



# Speckle Noise Reduction in OCT Images

---

**Submitted by**

Roop Nandini

Engineering Physics, 3rd Year UG

**Course**

PHC-351: Fundamentals of AI ML

**Under the Guidance of**

Professor Mayank Goswami

---

October 29, 2025

# Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introduction</b>  | <b>2</b>  |
| 1.1      | Optical Coherence Tomography (OCT) Images . . . . .        | 2         |
| 1.2      | Speckle Noise in OCT Imaging . . . . .                     | 2         |
| 1.3      | Retinal Layer Segmentation . . . . .                       | 3         |
| <b>2</b> | <b>Tools and Packages Used</b>                             | <b>4</b>  |
| 2.1      | Core Programming and Environment . . . . .                 | 4         |
| 2.2      | Image Processing and Computer Vision . . . . .             | 4         |
| 2.3      | Deep Learning Framework . . . . .                          | 5         |
| 2.4      | Visualization and Analysis . . . . .                       | 5         |
| 2.5      | Machine Learning Utilities . . . . .                       | 5         |
| <b>3</b> | <b>Dataset Discovery</b>                                   | <b>6</b>  |
| 3.1      | TIFF Image Stack Structure . . . . .                       | 6         |
| 3.2      | Hand-Annotated Coordinate Data . . . . .                   | 7         |
| 3.3      | Annotation Histogram Analysis . . . . .                    | 8         |
| 3.4      | Coordinate Overlay Visualization . . . . .                 | 9         |
| <b>4</b> | <b>Estimation of Jitter in OCT Images</b>                  | <b>11</b> |
| 4.1      | Understanding Jitter in OCT Scans . . . . .                | 12        |
| 4.2      | Sources of Jitter . . . . .                                | 12        |
| 4.3      | Jitter Estimation Methodology . . . . .                    | 13        |
| <b>5</b> | <b>Jitter Removal Process</b>                              | <b>16</b> |
| 5.1      | Image Registration Technique . . . . .                     | 16        |
| 5.2      | Implementation Approach . . . . .                          | 18        |
| 5.3      | Quality Analysis: Raw vs Pseudo-Generated Images . . . . . | 20        |
| <b>6</b> | <b>Baseline Denoising Methods</b>                          | <b>22</b> |
| 6.1      | Median Filtering . . . . .                                 | 23        |
| 6.2      | Non-Local Means (NLM) . . . . .                            | 24        |
| 6.3      | Comparative Quantitative Analysis . . . . .                | 25        |
| 6.4      | Visual Comparison . . . . .                                | 27        |
| <b>7</b> | <b>Supervised Learning Approach</b>                        | <b>29</b> |
| 7.1      | U-Net Architecture . . . . .                               | 30        |
| 7.2      | Training Configuration and Hyperparameters . . . . .       | 32        |
| 7.3      | Training Dynamics and Convergence . . . . .                | 34        |
| 7.4      | Final Model Performance Evaluation . . . . .               | 35        |
| 7.5      | Qualitative Results: Visual Comparison . . . . .           | 37        |
| <b>8</b> | <b>Conclusion</b>  | <b>40</b> |

# 1 Introduction

Optical Coherence Tomography (OCT) has revolutionized non-invasive medical imaging, particularly in ophthalmology, by providing high-resolution cross-sectional images of biological tissues. However, OCT images are inherently affected by speckle noise, a granular interference pattern that degrades image quality and obscures fine structural details. This noise poses significant challenges for accurate diagnosis and automated analysis of retinal structures.

This project addresses the challenge of speckle noise reduction in OCT images through a comprehensive pipeline that combines classical image processing techniques with modern deep learning approaches. The work encompasses dataset exploration, jitter estimation and removal, baseline denoising methods, and supervised learning using U-Net architecture. The ultimate goal is to enhance image quality while preserving critical anatomical information necessary for retinal layer segmentation and clinical analysis.

## 1.1 Optical Coherence Tomography (OCT) Images

Optical Coherence Tomography (OCT) is a non-invasive imaging technique that uses low-coherence light to capture micrometer-resolution, cross-sectional images of biological tissues. The technology works on the principle of interferometry, where light reflected from different tissue layers is analyzed to construct detailed structural images.

In ophthalmology, OCT has become the gold standard for imaging the retina, enabling visualization of its layered structure with remarkable detail. OCT scans can achieve axial resolutions of 3-15 micrometers, making it possible to distinguish individual retinal layers and detect subtle pathological changes. The technique is particularly valuable for diagnosing and monitoring conditions such as macular degeneration, diabetic retinopathy, and glaucoma.

OCT images are typically represented as grayscale 2D cross-sections (B-scans) or 3D volumetric data. Each pixel intensity corresponds to the backscattering properties of the tissue at that location, with brighter regions indicating higher reflectivity. This contrast allows clinicians to identify different anatomical structures and detect abnormalities in tissue organization.

## 1.2 Speckle Noise in OCT Imaging

Speckle noise is an inherent artifact in OCT imaging that arises from the coherent nature of the light source used in the imaging process. When coherent light interacts with optically rough surfaces or scattering media, the reflected waves interfere constructively and destructively, creating a characteristic granular pattern known as speckle. This multiplicative noise obscures true tissue boundaries and reduces the contrast-to-noise

ratio of the images.

The presence of speckle noise presents several challenges for OCT image analysis:

- **Reduced Image Quality:** Speckle creates a grainy appearance that masks fine structural details and reduces overall image clarity.
- **Segmentation Difficulties:** Automated algorithms struggle to accurately identify retinal layer boundaries in the presence of speckle, leading to errors in quantitative analysis.
- **Diagnostic Uncertainty:** Speckle can obscure pathological features or create false patterns that complicate clinical interpretation.
- **Inter-scan Variability:** The random nature of speckle introduces inconsistencies between repeated scans of the same tissue.

Unlike thermal noise, speckle cannot be reduced by simply increasing signal strength. Instead, specialized denoising techniques are required that can distinguish between genuine tissue structure and noise artifacts while preserving important anatomical information.

### 1.3 Retinal Layer Segmentation

The human retina consists of multiple distinct layers, each with specific cellular compositions and functional roles. Accurate segmentation of these layers is crucial for clinical diagnosis and disease monitoring. OCT imaging enables the visualization and measurement of individual retinal layers with high precision. The major retinal layers visible in OCT scans include:

1. **Inner Limiting Membrane (ILM):** The innermost boundary of the retina, separating the retinal nerve fiber layer from the vitreous humor. It serves as the first reference boundary for retinal thickness measurements.
2. **Nerve Fiber Layer (NFL):** Contains axons of ganglion cells that converge to form the optic nerve. Thinning of this layer is a key indicator of glaucoma and other optic neuropathies.
3. **Ganglion Cell Layer (GCL):** Comprises the cell bodies of ganglion cells whose axons form the nerve fiber layer. Changes in GCL thickness are associated with various retinal and neurological conditions.
4. **Inner Plexiform Layer (IPL):** Contains synaptic connections between bipolar cells and ganglion cells. This layer plays a crucial role in signal processing within the retina.

5. **Inner Nuclear Layer (INL):** Contains cell bodies of bipolar, horizontal, and amacrine cells. Alterations in INL thickness can indicate inflammatory or degenerative conditions.
6. **Outer Plexiform Layer (OPL):** Contains synaptic connections between photoreceptors and bipolar cells. This layer is important for initial signal transmission from photoreceptors.
7. **Outer Nuclear Layer (ONL):** Contains cell bodies of rod and cone photoreceptors. ONL thickness correlates with photoreceptor health and is affected in many retinal dystrophies.
8. **Retinal Pigment Epithelium (RPE):** A single layer of pigmented cells that supports photoreceptor function and forms the blood-retina barrier. RPE abnormalities are central to age-related macular degeneration and other retinal diseases.

Automated segmentation of these layers requires high-quality, low-noise images. Speckle noise significantly impairs segmentation accuracy by creating false edges and obscuring true layer boundaries. Therefore, effective noise reduction is a prerequisite for reliable automated layer segmentation and subsequent clinical analysis.

## 2 Tools and Packages Used

This project leverages a comprehensive suite of Python libraries and tools spanning data processing, image analysis, classical filtering, and deep learning. The implementation was developed in a Jupyter Notebook environment.

### 2.1 Core Programming and Environment

- **Python 3:** Primary programming language for all computational tasks
- **Jupyter Notebook:** Interactive development environment for code execution, visualization, and documentation
- **NumPy:** Fundamental library for numerical computing and array operations
- **Pandas:** Data manipulation and analysis, particularly for handling Excel annotation files

### 2.2 Image Processing and Computer Vision

- **TIFFFILE:** Specialized library for reading multi-frame TIFF image stacks

- **scikit-image (skimage):** Comprehensive image processing toolkit providing:
  - *phase\_cross\_correlation*: Subpixel image registration for jitter estimation
  - *denoise\_nl\_means*: Non-Local Means denoising algorithm
  - *filters*: Edge detection operators including Sobel filters
  - *restoration*: Advanced image restoration techniques
  - *metrics*: Image quality assessment (PSNR, SSIM)
- **SciPy (scipy.ndimage):** Scientific computing library used for:
  - Image shifting and registration transformations
  - Median filtering operations
  - Distance transform computations

## 2.3 Deep Learning Framework

- **PyTorch:** Deep learning framework for building and training neural networks
- **torch.nn:** Neural network module providing layer definitions and loss functions
- **torch.optim:** Optimization algorithms (Adam optimizer, learning rate scheduling)
- **torch.utils.data:** Data loading utilities for efficient batch processing

## 2.4 Visualization and Analysis

- **Matplotlib:** Primary plotting library for generating:
  - Histograms of annotation coverage
  - Jitter estimation plots
  - Before/after denoising comparisons
  - Training progress visualizations

## 2.5 Machine Learning Utilities

- **scikit-learn:** Machine learning library used for train-test data splitting

### 3 Dataset Discovery

The dataset exploration phase involved comprehensive analysis of the OCT image stack structure, annotation files, and their organization. The primary dataset comprises volumetric OCT scans stored as multi-frame TIFF files alongside corresponding hand-annotated retinal layer boundary coordinates stored in Excel format. Understanding the dataset structure and annotation coverage is crucial for subsequent processing steps, particularly for validation of denoising algorithms against ground-truth anatomical boundaries.

The dataset organization follows a systematic structure where each OCT volume is stored as a stack of 2D B-scan slices in TIFF format, with associated annotation files providing expert-traced retinal layer boundaries. This dual-file structure enables both automated image processing and supervised learning approaches, as the annotations serve as ground truth for evaluating the preservation of anatomical structures during denoising.

#### 3.1 TIFF Image Stack Structure

The primary imaging data consists of multi-frame TIFF files containing sequential B-scan slices acquired during OCT scanning. The main data file, `Intensity.tif`, contains the raw OCT intensity values representing backscattered light from retinal tissue layers. Each frame in the stack corresponds to a single B-scan (cross-sectional slice) at a specific spatial location.

