S. Schuman, Jay S. Duker, and James G. Fujimoto. "Ultrahigh speed 1050nm swept source/Fourier domain OCT retinal and anterior segment imaging at 100,000 to 400,000 axial scans per second." Optics express 18, no. 19 (2010): 20029-20048.
[20] Wieser, Wolfgang, Benjamin R. Biedermann, Thomas Klein, Christoph M. Eigenwillig, and Robert Huber. "Multi-megahertz OCT: High quality 3D imaging at 20 million A-scans and 4.5 GVoxels per second." Optics express 18, no. 14 (2010): 14685-14704.
[21] Klein, Thomas, Wolfgang Wieser, Lukas Reznicek, Aljoscha Neubauer, Anselm Kampik, and Robert Huber. "Multi-mhz retinal oct." *Biomedical optics express* 4, no. 10 (2013): 1890-1908.
[22] Attendu, Xavier, Roosje M. Ruis, Caroline Boudoux, Ton G. Van Leeuwen, and Dirk J. Faber. "Simple and robust calibration procedure for k-linearization and dispersion compensation in optical coherence tomography." *Journal of biomedical optics* 24, no. 5 (2019): 056001-056001.
[23] [48] Wu, Tong, Zhihua Ding, Ling Wang, and Minghui Chen. "Spectral phase based k-domain interpolation for uniform sampling in swept-source optical coherence tomography." Optics express 19, no. 19 (2011): 18430-18439.
[24] Szkulmowski, Maciej, Maciej Wojtkowski, Tomasz Bajraszewski, Iwona Gorczyńska, Piotr Targowski, Wojciech Wasilewski, Andrzej Kowalczyk, and Czesław Radzewicz. "Quality improvement for high resolution in vivo javascript:void(0)images by spectral domain optical coherence tomography with supercontinuum source." Optics communications 246, no. 4-6 (2005): 569-578.
[25] Chen, Yuping, Hong Zhao, and Zhao Wang. "Investigation on spectral-domain optical coherence tomography using a tungsten halogen lamp as light source." Optical review 16, no. 1 (2009): 26-29.
[26] Liang, Kaicheng, Zhao Wang, Osman O. Ahsen, Hsiang-Chieh Lee, Benjamin M. Potsaid, Vijaysekhar Jayaraman, Alex Cable, Hiroshi Mashimo, Xingde Li, and James G. Fujimoto. "Cycloid scanning for wide field optical coherence tomography endomicroscopy and angiography in vivo." Optica 5, no. 1 (2018): 36-43.
[27] Zhang, Jason, Tan Nguyen, Benjamin Potsaid, Vijaysekhar Jayaraman, Christopher Burgner, Siyu Chen, Jinxi Li et al. "Multi-MHz MEMS-VCSEL swept-source optical coherence tomography for endoscopic structural and angiographic imaging with miniaturized brushless motor probes." Biomedical Optics Express 12, no. 4 (2021): 2384-2403.
[28] Lee, Hwi Don, Gyeong Hun Kim, Jun Geun Shin, Boram Lee, Chang-Seok Kim, and Tae Joong Eom. "Akinetic swept-source optical coherence tomography based on a pulse-modulated active mode locking fiber laser for human retinal imaging." Scientific reports 8, no. 1 (2018): 17660.
[29] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241. Cham: Springer international publishing, 2015.
[30] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.
[31] Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla. "Segnet: A deep convolutional encoder-decoder architecture for image segmentation." IEEE transactions on pattern analysis and machine intelligence 39, no. 12 (2017): 2481-2495.

[32] Zhao, Hengshuang, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. "Pyramid scene parsing network." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2881-2890. 2017.
[33] Chen, Liang-Chieh, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. "Encoder-decoder with atrous separable convolution for semantic image segmentation." In Proceedings of the European conference on computer vision (ECCV), pp. 801-818. 2018.
[34] Ashish, Vaswani. "Attention is all you need." Advances in neural information processing systems 30 (2017): I.
[35] Gu, Shuhang, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. "Weighted nuclear norm minimization with application to image denoising." *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2862-2869. 2014.
[36] Zhang, Zhengxin, Qingjie Liu, and Yunhong Wang. "Road extraction by deep residual u-net." *IEEE Geoscience and Remote Sensing Letters* 15, no. 5 (2018): 749-753.

List of publications

Publication I M. Viqar, V. Madjarova, V. Baghel, and E. Stoykova, "Opto-UNet: optimized UNet for segmentation of varicose veins in optical coherence tomography," in *Proc. 2022 10th Eur. Workshop on Visual Information Processing (EUVIP)*, pp. 1–6, IEEE, 2022.

Publication II M. Viqar, V. Madjarova, E. Stoykova, D. Nikolov, E. Khan, and K. Hong, "Transfer learning-based approach for thickness estimation on optical coherence tomography of varicose veins," *Micromachines*, vol. 15, no. 7, 2024.

Publication III M. Viqar, E. Sahin, V. Madjarova, E. Stoykova, and K. Hong, "WAVE-UNET: wavelength-based image reconstruction method using attention UNET for OCT images," in *Proc. Medical Imaging 2024: Physics of Medical Imaging*, vol. 12925, pp. 839–847, SPIE, 2024.

Publication IV M. Viqar, E. Sahin, E. Stoykova, and V. Madjarova, "Reconstruction of optical coherence tomography images from wavelength space using deep learning," *Sensors*, vol. 25, no. 1, 2024.

