

UNIVERSITY OF TECHNOLOGY SYDNEY
Faculty of Engineering and Information Technology

**Modular Structure-Preserving OCT Image
Denoising Using Decorrelated Structural
Constraints**

by

Calvin D. Leighton

A THESIS SUBMITTED
IN FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Bachelors of Computing Science

Supervisor: A/Prof. Wenjing Jia

Sydney, Australia

Date: June 13, 2025

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Calvin D. Leighton, declare that this thesis is submitted in fulfilment of the requirements for Honours, within the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

Signature:

Date: June-10-2025

ABSTRACT

Modular Structure-Preserving OCT Image Denoising Using Decorrelated Structural Constraints

by

Calvin D. Leighton

Optical coherence tomography imagery is an in demand medical imaging modality. It is non-invasive and has the capability to capture images of highly sensitive and micro-sized parts of the human body such as the retina. Due to the process of interferometry, visual noise is introduced, creating a 'salt-and-pepper' effect which can make the diagnosis of disease more challenging. Machine learning has been used recently to address such an issue through denoising, a process that reduces the image's inherent visual artefacts.

By introducing a dedicated segmentation model named the "Structural-Preserving Denoising Network" (SPDN) to identify and preserve the most important information of an image, this research presents an alternative schema to denoise and improve upon reference-free image-evaluation metrics. This is a model which is modular in nature; that is, it can take the output of a baseline denoising model and compare it to the structural cohesion of the input to compare how much information was lost in important structural zones.

The modular design enables seamless integration with existing clinical workflows, while quantitative analysis establishes a framework for structure-preservation assessment in medical imaging applications. Evaluation was conducted across three diverse OCT datasets, revealing performance trade-offs between clinical utility metrics and traditional image quality measures, providing critical insights for medical image denoising evaluation methodologies.

SPDN was successfully integrated with three unsupervised frameworks (Noise2Noise,

Noise2Void, Noise2Self), demonstrating substantial signal-to-noise ratio improvements up to 9.19 dB (**reaching 29.4589 dB**). The N2S-SPDN variant achieved optimal performance in this metric, delivering state-of-the-art SNR improvements with preserved structural integrity essential for diagnostic accuracy. Clinicians can evaluate the trade-offs between image quality or diagnostic usability in their use cases, depending on what the situation calls for.

Dissertation directed by A/Prof. Wenjing Jia

School of Electrical and Data Engineering

Acknowledgements

First and foremost, I would like to acknowledge my supervisor for her continuous and endless supervision and encouragement.

Finally, I am particularly grateful to all people who directly or indirectly provided support on my research and the authors of the papers that I have cited.

Calvin D. Leighton
Sydney, Australia, 2025.

Contents

| | |
|-------------------|------|
| Abstract | iii |
| Acknowledgments | v |
| List of Figures | x |
| List of Tables | xi |
| List of Equations | xii |
| Abbreviation | xiii |
| Key Terminology | xiv |

1 Introduction 1

| | |
|------------------------------------------------------------------------|---|
| 1.0.1 Decorrelation Analysis in OCT Applications | 2 |
| 1.0.2 Motivation for Advanced Denoising | 3 |
| 1.0.3 Contextualising Speckle as Informative, Not Just Noisy | 4 |

2 Literature Review 6

| | |
|--------------------------------------------------------|---|
| 2.1 Overview of Optical Coherence Tomography | 6 |
| 2.1.1 Types of OCT Imaging | 6 |
| 2.1.2 Clinical Requirements | 7 |
| 2.1.3 Clinical Impact of Denoising | 7 |
| 2.2 Denoising Methods | 8 |
| 2.2.1 Traditional Approaches | 8 |
| 2.2.2 Deep Learning Approaches | 9 |

| | |
|------------------------------------------------------------------------|-----------|
| 2.3 Challenges and Gaps | 12 |
| 2.3.1 Evaluation Challenges in OCT Denoising | 13 |
| 2.3.2 Research Questions, Aims and Objectives | 14 |
| 2.3.3 Significance and Novelty | 15 |
| 3 Structural-Preserving Denoising Network | 17 |
| 3.1 Research Framework and Design | 17 |
| 3.2 Ethical and Practical Considerations | 18 |
| 3.3 Structural-Preserving Denoising Network Data-Preparation | 18 |
| 3.3.1 Datasets | 19 |
| 3.3.2 Dataset Configuration | 20 |
| 3.3.3 Standard Processing | 20 |
| 3.3.4 OCT Domain Preprocessing | 21 |
| 3.3.5 Background Thresholding | 22 |
| 3.4 SPDN Implementation | 24 |
| 3.4.1 Network Design | 24 |
| 3.4.2 SPDN Training Schema | 27 |
| 3.4.3 Integration with Unsupervised Frameworks | 30 |
| 3.5 Experimental Configuration | 35 |
| 3.5.1 Computational Infrastructure | 35 |
| 3.5.2 Development Environment | 35 |
| 3.5.3 Software Framework | 36 |
| 3.5.4 Dependencies and Libraries | 36 |
| 3.5.5 Experimental Design Framework | 36 |
| 3.6 Evaluation Protocol | 38 |

| | | |
|----------|--------------------------------------------------------------|-----------|
| 3.6.1 | Baseline Methods Selection | 38 |
| 3.6.2 | Qualitative Assessment | 38 |
| 3.6.3 | Objective Metrics for Image Quality Assessment | 38 |
| 3.6.4 | Domain-Specific Metrics for Medical Image Analysis | 39 |
| 3.7 | Methodological Constraints and Challenges | 41 |
| 4 | Experimental Results | 43 |
| 4.1 | N2N vs N2N-SPDN | 44 |
| 4.1.1 | Qualitative Analysis | 44 |
| 4.1.2 | Quantitative Analysis | 45 |
| 4.2 | N2V vs N2V-SPDN | 46 |
| 4.2.1 | Qualitative Analysis | 46 |
| 4.2.2 | Quantitative Analysis | 47 |
| 4.3 | N2S vs N2S-SPDN | 48 |
| 4.3.1 | Qualitative Analysis | 48 |
| 4.3.2 | Quantitative Analysis | 49 |
| 4.3.3 | Quantitative Summary | 49 |
| 4.4 | Ablation Studies | 50 |
| 4.4.1 | SPDN Ablations | 50 |
| 4.4.2 | N2x-SPDN Ablations | 53 |
| 4.4.3 | Duke Dataset Preprocessing Ablations Evaluation | 55 |
| 4.5 | Discussion | 56 |
| 4.5.1 | Performance Analysis Across Datasets | 56 |
| 4.5.2 | SPDN Integration Insights | 57 |
| 5 | Conclusion | 59 |

| | | |
|-----|-------------------------------------------------|-----------|
| 5.1 | Key Findings and Contributions | 59 |
| 5.2 | Limitations and Future Work | 60 |
| 5.3 | Final Remarks | 61 |
| | Bibliography | 62 |
| | Appendix | 69 |
| A | N2-Denoising Schemas | 69 |
| A.1 | Noise2Noise Framework | 69 |
| A.2 | Noise2Void Framework | 69 |
| A.3 | Noise2Self Framework | 70 |
| B | Dataset Statistics | 70 |
| B.1 | ICIP Dataset Sample-Image Statistics | 70 |
| B.2 | OCTDL Dataset Sample-Image Statistics | 71 |
| B.3 | Duke Dataset Sample-Image Statistics | 72 |
| C | Training Settings | 73 |
| C.1 | Parameters | 73 |
| C.2 | Evaluation Metrics | 73 |
| D | Ablation Studies | 74 |
| D.1 | Small SPDN U-Net | 74 |
| D.2 | Large SPDN U-Net | 75 |
| D.3 | SPDN U-Net without Attention | 76 |
| D.4 | U-Net without Attention Sample Output | 77 |

List of Figures

| | |
|-----------------------------------------------------------|----|
| 1.1 A Light-Speckle Sample | 2 |
| 3.1 U-Net Architecture for SPDN Segmentation | 25 |
| 3.2 SPDN Training | 27 |
| 3.3 SPDN Output | 28 |
| 3.4 N2xSPDN Training Schema | 29 |
| 3.5 U-Net-Architecture for N2-Denoising Schemas | 31 |
| 3.6 Regions of Interest (ROI) | 39 |
| 4.1 OCTDL Sample with Reference | 44 |
| 4.2 N2N Qualitative Output | 45 |
| 4.3 N2V Qualitative Output | 47 |
| 4.4 N2S Qualitative Output | 48 |
| 4.5 Small SPDN Output | 51 |
| 4.6 SPDN Output with Dice BCE Loss in SPDN | 52 |
| 4.7 SPDN Output with Different Losses | 52 |

List of Tables

| | | |
|-----|----------------------------------------------------------------------------------------------------|----|
| 4.1 | Comparison of N2N and N2N-SPDN Performance Metrics Across ICIP, OCTDL, and Duke Datasets | 46 |
| 4.2 | Comparison of N2V and N2V-SPDN Performance Metrics Across ICIP, OCTDL, and Duke Datasets | 47 |
| 4.3 | Comparison of N2S and N2S-SPDN Performance Metrics Across ICIP, OCTDL, and Duke Datasets | 49 |
| 4.4 | Comparison of N2S and N2S-SPDN Performance Metrics on ICIP Dataset | 53 |
| 4.5 | Comparison of N2S and N2S-SPDN Performance Metrics on OCTDL Dataset | 54 |
| 4.6 | N2S vs N2S-SPDN Performance on Duke Dataset Under Different Preprocessing Conditions | 55 |

List of Equations

| | |
|--------------------------------------------------------|----|
| 1.1 SSADA Decorrelation (Original) | 3 |
| 3.10 Channel Attention Mechanism | 26 |
| 3.11 Spatial Attention Computation | 26 |
| 3.12 Dilation Layer | 27 |
| 3.13 Structure Output Layer | 27 |
| 3.14 Mean Squared Error | 28 |
| 3.17 Loss n2n spdn | 33 |
| 3.21 Peak Signal to Noise Ratio | 38 |
| 3.22 Structural Similarity Index Measurement | 39 |
| 3.23 Signal to Noise Ratio | 40 |
| 3.24 Contrast-to-Noise Ratio | 40 |
| 3.25 Equivalent Number of Looks | 40 |
| 3.26 Edge Preservation Index | 40 |

Abbreviation

- **OCT:** Optical Coherence Tomography
- **SPDN:** Structural-Preserving Denoising Network
- **N2N:** Noise2Noise
- **N2V:** Noise2Void
- **N2S:** Noise2Self
- **N2N-SPDN:** Noise2Noise+Structural-Preserving Denoising Network
- **N2V-SPDN:** Noise2Void+Structural-Preserving Denoising Network
- **N2S-SPDN:** Noise2Self+Structural-Preserving Denoising Network
- **N2x-SPDN:** Noise2 Framework with Structural-Preserving Denoising Network
- **MSE:** Mean Squared Error
- **PSNR:** Peak Signal-to-Noise Ratio
- **SSIM:** Structural Similarity Index
- **CNR:** Contrast-to-Noise Ratio
- **SNR:** Signal-to-Noise Ratio
- **ENL:** Equivalent Number of Looks
- **EPI:** Edge Preservation Index
- **SSADA:** Split-Spectrum Amplitude-Decorrelation Angiography Algorithm
- **ROI:** Region-of-Interest

Key Terminology

- **B-scan:** A single cross-sectional OCT image slice
- **Speckle:** Granular noise pattern inherent to coherent imaging
- **Denoising:** The process or removing visual interference from an image.
- **Decorrelation:** Measure of temporal change between sequential images
- **Self-supervised:** Learning without requiring clean reference images
- **Unsupervised:** Denoising methods that learn from noisy data without clean ground truth images (e.g., N2N, N2V, N2S)
- **Supervised:** Learning methods that require paired clean-noisy training data for ground truth comparison
- **Reference-Required Metrics:** Image quality measures that need clean reference images for comparison (PSNR, SSIM)
- **Reference-Free Metrics:** Quality measures computed from the image alone without clean references (SNR, CNR, ENL)
- **Reference-Reduced Metrics:** Metrics requiring comparison between original and processed images but no clean reference (EPI)
- **Angiography OCT:** OCT imaging mode that detects blood flow through motion analysis
- **Structural OCT:** OCT imaging mode capturing static anatomical architecture and tissue boundaries
- **Blind-spot masking:** Training strategy where pixels are randomly masked and predicted (used in N2V)

- **Cross-sectional imaging:** Process generating slice views through biological imagery
- **Interferometry:** Optical technique using light interference patterns to measure structure
- **Patch-based processing:** Dividing images into smaller overlapping regions neural network training
- **Attention mechanisms:** Neural network components that selectively focus on important image features (channel and spatial)
- **Skip connections:** Direct pathways in U-Net architectures that preserve fine-scale details during reconstruction
- **Dilated convolution:** Convolution with expanded receptive fields
- **Ground truth:** Clean reference images used for supervised training (unavailable in clinical OCT)
- **Over-smoothing:** Denoising artefact where important diagnostic details are removed along with noise

Chapter 1

Introduction

Optical Coherence Tomography (OCT) is a medical imaging technique used to develop cross-sectional* images of a biological system [2]. Enhancing OCT denoising algorithms has significant potential to improve diagnostic accuracy and expand their application to other medical modalities. Currently, these algorithms face obstacles to clinical adoption due to their experimental nature, early-stage development, and challenges in producing tangible value. These algorithms are comprised of traditional methods and various supervised, semi-supervised and unsupervised deep learning methods, including: Generative Adversarial Networks (GANs) [3, 4, 5] and deep Convolutional Neural Networks (CNNs) [6, 7] all of which have their own training targets and denoising schemas.

OCT denoising methodologies have a primary challenge of inconsistent quality of raw OCT images and the outputs from denoising models and schemas. This concisely includes over-smoothing from denoising models, which results in a loss of diagnostic features or heavy computational costs, which are effectively unavailable to researchers without access to top-of-the-line graphical-processor-units (GPUs). There has been a growing trend in OCT denoising literature which moves away from traditional image quality metrics as they have become unrepresentative of what is actually important in a task as sensitive as medical imaging denoising; many researchers are asking: **What is actually important to denoise in an OCT**

*Cross-sectional imaging refers to some process which generates a view of a some biological system in slices [1]

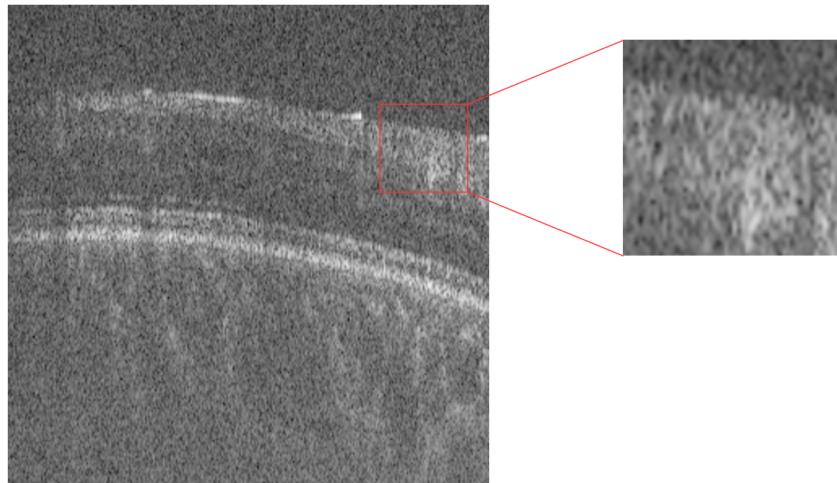


Figure 1.1 : A Light-Speckle Sample. The 'salt-and-pepper' artefacts are essentially extra information captured during the imaging process, which may impede or hide vital diagnostic information.

image? In a task so sensitive, it is of the utmost imperative to not remove any important information while removing uninformative information.

This research addresses the challenge of improving OCT image quality in unsupervised settings where ground truth clean images are unavailable. This is achieved from a dedicated speckle-splitting network called the '**Structural-Preserving Denoising Network**' (**SPDN**), which can be used during training for any baseline denoising schema as a dedicated constraint to preserve the quality and integrity of the medical images if sufficient data is available. This thesis is then split into two-sections: **segmenting based on decorrelation** and **Denoising using the segmentations as a constraint within the loss function**.

1.0.1 Decorrelation Analysis in OCT Applications

Jia et al. developed split-spectrum amplitude-decorrelation angiography (SSADA) for detecting blood flow within OCT acquisitions through speckle analysis between

sequential B-scans [8]. Originally designed for angiographic applications, the fundamental decorrelation principle can be adapted to identify structurally significant regions within standard OCT imaging. This quantifies intensity variation between consecutive frames, where high decorrelation typically indicates vascular locations due to frame-to-frame motion. This decorrelation can be used to preserve the important, high contrast information in SD-OCT images the same, which forms the basis of this research:

$$D(x, z) = 1 - \frac{1}{N-1} \sum_{n=1}^{N-1} \frac{A_n(x, z)A_{n+1}(x, z)}{\frac{1}{2}[A_n(x, z)^2 + A_{n+1}(x, z)^2]} \quad (1.1)$$

- D is the decorrelation value. The values closer to 1 indicate a higher variation.
- N is the total number of B-scans/slices/images.
- A_n is the amplitude / intensity at the pixel coordinates (x, z) at slice n .
- n is the current frame index.

Rather than measuring flow, the adaptation identifies high-contrast anatomical features. Subsequent developments in OCT angiography have demonstrated the clinical utility of decorrelation-based analysis. Gao et al. established comprehensive OCTA generation principles that enable non-invasive vascular assessment across multiple retinal layers [9]. Moult et al. advanced the technique through ultrahigh-speed swept-source implementations, reducing motion artefacts [10]. These developments validated decorrelation as a robust method for identifying speckle patterns with diagnostic significance.