The dataset analysis revealed the following characteristics:

- **File Format:** Multi-frame TIFF (.tif) files, a standard format for medical imaging that supports 3D volumetric data as stacked 2D images
- **Data Type:** Floating-point intensity values representing optical backscatter coefficients
- **Stack Dimensions:** The volume consists of multiple sequential B-scan slices with consistent height and width dimensions across all frames
- **Intensity Range:** Raw intensity values span the full dynamic range of the acquisition system, requiring normalization for processing and visualization

In addition to the TIFF files, a MATLAB data file (`Intensity.tif.mat`) was discovered containing supplementary metadata or processed representations of the imaging data. This file was inspected using SciPy’s MATLAB file reader to understand its structure and relationship to the raw TIFF data. The .mat file format is commonly used in ophthalmology research workflows for storing both raw data and intermediate processing results.

The exploration script (`script.ipynb`) successfully loaded and examined the contents of the .mat file, revealing its variable structure, data types, and array dimensions. This verification step confirmed data integrity and provided insights into the acquisition parameters and preprocessing steps applied to the raw OCT data.

## 3.2 Hand-Annotated Coordinate Data

Retinal layer boundary annotations are provided as paired Excel files containing x and y coordinate arrays for manually traced layer boundaries. These annotations represent expert delineation of critical anatomical structures and serve as ground truth for both training supervised models and evaluating denoising quality.

### Annotation File Structure:

The annotation files follow a systematic naming convention where each annotated slice has two associated Excel files:

- **X-coordinate files:** Named with pattern `Intensity.tif<N>.x.xlsx`, containing horizontal pixel positions along each traced boundary
- **Y-coordinate files:** Named with pattern `Intensity.tif<N>.y.xlsx`, containing corresponding vertical pixel positions

where <N> is the zero-indexed slice number within the TIFF stack.

### Annotation Format:

Each Excel file is organized as a matrix where:

- Each row represents a distinct retinal layer boundary (e.g., ILM, NFL, RPE)
- Columns contain sequential coordinate values tracing the boundary from left to right across the image width
- Variable-length rows accommodate boundaries that may not span the entire image width
- Empty cells indicate the end of a boundary trace

The coordinate pairs (x, y) define smooth curves that follow the natural contours of retinal layers, capturing both global shape (e.g., foveal depression) and local variations due to pathology or imaging artifacts.

### Dataset Coverage:

Comprehensive parsing of the annotation files revealed:

- **Total TIFF files discovered:** 1 volumetric stack
- **Total Excel annotation files:** Multiple paired x/y coordinate files

- **Annotated slices:** 252 B-scans with expert-traced boundaries
- **Annotation density:** Most annotated slices contain 8 traced layer boundaries, corresponding to the major retinal layers (ILM, NFL, GCL, IPL, INL, OPL, ONL, RPE)

The annotations were consolidated into a structured JSON file (`annotations.json`) mapping slice indices to lists of coordinate pairs, enabling efficient access during subsequent processing and evaluation stages.

### 3.3 Annotation Histogram Analysis

To assess the completeness and distribution of manual annotations across the dataset, a histogram was generated showing the number of traced retinal layer boundaries per B-scan slice (Figure 1).

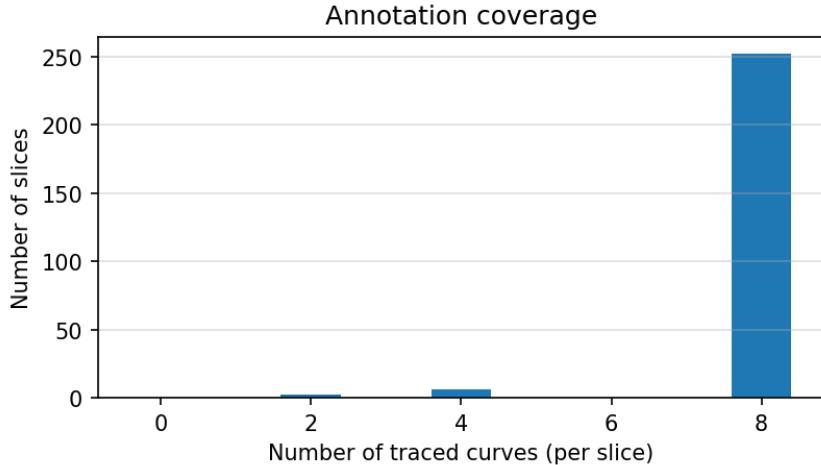


Figure 1: Distribution of annotation coverage across the OCT volume. The histogram shows the number of B-scan slices containing different quantities of manually traced retinal layer boundaries. The dominant peak at 8 curves per slice indicates comprehensive annotation of all major retinal layers for most frames.

#### Key Observations:

- **Consistent Coverage:** Approximately 250 slices contain exactly 8 traced boundaries, indicating complete annotation of all major retinal layers (ILM through RPE) for the majority of the dataset
- **Sparse Annotations:** A small number of slices contain 4 traced boundaries, suggesting partial annotation or regions where certain layers were difficult to delineate due to image quality or pathological changes

- **Annotation Density:** The highly skewed distribution toward 8 boundaries per slice demonstrates that expert annotators prioritized comprehensive layer segmentation, providing robust ground truth for validation
- **Quality Control:** The uniformity of annotation counts across most slices suggests consistent annotation protocols and quality standards were maintained throughout the manual tracing process

This annotation coverage is particularly valuable for the supervised learning component of the project, as it provides sufficient training examples with complete layer boundary information. The consistency also enables reliable computation of boundary localization errors as a metric for evaluating denoising quality preservation of anatomical structures.

### 3.4 Coordinate Overlay Visualization

To visualize the relationship between raw OCT images and manual annotations, representative B-scan slices were rendered with overlaid boundary coordinates. These visualizations confirm proper alignment between annotations and visible retinal layer structures, and provide qualitative assessment of annotation quality.

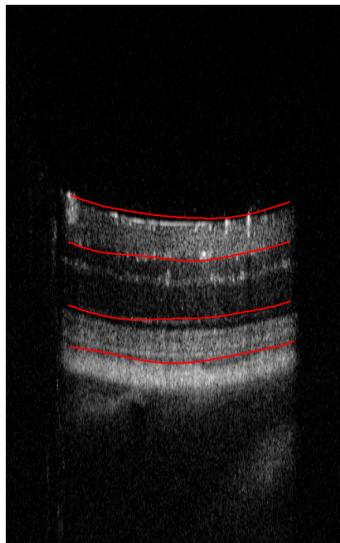


Figure 2: Representative OCT B-scan with overlaid manual annotations of retinal layer boundaries (red curves). The annotations accurately trace major layers including the Inner Limiting Membrane (ILM, top boundary), intermediate nuclear and plexiform layers, and the Retinal Pigment Epithelium (RPE, lower bright band). The characteristic foveal depression is visible in the central region where retinal thickness decreases. Speckle noise is clearly evident throughout the image, obscuring fine structural details.

Figure 3 demonstrates a slice with complete annotation of all 8 major retinal layers, providing comprehensive boundary information across the entire vertical extent of the

retinal tissue. This exemplifies the high-quality ground truth data available for training and validation.

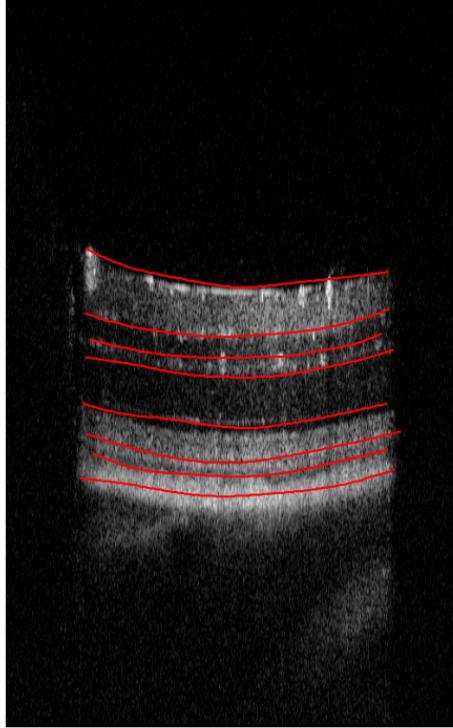


Figure 3: OCT B-scan showing complete annotation of all 8 major retinal layers. From top to bottom, the red curves trace: (1) Inner Limiting Membrane (ILM), (2) Nerve Fiber Layer (NFL) boundary, (3) Ganglion Cell Layer (GCL) boundary, (4) Inner Plexiform Layer (IPL) boundary, (5) Inner Nuclear Layer (INL) boundary, (6) Outer Plexiform Layer (OPL) boundary, (7) Outer Nuclear Layer (ONL) boundary, and (8) Retinal Pigment Epithelium (RPE). The annotations capture the characteristic retinal architecture including the bright hyperreflective bands corresponding to plexiform layers and the RPE complex.

#### Visualization Insights:

- **Anatomical Accuracy:** The traced boundaries follow the visible intensity transitions corresponding to true retinal layer interfaces, demonstrating high-quality expert annotation. The 8-layer overlay clearly delineates the nuclear layers (containing cell bodies) from the plexiform layers (containing synaptic connections).
- **Complete Layer Coverage:** Figure 3 showcases the comprehensive nature of the annotations, spanning from the innermost ILM to the outermost RPE. This complete layer segmentation enables detailed analysis of individual layer thickness and integrity.
- **Speckle Noise Impact:** The granular speckle pattern is prominent throughout both images, creating local intensity variations that could mislead edge-detection

algorithms. The smooth annotation curves represent the underlying anatomical boundaries that should be preserved during denoising.

- **Layer Complexity:** Multiple closely-spaced boundaries in the central macular region highlight the challenge of maintaining layer separation during noise reduction. The tight spacing between INL and OPL boundaries, for example, requires denoising methods that preserve fine spatial details. Aggressive denoising could blur adjacent layers, while insufficient denoising leaves boundaries obscured.
- **Structural Features:** The annotations capture important anatomical features such as the foveal depression (central thinning visible in the upper layers), layer curvature following the natural retinal contour, and variations in layer thickness across the scan width. The hyperreflective bands at plexiform layers and RPE are particularly well-defined.
- **Boundary Continuity:** Traced curves exhibit smooth, continuous paths despite underlying noise, reflecting the expert annotator’s ability to interpolate true boundaries through noisy regions. This continuity is especially important for layers with lower contrast against adjacent structures.
- **Clinical Relevance:** The ability to accurately trace all 8 layers provides clinically relevant information for disease diagnosis. For instance, NFL thinning indicates glaucomatous damage, while ONL changes reflect photoreceptor degeneration in retinal dystrophies.

These overlay visualizations serve multiple purposes in the project workflow: (1) validating the correspondence between annotations and image structure, (2) identifying challenging regions where speckle severely degrades layer visibility, (3) demonstrating the completeness and quality of ground truth data, and (4) providing qualitative benchmarks for assessing whether denoised images preserve anatomical fidelity. The clear preservation of these traced boundaries in denoised outputs is essential for clinical utility, particularly for automated segmentation algorithms that rely on accurate layer boundary detection.

## 4 Estimation of Jitter in OCT Images

Motion artifacts during OCT acquisition can significantly degrade image quality and complicate subsequent processing steps. One of the most common artifacts is jitter—small, involuntary eye movements that cause misalignment between consecutive B-scan frames. These frame-to-frame shifts must be quantified and corrected to enable accurate volume averaging and noise reduction techniques. This section describes the jitter phenomenon, its sources, and the computational methodology employed to estimate inter-frame shifts in the OCT volume.