List of presentations

Oral Presentations:

1. 10th European Workshop on Visual Information Processing (EUVIP), Lisbon, Portugal, 11-14 September, 2022.
2. PLENOPTIMA Webinar Series 1, 2022.
3. Presentation at Tampere University, Secondments, January, 2023.
4. PLENOPTIMA Webinar Series 2, 2023.
5. HISTRATE conference on Advanced Composites under High Strain Rates Loading: A Route to Certification-by-Analysis, Istanbul, 5-6 June, 2024.
6. PLENOPTIMA Webinar Series 3, 2024.
7. European Light Field Imaging Workshop (ELFI), Sozopol, Bulgaria, 25-27 June, 2024.
8. Bulgarian-Korean Workshop, Photonics and Sensorics, Sofia, Bulgaria, 10-11 October, 2024.

Poster Presentations:

9. Digital Holography and 3-D Imaging, Cambridge, United Kingdom, 1–4 August 2022.
10. Materials, Methods & Technologies Conference, Burgas, Bulgaria, 19- 22 August, 2022.
11. PLENOPTIMA Training School 2: Machine and Deep Learning and Workshop 3: Research Communication, Mid Sweden University, Sundsvall, Sweden, 19-23 September, 2022.
12. Medical Imaging 2024: Physics of Medical Imaging, San Diego, California, United States of America, 18-22 February, 2024.
13. HISTRATE conference on Advanced Composites under High Strain Rates Loading: A Route to Certification-by-Analysis, Istanbul, Turkiye, 5-6 June, 2024.
14. International Conference on Advances in Material Science & Photonics (AMSP'25), Bansko, Bulgaria, 15 - 18 July, 2025.

List of citations

Paper:

Viqar, Maryam, Violeta Madjarova, Vipul Baghel, and Elena Stoykova. "Opto-UNet: optimized UNet for segmentation of varicose veins in optical coherence tomography." In *2022 10th European Workshop on Visual Information Processing (EUVIP)*, pp. 1-6. IEEE, 2022.

Citations:

1. Rufer, Libor, Josué Esteves, Didace Ekeom, and Skandar Basrour. "Piezoelectric Micromachined Microphone with High Acoustic Overload Point and with Electrically Controlled Sensitivity." *Micromachines* 15, no. 7 (2024): 879.
2. Su, Junjia, Yihao Chen, Pengcheng Feng, Zhelong Jiang, Zhigang Li, and Gang Chen. "A high resolution and configurable 1T1R1C ReRAM Macro for Medical Semantic Segmentation." *IEICE Electronics Express* 21, no. 8 (2024): 20240071-20240071.
3. Mao, Rongzhi, Liangxu Xie, Xiaohua Lu, Jialu Pei, Xiaojun Xu, and Shan Chang. "Harnessing Multiple Level Features to Improve Segmentation Performance of Deep Neural Network: A Case Study in Magnetic Resonance Imaging of Nasopharyngeal Cancer." *IEEE Access* (2024).
4. Rajathi, V., A. Chinnasamy, and P. Selvakumari. "DUTC Net: A novel deep ulcer tissue classification network with stage prediction and treatment plan recommendation." *Biomedical Signal Processing and Control* 90 (2024): 105855.
5. Rathnam, KS Revanth Sai, V. Lokanadham, C. Vignesh, and Vijayakumar Peroumal. "Varicose Veins Disease Detection and Automated Treatment using Body Area Network." In *2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*, pp. 1-6. IEEE, 2024.
6. Arunkumar, M., A. Mohanarathinam, and Kamalraj Subramaniam. "Detection of varicose vein disease using optimized kernel Boosted ResNet-Dropped long Short term Memory." *Biomedical Signal Processing and Control* 87 (2024): 105432.
7. Das, M. Ashwin, I. Anand, C. Nihal, Kamalraj Subramaniyam, and A. Mohanarathinam. "Early Detection and Prevention of Varicose Veins using Embedded Automation and Internet of Things." In *2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp. 1476-1482. IEEE, 2023.

Viqar, Maryam, Violeta Madjarova, and Elena Stoykova. "Modified watershed approach for segmentation of complex optical coherence tomographic images." Materials, Methods & Technologies.

8. Arrabiyeh, Peter A., Moritz Bobe, Miro Duhovic, Maximilian Eckrich, Anna M. Dlugaj, and David May. "Implementing low budget machine vision to improve fiber alignment in wet fiber placement." *Journal of Reinforced Plastics and Composites* (2024): 07316844241278050.
9. Anzabi, Leila Chehreghani, Azizollah Shafiekhani, Elham Darabi, Mirosław Bramowicz, Sławomir Kulesza, Jamshid Sabbaghzadeh, and Shahram Solaymani. "Nanoscale 3D spatial analysis of FTO/ZnO/Ag-x films subjected to photocatalytic activity." *Scientific Reports* 15, no. 1 (2025): 7330.

Viqar, Maryam, Violeta Madjarova, Amit Kumar Yadav, Desislava Pashkuleva, and Alexander S. Machikhin. "Deep learning based segmentation of optical coherence tomographic images of human saphenous varicose vein." In *Digital Holography and Three-Dimensional Imaging*, pp. W2A-5. Optica Publishing Group, 2022.

10. Cui, Weixin, Shan Lou, Wenhan Zeng, Visakan Kadirkamanathan, Yuchu Qin, Paul J. Scott, and Xiangqian Jiang. "BlurRes-UNet: A novel neural network for automated surface characterisation in metrology." *Computers in Industry* 165 (2025): 104228.
11. MJ, Prasanna Kumar, and H. N. Prakash. "A Comprehensive Review on Varicose Veins and its Implications using AI Methods." *Grenze International Journal of Engineering & Technology (GIJET)* 10 (2024).

Acknowledgments

This doctoral dissertation has been prepared within the European Joint Doctoral Programme on Plenoptic Imaging (PLENOPTIMA) and received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956770.