1.0.2 Motivation for Advanced Denoising

The clinical applicability of OCT is directly linked to image quality, however, image-quality alone does not mean everything. Traditional techniques have been vastly

outperformed by denoising techniques in recent years and thus where this research is focused on. Advanced denoising methods are critical not only for improving the visual quality of OCT images but also for enhancing the reliability of quantitative analyses. Better noise suppression leads to more accurate separation of tissue boundaries and improved measurement of key clinical parameters, ultimately supporting more precise diagnoses and treatment monitoring.

1.0.3 Contextualising Speckle as Informative, Not Just Noisy

Speckle in OCT images arises from coherent light interference and can carry useful information about tissue structure and dynamics, not just noise [11, 12]. Traditional denoising methods often suppress all speckle, potentially discarding diagnostically relevant features, so treating all speckle as noise while denoising risks losing important information. SPDН aims to maintain the important information while removing uninformative speckle.

Hypothesis: A dedicated Structural Segmentation Network applied as a constraint will improve quantitative and qualitative results from existing denoising approaches (N2N, N2V, N2S) in clinically relevant signal preservation metrics, measured by signal-to-noise ratio (SNR) and edge preservation, while also revealing the fundamental limitation that current evaluation paradigms should not be emphasising traditional image quality metrics over clinical utility. SPDН improves existing methods by better preserving structure while reducing uninformative noise and preserving important structural information, which otherwise may be obstructed by visual speckle noise.

The inclusion of the SPDН network in the denoising schemas can be titled as N2N-SPDН; N2V-SPDН; N2S-SPDН (see footnote for link to publicly available training

code and model[†]). The classic denoising schemas which are widely acknowledged, are the noise-2-denoising networks: Noise2Noise (N2N) [13]; Noise2Void (N2V) [14]; Noise2Self (N2S) [15] which have shown to be effective in removing noise from images without a ground-truth target.

Structure of Paper: This thesis is structured as follows: Chapter 2 presents a comprehensive literature review examining OCT imaging fundamentals, traditional denoising methodologies, deep learning approaches for medical image enhancement, and identifies key research gaps to establish our research objectives. Chapter 3 details of the experimental framework, including data preprocessing protocols, the SPDН architecture implementation, evaluation metrics, ablation study design, and ethical considerations. Chapter 4 presents quantitative experimental findings and performance analysis across multiple evaluation criteria. Chapter 4.5 interprets these results, examining model performance characteristics, analysing successful approaches and limitations, synthesising insights from ablation studies, and acknowledging methodological constraints. Finally, a conclusion which summarises the key contributions, consolidates primary findings, and proposes directions for future research in OCT image enhancement.

[†]<https://github.com/nekoVIN/SPDN>

Chapter 2

Literature Review

2.1 Overview of Optical Coherence Tomography

OCT is a non-invasive imaging technique used primarily for capturing high-resolution images of biological tissues [2, 16]. OCT uses low-coherence interferometry, where light is directed into tissue and the back-scattered light from different depths is combined with a reference light beam. This creates interference patterns that allow the system to measure how far light has travelled into the tissue, enabling the reconstruction of cross-sectional images at different depths [17]. Due to the inherent nature of low-coherence interferometry, OCT images are frequently contaminated by noise, which can obscure subtle yet diagnostically critical features. Moreover, OCT systems acquire multiple frames in rapid succession, each of which potentially exhibits varying noise levels. This variability emphasises the need for robust denoising algorithms to enhance both image clarity and quantitative measurement reliability.

Speckle noise is a dominant artefact in OCT images, arising from the interaction of light waves during the imaging process. It is immediately characterisable by its ‘salt-and-pepper’ texture (Figure 1.1). Since these scatterers are randomly distributed within the tissue, the coherent light undergoes interference, speckle patterns that can cover important diagnostic information.

2.1.1 Types of OCT Imaging

Optical coherence tomography encompasses several imaging modalities, each with distinct technical approaches and clinical applications. The main categories include

structural OCT techniques and functional extensions that provide additional diagnostic information. OCT imagery started as Time-Domain OCT [16] (TD-OCT) which used mirrors, however, as technology advanced, so too did the OCT imaging process. Spectral-Domain OCT (SD-OCT) has become a clinical standard by using spectrometers to capture wavelengths [16] and capture images of the retinal layers; SD-OCT encompasses the primary data of this research.

- **Structural OCT:** Captures the static anatomical architecture of tissues (retinal layers, vessel walls, tissue boundaries)
- **OCT Angiography (OCTA):** Detects blood flow by analysing changes between sequential scans

This research primarily works with structural OCT images but leverages decorrelation analysis (originally developed for OCTA) to identify structurally important regions that should be preserved during denoising.

2.1.2 Clinical Requirements

OCT has revolutionised non-invasive imaging across multiple clinical domains. In ophthalmology, it is indispensable for diagnosing and monitoring conditions such as age-related macular degeneration [18], by providing detailed views of retinal layers and the optic nerve head. The high-resolution and non-invasive nature of OCT not only improves diagnostic precision but also enhances patient management by enabling regular, risk-free monitoring of disease progression.

2.1.3 Clinical Impact of Denoising

There is a heavy focus on traditional image quality metrics PSNR [19] and SSIM [20] (see equations 3.21 and 3.22) which evaluate the image purely on visual quality. Clinical OCT interpretation requires the preservation of the important features such

as vessel boundaries or detail of structure for abnormalities. A denoising method that achieves high PSNR and SSIM may be removing these features while providing little clinical value, highlighting a clear need for clinically-guided evaluation frameworks.

2.2 Denoising Methods

This thesis explores popular baseline approaches with the proposed algorithm for OCT image denoising. Each method aims to reduce noise but not all focus on the preservation of structural information in OCT images.

2.2.1 Traditional Approaches

Tensor Ring Decomposition Guided Dictionary Learning Liu et al. [21] propose an advanced method that leverages low-rank matrix approximation to separate noise from essential image content. Their approach groups similar patches from OCT images into 3-D tensors, capturing low-rank structures to enhance spatial and temporal correlations. By applying shared dictionaries across these tensors, the method achieves superior sparsity and more effective denoising than traditional approaches. This technique has demonstrated significant promise in denoising capabilities for OCT images.

BM3D [22] groups similar 2D image patches into 3D arrays, applies collaborative filtering in the transform domain using wavelet-shrinkage denoising, and then aggregates the results back to reconstruct the denoised image. BM3D became a gold standard for traditional (non-deep learning) image denoising due to its effectiveness in exploiting both local and non-local image redundancies. However, this has been greatly overshadowed due to deep learning alternatives.

2.2.2 Deep Learning Approaches

Recent advances in machine learning have improved OCT denoising, though many methods struggle to achieve clinical-grade reference-required benchmarks (Appendix C.2) ($\text{PSNR} \geq 30$ and $\text{SSIM} \geq 0.90$) with minimal standard metric exploration on reference-free metric images (Appendix C.2). Signal-to-noise ratio (SNR), routinely used in clinical practice, demonstrates potential as a reference-free baseline indicator for medical image quality assessment [23].

2.2.2.1 Convolutional Neural Networks

Spectrogram-Based Deep Learning [24]: This method employs a simple architecture comprising two fully connected layers, convolution layers, and one deconvolution layer to predict noise patterns. This recent approach achieves high PSNR and SNR scores (26.2 and 35.0), indicating a promising direction for OCT denoising.

2.2.2.2 Noise-to-Noise Family Methods (Noise2X)

The noise-to-noise paradigm represents a breakthrough in training denoising networks without clean reference images. These methods exploit the principle that networks can learn to map between different noisy observations of the same underlying signal.

Noise2Noise (N2N): This foundational method trains on pairs of noisy images without requiring clean references and has been used in OCT denoising previously [13, 6, 25]. Built on the U-Net architecture [26], N2N is particularly well-suited for medical imaging, where clean data is often unavailable. The method leverages the insight that the expected value of multiple noisy observations converges to the clean signal (Appendix A.1).

Noise2Void (N2V): This method extends the noise-to-noise concept by training

on single noisy images [14]. N2V masks portions of the input during training and attempts to reconstruct them from the surrounding context, eliminating the need for paired noisy images (Appendix A.2).

Split-Spectrum Noise2Void (SSN2V): Building on N2V, this variant predicts missing image regions by comparing with other images to fill voids [7]. Schottenhamml et al provide a method that employs an elegant self-supervised learning strategy that makes it particularly suitable for scenarios where paired data is unavailable. The weakness in this research was the reliance on a computationally heavy model and the two step denoising process which is extremely computationally taxing. The biggest weakness of this paper is it's lack of exploration in quantitative results; Schottenhamml et al focus on clinical professional's preference scores which although makes sense, takes much of the scientific rigour away from the otherwise impressive results.

Noise2Self N2S uses image partitioning. Rather than needing paired noisy images or known noise characteristics, the method divides pixels into subsets and trains the network to predict one subset from another within the same noisy image[15] (Appendix A.3).

Other N2X Variants: **Neighbour2Neighbour** trains on random neighbouring pixels within a single noisy image, exploiting spatial redundancy for self-supervised denoising without clean or paired data [27]. **Patch2Self** performs denoising by predicting each voxel from surrounding non-overlapping patches, explicitly excluding the target voxel to prevent signal leakage, making it particularly effective for anisotropic biomedical data [28]. Because these methods all follow a similar design pattern to N2N, N2V and N2s, the SPDM model should theoretically work with each of them as well, however, time and computational resources restricted this thesis.

2.2.2.3 Transformers and GANS

Transformers generally outshine CNN approaches and GANs offer a powerful framework for OCT denoising by leveraging adversarial training; however, the computational cost are greater.

Transformer Enhanced Autoencoder Rendering (TEAR): Combines autoencoders and vision transformers for adaptive OCT cleaning, reaching PSNR, SSIM, CNR and ENL scores of 24.6, 0.74, 14.2 and 13807. These are extremely impressive scores and leaves this as a state-of-the-art method. Reaching scores like these with a U-Net is still achievable.

DA-TransUNet: represents a significant advancement in transformer-based medical image processing, integrating spatial and channel dual attention mechanisms with the U-Net architecture [29].

CycleGAN [5]: This unsupervised method employs cycle-consistent adversarial networks to remove speckle noise from OCT images. By learning mappings between noisy and clean image domains without paired examples, CycleGAN addresses the challenge of obtaining matched training data. The unsupervised nature means the model may learn to map denoised regions based on superficial visual similarities rather than true noise characteristics, potentially oversmoothing or creating new artefacts.

Non-Local GAN [3]: This approach incorporates real noise samples from OCT background regions to guide the generator. A non-local means layer enhances performance by capturing long-range dependencies in the image structure.

2.2.2.4 Attention Mechanisms in Medical Image Processing

Attention U-Net: Oktay et al. (2018) presented a U-Net model that incorporates attention gates to adaptively identify the important shapes and features while

suppressing irrelevant ones, called Attention U-Net [30]. The authors introduced attention gates to the standard U-Net architecture, enabling the model to focus on target structures while ignoring irrelevant regions.

CBAM: Woo et al showed that the CBAM [31] (Convolutional Block Attention Module) to improve convolutional neural networks, is easy to incorporate and improve adaptive feature refinement and is publicly*.

While recent advances in transformers and GANs models dominate more modern OCT denoising, many clinical and research settings still rely on lightweight self-supervised architectures due to real-time constraints and interpretability. This work proposes a modular enhancement block designed to integrate seamlessly with foundational Noise2X architectures (N2N, N2V, N2S), yielding performance gains without altering their core structure. This design allows legacy and future denoising methods to benefit from improved speckle suppression and structural preservation through architectural retrofitting.

2.3 Challenges and Gaps

The field of OCT image denoising faces several critical challenges that limit clinical adoption and research progress. First, the acquisition of clean reference OCT images remains practically infeasible due to the inherent nature of the imaging technique. Unlike natural images, where multiple exposures can be averaged to obtain ground truth, OCT imaging is particularly challenging to obtain a clean reference image.

No Clean Data: A primary challenge is that OCT data is hard to find- it's even harder to find verified, denoised, and quantified OCT datasets.

Poor Denoising Quality: OCT images are notoriously difficult to denoise, which is why there is so much research on trying to find the most effective method to

*Module available: <https://github.com/Jongchan/attention-module>

denoise them.

Segmentation Difficulties: Conventional edge-detection fails in OCT images due to speckle-induced false edges, requiring advanced segmentation approaches. Attempts made with regular segmentation attempts are normally fruitless without some form of denoising or image processing. This thesis aims to address this challenge in an elegant way by approximating the most important structures to focus on in an image using OCT-domain processing methods, in this case, using decorrelation to approximate the most important parts.

Insufficient differentiation between informative and non-informative speckle: Many existing methods treat all speckle as undesirable noise, overlooking the important domain-specific fact that some speckle patterns carry valuable information. Current approaches lack mechanisms to selectively preserve structural-related speckle while removing uninformative noise, potentially discarding clinically relevant information.

Limited exploration of self-supervised methods for OCT-specific constraints: While self-supervised approaches like Noise2Noise and Noise2Void have shown promise, their application to OCT has not fully exploited the unique properties of this imaging modality. Specifically, the relationship between structural OCT and OCT angiography remains underutilised in denoising frameworks.

2.3.1 Evaluation Challenges in OCT Denoising

The absence of ground truth data creates fundamental evaluation limitations beyond data acquisition challenges. Standard image quality metrics (PSNR, SSIM) cannot be computed without reference images, forcing reliance on no-reference metrics that may not correlate with clinical utility.

2.3.2 Research Questions, Aims and Objectives

Some research questions stemming from these gaps include:

1. Can a dedicated decorrelation-based segmentation constraint in deep learning preserve structural-related information whilst maintaining competitive denoising performance compared to standard unsupervised methods?
2. How effectively can decorrelation-based segmentation distinguish between diagnostically relevant structural information and uninformative noise artefacts in OCT images?

Aims: Explicitly, the aims of this research are:

1. To develop a novel segmentation constraint that preserves structural-related information within unsupervised denoising frameworks for OCT imagery.
2. Achieve measurable improvements among reference-based metrics (PSNR, SSIM).
3. Achieve measurable improvements among reference-free metrics (CNR, SNR, ENL).
4. Achieve measurable improvements among reference-reduced[†] metrics(EPI).

Objectives: The objectives were completed in order to meet the aims are:

1. **Obtained sufficient dataset:** It has been shown time and time again that using a sufficient dataset can be the only difference between a well-trained model or an under-trained one. A substantial dataset from participation in

[†]reference-reduced metrics refers to metrics which requires comparison between the original and the denoised but not a clean reference

the VIP Cup [32] for OCT denoising was obtained; however, to truly make this an impactful method, as much data as possible should be acquired, therefore, two other datasets were acquired for validation and testing: OCTDL [33] and Duke (2011) [34].

2. **Developed denoising algorithms:** During the research process, successful implementation of various denoising algorithms were implemented. Although some state-of-the-art denoising methods were originally developed, it took away the focus from the modular design of the work; so, instead the three-most popular N2x (N2N, N2V, N2S) denoising methods were developed with their N2x-SPDN variant to demonstrate that SPDN is flexible and modular. The results were recorded and relevant ablation studies were also recorded and included.
3. **Validated algorithmic performance:** These implementations were all fairly evaluated through quantitative image-quality scores (PSNR, SSIM) and clinically-relevant image scores (CNR, SNR, ENL, EPI).

2.3.3 Significance and Novelty

The N2x-SPDN approach seeks to overcome limitations in existing methods. This thesis offers a novel framework for future research and the progression of the OCT domain denoising. Quantitatively some important clinical measurements were improved, and although clinical evaluation could not be done by a registered ophthalmologist, the results lead to believe that the model improves the clinical viability of the images.

This research addresses several limitations in existing OCT segmentation approaches. Traditional denoising methods often compromise segmentation accuracy by over-smoothing fine-scale features critical for quantitative analysis. Existing deep learn-

ing approaches typically require paired clean-noisy training data, which is unavailable in clinical OCT acquisition.

SPDN represents a paradigm shift from conventional pipeline approaches. Rather than treating denoising and segmentation as separate sequential tasks, the proposed method jointly optimises both objectives through a unified architecture. This approach preserves speckle patterns essential for tissue characterisation whilst enabling precise boundary delineation. This approach offers improvements in OCT clinical image viability and can be extended to other medical imaging fields. While this method is specifically tailored to OCT image denoising, its nature makes it flexible enough to extend to other medical imaging formats, such as MRI or CT scans.

The key contributions made in this research include:

- Development of an unsupervised learning framework that operates without clean reference images
- Demonstration of improved quantitative clinical measurements compared to conventional preprocessing pipelines

Chapter 3

Structural-Preserving Denoising Network

3.1 Research Framework and Design

The research methodology of this thesis is highly experimental. This aligns with the objectives as they require lots of testing and quantifiable verification instead of only qualitative observation. To achieve the goals for a successful thesis, this thesis follows clear steps for reproducibility:

1. **Data Acquisition and Preprocessing:** Collected OCT data. Preprocessing was standardised across all data
2. **Denoising Algorithm Implementation:** The algorithms were implemented and the code has been publicly published for complete transparency*. The parameters of this schema was tuned through scrutiny and trial-and-error.
3. **Evaluation:** The algorithms were fairly evaluated and the hypothesis was tested. The image-reconstruction scores were used (PSNR, SSIM) as well as more clinically relevant measures (CNR, SNR, ENL, EPI)
4. **Challenges and Risks:** Challenges arose from re-implementation of baseline, state-of-the-art methods and the SPDN network.

This experimental research design is set in such a way to allow for duplication for future research.

*All code found here: <https://github.com/nekovin/FPSS>

3.2 Ethical and Practical Considerations

The ethical considerations of this research is essential. The main dataset contains 100 percent anonymised OCT images sourced from the 2024 VIP Cup competition [32]. The goal of this work is to provide zero misleading results, by providing explicit output scores with little focus on qualitative scores.

3.3 Structural-Preserving Denoising Network Data-Preparation

The preprocessing pipeline operates on the ICIP image sequences collected from patients across diabetes categories (types 0, 1, and 2), ensuring representative clinical sampling for training. Then the OCTDL and Duke for validation and testing.