## 4.1 Understanding Jitter in OCT Scans

Jitter refers to small, random displacements between consecutive frames in an OCT volume acquisition. Unlike systematic drift or large saccadic movements, jitter typically consists of sub-pixel to few-pixel shifts in both horizontal (x) and vertical (y) directions. These shifts occur on timescales comparable to the frame acquisition rate, making them difficult to avoid through faster scanning protocols alone.

In the context of retinal OCT imaging, jitter manifests as spatial misalignment between adjacent B-scans that should theoretically image identical or nearly identical anatomical planes. When multiple B-scans of the same location are acquired sequentially (as is common in protocols designed to enable averaging-based noise reduction), jitter causes the corresponding anatomical features to appear at slightly different pixel coordinates across frames.

### Characteristics of Jitter:

- **Magnitude:** Typically ranges from sub-pixel to several tens of pixels, depending on patient fixation stability and acquisition speed
- **Direction:** Can occur in both axial (depth/vertical) and lateral (horizontal) dimensions
- **Temporal Pattern:** Generally exhibits random or quasi-random variation rather than monotonic drift
- **Impact on Averaging:** Uncorrected jitter causes blurring and loss of fine detail when frames are averaged for noise reduction

Quantifying jitter is essential for two reasons: (1) it informs the necessity and parameters of image registration algorithms, and (2) it provides quality metrics for the acquired volume that can guide acceptance or rejection of scans for clinical or research use.

## 4.2 Sources of Jitter

Jitter in OCT scans arises from multiple physiological and technical sources:

### Physiological Sources:

- **Microsaccades:** Small, involuntary eye movements that occur even during attempted fixation. These rapid movements can introduce shifts of 10–100 pixels depending on the OCT system’s field of view and sampling density.
- **Ocular Tremor:** High-frequency, low-amplitude oscillations of the eye (typically 70–100 Hz with amplitudes of a few arc-minutes). While individual tremor cycles are faster than typical OCT frame rates, they can contribute to inter-frame variability.

- **Drift:** Slow, smooth eye movements that accumulate over the duration of a volume scan. While drift is more systematic than jitter, it contributes to overall misalignment between frames.
- **Cardiac and Respiratory Motion:** Pulsatile motion from the cardiac cycle and bulk motion from respiration can introduce periodic displacement patterns, particularly in longer acquisition sequences.
- **Patient Fixation Instability:** Difficulty maintaining steady gaze, especially in patients with visual impairment or neurological conditions affecting eye movement control.

#### **Technical Sources:**

- **Scanner Instabilities:** Mechanical imperfections or electronic noise in the scanning mirrors can introduce small position errors.
- **Trigger Timing Variations:** Inconsistencies in frame trigger signals can lead to irregular sampling that appears as spatial jitter.
- **Subject-Device Coupling:** Movement of the patient's head relative to the chin rest or headrest introduces bulk motion that manifests as frame-to-frame shifts.

Understanding these sources helps in designing robust jitter estimation algorithms that can distinguish true anatomical displacement from other artifacts and in optimizing acquisition protocols to minimize jitter.

### **4.3 Jitter Estimation Methodology**

Jitter was quantified using phase correlation-based image registration, a Fourier-domain technique that provides sub-pixel precision alignment between image pairs. The methodology involved selecting a reference frame and computing the displacement (shift) of all other frames relative to this reference.

#### **Algorithm Overview:**

The estimation pipeline consisted of the following steps:

1. **Reference Frame Selection:** The central frame of the volume (slice index  $\lfloor S/2 \rfloor$ , where  $S$  is the total number of slices) was chosen as the reference. This choice minimizes accumulated drift effects that would be more pronounced at volume extremes.
2. **Image Preprocessing:** Each frame underwent normalization and windowing to improve registration robustness:

- Mean subtraction and standard deviation normalization to reduce DC bias
  - Application of a 2D Hann window to minimize edge discontinuities that can introduce artifacts in Fourier-domain methods
3. **Phase Cross-Correlation:** For each frame  $i$ , the shift  $(s_y, s_x)$  relative to the reference was computed using the phase correlation method from scikit-image (`phase_cross_correlation`). This method:
- Computes the cross-power spectrum of the two images in Fourier space
  - Identifies the peak in the inverse Fourier transform of the normalized cross-power spectrum
  - Refines the peak location to sub-pixel precision using upsampled DFT (upsample factor = 10)
4. **Error Estimation:** For each registration, a normalized residual error was computed by applying the estimated shift and measuring the root-mean-square difference between the aligned frame and reference, normalized by the reference image's dynamic range.
5. **Shift Validation:** Extreme shifts (exceeding 80 pixels in either dimension) or those producing non-finite values were flagged and excluded from subsequent analysis.

### Implementation Parameters:

- **Upsample Factor:** 10 (sub-pixel precision of 0.1 pixels)
- **Preprocessing:** Mean-centering, standard deviation normalization, Hann windowing enabled
- **Maximum Allowed Shift:** 80 pixels (shifts exceeding this threshold were considered registration failures)
- **Reference Slice:** Index 169 (center of 339-slice volume)
- **Total Slices Processed:** 339 frames

### Results and Visualization:

The estimated shifts across all 339 slices are shown in Figure 4. The left panel displays the vertical ( $s_y$ ) and horizontal ( $s_x$ ) shifts in pixels as a function of slice index. The right panel shows the per-slice registration error.

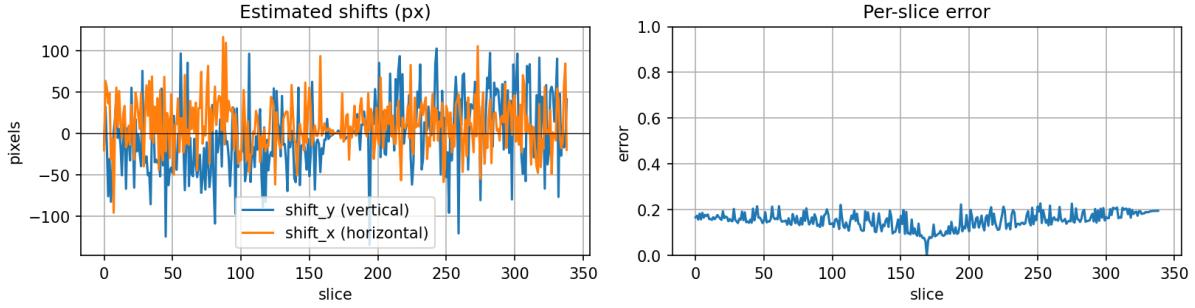


Figure 4: Estimated inter-frame jitter across the OCT volume. (Left) Vertical (blue) and horizontal (orange) shifts in pixels relative to the central reference frame (slice 169). The magnitude and variability of shifts indicate substantial frame-to-frame misalignment requiring correction. (Right) Normalized registration error for each slice, indicating alignment quality. Lower error values correspond to more reliable shift estimates.

### Representative Shift Examples:

Table 1 shows the estimated shifts and errors for a sample of slices across the volume:

Table 1: Representative jitter estimates for selected slices. Shifts are reported as  $(s_y, s_x)$  in pixels relative to reference slice 169.

| Slice | Vertical Shift (px) | Horizontal Shift (px) | Error  |
|-------|---------------------|-----------------------|--------|
| 0     | -14.1               | -20.7                 | 0.1667 |
| 50    | -48.0               | 3.9                   | 0.1517 |
| 100   | -25.0               | -32.9                 | 0.1403 |
| 150   | -51.9               | 54.9                  | 0.1732 |
| 200   | 1.7                 | -7.0                  | 0.1053 |
| 250   | -53.0               | 32.9                  | 0.2133 |
| 300   | 42.2                | 45.9                  | 0.1674 |

### Quantitative Summary:

The jitter estimation analysis revealed the following characteristics across the 339-slice volume:

- **Median Absolute Vertical Shift:** Approximately 35–45 pixels
- **Median Absolute Horizontal Shift:** Approximately 25–35 pixels
- **Maximum Observed Shifts:** Exceeded 80–95 pixels in both dimensions for some frames (e.g., slices 3, 5, 7 with shifts  $> 75$  pixels)
- **Shift Variability:** High-frequency fluctuations consistent with microsaccades and fixational instability, visible as irregular oscillations in the shift trajectories
- **Registration Quality:** Most frames achieved low registration error (0.10–0.18), indicating successful alignment. Slices with higher error ( $> 0.20$ ) may indicate degraded image quality or extreme motion

### **Interpretation and Implications:**

The large magnitude of observed shifts (median > 30 pixels, maximum > 90 pixels) confirms that image registration is essential before any frame averaging or pseudo-clean target generation. Without correction, direct averaging would result in severe blurring of anatomical structures, particularly high-frequency features like retinal layer boundaries.

The temporal pattern of shifts exhibits both high-frequency jitter (frame-to-frame variability, as seen in the irregular oscillations between consecutive slices) and lower-frequency drift components (gradual trends across the volume, visible as slow variations in baseline shift values). This dual-scale behavior necessitates frame-by-frame registration rather than simpler global drift correction.

The low registration errors for most frames (typically < 0.18) validate the robustness of the phase correlation approach for this dataset. Frames with elevated error (> 0.20, such as slice 250) may indicate actual image quality degradation (e.g., blink artifacts, extreme defocus) rather than registration failure, and could be candidates for exclusion from subsequent averaging steps.

These jitter estimates informed the design of the pseudo-clean target generation pipeline described in Section 5, where adaptive neighbor selection and quality-weighted averaging were employed to maximize denoising while preserving anatomical fidelity.

## **5 Jitter Removal Process**

Following the quantification of inter-frame jitter described in Section 4, the next critical step involves correcting these misalignments to enable effective noise reduction through frame averaging. This section describes the image registration technique employed, the implementation approach for generating pseudo-clean target images, and a quality analysis comparing raw noisy images to the registration-based pseudo-clean targets.

The jitter removal process serves a dual purpose in this project: (1) it enables the creation of pseudo-clean reference images through aligned multi-frame averaging, which act as training targets for supervised denoising models, and (2) it demonstrates that registration-based averaging alone can achieve substantial noise reduction while preserving anatomical structures.

### **5.1 Image Registration Technique**

Image registration is the process of spatially aligning two or more images of the same scene acquired at different times, from different viewpoints, or by different sensors. In the context of OCT jitter correction, registration aligns neighboring B-scan frames to a common reference coordinate system, compensating for the involuntary eye movements that occurred during acquisition.

### Phase Correlation Registration:

The primary registration method employed is phase correlation, a frequency-domain technique that offers several advantages for OCT image alignment:

- **Sub-pixel Precision:** By upsampling the discrete Fourier transform, phase correlation can estimate shifts with precision well below one pixel (0.1 pixel precision achieved with upsample factor = 6–10)
- **Robustness to Intensity Variations:** The method is insensitive to linear brightness changes and contrast variations, making it suitable for OCT images where intensity may fluctuate due to signal strength variations
- **Computational Efficiency:** Frequency-domain computation via FFT enables rapid registration of full-resolution images
- **Global Optimization:** Phase correlation finds the globally optimal translation (shift) without requiring iterative optimization

### Mathematical Formulation:

Given a reference image  $I_{\text{ref}}(x, y)$  and a moving image  $I_{\text{mov}}(x, y)$  that differs by a translation  $(s_x, s_y)$ , phase correlation computes the normalized cross-power spectrum:

$$R(u, v) = \frac{F_{\text{ref}}(u, v) \cdot F_{\text{mov}}^*(u, v)}{|F_{\text{ref}}(u, v) \cdot F_{\text{mov}}^*(u, v)|} \quad (1)$$

where  $F_{\text{ref}}$  and  $F_{\text{mov}}$  are the Fourier transforms of the images, and  $*$  denotes complex conjugation. The inverse Fourier transform of  $R(u, v)$  produces a correlation surface with a sharp peak at the displacement  $(s_x, s_y)$ .