Dataset construction follows a structured methodology: patient data collection with balanced diabetes representation, sequential OCT image loading with configurable neighbour frame selection ($n_neighbours = 4$), structural prediction using correlation-based detection ($threshold = 99$), binary thresholding transformation, and speckle noise removal through morphological filtering ($post_process_size = 2$) (Appendix C.1 for full parameter list used) which is explored in.

All the used datasets are **structural OCT images**; decorrelation was developed for angiography data but this thesis has implemented the same mathematical formulation to approximate the structures instead.

The pipeline maintains spatial consistency at $256 \times 256 \times 1$ image dimensions, with each patient contributing configurable image pairs. OCT images serve as network inputs whilst processed OCT binary masks function as target outputs. These binary masks act as supervision for the target manifestation; that is the target is the data which is located within the mask.

The training pipeline employs splitting, implementing 80:20 train-validation division

with fixed random seeding for reproducibility[†]. The pipeline utilised configurable batch sizes and shuffled training data with deterministic validation ordering, while ensuring Pytorch compatibility, and device-specific memory allocation for CUDA acceleration.

3.3.1 Datasets

No machine-learning project can start without obtaining the proper data. Three different datasets for training and validation purposes were used:

- **ICIP-OCT**: Contains structural OCT B-scans from 100 patients [32]. Used for SPDN training by generating decorrelation maps from sequential frames. This OCT dataset was collected from the 2024 VIP (Video and Image Processing) Cup. This is a high-quality dataset, containing 100 different volumes from patients with diagnosed “normal”, “diabetic”, or “non-diabetic” retinas. The dataset is of high-quality, raw images from the patients without accompanying ‘clean’ images.
- **OCTDL**: Contains both noisy structural OCT images and averaged (cleaner) versions [33]. Used for quantitative evaluation with reference metrics. Due to the private acquisition of this dataset and the unverified statistical significance, this will only be used for training. Therefore, the publicly available datasets SD-OCTDL[‡] [33] and Duke Dataset [34].
- **Duke 2011**: Contains structural OCT images with manual layer segmentations but no clean references. Used for validation with no-reference metrics. The most frequent and widely accepted statistical dataset is the Duke

[†]Machine learning projects tend to use the seed at 42 which is what was used here too.

[‡]OCTDL found here: <https://github.com/MikhailKulyabin>

dataset [34][§]. This dataset is accompanied with automatic segmentations of the layers, which can be fairly used to segment regions of interest (ROI).

3.3.2 Dataset Configuration

To ensure computational feasibility while maintaining statistical validity, a stratified random sampling approach was taken. For the SPDN training, 50 patients were taken with 20 OCT images C.1. For each of these patients the data-pipeline processed; considering each image, taking n-neighbours and calculating the decorrelation. This configuration balanced computational overhead constraints with sufficient sample size for robust statistical analysis and model validation procedures.

For the denoising stage, 100 patients were considered, 20 images per each. The varying patient sizes were considered like this because the entirety of the denoising dataset should not work with the SPDN model trained on the same data to ensure fair validation.

3.3.3 Standard Processing

The preprocessing of these images is the first critical step for the data pipeline. In the pipeline, there are general image-processing steps that are performed on all images. Then there is a separate preprocessing for the SPDN and the denoising process. The general preprocessing steps include normalisation and resizing for convolution layers into perfect squares.

1. **Normalisation:** The images were normalised between pixel ranges of 0-1 through min-max normalisation.
2. **Resizing:** All images during training are resized to 256x256; this is for smooth U-Net compatibility for adjustability and quick fixes.

[§]Duke Dataset publicly available at http://www.duke.edu/sf59/Chiu_BOE_2014_dataset.htm

3.3.4 OCT Domain Preprocessing

OCT image processing can undergo unique image-processing techniques which are only applicable within this domain. One such application is decorrelation.

3.3.4.1 Decorrelation Analysis

This research implemented decorrelation analysis to generate decorrelated images from structural OCT data, enabling the creation of paired datasets for training the split-speckle segmentation network. The decorrelation algorithm computes temporal changes between consecutive OCT amplitude frames acquired at the same spatial location [8]. For two amplitude frames A_n and A_{n+1} , the correlation coefficient is calculated as before in equation 1.0.1; where the decorrelation values approaches 1 for high motion and approaches 0 for static tissue [8]. To enhance signal-to-noise ratio and reduce bulk motion artefacts, this thesis has implemented multi-frame averaging using n neighbouring B-scans. For each central frame i , decorrelation values are computed with n preceding and n following frames:

$$D_{avg}(i) = \frac{1}{2n} \sum_{j=i-n, j \neq i}^{i+n} D(i, j). \quad (3.1)$$

- $D_{avg}(i)$: averaged decorrelation value for slice i
- i : slice being processed
- n : neighbouring slice
- j : index of neighbouring slice
- $D(i, j)$: decorrelation between slice i and j

This approach, similar to SSADA [8], improves vessel contrast while maintaining spatial resolution. SSN2V [7] implements a version with pairwise implementation

instead of n-nearest neighbours. This choice was made as it effectively suppresses decorrelation artefacts arising from residual speckle or involuntary eye movements.

Why Use Decorrelation on Structural Data? Although decorrelation analysis was originally designed to detect blood flow in OCTA, it effectively identifies high-contrast structural features in regular OCT images. By comparing sequential B-scans of the same location, decorrelation highlights:

- Vessel boundaries (due to blood cell movement)
- High-contrast tissue interfaces
- Regions with significant structural information

This allows SPDN to learn which image regions contain diagnostically important information that should be preserved during denoising.

3.3.5 Background Thresholding

An thresholding algorithm based on background tissue statistics was implemented to enhance structural contrast by suppressing these artefacts while preserving genuine structural signals. Background regions are identified using intensity-based segmentation, as low-intensity regions in OCT typically correspond to areas with minimal tissue structure:

$$M_{bg} = I_{OCT} < P_{th}(I_{OCT}) \quad (3.2)$$

where P_{th} represents the th -percentile threshold of the OCT intensity distribution [35]. Background statistics are computed from identified regions to establish baseline noise characteristics:

$$\mu_{bg} = \mathbb{E}[I_{OCT}|M_{bg}] \quad (3.3)$$

$$\sigma_{bg} = \sqrt{\text{Var}[I_{OCT}|M_{bg}]} \quad (3.4)$$

The intensity threshold separates the noise from the signal:

$$T_{intensity} = \mu_{bg} + 2\sigma_{bg} \quad (3.5)$$

Then, a signal mask is made:

$$M_{signal} = \text{clip}\left(\frac{I_{OCT} - T_{intensity}}{2\sigma_{bg}}, 0, 1\right) \quad (3.6)$$

The final structural prediction combines the decorrelatated pixels with the signal mask.

$$I_{Structural} = D_{avg} \cdot M_{signal} \quad (3.7)$$

This approach ensures that SPDN receives clean structural targets focused on genuine tissue architecture rather than noise artefacts, improving the reliability of structural preservation during denoising.

3.3.5.1 Pixel-wise Noise Removal

The final step in processing is the removal of small pixel groupings after decorrelation. These small groupings were found to be isolated and far from meaningful features, which simply added confusion to the model when training. Binary masks are created from thresholded decorrelated images, and connected components with an area smaller than a minimum threshold (5-10 pixels) are eliminated:

$$M_{cleaned} = \bigcup_i C_i \quad \text{where} \quad |C_i| \geq A_{min} \quad (3.8)$$

where big-cup is the Union of the components, where C_i represents individual connected components and A_{min} is the minimum area threshold.

3.3.5.2 Paired Dataset Construction

The preprocessing pipeline generates paired OCT-OCT datasets suitable for training. For each patient volume N B-scans, we extract $(N - 2n)$ valid OCT images,

where n is the neighbourhood size for decorrelation computation. Each OCT image is paired with its neighbour and corresponding decorrelated image that was generated from them to create input-target pairs for network training. The pairing strategy ensures spatial correspondence between structural and functional information:

$$\left\{ \left(I_{\text{OCT}}^{(i)}, I_{\text{OCT}}^{(i+k)} \right) \mid i = n+1, \dots, N-n; k \in [-n, n] \setminus \{0\} \right\} \quad (3.9)$$

The resulting paired datasets enable training of segmentation networks that can identify vascular structures in single OCT B-scans without requiring repeated acquisitions.

3.4 SPDN Implementation

3.4.1 Network Design

There are two U-Nets presented in this thesis. This thesis presents the SPDN architecture that addresses the fundamental challenge of preserving diagnostically relevant information whilst removing uninformative noise in OCT imagery and is the first U-Net. The proposed method extends the traditional U-Net framework through integrated attention mechanisms and speckle-specific architectural modifications tailored for OCT angiography applications.

3.4.1.1 Architectural Foundation

The SPDN employs a modified encoder-decoder architecture (Figure 3.1) based on the U-Net [26], enhanced with CBAM [31] mechanisms to improve feature selectivity in OCT speckle separation tasks. The network processes single-channel grayscale OCT images and produces outputs representing extracted structural (high contrast) components, which achieves segmentation while maintaining intensity levels of the features (Figure 3.1). The architectural design incorporates several key inclusions:

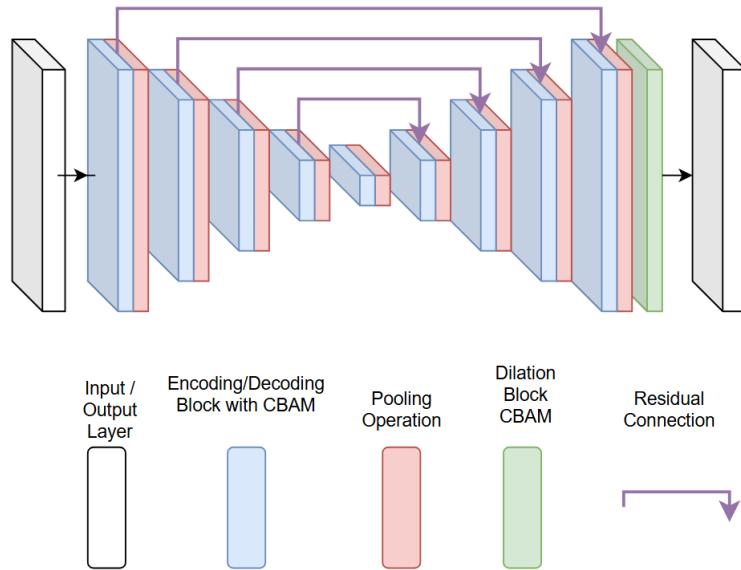


Figure 3.1 : U-Net Architecture for SPDN Segmentation. This is a 5-layer encoder-decoder network with attention mechanisms applied after each block, featuring progressively increasing feature dimensions (32,64,128,256,256 channels) and skip connections for multi-scale feature preservation.

- Configurable depth encoding blocks with progressive feature dimension scaling
- Integrated attention mechanisms at multiple hierarchical levels of the model
- Dilated convolution modules for expanded receptive field coverage
- Regularisation through strategic dropout placement

3.4.1.2 Encoder Architecture

The encoder pathway consists of d encoding blocks, where d represents the network depth (5). Each encoding block contains b convolutional layers (default: 3), implementing a sequential structure of 3×3 convolution with padding, batch normalisation, and ReLU activation. Feature dimensions progress geometrically from

an initial dimension f (default: 32) to $f \times 2^{\min(i,3)}$ at the i -th level, with growth capped to prevent excessive memory consumption. Max pooling with 2×2 kernels provides spatial downsampling between encoding levels.

3.4.1.3 Attention Mechanisms

The architecture incorporates dual attention modules applied at each encoder and decoder level. Following the CBAM methodology [31], channel attention and spatial attention have been employed:

Channel Attention: Channel attention employs both average and max pooling operations followed by a shared multi-layer perceptron (MLP):

$$\mathcal{M}_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \quad (3.10)$$

Spatial Attention: Spatial attention generates attention maps through channel-wise pooling operations:

$$\mathcal{M}_s(F) = \sigma(f^{7 \times 7}([\text{AvgPool}_c(F); \text{MaxPool}_c(F)])) \quad (3.11)$$

where $f^{7 \times 7}$ represents a 7×7 convolution operation, σ denotes the sigmoid activation function, F represents input feature maps, and $[; ;]$ indicates channel concatenation.

3.4.1.4 Decoder Implementation

The decoder pathway reconstructs spatial resolution through bilinear upsampling, incorporating skip connections from corresponding encoder levels. Each decoder block processes concatenated features from the upsampled path and encoder skip connections, maintaining the same convolutional block structure with integrated attention mechanisms.

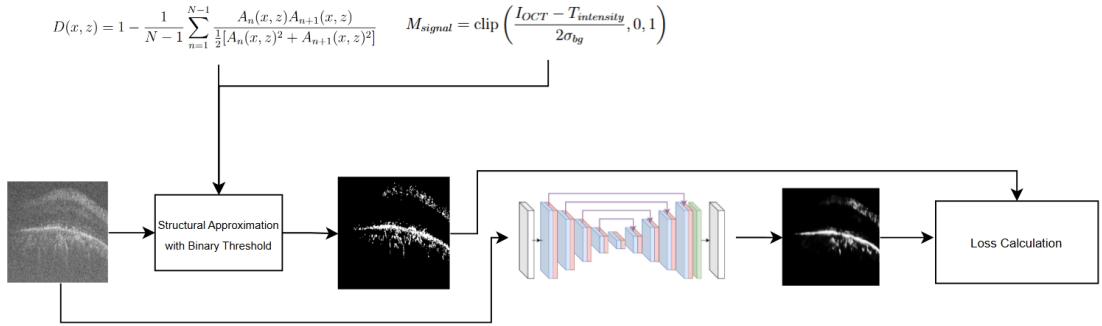


Figure 3.2 : SPDN Training: From the original slice of an OCT volume, process the image to get the decorrelated estimate and predict it using the U-Net (Figure 3.1).

3.4.1.5 Dilated Convolution Module

To capture multi-scale contextual information without increasing computational complexity, the architecture incorporates dilated convolutions with rates $\{1, 2, 4\}$:

$$\text{Dilation}(F) = \text{Conv}_{d=4}(\text{Conv}_{d=2}(\text{Conv}_{d=1}(F))) \quad (3.12)$$

where d represents the dilation rate and each convolution includes batch normalisation and ReLU activation.

3.4.1.6 Output Generation

The network produces dual output branches optimised for component separation:

$$F_{\text{Structure}} = \sigma(\text{Conv}_{1 \times 1}(\text{Conv}_{3 \times 3}(\text{Attention}(\text{Dilation}(F_{\text{decoder}}))))) \quad (3.13)$$

where σ represents sigmoid activation ensuring output values in $[0, 1]$.

3.4.2 SPDN Training Schema

The SPDN employs a supervised segmentation approach where the network learns to predict corresponding areas to decorrelated regions (Figure 3.2).

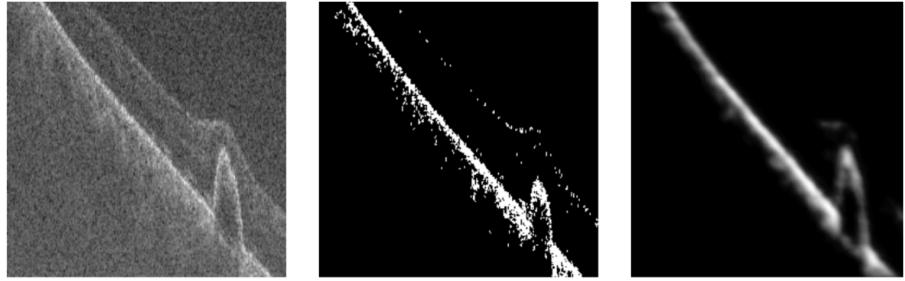


Figure 3.3 : SPDN Output using MSE as the loss function: raw image (left); decorrelated and processed binary structural prediction (middle); SPDN output (right). It can be seen that the threshold (middle) does not perfectly align with the original (left), this is because it is weighing against neighbours, but then the SPDN network takes over and more accurately captures the important structural features of the image (right).

3.4.2.1 Loss Function

Network optimisation employs **Mean Squared Error (MSE)** loss between predicted structural components and OCT-derived targets. MSE quantifies the average squared differences between original and reconstructed pixel intensities [36]. Lower MSE values indicate superior reconstruction quality:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^N (y_i^{\text{true}} - y_i^{\text{pred}})^2 \quad (3.14)$$

Figure 3.3 demonstrates the model’s input processing and prediction outputs. The network performs structural prediction through decorrelation analysis, subsequently optimising intensity distributions at corresponding spatial locations via MSE loss, where the binary calculation is present.

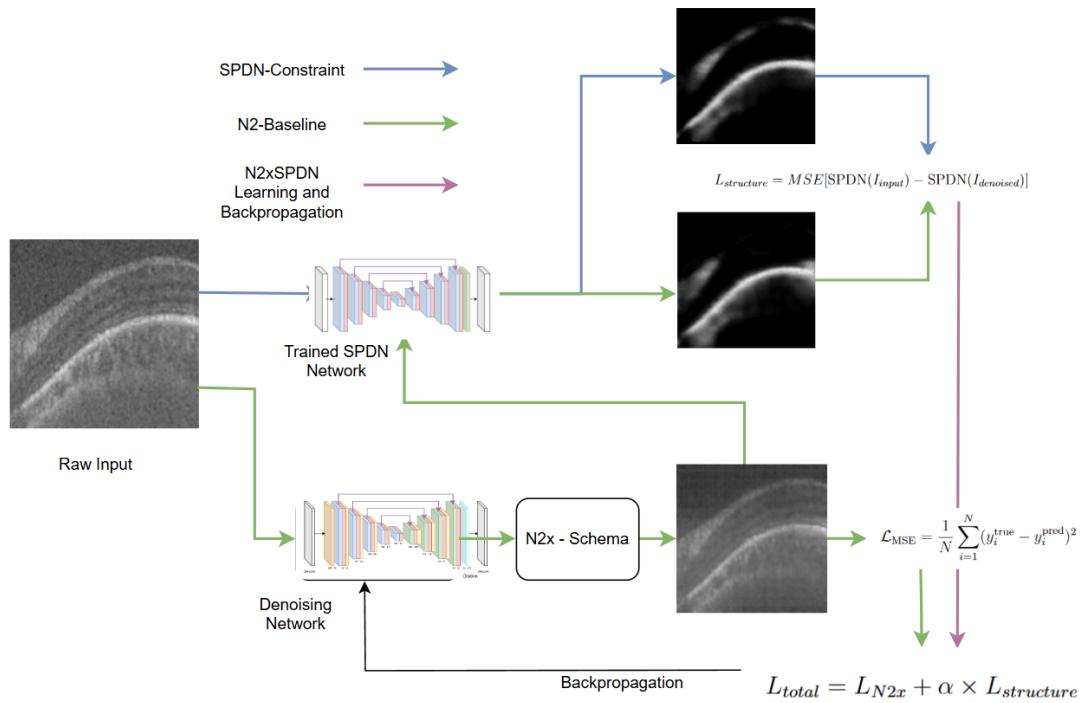


Figure 3.4 : N2xSPDN Training Schema: Raw image is passed into a trained SPDN network and an untrained denoising U-Net. The structure loss is calculated between the output from the raw input and the output of the denoising network. The structure loss and the N2x-Schema loss is cumulated with an adaptive alpha parameter as the loss function for the entire schema.

3.4.2.2 N2x-SPDN Schema

The protocol (shown in Figure 3.4) supports configurable hyperparameters through structured configuration dictionaries, enabling systematic experimental reproducibility. Training phases enable forward propagation, loss computation, backpropagation, and parameter updates, whilst validation phases operate with frozen gradients for performance evaluation without optimisation. Within this protocol, a dedicated batch processing function manages both phases through mode-dependent execution paths, incorporating comprehensive checkpoint management, preserving model states, achieving the lowest validation loss (best checkpointing) and maintaining recent training states for resumption capability (latest checkpointing).

3.4.2.3 Adaptive Training Features

Implementation incorporates learning rate scheduling through optimiser configuration, validation-based early stopping, debug mode for gradient flow analysis during initial epochs, and optional visualisation capabilities for training progress monitoring. This methodology ensures robust model development whilst maintaining experimental flexibility across diverse OCT imaging conditions.

3.4.3 Integration with Unsupervised Frameworks

3.4.3.1 SPDN Denoising Schema

The SPDN architecture demonstrates compatibility with established unsupervised denoising methodologies, including Noise2Noise (N2N), Noise2Void (N2V), and Noise2Self (N2S) frameworks. This integration enables training without clean reference images whilst maintaining speckle-specific separation capabilities.

The training methodology employs a pass-through baseline approach: baseline method output serves as twin input for the SPDN network. Following SPDN output generation from original and denoised images, average difference calculation between

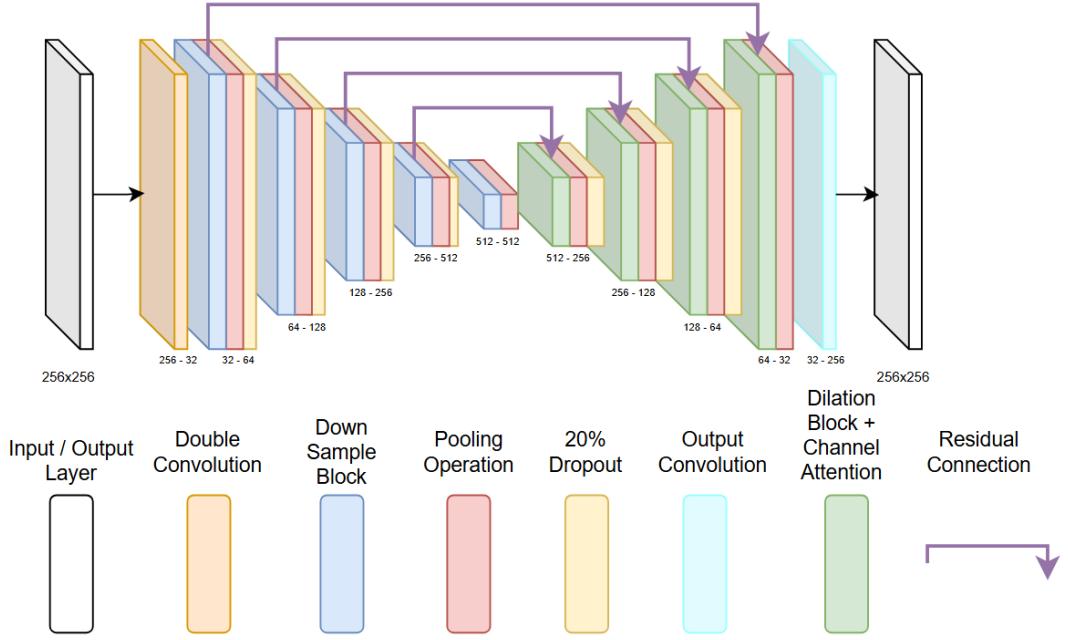


Figure 3.5 : U-Net-Architecture for N2-Denoising Schemas

structural predictions using MSE contributes to the total loss via adaptive alpha weighting. The modular design facilitates seamless integration whilst preserving core architectural innovations specific to OCT angiography applications, demonstrating network versatility through three implementation variations for comprehensive speckle separation across noisy-to-clean image transformations (Figure ??).

3.4.3.2 SPDN Constraint Mechanism

The SPDN constraint operates through a dual-loss framework that preserves structural information during denoising. The integration mechanism ensures that denoised outputs maintain the same structural patterns identified in the original noisy input. The total loss function combines reconstruction objectives with structural preservation:

$$L_{total} = L_{N2x} + \alpha \times L_{structure} \quad (3.15)$$

where L_{N2x} represents the reconstruction loss on masked pixels and L_{flow} quantifies structural preservation. The structural constraint is defined as:

$$L_{structure} = MSE[\text{SPDN}(I_{input}) - \text{SPDN}(I_{denoised})] \quad (3.16)$$

This constraint forces the denoising network to preserve structural patterns by minimising the difference between SPDN predictions on the original noisy image and the denoised output.

3.4.3.3 N2N and N2N-SPDN Framework Implementation

N2N networks learn by taking one noisy image as input and attempting to predict a different noisy image of the same scene as the target [13]. The key insight is that when the network minimises the difference between its prediction and the second noisy image, it learns to output the underlying clean signal since noise between the two images is uncorrelated (Appendix A.1 for N2S diagram).

The N2N framework integration with SPDN (N2N-SPDN) employs paired noisy image training methodology, where the network learns to map between different noise realisations of the same underlying signal. This approach leverages the statistical independence of noise between paired acquisitions whilst preserving underlying structural information through the integrated SPDN structural preservation mechanism.

The N2N-SPDN implementation utilises patch-based processing to accommodate GPU memory constraints whilst maintaining spatial coherence. The training protocol employs configurable patch sizes ($patch_size = 96$) with overlapping stride patterns ($stride = 48$) to ensure comprehensive spatial coverage during reconstruction (Appendix C.1).

The N2N-SPDN implementation demonstrates this constraint mechanism in practice. During training, both the noisy input and denoised output are processed

through the SPDN network to generate structural predictions. The absolute difference between these predictions forms the structural constraint term, ensuring that the denoising process preserves diagnostically relevant structural information identified in the original image.

The integrated loss function combines N2N reconstruction objectives with structural component preservation:

$$\mathcal{L}_{\text{N2N-SPDN}} = \mathcal{L}_{\text{N2N}} + \alpha_{\text{adaptive}} \cdot \mathcal{L}_{\text{structural}} \quad (3.17)$$

where $\mathcal{L}_{\text{reconstruction}}$ represents the base N2N criterion loss, $\mathcal{L}_{\text{structural_abs}}$ denotes absolute structural component preservation loss, and $\mathcal{L}_{\text{structural_mse}}$ represents mean squared error structural preservation loss.

3.4.3.4 N2V and N2V-SPDN Framework Implementation

The N2V framework integration employs blind spot training methodology, where a subset of pixels (10%) are masked during training [14]. The network learns to reconstruct masked pixels using surrounding contextual information, eliminating the requirement for paired clean-noisy training data (Appendix A.2 for N2S diagram). The implementation utilises Bernoulli sampling with a configurable mask ratio to generate blind spot inputs:

$$\text{blind input} = \text{mask} \odot \text{noise} + (1 - \text{mask}) \odot \text{input} \quad (3.18)$$

where \odot denotes element-wise multiplication and the mask follows Bernoulli distribution with mask ratio p . The combined loss function integrates N2V reconstruction loss with structural preservation:

$$\mathcal{L}_{\text{N2V-SPDN}} = \mathcal{L}_{\text{N2V}} + \alpha_{\text{adaptive}} \cdot \mathcal{L}_{\text{structural}} \quad (3.19)$$

where \mathcal{L}_{N2V} represents the reconstruction loss on masked pixels only.

For N2V-SPDN, the input image and reconstructed output are then passed into the SPDN model; the difference between the outputs of these is then compared with MSE.

3.4.3.5 N2S and N2S-SPDN Framework Implementation

N2S networks learn to denoise using only single noisy images by exploiting the self-similarity and redundancy present within the image itself. It learns to distinguish between coherent structure (which appears consistently) and random noise (which varies spatially). It does this by partitioning the image into defined patches in order to predict a subset of partitions in the same image (Appendix A.3 for N2S diagram).

The self-supervised approach eliminates requirements for paired training data by exploiting statistical properties of noise independence across spatial neighbourhoods. The integrated SPDN module ensures preservation of diagnostically relevant structural information throughout the self-supervised denoising process.

$$\mathcal{L}_{\text{N2S-SPDN}} = \mathcal{L}_{\text{N2S}} + \alpha_{\text{adaptive}} \cdot \mathcal{L}_{\text{structural}} \quad (3.20)$$

3.4.3.6 Patch-Based Processing Architecture

All N2x-SPDN framework implementations employ patch-based processing methodologies to address computational constraints whilst maintaining spatial coherence. The patch extraction protocol utilises overlapping sliding window approaches with configurable parameters, enabling memory management. The processing pipeline implements sub-batch processing with configurable sub-batch sizes to accommodate varying GPU memory configurations.

3.4.3.7 Checkpoint Management and Performance Tracking

The N2x-SPDN training protocol implements a checkpoint management strategies incorporating multiple performance criteria. The system maintains three distinct

checkpoint categories: best validation loss preservation and latest training state continuity.

3.4.3.8 Loss Function Analysis

The mathematical derivation of the combined N2x-SPDN loss function addresses the fundamental challenge of balancing reconstruction fidelity with structural preservation. The theoretical justification for MSE choice in structural preservation stems from its behaviour as an unbiased estimator for signal reconstruction quality. MSE provides differentiable loss landscapes conducive to gradient-based optimisation (Equation 3.20)

3.4.3.9 Overfitting Prevention Strategy

Overfitting prevention measures incorporate multiple regularisation strategies, including dropout placement within attention mechanisms, gradient clipping during backpropagation, and early stopping based on validation performance plateaus. Checkpoint management system was developed which enables model selection based on multiple performance criteria, ensuring optimal balance between training convergence and generalisation capability.

3.5 Experimental Configuration

3.5.1 Computational Infrastructure

Experimental procedures were conducted using an NVIDIA RTX 2080 GPU equipped with 8GB VRAM and supported by 32GB system RAM.

3.5.2 Development Environment

The primary development environment utilised Visual Studio Code for local implementation, debugging, and iterative model development. Virtual environments were

employed for training architectural variants and conducting extensive hyperparameter optimisation experiments requiring substantial computational resources.

3.5.3 Software Framework

Python 3.8+ served as the primary programming language, selected for its comprehensive machine learning ecosystem, extensive library support, and established precedent in medical image processing research. The development environment employed virtual environments to ensure reproducible dependency management.

3.5.4 Dependencies and Libraries

The experimental implementation used the following Python libraries:

- **PyTorch 1.12+:** Primary deep learning framework for model architecture implementation, training procedures, and GPU acceleration [37]
- **NumPy:** Fundamental numerical computing operations and multi-dimensional array processing [38]
- **OpenCV:** Computer vision operations including image preprocessing, augmentation, and quality assessment [39]
- **Matplotlib/Seaborn:** Statistical visualisation, result presentation, and publication-quality figure generation [40, 41]
- **Scikit-learn:** Traditional machine learning implementations for baseline method comparisons and evaluation metrics [42]
- **Torchvision:** Image transformation utilities and standard computer vision preprocessing pipelines [43]

3.5.5 Experimental Design Framework

The experimental design considers the variables and the experimental groups.

3.5.5.1 Variable Classification

- **Independent Variables:** SPDN architectural variants (N2N-SPDN, N2V-SPDN, N2S-SPDN with modular constraint integration), attention mechanism configurations, loss function formulations, training protocol parameters (learning rates 1e-4 to 1e-3, batch sizes 8-16, epoch counts 50-200, Adam optimiser with a weight decay of 1e-4 and scheduler with ReduceLROnPlateau with a patience of 5 and a factor of 0.5)
- **Dependent Variables:** Quantitative image quality metrics including PSNR (dB, pixel-level similarity), SSIM (0-1 scale, structural similarity), CNR (dB, layer contrast preservation), SNR (dB, signal clarity), ENL (equivalent number of looks, speckle reduction), and EPI (edge preservation fidelity), alongside computational metrics (inference time, memory usage, training convergence)
- **Controlled Variables:** Dataset composition (ICIP: 100 patients, 20 images per patient; OCTDL: B-scans; Duke 2011: segmented retinal data), preprocessing standardisation, evaluation methodologies, and hardware configurations (NVIDIA RTX GPU, CUDA 11.8, PyTorch 1.12+, reproducible random seeds)

3.5.5.2 Experimental Groups

- **Treatment Groups:** SPDN-enhanced unsupervised denoising implementations (N2N-SPDN, N2V-SPDN, N2S-SPDN) incorporating modular structural constraint modules, decorrelated feature preservation mechanisms, and adaptive attention-based architectural modifications for OCT-specific noise characteristics
- **Control Groups:** Standard unsupervised frameworks without SPDN integration (N2N, N2V, N2S) using identical training protocols, hyperparameters,

and evaluation procedures to ensure fair comparison and isolation of SPDN contributions

3.6 Evaluation Protocol

3.6.1 Baseline Methods Selection

The N2-baseline frameworks (N2N, N2V, N2S), as introduced in [13, 14, 15] and refined for OCT denoising in [6, 7], represents the more well-established unsupervised denoising methods and will serve as a primary baseline for this research. These baseline methods have been selected due to their widespread adoption in medical image denoising and availability of pre-trained models for reproducible evaluation.

3.6.2 Qualitative Assessment

A qualitative assessment ensures that quantitative improvements translate to practical clinical utility.

3.6.3 Objective Metrics for Image Quality Assessment

Rigorous quantitative evaluation is essential to objectively measure the effectiveness of the proposed SPDN. Multiple image quality metrics will be employed to provide comprehensive assessment of denoising performance. These metrics offer measurable, reproducible, and unbiased evaluation of image quality improvements.

1. **Peak Signal-to-Noise Ratio (PSNR):** PSNR measures the ratio between maximum signal power and corrupting noise power [19, 36]. Higher PSNR values indicate better reconstruction quality:

$$PSNR = 10 \cdot \log_{10} \left(\frac{R^2}{MSE} \right) \quad (3.21)$$

where R represents the maximum pixel value (255 for 8-bit images).

2. **Structural Similarity Index (SSIM)**: SSIM evaluates perceptual similarity by considering luminance, contrast, and structural information [20, 36]:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + K)(2\sigma_{xy} + K)}{(\mu_x^2 + \mu_y^2 + K)(\sigma_x^2 + \sigma_y^2 + K)} \quad (3.22)$$

where μ_x and μ_y represent mean intensities, σ_x^2 and σ_y^2 denote variances, σ_{xy} is the covariance, and K represents stabilising constants. SSIM values closer to 1 indicate higher structural similarity.

3.6.4 Domain-Specific Metrics for Medical Image Analysis

To calculate unreference metrics, regions of interest need to be considered (Figure 3.6 for selected regions).

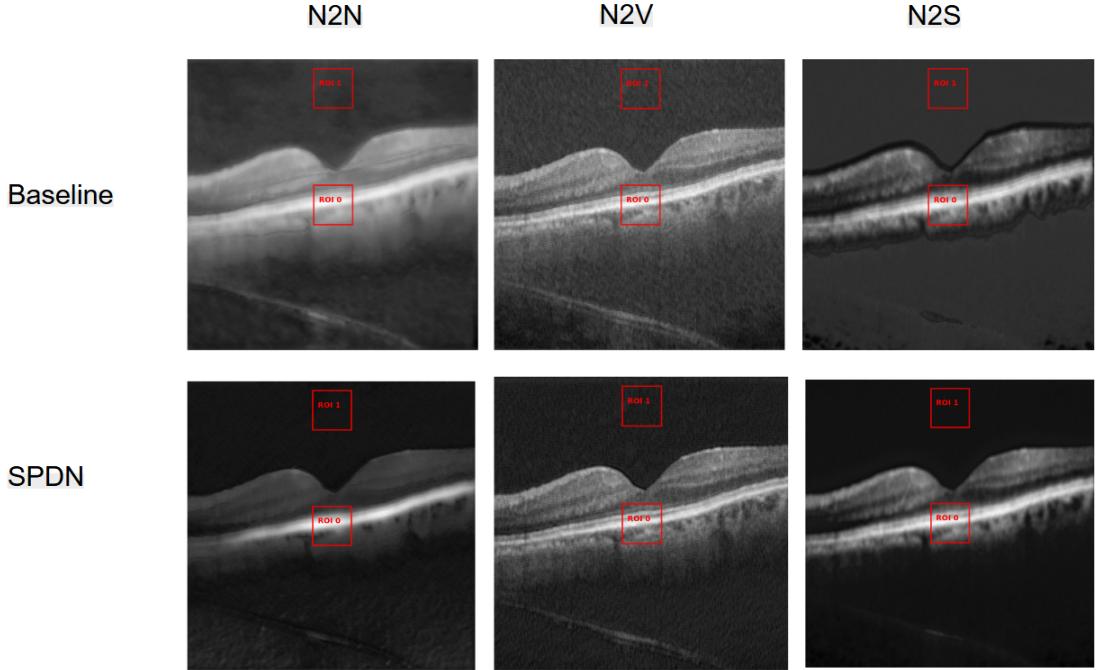


Figure 3.6 : Regions of Interest (ROI) which encompass areas which are definitively background with noise and foreground with noise. These regions are used for reference free measurements.