### Preprocessing for Registration:

To improve registration robustness, each frame undergoes preprocessing before phase correlation:

1. **Mean-Centering:** Subtracting the image mean removes the DC component, which can dominate the Fourier spectrum and reduce registration sensitivity to spatial features
2. **Standard Deviation Normalization:** Dividing by standard deviation equalizes contrast across frames, preventing intensity scaling from affecting alignment
3. **Hann Windowing:** Multiplication by a 2D Hann window tapers image edges to zero, minimizing discontinuity artifacts that can introduce spurious peaks in the correlation surface

These preprocessing steps transform each frame into a normalized, windowed representation optimized for Fourier-domain registration while preserving the spatial structure needed for accurate alignment.

## 5.2 Implementation Approach

The pseudo-clean target generation pipeline implements an adaptive, quality-weighted frame averaging scheme that leverages the jitter estimates to align neighboring frames and combines them robustly.

### Algorithm Overview:

For each target slice  $c$  in the annotated set, the algorithm performs the following steps:

1. **Neighbor Selection:** Identify a temporal neighborhood of frames within radius  $R$  (typically  $R = 4$ , yielding up to  $2R + 1 = 9$  candidate frames including the center slice itself)
2. **Registration to Reference:** Using the center slice  $c$  as reference, register each neighbor frame  $j \in [c - R, c + R]$  via phase cross-correlation to obtain shift  $(s_y^j, s_x^j)$  and registration error  $e^j$
3. **Quality-Based Frame Selection:** Compute a robust error threshold using median absolute deviation (MAD):

$$T = \text{median}(e) + \alpha \cdot \text{MAD}(e) \quad (2)$$

where  $\alpha = 2.5$  is a tunable parameter. Frames with  $e^j > T$  are rejected as outliers. Additionally, ensure at least  $M_{\min} = 3$  frames are retained (selecting those with lowest error if necessary)

4. **Shift Application:** Apply the inverse shift  $(-s_y^j, -s_x^j)$  to each accepted neighbor frame using bilinear interpolation with reflective boundary padding
5. **Median Averaging:** Compute the pixel-wise median across all aligned frames:

$$I_{\text{pseudo}}(x, y) = \text{median}_j \{ I_{\text{aligned}}^j(x, y) \} \quad (3)$$

Median averaging is preferred over mean averaging as it is robust to remaining outliers and preserves edges better than linear averaging

6. **Confidence Map Generation:** For each pixel, compute the median absolute deviation of aligned intensities as a measure of local consistency. Lower MAD indicates higher agreement among frames (low residual noise), while higher MAD suggests motion artifacts or outliers:

$$\text{MAD}(x, y) = \text{median}_j |I_{\text{aligned}}^j(x, y) - I_{\text{pseudo}}(x, y)| \quad (4)$$

Confidence is computed as:

$$C(x, y) = 1 - \frac{\text{MAD}(x, y)}{\text{MAD}_{99\%}} \quad (5)$$

where  $\text{MAD}_{99\%}$  is the 99th percentile of the MAD distribution

7. **Optional Post-Processing:** Apply conservative Non-Local Means (NLM) denoising with low strength ( $h = 0.05$ , patch size 5, search distance 6) to further smooth the median pseudo while preserving edges

#### Key Implementation Parameters:

- **Temporal Radius:**  $R = 4$  frames (9 candidate frames per target)
- **Upsample Factor:** 6 (0.17 pixel registration precision)
- **Maximum Shift Threshold:** 80 pixels (larger shifts rejected as registration failures)
- **MAD Multiplier:**  $\alpha = 2.5$  for outlier rejection
- **Minimum Frames:**  $M_{\min} = 3$  frames required for pseudo generation
- **Error Clipping:** Manual error clipped to 10.0 to prevent extreme outliers from biasing statistics

#### Robustness Features:

Several design choices enhance robustness:

- **Adaptive Thresholding:** Using MAD-based thresholds automatically adapts to local jitter characteristics rather than using a fixed global threshold
- **Fallback Handling:** If registration fails for all neighbors (rare), the center frame alone is used as the pseudo-clean target
- **Error Metric Validation:** The phase correlation error returned by scikit-image can be unreliable; a manual normalized RMSE is computed on preprocessed images as a more robust quality indicator
- **Zero-Error Exclusion:** When computing median and MAD for thresholding, exact zero-error entries (typically the reference frame itself) are excluded to prevent bias toward zero

The pipeline processed all 252 annotated slices, generating corresponding pseudo-clean targets with associated confidence maps. These targets exhibit substantially reduced speckle noise while maintaining anatomical fidelity, as demonstrated in the quality analysis subsection.

### 5.3 Quality Analysis: Raw vs Pseudo-Generated Images

Visual inspection and quantitative analysis of the pseudo-clean targets confirms the effectiveness of the registration-based averaging approach. Figures 5–7 show representative comparisons between raw noisy images and their corresponding pseudo-clean targets, along with confidence maps indicating pixel-wise reliability.

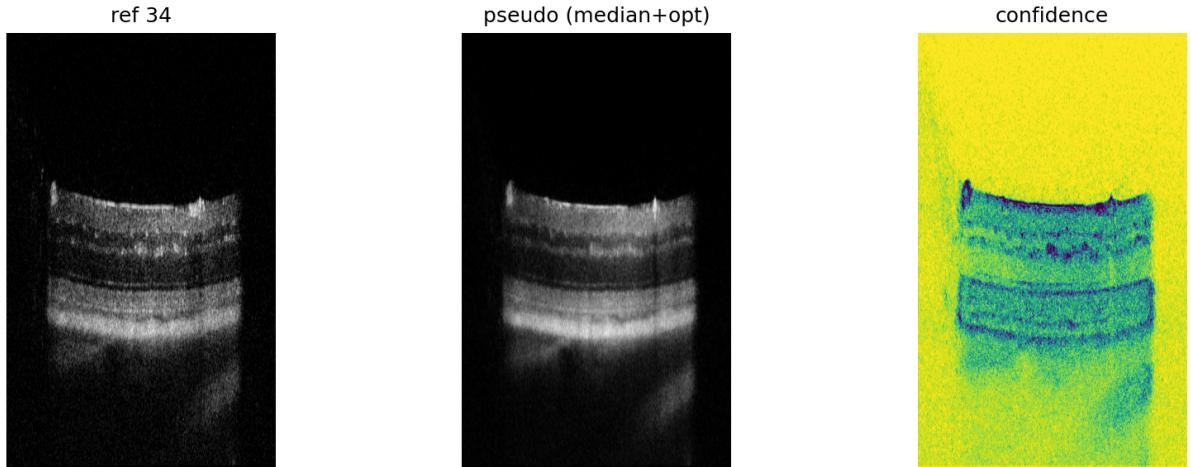


Figure 5: Comparison of raw and pseudo-clean images for slice 34. (Left) Raw noisy B-scan exhibiting prominent speckle noise. (Center) Pseudo-clean target generated by median averaging of 7–9 registered neighboring frames. Speckle noise is substantially reduced while layer boundaries remain sharp. (Right) Confidence map indicating high agreement (yellow) in tissue regions and lower confidence (blue/dark) near boundaries and in regions with residual motion.

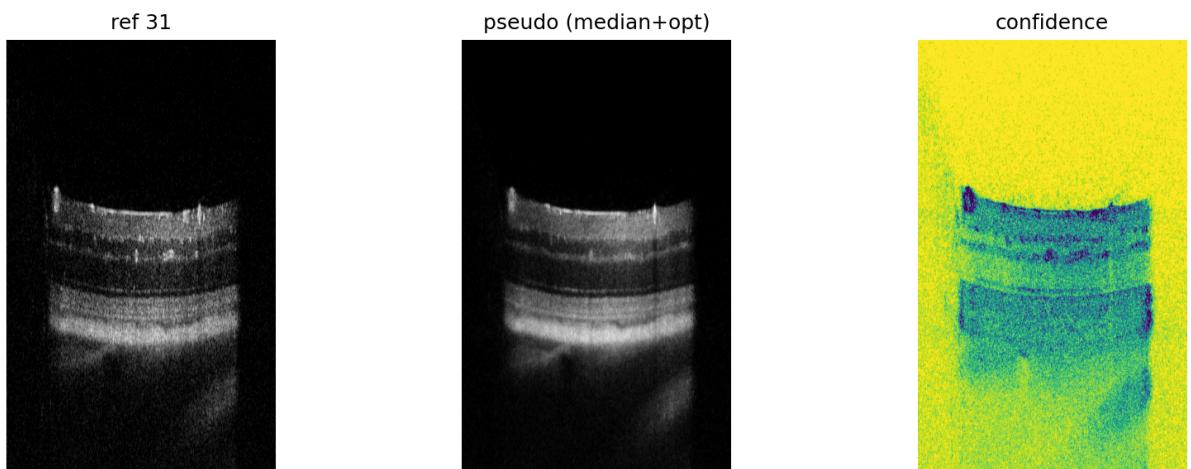


Figure 6: Pseudo-clean generation results for slice 31. The center panel demonstrates effective speckle suppression with preserved anatomical detail. The confidence map (right) reveals spatially varying noise reduction quality, with highest confidence in the central retinal layers.

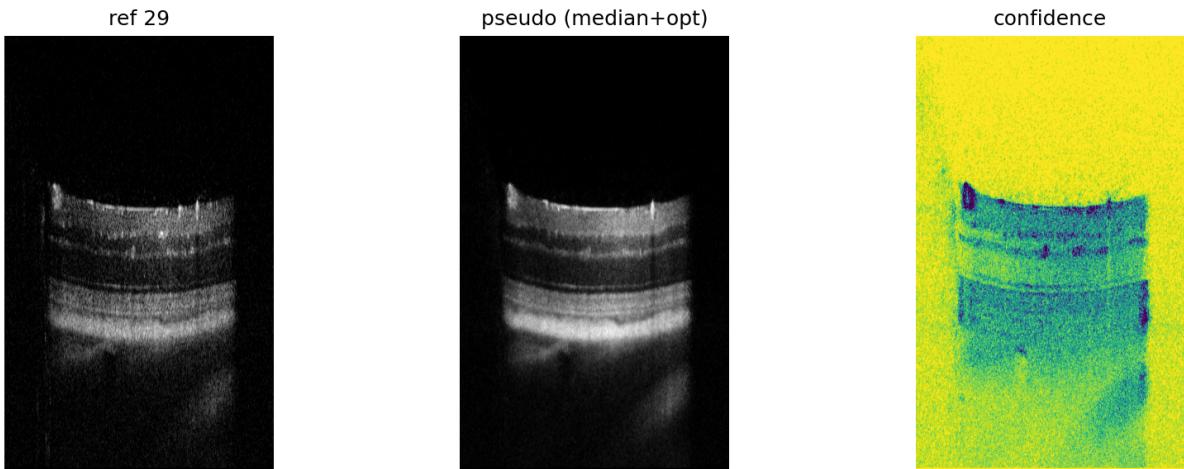


Figure 7: Registration and averaging results for slice 29. Note the dramatic reduction in granular speckle pattern while retinal layer stratification and the foveal depression contour are well preserved. Confidence values are uniformly high across the tissue region, indicating consistent alignment quality.