1. **Signal-to-Noise Ratio (SNR):** SNR quantifies the ratio between signal magnitude and noise level, providing insight into image clarity [44]:

$$SNR = \frac{\mu_{\text{signal}}}{\sigma_{\text{noise}}} \quad (3.23)$$

where μ_{signal} represents mean signal intensity and σ_{noise} denotes noise standard deviation.

2. **Contrast-to-Noise Ratio (CNR):** CNR measures the ability to distinguish between different tissue regions, which is crucial for clinical interpretation [45]:

$$CNR = \frac{|\mu_1 - \mu_2|}{\sigma_{\text{noise}}} \quad (3.24)$$

where μ_1 and μ_2 represent mean intensities of regions of interest and background, respectively, and σ_{noise} denotes background noise standard deviation.

3. **Equivalent Number of Looks (ENL):** ENL quantifies speckle suppression effectiveness, with higher values indicating better noise reduction:

$$ENL = \frac{\mu^2}{\sigma^2} \quad (3.25)$$

where μ represents mean intensity and σ^2 denotes variance of the selected homogeneous region.

4. **Edge Preservation Index (EPI):** EPI measures retention of edge information during denoising, critical for maintaining diagnostic features [46]:

$$EPI = \frac{\sum_{i,j} (E_o(i,j) - \bar{E}_o)(E_d(i,j) - \bar{E}_d)}{\sqrt{\sum_{i,j} (E_o(i,j) - \bar{E}_o)^2 \sum_{i,j} (E_d(i,j) - \bar{E}_d)^2}} \quad (3.26)$$

where E_o and E_d represent edge maps of original and denoised images respectively, computed using Sobel operators.

These domain-specific metrics provide clinically relevant assessment criteria, complementing traditional image quality measures to ensure both technical excellence and diagnostic utility.

3.7 Methodological Constraints and Challenges

This thesis acknowledges various methodological limitations:

- **Experimental Uncertainty:** The technique is largely novel and unexplored, with limited literature validating its effectiveness besides the established baseline methods. Comparative analyses with existing techniques will be conducted to assess whether it delivers measurable improvements in key metrics like PSNR, SSIM, SNR, CNR, ENL and EPI.
- **No True Clean Images:** The lack of truly clean reference images in OCT denoising adds uncertainty to the benchmarking process. This is a common challenge among all OCT denoising literature. This research addresses the challenge by exploring methods that work within noisy data constraints and aims to contribute solutions for cleaner ground-truth images.
- **Hardware Constraints:** The research is limited by the NVIDIA 2080 GPU, restricting the scale of experimentation and real-time application testing. Lack of clinical access to OCT imaging further limits the scope, but future work could benefit from better resources.
- **Conservatism in Medical Technology:** The medical field is slow to adopt new technologies, requiring extensive validation and testing before integration. AI-based techniques face significant barriers to acceptance in clinical environments.
- **Lack of Extensive Validation:** While the denoising methods show promise, their clinical applicability is not yet fully validated. More clinical trials and real-world testing are needed to confirm their effectiveness in medical environments.

- **Interpretability of Neural Networks:** Neural networks function as black boxes, making it difficult to interpret their decisions, which is a limitation in medical applications. Improving explainability is critical for gaining clinician trust and broader adoption.

Chapter 4

Experimental Results

This chapter presents an evaluation of the SPDN framework across three OCT datasets for statistical certainty and reproducibility. The framework demonstrates improvements in clinically relevant metrics while revealing important trade-offs.

Due to the data being sourced from three different datasets, each serves a different purpose in training and validation and so must be considered independently. Since the Duke 2011 dataset comes with layer segmentations but no reference image, the unreference metrics are strongly considered here, with SNR and CNR being calculable by taking the area within the segmentation for calculation. The OCTDL dataset comes with averaged versions of the noisy version, making this more appropriate for quantified evaluation compared to the reference metrics. Figure 4.1 shows a random sample from the OCTDL dataset, which is used for qualitative results.

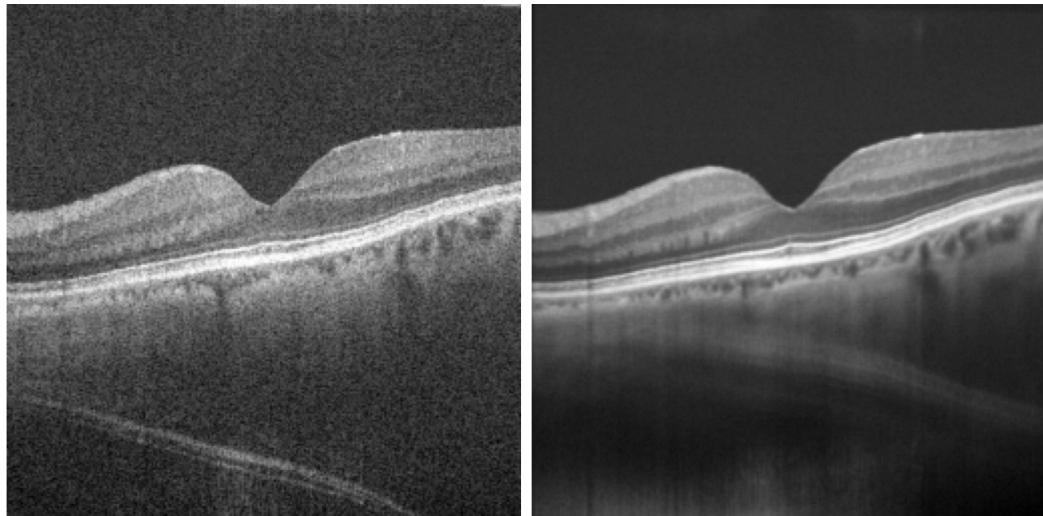


Figure 4.1 : OCTDL Sample with a raw sample (left) and an averaged image (reference) from the original data (right), which is given with the dataset download. This is used for reference metrics (PSNR and SSIM) as well as unreferenced metrics.

4.1 N2N vs N2N-SPDN

4.1.1 Qualitative Analysis

Qualitatively, N2N-SPDN is generally the most aggressive when it comes to denoising. The smoothing is overall very impressive, the information loss is generally still high. However, the inclusion of the SPDN module definitely retains lots of detail.

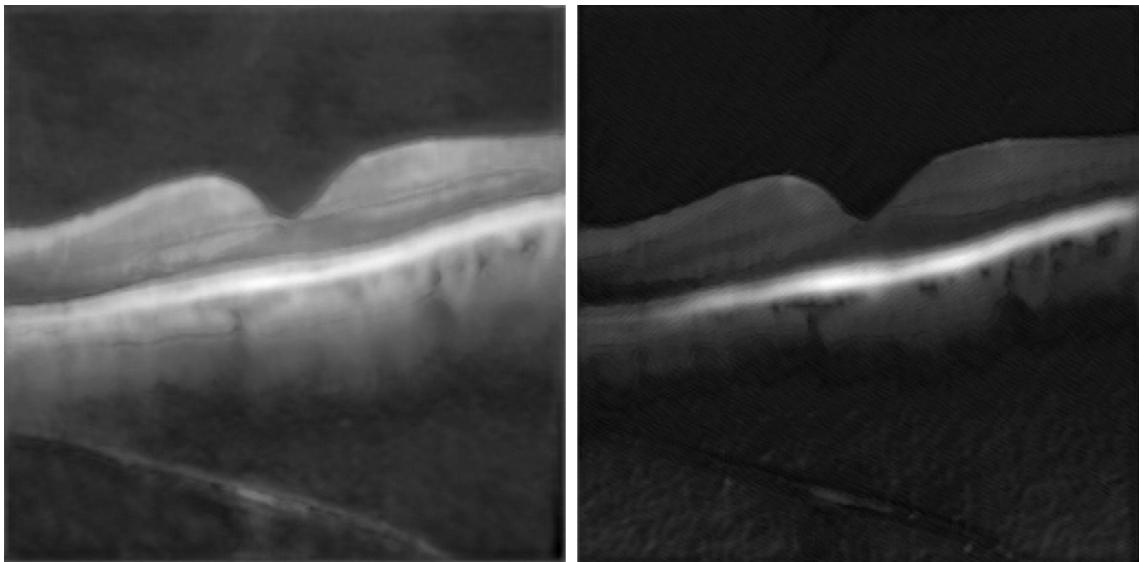


Figure 4.2 : N2N Qualitative Output: Baseline N2N output (left) and N2N-SPDN output (right).

4.1.2 Quantitative Analysis

The foundational approach for learning denoising mappings from paired noisy observations. N2N achieved moderate performance with stable training convergence and computational efficiency. Integration with N2N yielded measurable improvements across important clinical metrics as shown in Table C.1 and reveals significant improvements in important metrics. The proposed denoising method demonstrated effective noise reduction, with SNR improvements of **3.74 dB** (ICIP), 0.84 dB (OCTDL), and 0.35 dB (Duke), indicating strong noise reduction.

| N2N Metric | ICIP Dataset | | OCTDL Dataset | | Duke Dataset | |
|------------|----------------|------------------|----------------|----------------|---------------|---------------|
| | N2N | N2N-SPDN | N2N | N2N-SPDN | N2N | N2N-SPDN |
| PSNR | 10.2157 | 9.3821 | 22.3622 | 15.8721 | — | — |
| SSIM | 0.2980 | 0.2991 | 0.7085 | 0.5632 | — | — |
| SNR | 15.855821 | 19.592494 | 24.1928 | 25.0363 | 18.806 | 19.156 |
| CNR | 3.64 | 2.24 | 3.5865 | 0.2285 | -1.282 | -1.543 |
| ENL | 27.820 | 13.457 | 65.4700 | 33.4410 | 35.060 | 21.752 |
| EPI | 0.42303 | 0.347 | 0.5451 | 0.6011 | 0.335 | 0.348 |

Table 4.1 : Comparison of N2N and N2N-SPDN Performance Metrics Across ICIP, OCTDL, and Duke Datasets

These results demonstrate that N2N-SPDN successfully prioritises the preservation of diagnostically relevant information while achieving substantial noise reduction. This aligns with clinical requirements over the traditional imaging benchmarks.

4.2 N2V vs N2V-SPDN

4.2.1 Qualitative Analysis

Consistent with the the N2x-SPDN paradigm, the contrast difference between the background and the foreground are effectively separated with the background suppression be much more aggressive than the baseline N2V.

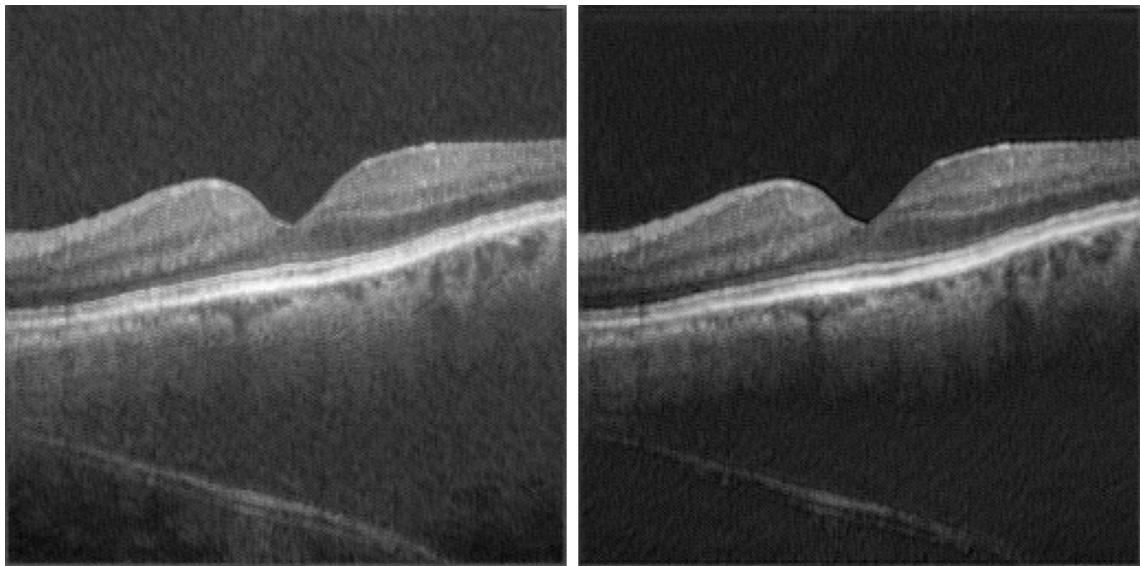


Figure 4.3 : N2V Qualitative Output: Baseline N2V output (left) and N2S-SPDN output (right).

4.2.2 Quantitative Analysis

| N2V Metric | ICIP Dataset | | OCTDL Dataset | | Duke Dataset | |
|------------|----------------|----------------|----------------|----------------|---------------|----------------|
| | N2V | N2V-SPDN | N2V | N2V-SPDN | N2V | N2V-SPDN |
| PSNR | 18.1479 | 21.8290 | 22.5105 | 21.3141 | – | – |
| SSIM | 0.7423 | 0.7744 | 0.4982 | 0.5322 | – | – |
| SNR | 19.4189 | 20.6375 | 20.6436 | 23.5721 | 18.8685 | 19.0885 |
| CNR | -4.0450 | -4.3216 | 3.6980 | 3.4738 | -1.2830 | -1.6910 |
| ENL | 12.6895 | 19.2198 | 37.7231 | 18.5303 | 28.5051 | 35.8304 |
| EPI | 0.8291 | 0.8259 | 0.7377 | 0.7820 | 0.4443 | 0.2011 |

Table 4.2 : Comparison of N2V and N2V-SPDN Performance Metrics Across ICIP, OCTDL, and Duke Datasets

N2V-SPDN improved the quantitative image metrics (21.8290 dB) in the ICIP dataset while staying competitive in OCTDL (21.3141 dB). N2V-SPDN improved SNR across all metrics as well. ENL was more inconsistent while CNR stayed competitive.

This model may be the most suitable for clinical deployment where image quality metrics are just as important as diagnostic utility.

4.3 N2S vs N2S-SPDN

4.3.1 Qualitative Analysis

N2S-SPDN achieved very high qualitative scores. The N2S converged very quickly compared to baseline N2S. Not only the N2S was over-smoothing across the board, albeit, maintaining important information compared to the other N2-family schemas, but the N2S-SPDN really focused on reducing background noise and getting through the noise on the important features.

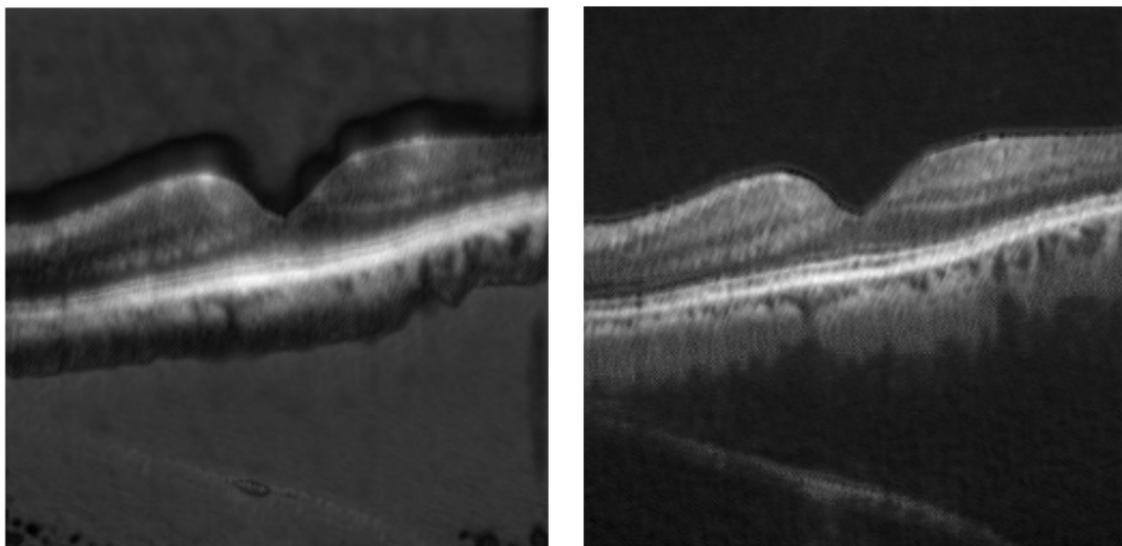


Figure 4.4 : N2S Qualitative Output: Baseline N2S output (left) and N2S-SPDN output (right).

4.3.2 Quantitative Analysis

| N2S Metric | ICIP Dataset | | OCTDL Dataset | | Duke Dataset | |
|------------|----------------|----------------|------------------|----------------|--------------|---------------|
| | N2S | N2S-SPDN | N2S | N2S-SPDN | N2S | N2S-SPDN |
| PSNR | 12.7865 | 11.1153 | 18.8636 | 17.2764 | — | — |
| SSIM | 0.2174 | 0.2355 | 0.7086 | 0.6363 | — | — |
| SNR | 17.0759 | 25.7331 | 20.2655 | 29.4589 | 17.924 | 24.564 |
| CNR | -4.6540 | -1.1315 | 1.8895 | 3.0894 | -5.753 | -2.693 |
| ENL | 18.7836 | 13.9615 | 1424.1078 | 312.6385 | 8.576 | 4.460 |
| EPI | 0.4023 | 0.5097 | 0.5584 | 0.6712 | 0.301 | 0.395 |

Table 4.3 : Comparison of N2S and N2S-SPDN Performance Metrics Across ICIP, OCTDL, and Duke Datasets

The most obvious quantitative advantage to this method is the outstanding SNR improvements:

1. ICIP: 8.66 dB improvement
2. OCTDL: 9.19 dB improvement
3. Duke: 6.64 dB improvement

These were much higher than the rest of the N2x-SPDN-Family; potentially indicating that this is the best model in clinical usage. The ablations study focus on N2S-SPDN for this reason.

4.3.3 Quantitative Summary

The SPDN framework demonstrates consistent effectiveness across the datasets.

1. **Best Qualitative:** N2S-SPDN with defined edges and clear feature retention within the structures.
2. **Best Performance:** N2S-SPDN with SNR improvements of up to 9.19.
3. **Most Balanced:** N2V-SPDN with improved image-metrics scores (PSNR and SSIM) when evaluated on the ICIP dataset, and close PSNR scores on the OCTDL while having higher SSIM and having higher SNR scores overall.
4. **Highest Potential Clinical Utility:** All methods show enhanced edge preservation and structural to background separation.

4.4 Ablation Studies

Ablation studies have been separated into the segmenting SPDН network and the denoising N2x-SPDN networks.

4.4.1 SPDН Ablations

The SPDН training selections were carefully selected after extensive testing and experimentation. This includes the architecture depth and its components.

Analysis of the Architecture Depth: Any smaller and the results were unquantifiable, as it simply did not capture enough information (Appendix D.1). Any higher and the computational overhead increases substantially, reducing or even fully nullifying the potential usefulness of this methodology. We can see by the output in Figure 4.5 that the output is fine, but the sharp edges and defined intensities and features are missing because the model is not capable of picking them up with such shallow depth.

A larger model was also used (Appendix D.2), however, the ICIP images are natively 300×300 in resolution, meaning that it is ineffective to use larger U-Nets as they are designed in powers of 2 (32, 64, 128, 256). Having a 5-layer architecture would result

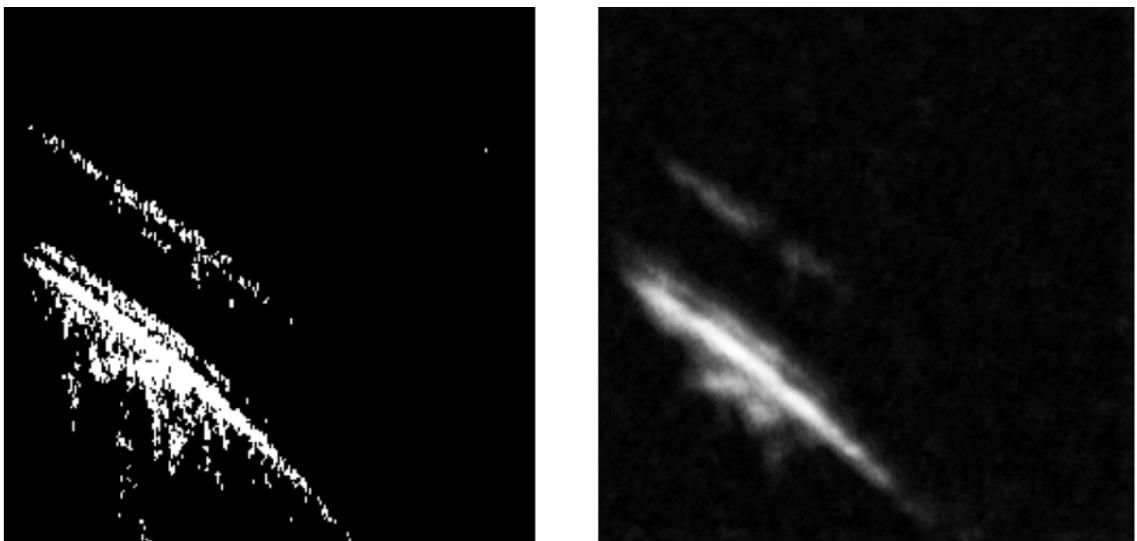


Figure 4.5 : Small SPDN output with binary calculation from decorrelation (left) and output from 3-layered U-Net (right).

in channel dimensions of 32, 64, 128, 256, and then 256 again, which provides little benefit as the final two layers have identical channel capacity (see Appendix D.2 for model). The outputs are identical to the 4-layer U-Net, leading to no additional benefit for the increased computational overhead.

Effectiveness of the Attention Mechanisms: Removing attention had a detrimental qualitative effect on the segmentation. The output was nowhere near as fine. As seen in the figure in Appendix D.3, the output of the segmentation simply does not work without the attention.



Figure 4.6 : SPDN Output with Dice BCE Loss in SPDN: Mid-training output using Dice BCE. (From left to right) 1. Raw input; 2. SPDN output from raw input; 3. SPDN output from denoised output; 4. denoised output from N2S-SPDN.

Impact of the Loss Functions: When applying it in the N2x-SPDN schema (4.4.1), the results are actually quite good, however, the SNR is reduced- and from the extensive testing and the literature, this was deemed the most important quantitative measure to preserve. It is apparent in the qualitative findings (Figures 4.7, 4.6) the SPDN module is quite aggressive with the segmentation task and loses much of the important information and losing the intensity scaling which is so important in this task.

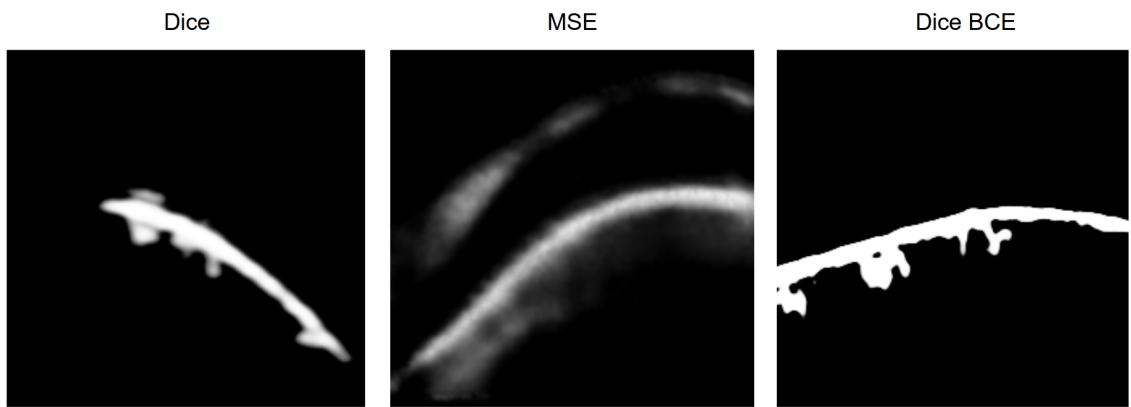


Figure 4.7 : SPDN Output with Different Losses: Dice (left), MSE (middle) and Dice BCE (right). The Dice losses lose much of the intensities and threshold too aggressively and don't capture the nuance of the structural detail in an OCT image.

| N2S-SPDN | MSE | Dice BCE |
|----------|----------------|---------------|
| PSNR | 11.1153 | 18.8226 |
| SSIM | 0.2355 | 0.6854 |
| SNR | 25.7331 | 23.6095 |
| CNR | -1.1315 | 3.4562 |
| ENL | 13.9615 | 240.8354 |
| EPI | 0.5097 | 0.4838 |

Table 4.4 : Comparison of N2S and N2S-SPDN Performance Metrics on ICIP Dataset

4.4.2 N2x-SPDN Ablations

For denoising ablation studies, as the N2S was the most effective model, the ablations focused on this.

Architecture Modifications: Three variant models were developed for the denoising component. A small 3-layer equivalent with the same attention CBAM attention mechanisms (Appendix D.1); a larger model with the same attention mechanisms (Appendix D.2) and an equivalent depth model without any attention (Appendix D.3)

When training on a simple model without attention, the model was simply not capable of capturing enough detail on smaller datasets; it simply did not capture nearly enough detail to even be quantifiable and relevant. What is required is sufficient depth and attention mechanisms for limited data; however, if data is properly supplied, the simple model is also sufficient, **indicating that it is less important on what model is used but the schema in what is being used and the quantity and quality of the dataset**. The third model which was attempted to

train on was a 3-layer model with 64 input-output channels; however, this was far too computationally expensive. [7] used a similar model which was not possible to replicate due to hardware constraints, indicating a need for further research in this area with less computational constraints.

Alpha Weighting or Adaptive Weighting: Two weighting approaches are implemented to balance the dual objectives:

Fixed weighting: α remains constant throughout training, providing consistent balance between denoising and structural preservation.

Adaptive weighting: α adjusts dynamically based on loss magnitudes to address convergence rate disparities between loss components:

$$\alpha_{adaptive} = \frac{L_{N2V}}{L_{flow} + \epsilon} \quad (4.1)$$

where $\epsilon = 1 \times 10^{-8}$ prevents division by zero. This adaptive approach automatically balances the dual objectives when loss scales differ significantly, ensuring balanced optimisation between denoising performance and structural preservation throughout training.

| N2S-SPDN Alpha Ablation | | | | |
|-------------------------|----------------|-----------------|----------------|--|
| OCTDL Metric / Alpha | Adaptive | 1.0 | 5.0 | |
| PSNR | 17.2764 | 16.8918 | 14.7103 | |
| SSIM | 0.6363 | 0.6290 | 0.4415 | |
| SNR | 29.458 | 27.9221 | 30.0604 | |
| CNR | 3.0894 | 2.7281 | 3.0233 | |
| ENL | 312.6385 | 491.3309 | 107.6646 | |
| EPI | 0.6712 | 0.6174 | 0.5816 | |

Table 4.5 : Comparison of N2S and N2S-SPDN Performance Metrics on OCTDL Dataset

Adaptive held the overall best results, hitting a good middle ground.

4.4.3 Duke Dataset Preprocessing Ablations Evaluation

The Duke dataset cannot be fairly evaluated with image quality scores such as PSNR and SSIM which require a reference image. An averaged score can be computed, however, for the sake of near-zero bias, this thesis has elected to use only unreference/reduced-reference image metric scores for this dataset (SNR, CNR, ENL, EPI).

SPDN was evaluated against baselines approaches using the Duke 2011 dataset with pre-segmented retinal layers. Results were obtained under two preprocessing conditions: images resized to 256×256 pixels (4.6) and unresized images with normalisation. This is because the images from this dataset natively come in 512x1000-sized slices, whereas the ICIP training data is made up natively of 300x300-sized slices.

| Metric | Resized to 256×256 | | Unresized but Normalised | |
|--------|-----------------------------|---------------|--------------------------|---------------|
| | N2S | N2S-SPDN | N2S | N2S-SPDN |
| SNR | 17.924 | 24.564 | 18.188 | 21.644 |
| CNR | -5.753 | -2.693 | -1.721 | -0.751 |
| ENL | 8.576 | 4.460 | 8.978 | 6.176 |
| EPI | 0.301 | 0.395 | 0.207 | 0.230 |

Table 4.6 : N2S vs N2S-SPDN Performance on Duke Dataset Under Different Pre-processing Conditions

The results in Table 4.6 highlight the importance of preprocessing steps, with unresized approaches better preserving structural information at the expense of computational efficiency, whilst resized approaches enable direct application of pre-trained

models but may compromise important anatomical details essential for clinical interpretation. However, in this test-case, N2S-SPDN outperformed the baseline variant in all clinically relevant metrics.

4.5 Discussion

The results demonstrate that SPDН integration produces complex, method-dependent effects on unsupervised OCT denoising approaches, with significant trade-offs between different image quality metrics. While SNR improvements are consistently observed, these gains often come at the cost of degraded performance in traditional image quality measures.

4.5.1 Performance Analysis Across Datasets

The results were mostly consistent across the datasets when prepared in the same way. As mentioned prior, each dataset has a different noise distribution which was likely due to the machine or process in how the images were taken; the models and methodologies being trained on ICIP data was still as applicable on the other datasets, however, less so.

ICIP Dataset Results: A tiny subset of the data (5-patients, 20 images from each) were used randomly as a validation set from the exact data so evaluate how it does on the exact data it was trained on. The results were definitely more consistent; however, this is still a statistically invalid set due to the uncleanness on the processing details and private nature. This shows that with the SPDН schema, if the data is available, this is a highly appropriate method if PSNR and SSIM are not required to be extremely high.

OCTDL Dataset Results: The OCTDL dataset results highlight the fundamental trade-offs inherent in the SPDН approach. While all methods achieved positive SNR improvements, these came at cost to structural fidelity metrics. N2N-SPDN showed

degradation in both PSNR and SSIM, indicating of image quality despite noise reduction, however, the qualitative output does not reflect the significance of this.

Duke Dataset Results: The consistently negative CNR values across most methods indicate challenges in maintaining contrast between retinal layers, with N2S methods showing particularly poor contrast preservation when images are resized; once again this emphasises the need for consistent processing methods for the best results.

Distribution Analysis: Distribution analysis B.1 revealed that this dataset followed a Weibull noise distribution while the B.2 B.3 followed log-normal patterns.

The N2x-SPDN still performed admirably across all datasets, demonstrating its versatility and adaptability to similar but not exact noise patterns.

4.5.2 SPDN Integration Insights

The results demonstrate that SPDN integration successfully achieves its primary objectives of enhanced SNR performance and superior feature retention across all tested architectures, with consistent improvements in clinically relevant structural preservation. While traditional image quality metrics such as PSNR and SSIM showed degradation in some cases, these pixel-level similarity measures have limited clinical relevance compared to the achieved improvements in noise reduction and anatomical structure clarity that directly impact diagnostic utility. With SNR improvements up to 9.19 dB representing clinically meaningful enhancements. The improved SNR qualitatively translates into better issue contrast and clearer boundaries between structure and background. This improvement may lead to earlier identification of malignant or harmful ailments such as age-based macular disease [18].

The consistent SNR improvements across methods validate the core hypothesis that

structural information can guide noise reduction processes. However, the accompanying degradation in other metrics highlights the need for more sophisticated integration strategies. The consistent SNR improvements are up across all datasets and translate to measurable clinical benefits.

Chapter 5

Conclusion

This research investigated the integration of structural-preservation constraints into unsupervised OCT denoising through the SPDN framework. The study evaluated SPDN integration across N2N, N2V, and N2S architectures, revealing complex trade-offs between different image quality metrics.

5.1 Key Findings and Contributions

The SPDN framework delivers substantial improvements in clinically relevant metrics, with SNR enhancements up to 9.19 dB while maintaining diagnostic structural information. The modular design successfully integrates with established unsupervised methods (N2N, N2V, N2S), providing immediate applicability to existing clinical workflows.

This work provides the first systematic analysis of decorrelation-based structural constraints in unsupervised medical image denoising, revealing fundamental trade-offs between clinical utility and traditional image quality metrics. The quantitative framework for structure-preservation assessment establishes new evaluation paradigms for medical imaging applications. The investigation establishes a foundation for future research into more sophisticated integration strategies that could potentially achieve a better balance between noise reduction and image quality preservation.

The substantial noise reduction directly addresses clinical requirements for enhanced diagnostic accuracy. The performance of N2V-SPDN and the optimal clinical utility

of N2S-SPDN may provide deployment options for optical diagnostic scenarios.

5.2 Limitations and Future Work

The current approach demonstrates significant limitations that must be addressed before practical deployment:

- The trade-off between SNR improvement and traditional image quality metrics requires better optimisation strategies
- The modular integration approach may be too simplistic for effective constraint incorporation
- Dataset-specific performance variations indicate the need for more robust, generalisable architectures
- The dramatic changes in training dynamics suggest fundamental compatibility issues between structural constraints and existing frameworks
- The datasets have clearly been processed without much extrapolation on how. Future research may wish to model device-specific noise characteristics to accommodate how various different devices take the images.

Future research should focus on developing more sophisticated integration methodologies that can achieve SNR improvements without compromising overall image quality, potentially through multi-objective optimisation or adaptive constraint weighting strategies. Potential directions include involving ophthalmologists for clinical validation; extension into other medical imaging modalities and seeing the differences with improved hardware.

5.3 Final Remarks

This research demonstrates that unsupervised medical image denoising can be significantly enhanced through domain-specific structural constraints. The SPDN framework’s success in preserving diagnostically important information while reducing noise represents a meaningful step toward clinically viable automated image enhancement. The substantial SNR improvements achieved (up to 9.19 dB) suggest that structure-preserving denoising can deliver measurable clinical value. If the clinician values that more so, then this modular SPDN design should be considered.