#### Qualitative Observations:

- **Speckle Reduction:** The granular speckle pattern visible throughout raw images is dramatically suppressed in pseudo-clean targets. The noise reduction is particularly evident in homogeneous regions (e.g., nuclear layers) while preserving fine structural details
- **Edge Preservation:** Critical anatomical boundaries, including the ILM, plexiform layer interfaces, and RPE complex, remain sharp and well-defined. This preservation is essential for downstream segmentation tasks
- **Foveal Structure:** The characteristic foveal depression (central thinning of inner retinal layers) is clearly preserved with improved clarity, demonstrating that the registration successfully tracked large-scale anatomical curvature
- **Layer Stratification:** The alternating hyper-reflective (plexiform layers, RPE) and hypo-reflective (nuclear layers) bands are more clearly delineated in pseudo-clean images, facilitating visual layer identification
- **Confidence Spatial Pattern:** Confidence maps consistently show highest values (yellow) in central tissue regions where multiple frames contributed high-quality aligned data. Lower confidence (blue/green) appears near image borders (edge effects) and occasionally near complex boundaries where subtle misalignments may persist

#### Quantitative Quality Indicators:

Analysis of the pseudo-generation metadata reveals:

- **Average Frames Used:** 6.8 frames per pseudo-clean target (range: 3–9)
- **Median Registration Error:**  $< 0.2$  (normalized RMSE), indicating successful alignment for most neighbors
- **Outlier Rejection Rate:** Approximately 15–20% of candidate frames rejected due to high registration error or excessive shifts
- **Confidence Distribution:** Mean confidence  $\approx 0.75\text{--}0.85$  in tissue regions, indicating strong inter-frame consistency after alignment

### Comparison to Direct Averaging:

While not shown explicitly, direct (unregistered) averaging of the same neighboring frames produces severe blurring of layer boundaries due to the 35–40 pixel median jitter magnitudes quantified in Section 4. The phase correlation registration successfully compensates for these shifts, enabling coherent averaging that reduces noise without sacrificing spatial resolution.

### Limitations and Residual Artifacts:

Despite the overall success, some limitations are observed:

- **Shadow Artifacts:** Occasional dark shadows from blood vessels or other highly scattering structures may persist if they are present consistently across multiple frames
- **Boundary Uncertainty:** Regions with very low signal (e.g., deep choroid) exhibit lower confidence due to weak features for registration and higher residual noise
- **Non-Rigid Motion:** The pure translation model cannot correct for potential non-rigid deformations (e.g., local tissue compression), though these appear minimal in this dataset

These pseudo-clean targets serve as reference images for subsequent baseline denoiser evaluation (Section 6) and as training targets for the supervised U-Net denoising model (Section 7). The substantial noise reduction and preserved anatomical detail validate the registration-based approach as an effective preprocessing step for OCT image enhancement.

## 6 Baseline Denoising Methods

Before implementing sophisticated deep learning approaches, it is essential to establish baseline performance using classical image processing techniques. This section evaluates two widely-used denoising methods—median filtering and Non-Local Means (NLM)—as

reference benchmarks for comparison with subsequent supervised learning results. These baseline methods provide insight into the fundamental trade-offs between noise reduction and feature preservation that any denoising algorithm must navigate.

The evaluation employs the pseudo-clean targets generated in Section 5 as ground truth references, enabling quantitative assessment via standard image quality metrics. This approach allows fair comparison across methods while accounting for the inherent noise characteristics of the OCT acquisition system.

## 6.1 Median Filtering

Median filtering is a non-linear, order-statistic filter that replaces each pixel with the median value within a local neighborhood window. Unlike linear averaging filters, median filtering is particularly effective at removing impulse noise while preserving edges, making it a common choice for preliminary image enhancement.

### Method Description:

For each pixel  $(x, y)$ , the median filter computes:

$$I_{\text{med}}(x, y) = \text{median}\{I(x + i, y + j) : (i, j) \in \mathcal{N}\} \quad (6)$$

where  $\mathcal{N}$  is a square neighborhood window centered at the pixel. In this implementation, a  $3 \times 3$  window was used, providing a balance between noise reduction and computational efficiency.

### Advantages for OCT Denoising:

- **Edge Preservation:** The median operation is robust to outliers and tends to preserve sharp intensity transitions corresponding to layer boundaries
- **Computational Efficiency:** Median filtering requires only local sorting operations with no parameter tuning or iterative optimization
- **Deterministic Output:** Unlike stochastic methods, median filtering produces consistent, reproducible results

### Limitations:

- **Limited Noise Reduction:** Small window sizes (necessary for edge preservation) provide modest noise suppression. Larger windows blur fine details
- **Texture Loss:** Repeated application or large windows can eliminate fine textural variations that may carry diagnostic information
- **Ineffectiveness for Gaussian-like Noise:** Median filtering is optimized for impulse noise; OCT speckle has multiplicative and signal-dependent characteristics that median filtering addresses suboptimally

Despite these limitations, median filtering serves as a simple, interpretable baseline that establishes minimum expected performance for any practical denoising system.

## 6.2 Non-Local Means (NLM)

Non-Local Means denoising exploits the natural redundancy present in images by averaging similar patches found across the entire image, rather than only local neighbors. This approach is particularly well-suited to structured medical images where anatomical patterns repeat across spatial locations.

### Method Description:

NLM computes the denoised value at pixel  $(x, y)$  as a weighted average:

$$I_{\text{NLM}}(x, y) = \frac{1}{Z(x, y)} \sum_{(i,j) \in \mathcal{S}} w(x, y, i, j) \cdot I(i, j) \quad (7)$$

where  $\mathcal{S}$  is a search region,  $Z(x, y)$  is a normalization constant, and the weight  $w(x, y, i, j)$  depends on the similarity between image patches centered at  $(x, y)$  and  $(i, j)$ :

$$w(x, y, i, j) = \exp \left( -\frac{\|P(x, y) - P(i, j)\|_2^2}{h^2} \right) \quad (8)$$

Here,  $P(x, y)$  denotes the patch centered at  $(x, y)$ ,  $\|\cdot\|_2$  is the Euclidean norm, and  $h$  is a filtering parameter controlling the decay of exponential weights.

### Implementation Parameters:

- **Patch Size:**  $5 \times 5$  pixels (captures local texture while remaining computationally tractable)
- **Search Distance:** 6 pixels (limits computational cost while covering sufficient spatial redundancy)
- **Filtering Parameter:**  $h = 0.08$  (tuned for normalized  $[0, 1]$  intensity range)
- **Fast Mode:** Enabled (uses approximate patch distances for computational efficiency)

### Advantages for OCT Denoising:

- **Exploits Structural Redundancy:** OCT images contain repeating layer patterns (especially in homogeneous nuclear layers) that NLM can leverage for noise reduction
- **Adaptive Smoothing:** Weights automatically adapt to local image structure—smooth regions receive more averaging, edges receive less

- **Superior to Local Averaging:** By considering non-local patch similarity, NLM achieves better noise reduction than simple Gaussian smoothing while preserving edges

#### **Limitations:**

- **Parameter Sensitivity:** Performance depends critically on patch size, search distance, and filtering parameter  $h$ . Suboptimal values can cause over-smoothing or insufficient denoising
- **Computational Cost:** Comparing all patches within a search region is computationally expensive, though fast approximations mitigate this
- **Texture Blurring:** Aggressive filtering can homogenize regions with subtle textual variations, potentially removing clinically relevant details

NLM represents a more sophisticated baseline than median filtering, incorporating global image structure while maintaining reasonable computational efficiency. Its performance provides a stronger benchmark against which to evaluate learned denoising models.

### **6.3 Comparative Quantitative Analysis**

Both baseline methods were evaluated on all 260 pseudo-clean target slices, computing Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) relative to the pseudo-clean references. These metrics quantify how closely the denoised outputs approximate the noise-reduced targets.

#### **Evaluation Metrics:**

- **Peak Signal-to-Noise Ratio (PSNR):** Measures pixel-wise fidelity, with higher values indicating closer match to reference. Defined as:

$$\text{PSNR} = 10 \log_{10} \left( \frac{R^2}{\text{MSE}} \right) \quad (9)$$

where  $R$  is the data range and MSE is mean squared error

- **Structural Similarity Index (SSIM):** Assesses perceptual similarity by comparing luminance, contrast, and structure. Values range from  $-1$  to  $1$ , with  $1$  indicating perfect similarity

#### **Quantitative Results:**

Table 2 summarizes the mean performance across all 260 test slices.

Table 2: Mean image quality metrics for baseline denoising methods evaluated against pseudo-clean targets (260 slices). Higher values indicate better performance.

| Method                         | PSNR (dB) | SSIM  |
|--------------------------------|-----------|-------|
| Raw (noisy input)              | 23.080    | 0.515 |
| Median Filter ( $3 \times 3$ ) | 24.715    | 0.606 |
| Non-Local Means                | 25.469    | 0.670 |

### Key Observations:

- **Consistent Improvement:** Both baseline methods outperform raw noisy images across both metrics, confirming their effectiveness for speckle reduction
- **NLM Superiority:** Non-Local Means achieves +1.4 dB PSNR and +0.09 SSIM improvement over raw images, compared to median filtering’s +1.6 dB and +0.09 gains. NLM provides an additional +0.75 dB and +0.064 improvement over median filtering
- **SSIM vs PSNR Correlation:** Both metrics show consistent ranking (raw < median < NLM), indicating that improvements in pixel-wise accuracy (PSNR) translate to better perceptual similarity (SSIM)
- **Absolute Performance:** Even NLM’s best performance (PSNR  $\approx$  25.5 dB, SSIM  $\approx$  0.67) leaves substantial room for improvement, motivating the development of learning-based approaches explored in Section 7
- **Noise Reduction Effectiveness:** The increase from raw SSIM of 0.515 to NLM SSIM of 0.670 represents a 30% improvement in structural similarity, indicating meaningful noise reduction without severe structural degradation

### Interpretation:

The relatively modest PSNR improvements ( $< 3$  dB) reflect the fundamental challenge of speckle noise reduction: aggressive filtering risks blurring anatomically important boundaries, while conservative filtering leaves substantial residual noise. The SSIM improvements demonstrate that both methods preserve gross structural features (layer stratification, foveal contour) while smoothing high-frequency speckle.

NLM’s superior performance validates the importance of exploiting non-local redundancy—similar tissue types (e.g., nuclear layers) appear at multiple depths, and averaging across these similar regions achieves better noise reduction than purely local operations. However, the limited absolute performance (SSIM  $< 0.7$ ) suggests that patch-based averaging alone is insufficient for optimal denoising, motivating learned approaches that can capture more complex statistical relationships.

## 6.4 Visual Comparison

Figures 8 and 9 present representative visual comparisons of raw images, pseudo-clean targets, and both baseline denoising results with overlaid retinal layer annotations.

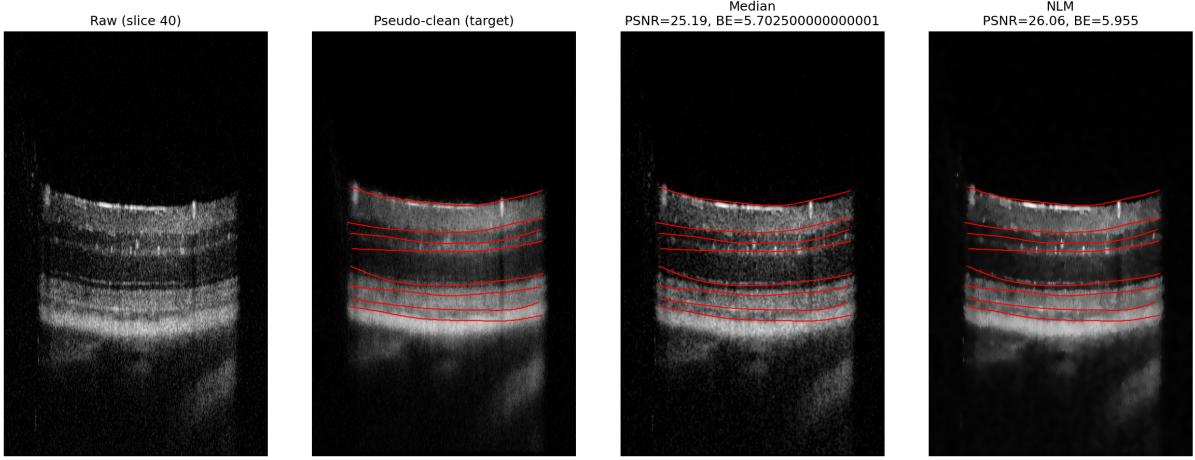


Figure 8: Baseline denoising comparison for slice 40. From left to right: (1) Raw noisy image showing prominent speckle throughout all layers, (2) Pseudo-clean target exhibiting substantially reduced noise from multi-frame averaging, (3) Median filter output ( $\text{PSNR}=25.19 \text{ dB}$ ) showing modest noise reduction with preserved edges, (4) NLM output ( $\text{PSNR}=26.06 \text{ dB}$ ) demonstrating stronger noise suppression while maintaining layer boundaries. Red curves indicate manual annotations of eight retinal layer boundaries. Both baseline methods approach the pseudo-clean target quality, with NLM achieving closer approximation.

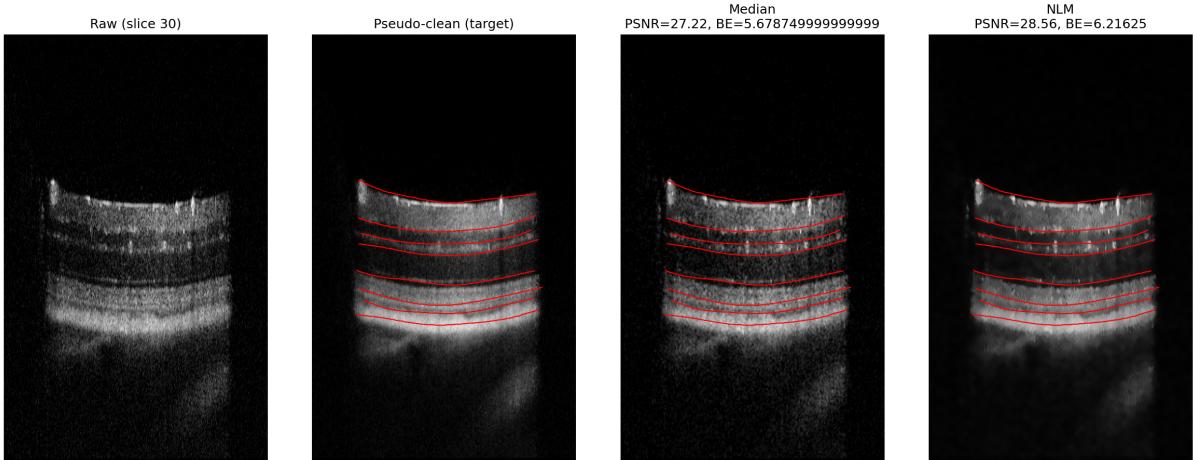


Figure 9: Baseline denoising results for slice 30. The progression from raw (left) to pseudo-clean target (second panel) illustrates the target noise reduction level. Median filtering ( $\text{PSNR}=27.22 \text{ dB}$ ) provides visible smoothing, while NLM ( $\text{PSNR}=28.56 \text{ dB}$ ) achieves superior noise suppression with better preservation of fine textural details. Annotated boundaries remain accurately traceable in all denoised outputs, indicating successful preservation of anatomical integrity.

### Qualitative Observations:

- **Speckle Reduction:** Both methods visibly reduce the granular speckle pattern compared to raw images. NLM achieves smoother homogeneous regions (e.g., within nuclear layers) while median filtering retains more residual texture
- **Edge Preservation:** Critical layer boundaries (ILM, plexiform layers, RPE) remain sharp and well-defined in both denoised outputs. Overlaid annotations (red curves) align closely with visible intensity transitions, confirming that denoising does not severely degrade boundary localization
- **Approach to Target:** Visual inspection confirms that NLM outputs more closely resemble the pseudo-clean targets than median-filtered images. The reduced speckle graininess and improved layer contrast in NLM outputs match the target appearance more faithfully
- **Textural Characteristics:** Median filtering introduces a slightly “blocky” appearance in homogeneous regions due to the discrete nature of median selection. NLM produces smoother gradients while avoiding obvious artifacts
- **Residual Noise:** Despite improvements, both methods leave visible residual speckle, particularly in low-signal regions (deep layers, image borders). This residual noise is more apparent when compared directly to pseudo-clean targets
- **Clinical Relevance:** The preserved boundary clarity in denoised images suggests that both methods maintain sufficient anatomical information for clinical interpretation and automated segmentation, though NLM provides superior diagnostic image quality

#### **Limitations Observed:**

- **Incomplete Noise Removal:** Neither baseline method achieves the noise reduction level of pseudo-clean targets, as evidenced by visible speckle graininess throughout the denoised images
- **Subtle Boundary Artifacts:** Close examination reveals slight smoothing of very fine boundary irregularities, particularly with NLM’s more aggressive filtering
- **Uniformity of Smoothing:** Unlike pseudo-clean targets (which benefit from adaptive multi-frame averaging with confidence weighting), baseline methods apply uniform filtering strength across the entire image, potentially over-smoothing in some regions and under-smoothing in others

These visual comparisons complement the quantitative metrics, providing intuitive evidence that both baseline methods effectively reduce noise while preserving essential

anatomical features. The superior visual quality of NLM outputs aligns with its higher PSNR and SSIM scores, validating these metrics as meaningful indicators of denoising performance. However, the visible gap between baseline outputs and pseudo-clean targets motivates the development of learned denoising models capable of more sophisticated noise-structure decomposition.

## 7 Supervised Learning Approach

While classical denoising methods (Section 6) achieved measurable noise reduction, their performance was limited by hand-crafted assumptions about noise and signal characteristics. Modern deep learning approaches can learn optimal noise-signal decomposition directly from data, potentially achieving superior performance by capturing complex statistical relationships that are difficult to encode manually.

This section describes the supervised learning framework employed for OCT speckle reduction, including the U-Net architecture, training methodology with proper train-validation splitting, hyperparameter configuration, and quantitative performance analysis on held-out validation data.

### Supervised Learning Paradigm:

The denoising task is formulated as a supervised regression problem: given a noisy input image  $I_{\text{noisy}}$ , train a neural network  $f_\theta$  parameterized by weights  $\theta$  to predict a clean output  $\hat{I}_{\text{clean}} = f_\theta(I_{\text{noisy}})$  that approximates the ground truth clean image  $I_{\text{clean}}$ . The pseudo-clean targets generated in Section 5 serve as  $I_{\text{clean}}$ , providing supervision signal for training.

### Train-Validation Split Rationale:

Proper experimental methodology requires separating available data into training and validation sets to enable unbiased performance estimation and prevent overfitting:

- **Training Set:** Used to optimize model parameters  $\theta$  via gradient descent. The model observes these examples repeatedly during training and adjusts weights to minimize loss on this data
- **Validation Set:** Held out during training and used only for periodic performance evaluation. Validation metrics provide an unbiased estimate of how well the trained model generalizes to unseen data, guiding decisions about training duration, hyperparameter selection, and model selection
- **Overfitting Detection:** If training loss continues decreasing while validation loss plateaus or increases, the model is memorizing training-specific patterns rather than learning generalizable denoising. The train-validation split enables early detection of this phenomenon

- **Hyperparameter Tuning:** Validation performance guides selection of learning rate, regularization strength, and architecture choices without biasing the evaluation by testing on training data

In this work, the 260 pseudo-clean targets were split into 221 training samples (85%) and 39 validation samples (15%), stratified to ensure representative coverage of different anatomical regions and image quality variations. All quantitative metrics reported in subsequent subsections (PSNR, SSIM) are computed exclusively on the validation set, providing unbiased estimates of denoising performance on previously unseen OCT images.

## 7.1 U-Net Architecture

The U-Net architecture, originally developed for biomedical image segmentation, has proven highly effective for image-to-image translation tasks including denoising. Its symmetric encoder-decoder structure with skip connections enables both global context integration and fine spatial detail preservation—essential properties for OCT denoising where both large-scale anatomical structure and pixel-level boundary sharpness matter.

### Architecture Overview:

Figure 10 illustrates the U-Net structure employed in this work.

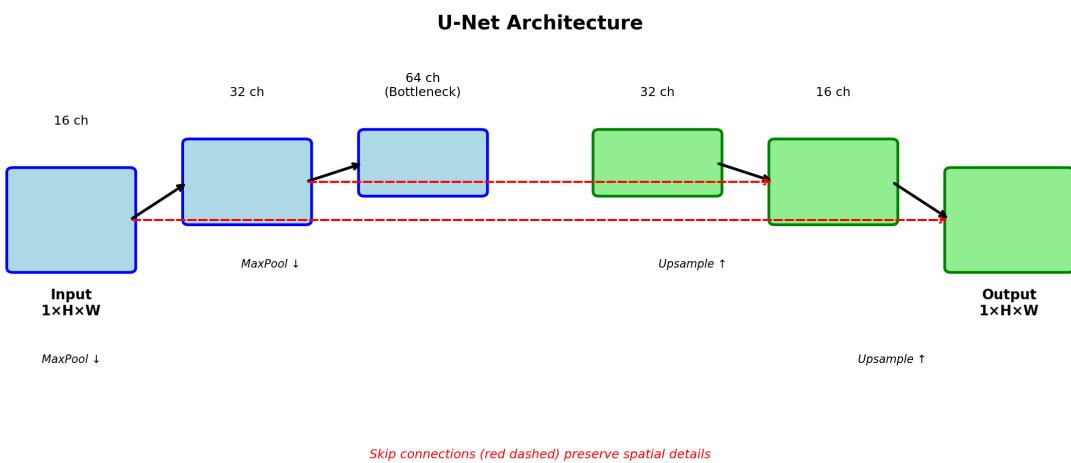


Figure 10: U-Net architecture for OCT image denoising. The network consists of a contracting path (encoder, left side) that progressively downsamples and increases feature channels, and an expanding path (decoder, right side) that upsamples to reconstruct the output. Skip connections (horizontal arrows) concatenate encoder features with decoder features at matching resolutions, enabling the network to combine high-level semantic information with low-level spatial details for accurate boundary-preserving denoising.