Bibliography

- [1] Radiopaedia, “Cross-sectional imaging,” Radiopaedia, accessed: 2025. [Online]. Available: <https://radiopaedia.org/articles/cross-sectional-imaging-1>
- [2] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and J. G. Fujimoto, “Optical coherence tomography,” *Science*, 254(5035), 1178-1181, 1991.
- [3] A. Guo, L. Fang, M. Qi, and S. Li, “Unsupervised denoising of optical coherence tomography images with nonlocal-generative adversarial network,” *IEEE*, 2020. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9174804>
- [4] Y. Ma, X. Chen, W. Zhu, X. Cheng, D. Xiang, and F. Shi, “Speckle noise reduction in optical coherence tomography images based on edge-sensitive cgan,” *Biomedical Optics Express*, vol. 9, no. 11, pp. 5129–5146, 2018.
- [5] K. Xie, M. Luo, H. Chen, M. Yang, Y. He, P. Liao, and Y. Zhang, “Speckle denoising of optical coherence tomography image using residual encoder-decoder cyclegan,” *SIViP* 17, 2023. [Online]. Available: <https://doi.org/10.1007/s11760-022-02361-6>
- [6] B. Qiu, Y. You, Z. Huang, X. Meng, Z. Jiang, C. Zhou, G. Liu, K. Yang, Q. Ren, and Y. Lu, “N2nsr-oct: Simultaneous denoising and super-resolution in optical coherence tomography images using semisupervised deep learning,” *Journal of Biophotonics*, 2020. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jbio.202000282>

- [7] J. Schottenhamml, T. Würfl, S. B. Ploner, L. Husvogt, B. Hohberger, J. G. Fujimoto, and A. Maier, “SSN2V: unsupervised OCT denoising using speckle split,” *Scientific Reports*, vol. 13, no. 1, p. 10382, Jun 2023. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/37369731/>
- [8] Y. Jia, O. Tan, J. Tokayer, B. Potsaid, Y. Wang, J. J. Liu, M. F. Kraus, H. Subhash, J. G. Fujimoto, J. Horngger, and D. Huang, “Split-spectrum amplitude-decorrelation angiography with optical coherence tomography,” *Opt. Express*, vol. 20, no. 4, pp. 4710–4725, Feb 2012. [Online]. Available: <https://opg.optica.org/oe/abstract.cfm?URI=oe-20-4-4710>
- [9] S. S. Gao, Y. Jia, M. Zhang, J. Su, G. Liu, T. S. Hwang, S. T. Bailey, and D. Huang, “Optical Coherence Tomography Angiography,” *Investigative Ophthalmology & Visual Science*, vol. 57, no. 9, pp. OCT27–OCT36, Jul 2016. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/27409483/>
- [10] W. Choi, E. M. Moult, N. K. Waheed, M. Adhi, B. Lee, C. D. Lu, T. E. de Carlo, V. Jayaraman, P. J. Rosenfeld, J. S. Duker, and J. G. Fujimoto, “Ultrahigh-speed, swept-source optical coherence tomography angiography in nonexudative age-related macular degeneration with geographic atrophy,” *Ophthalmology*, vol. 122, no. 12, pp. 2532–2544, 2015. [Online]. Available: <https://doi.org/10.1016/j.ophtha.2015.08.029>
- [11] V. B. Silva, D. Andrade De Jesus, S. Klein, T. van Walsum, J. Cardoso, L. S. Brea, and P. G. Vaz, “Signal-carrying speckle in optical coherence tomography: a methodological review on biomedical applications,” *Journal of Biomedical Optics*, vol. 27, no. 3, p. 030901, 2022.
- [12] J. M. Schmitt, S. Xiang, and K. M. Yung, “Speckle in optical coherence tomography,” *Journal of biomedical optics*, vol. 4, no. 1, pp. 95–105, 1999.

- [13] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, “Noise2Noise: Learning image restoration without clean data,” in Proceedings of the 35th International Conference on Machine Learning, ser. Proceedings of Machine Learning Research, J. Dy and A. Krause, Eds., vol. 80. PMLR, 10–15 Jul 2018, pp. 2965–2974. [Online]. Available: <https://proceedings.mlr.press/v80/lehtinen18a.html>
- [14] A. Krull, T.-O. Buchholz, and F. Jug, “Noise2void - learning denoising from single noisy images,” in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 2124–2132.
- [15] J. Batson and L. Royer, “Noise2self: Blind denoising by self-supervision,” 2019.
- [16] E. Enaholo, M. Mutali, and M. Zeppieri, Optical Coherence Tomography. ResearchGate, Oct. 2024. [Online]. Available: https://www.researchgate.net/publication/384074247_Optical_Coherence_Tomography
- [17] S. Aumann, S. Donner, J. Fischer, and F. Müller, “Optical coherence tomography (oct): Principle and technical realization,” in High Resolution Imaging in Microscopy and Ophthalmology: New Frontiers in Biomedical Optics, J. F. Bille, Ed. Cham: Springer, 2019, ch. 3. [Online]. Available: <https://www.ncbi.nlm.nih.gov/books/NBK554044/>
- [18] M. Elsharkawy, M. Elrazzaz, M. Ghazal, M. Alhalabi, A. Soliman, A. Mahmoud, E. El-Daydamony, A. Atwan, A. Thanos, H. Sandhu, G. Giridharan, and A. El-Baz, “Role of optical coherence tomography imaging in predicting progression of age-related macular disease: A survey,” Diagnostics, vol. 11, no. 12, p. 2313, Dec 2021.
- [19] D. Sethi, S. Bharti, and C. Prakash, “A comprehensive survey on gait analysis: History, parameters, approaches, pose estimation, and

- future work,” *Artificial Intelligence in Medicine*, 2021. [Online]. Available: <https://doi.org/10.1016/j.artmed.2022.102314>
- [20] D. Otero, D. L. Torre, O. Michailovich, and E. R. Vrscay§, “Optimization of structural similarity in mathematical imaging,” 2020. [Online]. Available: <https://doi.org/10.48550/arXiv.2002.02657>
- [21] Z. Liu, L. Yu, and H. Sun, “Image denoising via nonlocal low rank approximation with local structure preserving,” *IEEE*, 2019. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8598927>
- [22] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3d transform-domain collaborative filtering,” https://webpages.tuni.fi/foi/GCF-BM3D/BM3D_TIP_2007.pdf, 2007, accessed June 2025.
- [23] Z. Zhang, G. Dai, X. Liang, S. Yu, L. Li, and Y. Xie, “Can signal-to-noise ratio perform as a baseline indicator for medical image quality assessment,” *IEEE Access*, vol. 6, pp. 11 534–11 543, 2018.
- [24] X. Wu, W. Gao, and H. Bian, “Self-denoising method for oct images with single spectrogram-based deep learning,” *Optics Letters*, 2023. [Online]. Available: <https://opg.optica.org/ol/abstract.cfm?uri=ol-48-19-4945>
- [25] Y. Huang, N. Zhang, , and Q. Hao, “Real-time noise reduction based on ground truth free deep learning for optical coherence tomography,” *Biomedical Optics Express*, 2021. [Online]. Available: <https://opg.optica.org/boe/fulltext.cfm?uri=boe-12-4-2027id=449154>
- [26] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” 2015. [Online]. Available: <https://doi.org/10.48550/arXiv.1505.04597>

- [27] T. Huang, S. Li, X. Jia, H. Lu, and J. Liu, “Neighbor2neighbor: Self-supervised denoising from single noisy images,” 2021. [Online]. Available: <https://arxiv.org/abs/2101.02824>
- [28] S. Fadnavis, J. Batson, and E. Garyfallidis, “Patch2self: Denoising diffusion mri with self-supervised learning,” 2020. [Online]. Available: <https://arxiv.org/abs/2011.01355>
- [29] G. Sun, Y. Pan, W. Kong, Z. Xu, J. Ma, T. Racharak, L.-M. Nguyen, and J. Xin, “Da-transunet: integrating spatial and channel dual attention with transformer u-net for medical image segmentation,” *Frontiers in Bioengineering and Biotechnology*, vol. 12, p. 1398237, 2024.
- [30] O. Oktay, J. Schlemper, L. L. Folgoc, M. C. H. Lee, M. P. Heinrich, K. Misawa, K. Mori, S. G. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, “Attention u-net: Learning where to look for the pancreas,” *CoRR*, vol. abs/1804.03999, 2018. [Online]. Available: <http://arxiv.org/abs/1804.03999>
- [31] S. Woo, J. Park, J. Lee, and I. S. Kweon, “CBAM: convolutional block attention module,” *CoRR*, vol. abs/1807.06521, 2018. [Online]. Available: <http://arxiv.org/abs/1807.06521>
- [32] VIP Cup 2024. [Online]. Available: <https://signalprocessingsociety.org/community-involvement/video-image-processing-cup>
- [33] M. Kulyabin, A. Zhdanov, A. Nikiforova, A. Stepichev, A. Kuznetsova, M. Ronkin, V. Borisov, A. Bogachev, S. Korotkikh, P. A. Constable, and A. Maier, “Octdl: Optical coherence tomography dataset for image-based deep learning methods,” *Scientific Data*, vol. 11, p. 365, 2024.

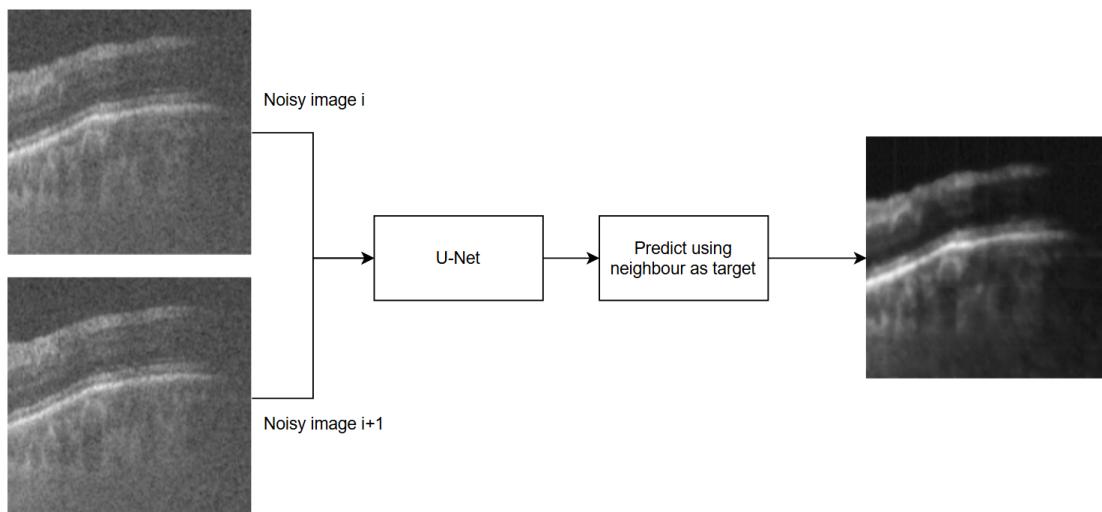
- [34] S. J. Chiu, J. A. Izatt, R. V. O'Connell, K. P. Winter, C. A. Toth, and S. Farsiu, "Validated automatic segmentation of amd pathology including drusen and geographic atrophy in sd-oct images," *Investigative Ophthalmology Visual Science*, vol. 53, no. 1, pp. 53–61, 01 2012. [Online]. Available: <https://doi.org/10.1167/iovs.11-7640>
- [35] H. Li, K. Liu, T. Cao, L. Yao, Z. Zhang, X. Deng, C. Du, and P. Li, "High performance octa enabled by combining features of shape, intensity, and complex decorrelation," *Optics Letters*, vol. 46, no. 2, pp. 368–371, 2021.
- [36] U. Sara, M. Akter, and M. S. Uddin, "Image quality assessment through fsim, ssim, mse and psnr—a comparative study," *Journal of Computer and Communications*, 2019. [Online]. Available: <https://doi.org/10.4236/jcc.2019.73002>
- [37] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," *Advances in Neural Information Processing Systems*, vol. 32, pp. 8024–8035, 2019.
- [38] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith *et al.*, "Array programming with numpy," *Nature*, vol. 585, no. 7825, pp. 357–362, 2020.
- [39] "The opencv library," 2000. [Online]. Available: <https://opencv.org/>
- [40] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

- [41] M. L. Waskom, “seaborn: statistical data visualization,” *Journal of Open Source Software*, vol. 6, no. 60, p. 3021, 2021. [Online]. Available: <https://doi.org/10.21105/joss.03021>
- [42] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, “Scikit-learn: Machine learning in python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [43] S. Marcel and Y. Rodriguez, “Torchvision: Pytorch’s computer vision library,” <https://github.com/pytorch/vision>, 2016.
- [44] S. Huang, R. Wang, R. Wu, J. Zhong, X. Ge, Y. Liu, and G. Ni, “Snr-net oct: brighten and denoise low-light optical coherence tomography images via deep learning,” *Optics Express*, 2023. [Online]. Available: <https://opg.optica.org/oe/fulltext.cfm?uri=oe-31-13-20696id=531252>
- [45] I. Young, “Mri theory,” *Encyclopedia of Spectroscopy and Spectrometry* (Third Edition), 2017. [Online]. Available: <https://www.sciencedirect.com/topics/immunology-and-microbiology/contrast-to-noise-ratio>
- [46] L. Chen, F. Jiang, H. Zhang, S. Wu, S. Yu, and Y. Xie, “Edge preservation ratio for image sharpness assessment,” in *2016 12th World Congress on Intelligent Control and Automation (WCICA)*, 2016, pp. 1377–1381.

Appendix

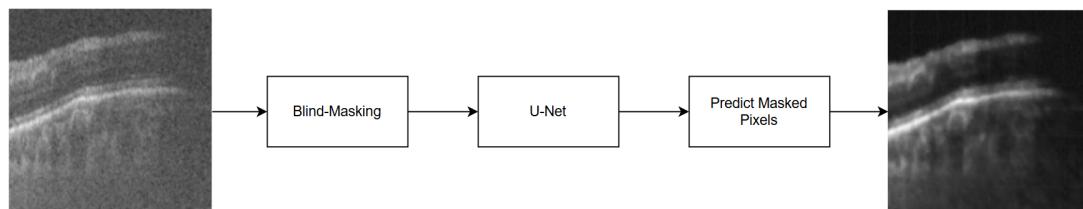
A N2-Denoising Schemas

A.1 Noise2Noise Framework



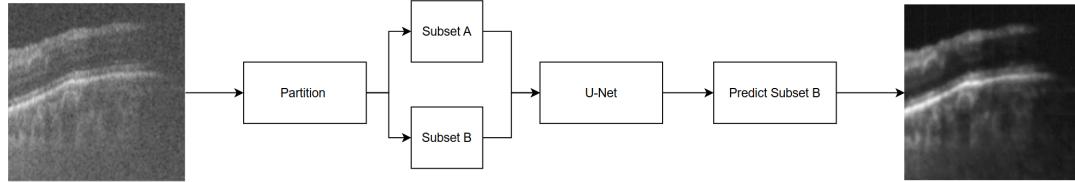
A brief look at N2N which takes an image and compares itself to a neighbouring slice to approximate what is underneath the salt-and-pepper speckling.

A.2 Noise2Void Framework



A brief look at N2N which masks itself and attempts to predict what is underneath.