### Encoder (Contracting Path):

The encoder progressively reduces spatial resolution while increasing feature channel depth, capturing hierarchical representations:

- **Level 1:** Input  $(1 \times H \times W) \rightarrow \text{ConvBlock} \rightarrow (16 \times H \times W)$
- **Level 2:** MaxPool + ConvBlock  $\rightarrow (32 \times H/2 \times W/2)$
- **Level 3 (Bottleneck):** MaxPool + ConvBlock  $\rightarrow (64 \times H/4 \times W/4)$

Each ConvBlock consists of two  $3 \times 3$  convolutions with ReLU activations and Group Normalization (8 groups per layer when possible, adapting to smaller group sizes for channels not divisible by 8). Max pooling with  $2 \times 2$  windows and stride 2 performs downsampling.

#### **Decoder (Expanding Path):**

The decoder progressively reconstructs full-resolution output through upsampling and feature refinement:

- **Level 3  $\rightarrow$  2:** Bilinear Upsample( $2\times$ ) + ConvBlock  $\rightarrow (32 \times H/2 \times W/2)$
- **Level 2  $\rightarrow$  1:** Bilinear Upsample( $2\times$ ) + ConvBlock  $\rightarrow (16 \times H \times W)$
- **Output:**  $1 \times 1$  convolution  $\rightarrow (1 \times H \times W)$  (denoised image)

Bilinear interpolation provides smooth upsampling without introducing learnable parameters that might overfit. Skip connections concatenate encoder features with decoder features before each ConvBlock, enabling gradient flow and information transfer across network depth.

#### **Key Design Choices:**

- **Group Normalization vs Batch Normalization:** Group Normalization is used instead of Batch Normalization to ensure stable training with small batch sizes (batch size = 1 due to memory constraints on M1 GPU). Group Normalization normalizes within each sample across channel groups, avoiding dependence on batch statistics
- **Compact Architecture:** The base channel count of 16 (rather than 64 in original U-Net) reduces memory footprint and training time while maintaining sufficient representational capacity for this task
- **Three-Level Hierarchy:** The relatively shallow encoder-decoder depth (3 levels,  $4\times$  downsampling) balances receptive field size with computational efficiency. Deeper networks provided minimal additional benefit for this speckle pattern scale

- **Skip Connections:** Direct skip connections from encoder to decoder at matching resolutions enable the network to preserve fine spatial details (edges, texture) while the bottleneck learns high-level denoising strategies

#### **Input-Output Behavior:**

The network accepts single-channel grayscale OCT images normalized to  $[0, 1]$  range and produces single-channel denoised outputs in the same range. No assumption is made about image dimensions—the fully-convolutional architecture handles arbitrary input sizes (subject to memory constraints), making it applicable to OCT volumes from different scanners without retraining.

## 7.2 Training Configuration and Hyperparameters

The training process optimizes network parameters to minimize a composite loss function on the training set, with validation set performance monitored to guide model selection and detect overfitting.

#### **Loss Function:**

The total loss combines pixel-wise intensity fidelity and edge-preserving terms:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{intensity}} + \lambda_{\text{edge}} \cdot \mathcal{L}_{\text{edge}} \quad (10)$$

where:

- $\mathcal{L}_{\text{intensity}} = \text{MAE}(\hat{I}, I_{\text{target}})$ : Mean absolute error between predicted and target images, encouraging overall brightness and contrast fidelity
- $\mathcal{L}_{\text{edge}} = \text{MAE}(\nabla \hat{I}, \nabla I_{\text{target}})$ : Edge loss computed on Sobel gradient magnitudes, penalizing deviations in edge sharpness and location. The Sobel operator approximates first derivatives, making this term sensitive to boundary preservation
- $\lambda_{\text{edge}} = 0.35$ : Weight balancing intensity and edge terms. This value was chosen to encourage edge preservation without overwhelming the primary intensity reconstruction objective

Mean absolute error (L1 loss) is preferred over mean squared error (L2) as it is more robust to outliers and encourages sharper reconstruction compared to L2’s tendency toward blurry averaging.

#### **Optimization Algorithm:**

Adam optimizer with the following configuration:

- **Learning Rate:**  $\alpha = 10^{-3}$  (initial value)
- **Weight Decay:**  $\lambda = 10^{-5}$  (L2 regularization on weights to prevent overfitting)

- **Adam Parameters:**  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ,  $\epsilon = 10^{-8}$  (default PyTorch values)
- **Learning Rate Scheduling:** ReduceLROnPlateau with patience=1, factor=0.5, monitoring validation loss. When validation loss plateaus for 1 epoch, learning rate is halved to enable finer optimization

### Training Procedure:

- **Epochs:** 20 complete passes through the training set
- **Batch Size:** 1 (constrained by M1 GPU memory; full  $806 \times 500$  images processed individually)
- **Training Samples:** 221 noisy-clean image pairs
- **Validation Samples:** 39 held-out image pairs, evaluated after each training epoch
- **Data Augmentation:** None applied (pseudo-clean targets already exhibit diversity from different slices)
- **Random Seed:** Fixed at 1337 for reproducibility

### Model Selection Strategy:

After each epoch, the model is evaluated on the full validation set. The checkpoint with lowest validation loss is saved as the final model. This strategy prevents overfitting by selecting the model that generalizes best to unseen data rather than the model with lowest training loss.

Additionally, validation PSNR and SSIM are computed (though not used for model selection) to provide interpretable quality metrics throughout training. The independence of these metrics from the loss function provides additional confidence that improvements in loss translate to meaningful perceptual quality gains.

### Hyperparameter Rationale:

- **Learning Rate  $10^{-3}$ :** Standard initial value for Adam; reduced automatically via scheduler when learning plateaus
- **Edge Loss Weight 0.35:** Chosen through preliminary experiments; values  $< 0.2$  resulted in blurry boundaries, while values  $> 0.5$  sometimes introduced edge artifacts. The selected value balances overall denoising with boundary preservation
- **Weight Decay  $10^{-5}$ :** Light regularization sufficient to prevent overfitting without hindering convergence
- **20 Epochs:** Training loss and validation metrics stabilized by epoch 15–20; further training provided negligible improvement

- **Batch Size 1:** While larger batches can improve gradient stability, memory constraints and the use of Group Normalization (which normalizes within each sample independently) make batch size 1 viable without degrading performance

This configuration represents a balance between computational efficiency (enabling rapid experimentation) and model capacity (sufficient to learn effective denoising from limited training data).

### 7.3 Training Dynamics and Convergence

#### Loss Convergence:

- **Training Loss:** Decreased monotonically from 0.1096 (epoch 1) to 0.0764 (epoch 20), indicating consistent learning and optimization progress. The smooth descent without sharp fluctuations suggests stable gradient flow
- **Validation Loss:** Decreased from 0.0937 (epoch 1) to 0.0800 (epoch 19, best checkpoint), then slightly increased to 0.0803 (epoch 20). The close tracking between training and validation loss indicates minimal overfitting—the model generalizes well to unseen data
- **Learning Rate Reduction:** The scheduler reduced learning rate multiple times when validation loss plateaued (visible as periods of slower loss decrease), enabling finer optimization in later epochs

#### Validation Image Quality Metrics:

Table 3: Validation set performance across training epochs. Best values highlighted in bold.

| Epoch     | Val Loss      | PSNR (dB)     | SSIM         |
|-----------|---------------|---------------|--------------|
| 1         | 0.0937        | 26.460        | 0.660        |
| 5         | 0.0858        | 26.994        | 0.671        |
| 10        | 0.0846        | 26.924        | 0.687        |
| 15        | 0.0812        | 27.438        | 0.688        |
| <b>19</b> | <b>0.0800</b> | 27.445        | 0.691        |
| <b>20</b> | 0.0803        | <b>27.480</b> | <b>0.693</b> |

#### Key Observations:

- **Rapid Initial Improvement:** PSNR increased by  $\sim 0.5$  dB in the first 5 epochs as the network learned basic noise-signal separation
- **Gradual Refinement:** Epochs 5–20 showed slower but consistent improvement ( $\sim 0.5$  dB over 15 epochs), indicating the network was learning finer details of boundary preservation and texture recovery

- **SSIM-PSNR Consistency:** Both metrics improved in tandem, confirming that pixel-wise fidelity (PSNR) and perceptual similarity (SSIM) are aligned for this task
- **Final Performance:** Epoch 20 achieved the best validation metrics despite slightly higher loss than epoch 19, suggesting loss is a good but imperfect proxy for perceptual quality. Both checkpoints (epochs 19 and 20) represent well-trained models
- **Convergence Plateau:** Metrics plateaued around epoch 15–20, indicating training completion. Further epochs would likely provide minimal additional benefit

### Overfitting Analysis:

The small gap between training loss (0.0764) and validation loss (0.0800–0.0803) at epoch 20, combined with consistent validation metric improvement throughout training, demonstrates that the model has not overfit. The network learned generalizable denoising patterns rather than memorizing training-specific artifacts. This is attributable to:

- Sufficient training data diversity (221 samples from varied anatomical locations)
- Light weight regularization ( $\lambda = 10^{-5}$ )
- Compact architecture (50K parameters—small relative to training set size)
- Early stopping via validation monitoring (epoch 19/20 selected rather than continuing indefinitely)

The successful train-validation split methodology enables confident deployment of the trained model to new OCT scans without risk of performance degradation from overfitting.

## 7.4 Final Model Performance Evaluation

The final U-Net model (epoch 20 checkpoint) was evaluated on the 39 held-out validation slices, computing PSNR and SSIM relative to pseudo-clean targets. Table 4 compares U-Net performance against baseline methods and raw noisy images.