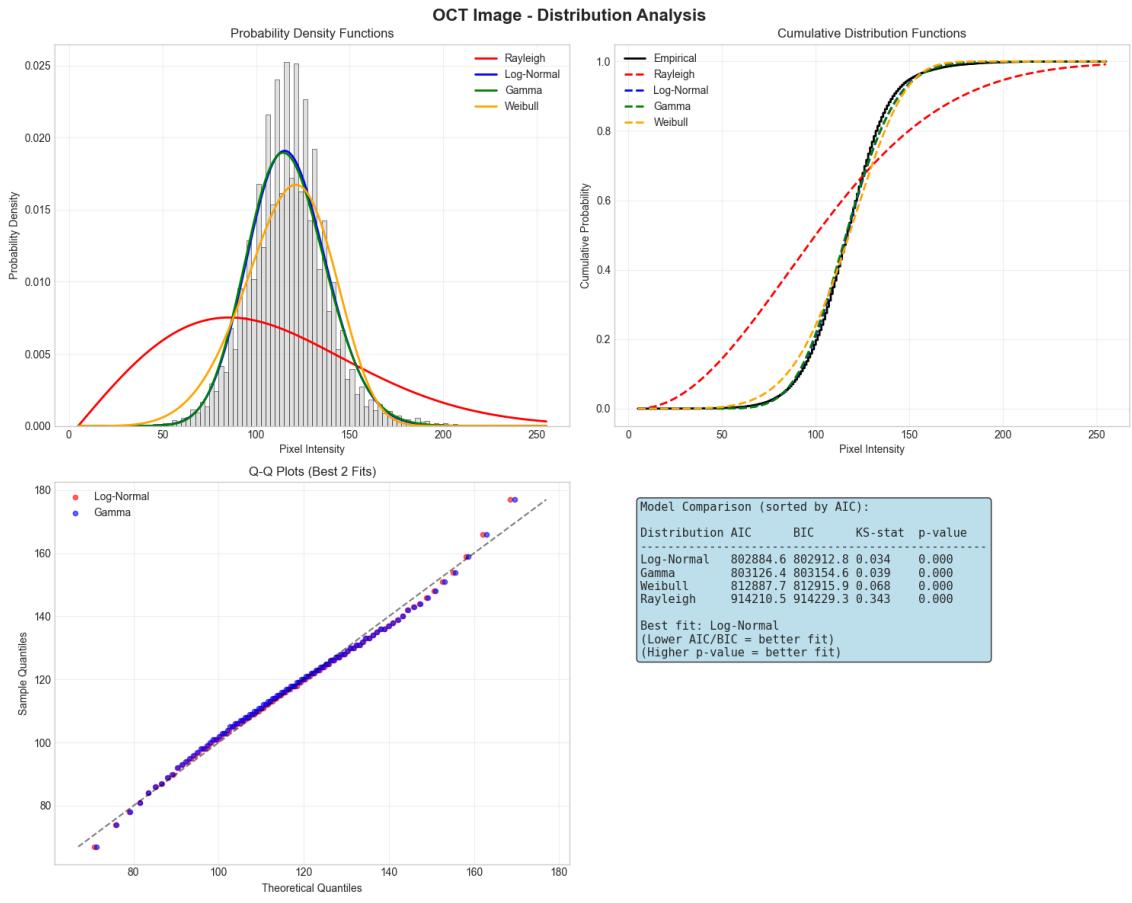
A.3 Noise2Self Framework



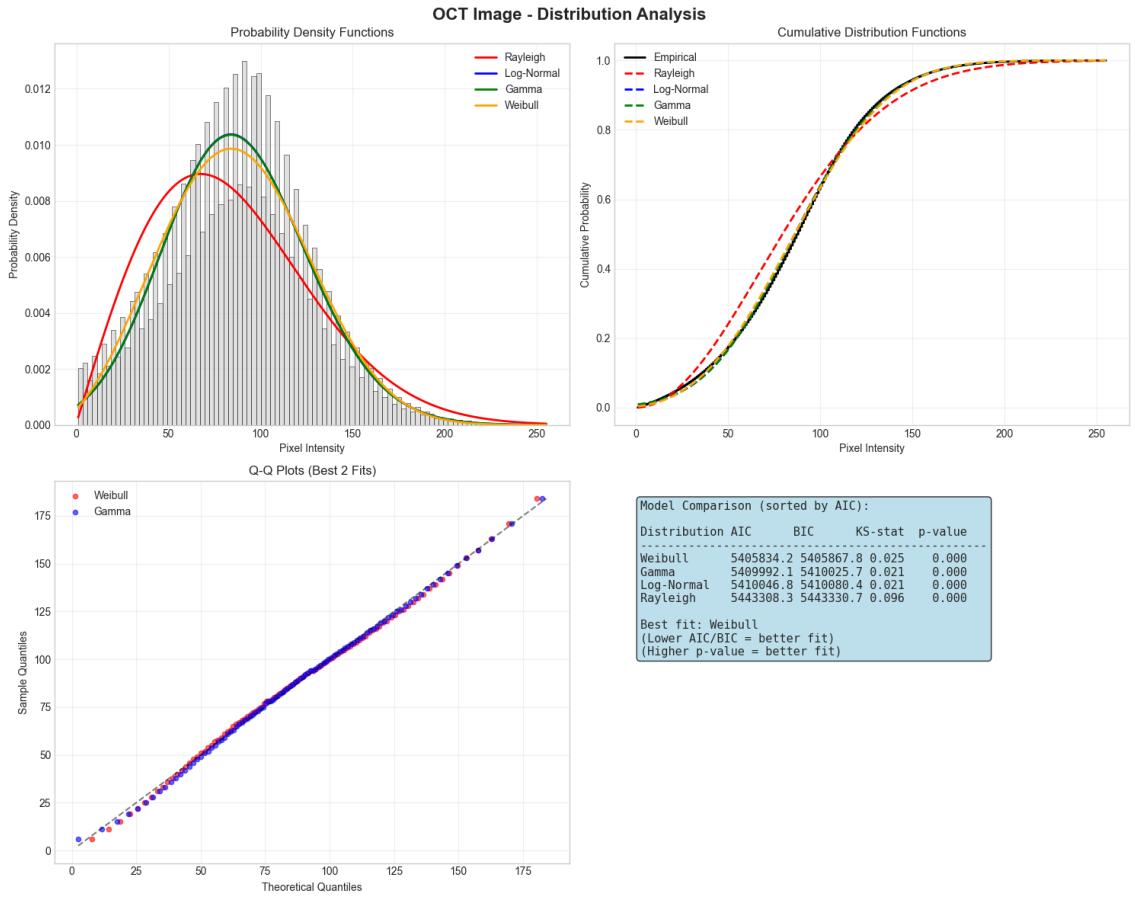
A brief look at N2S which takes an image and partitions itself into subpatches of the same image, which then tries to predict one-another.

B Dataset Statistics

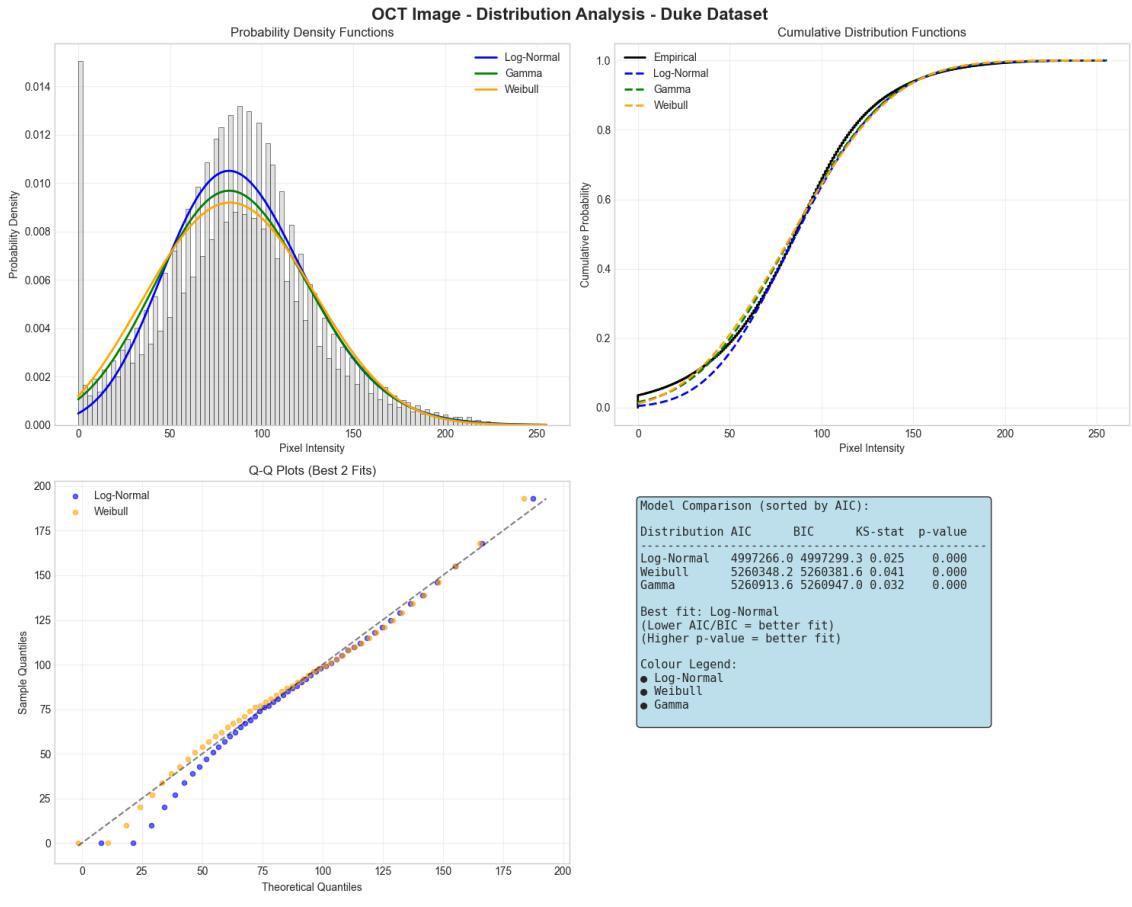
B.1 ICIP Dataset Sample-Image Statistics



B.2 OCTDL Dataset Sample-Image Statistics



B.3 Duke Dataset Sample-Image Statistics



[H]

C Training Settings

C.1 Parameters

| SPDN Parameter | Minimum | Maximum | Optimal |
|----------------------|---------|---------|------------|
| Alpha | 1.0 | 1.0 | 1.0 |
| N-Neighbours | 1.0 | 10.0 | 4.0 |
| Threshold Percentile | 0.9 | 0.99 | 99 |
| Post-process size | 1.0 | 5.0 | 2.0 |
| Batch-Size | 4 | 32 | 16 |
| Patch-Size | 48 | 128 | 96 |
| Stride | 32 | 64 | 48 |

The list of parameters which were used throughout training. The alpha parameter was an adaptable coefficient which adjusted itself based on the current loss-magnitude; if used statically, 1.0 was found to be viable.

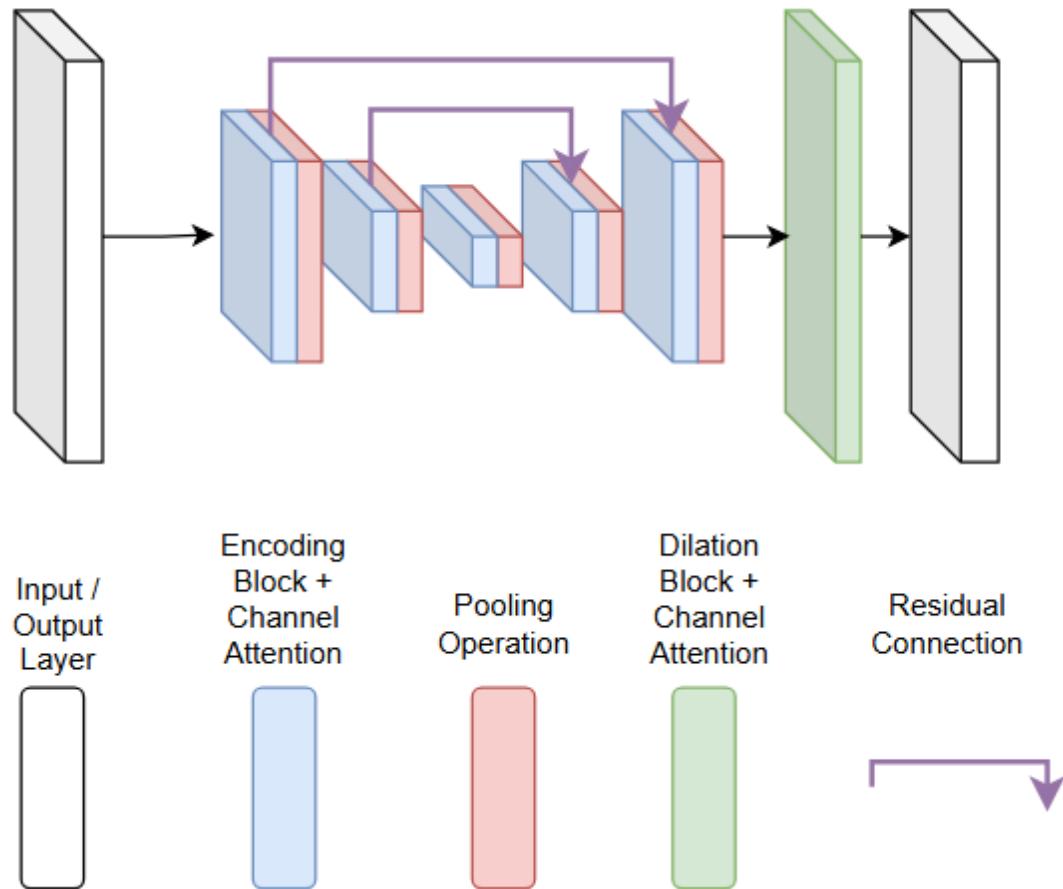
C.2 Evaluation Metrics

| Metric | Required Reference Image |
|--------|--------------------------|
| PSNR | True |
| SSIM | True |
| SNR | False |
| CNR | False |
| ENL | False |
| EPI | False |

This table is a reference to show which metrics required references and those which did not. The ones without a required reference image are clinically more relevant as they only compare within themselves and not an averaged equivalent.

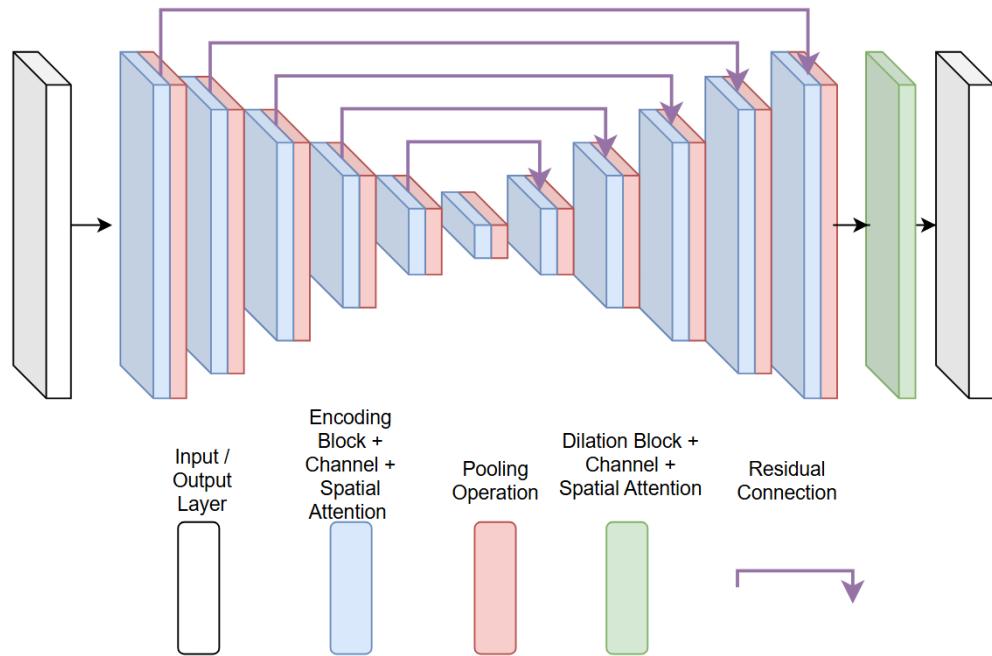
D Ablation Studies

D.1 Small SPDN U-Net



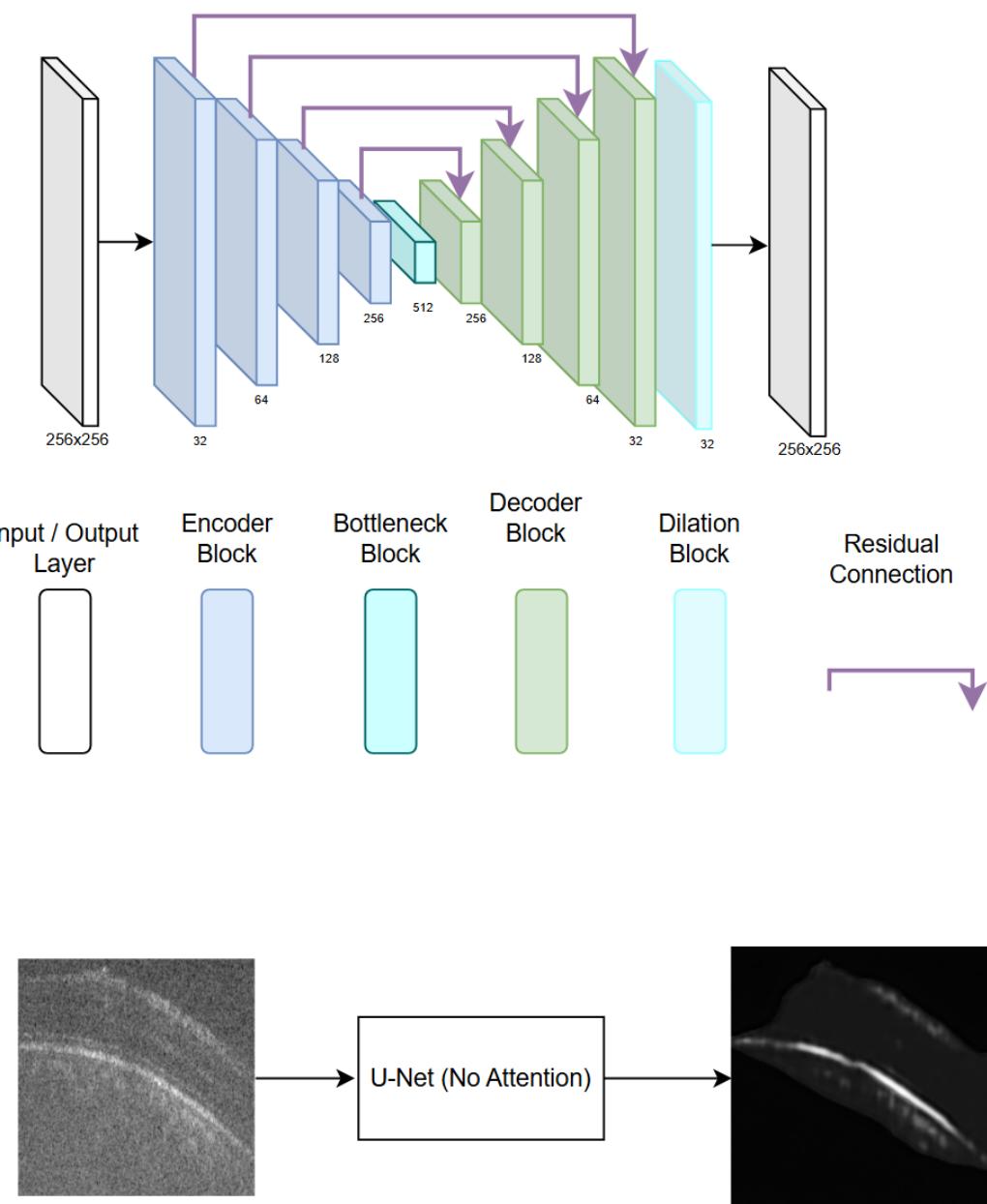
The small-SPDN module increased the input-output channels to 64-64, substantially increasing the computational overhead, leading to hardware limitations.

D.2 Large SPDN U-Net



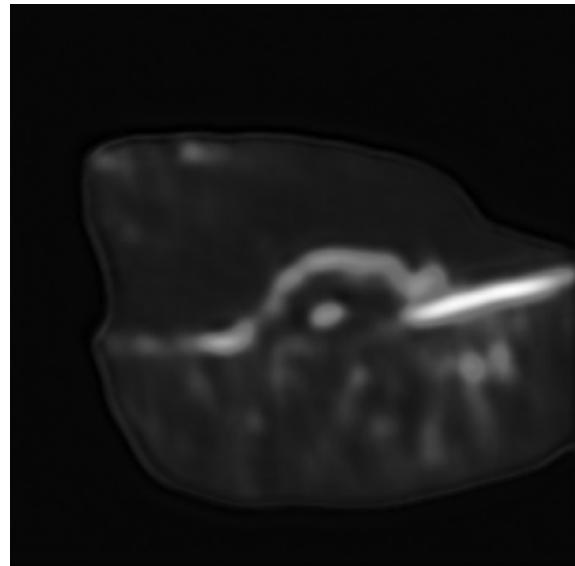
The larger SPDN module still had 32 input-output channels, leading to a redundancy in computational overhead.

D.3 SPDN U-Net without Attention



An equal in depth U-Net without CBAM node showed to have diminished effects on the output.

D.4 U-Net without Attention Sample Output



The output of the model without attention was simply unusable. Indicating a strong reliance on the CBAM methodology to make SPDN work.