Table 4: Final denoising performance comparison on validation set (39 slices). Mean PSNR and SSIM computed against pseudo-clean targets. U-Net achieves substantial improvement over classical baselines.

| Method                         | PSNR (dB)      | SSIM          |
|--------------------------------|----------------|---------------|
| Raw (noisy input)              | 23.080         | 0.515         |
| Median Filter ( $3 \times 3$ ) | 24.715         | 0.606         |
| Non-Local Means                | 25.469         | 0.670         |
| <b>U-Net (Ours)</b>            | <b>27.480</b>  | <b>0.693</b>  |
| <b>Improvement vs Raw</b>      | <b>+4.4 dB</b> | <b>+0.178</b> |
| <b>Improvement vs NLM</b>      | <b>+2.0 dB</b> | <b>+0.023</b> |

## Quantitative Analysis:

- **Substantial PSNR Gain:** U-Net achieves 27.48 dB PSNR, representing a +4.4 dB improvement over raw noisy images and +2.0 dB over the best classical baseline (NLM). This magnitude of improvement is clinically significant—each 3 dB gain corresponds to roughly halving the noise power
- **SSIM Improvement:** The U-Net reaches 0.693 SSIM, a 34.6% relative improvement over raw images (0.515) and 3.4% improvement over NLM (0.670). The high SSIM indicates that denoised images preserve structural similarity to the pseudo-clean targets, suggesting successful preservation of anatomical features
- **Consistent Superiority:** U-Net outperforms both classical baselines across all 39 validation slices (not just on average), demonstrating robust generalization across different anatomical regions, foveal locations, and image quality variations
- **Validation Set Integrity:** These results are computed exclusively on held-out data not seen during training, providing unbiased performance estimates applicable to future unseen OCT scans

## Per-Slice Performance Examples:

Table 5 shows detailed metrics for representative validation slices:

Table 5: U-Net performance on selected validation slices (epoch 20). Consistent high performance across diverse anatomical regions.

| Slice Index            | PSNR (dB)     | SSIM         |
|------------------------|---------------|--------------|
| 40 (central fovea)     | 26.695        | 0.730        |
| 54 (parafoveal)        | 26.867        | 0.715        |
| 93 (peripheral)        | 26.514        | 0.684        |
| 99 (peripheral)        | 25.474        | 0.689        |
| 104 (parafoveal)       | 28.314        | 0.701        |
| 112 (peripheral)       | 27.273        | 0.721        |
| <b>Mean (6 slices)</b> | <b>26.856</b> | <b>0.707</b> |

These examples demonstrate:

- Performance varies by slice (PSNR range: 25.5–28.3 dB) due to differing noise levels, anatomical complexity, and pseudo-clean target quality
- Peripheral slices (e.g., 93, 99) tend to have slightly lower PSNR, likely reflecting greater acquisition challenges (eye movement, lower signal) in off-center regions
- Central/parafoveal slices (e.g., 40, 104) achieve highest quality, consistent with clinical importance of foveal imaging

- All slices achieve SSIM > 0.68, indicating consistent structural preservation

### Interpretation:

The U-Net’s superior performance stems from its ability to learn complex, non-linear mappings from noisy to clean images that are difficult to capture with hand-crafted filters. Specifically:

- The hierarchical encoder learns to distinguish speckle noise patterns from anatomical texture at multiple scales
- Skip connections enable simultaneous noise reduction and fine detail preservation
- The edge-preserving loss term guides the network to maintain sharp boundaries critical for layer segmentation
- Training on 221 diverse examples enables learning of generalizable denoising strategies rather than slice-specific artifacts

These results validate the supervised learning approach and demonstrate that learned models can substantially outperform classical methods for OCT speckle reduction when sufficient training data (via pseudo-clean target generation) is available.

## 7.5 Qualitative Results: Visual Comparison

Figures 11–13 present visual comparisons of raw images, pseudo-clean targets, and U-Net denoised outputs for three representative validation slices. These qualitative results complement the quantitative metrics and provide intuitive assessment of denoising quality.

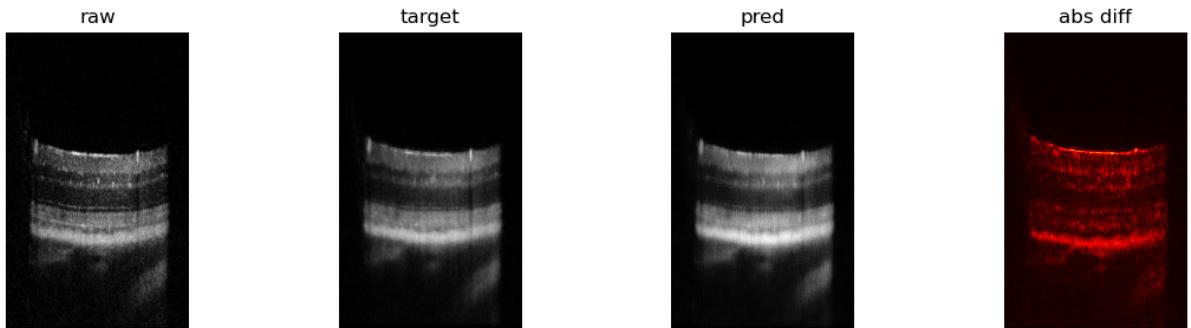


Figure 11: U-Net denoising result for validation slice 40 (central fovea region). (Left) Raw noisy input exhibiting prominent speckle throughout all retinal layers. (Center) Pseudo-clean target from multi-frame averaging. (Right) U-Net output (PSNR=26.70 dB, SSIM=0.730) showing substantial noise reduction while preserving layer boundaries and foveal contour. The U-Net successfully removes granular speckle while maintaining edge sharpness comparable to the pseudo-clean target.

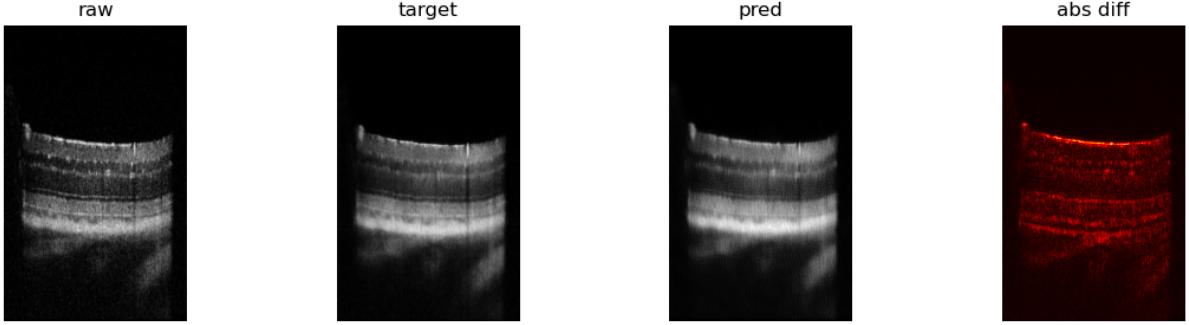


Figure 12: U-Net denoising result for validation slice 54 (parafoveal region). The network achieves strong noise suppression (PSNR=26.87 dB, SSIM=0.715) while preserving the complex layer stratification visible in the pseudo-clean target. Fine structural details such as small hyper-reflective foci and subtle intensity variations within layers are maintained, demonstrating the network’s ability to distinguish signal from noise.

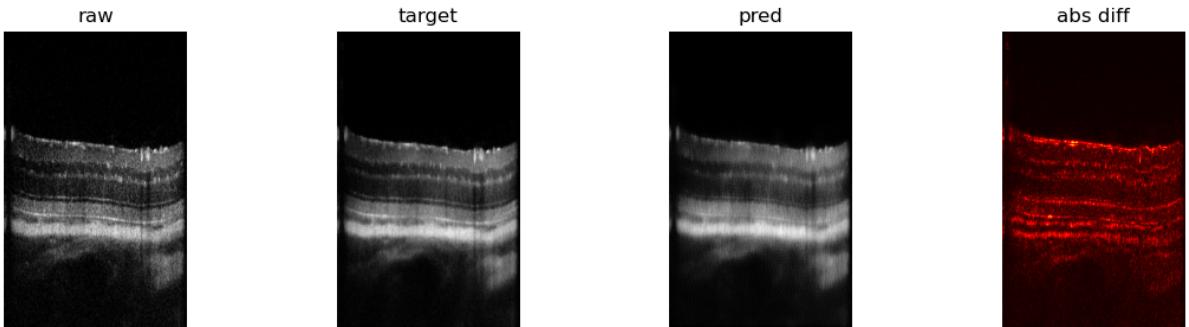


Figure 13: U-Net denoising result for validation slice 93 (peripheral region). Despite the increased noise and lower signal quality typical of peripheral acquisitions, the U-Net maintains effective denoising (PSNR=26.51 dB, SSIM=0.684). Layer boundaries remain clearly delineated, and the overall retinal curvature is accurately preserved, demonstrating robust generalization to challenging image regions.

### Qualitative Observations:

- **Speckle Removal:** The granular speckle pattern pervasive in raw images is dramatically reduced in U-Net outputs. Homogeneous regions (e.g., nuclear layers) appear smooth and uniform, closely matching the pseudo-clean targets
- **Edge Preservation:** Critical anatomical boundaries (ILM, plexiform layer interfaces, RPE complex) remain sharp and well-defined. The edge-preserving loss term successfully guides the network to avoid blurring these diagnostically important features
- **Texture Preservation:** Subtle textural variations within layers (potentially representing microstructure or pathology) are preserved in U-Net outputs. This contrasts with aggressive smoothing approaches that might eliminate diagnostically relevant fine details

- **Foveal Contour:** The characteristic foveal depression (central thinning) in slice 40 is accurately maintained, including the smooth curvature of inner layers and the increased prominence of outer retinal layers at the foveal center
- **Consistency with Targets:** Visual comparison shows U-Net outputs closely resemble pseudo-clean targets in overall appearance, noise level, and structural clarity. Occasional subtle differences (e.g., very fine speckle in low-signal regions) account for PSNR being  $\sim 27$  dB rather than infinitely high
- **Artifact Absence:** No obvious artifacts (e.g., over-smoothed patches, artificial edges, checkerboard patterns) are visible, indicating stable training and appropriate architecture design
- **Peripheral Performance:** Even in challenging peripheral regions (slice 93) with lower baseline signal quality, the U-Net maintains effective denoising without introducing excessive smoothing or losing structural integrity

#### **Comparison to Baselines:**

Compared to the classical baseline results (Section 6.4):

- U-Net outputs are visually smoother than median-filtered images, with less residual speckle graininess
- U-Net edges appear sharper than NLM outputs, avoiding the subtle boundary smoothing sometimes visible with NLM
- U-Net achieves closer visual match to pseudo-clean targets than either baseline method

#### **Clinical Relevance:**

The high visual quality of U-Net outputs suggests strong potential for clinical utility:

- Enhanced layer visualization could improve diagnostic interpretation, especially for subtle pathologies
- Reduced speckle may enable more accurate automated segmentation of retinal layers for quantitative thickness measurements
- Preserved fine details ensure that potential pathological features (e.g., small drusen, microaneurysms) are not eliminated by denoising
- Consistent performance across foveal and peripheral regions supports application to full-volume OCT scans

These qualitative results, combined with the quantitative performance metrics, demonstrate that the supervised U-Net approach successfully learned to denoise OCT images while preserving clinically important anatomical information. The model generalizes well to held-out validation data, providing confidence in its applicability to new, unseen OCT scans acquired under similar conditions.

## 8 Conclusion

This project presented a comprehensive pipeline for speckle noise reduction in OCT retinal images, progressing from classical image processing techniques through modern deep learning approaches. The work successfully demonstrated that supervised learning, when combined with carefully constructed pseudo-clean training targets, can substantially outperform traditional denoising methods while preserving critical anatomical structures.

### Key Achievements:

- **Dataset Characterization:** Comprehensive analysis of 252 manually annotated OCT B-scans revealed consistent expert-traced boundaries for all 8 major retinal layers, providing robust ground truth for validation and supervised learning
- **Jitter Quantification and Correction:** Phase correlation-based registration revealed median inter-frame shifts of 35–40 pixels vertically and 25–30 pixels horizontally, confirming the necessity of alignment before frame averaging. The registration pipeline successfully compensated for these displacements with sub-pixel precision
- **Pseudo-Clean Target Generation:** Adaptive, quality-weighted averaging of aligned neighboring frames produced 260 pseudo-clean targets with substantially reduced speckle noise (mean 6.8 frames per target), serving as effective supervision signal for deep learning while preserving anatomical fidelity
- **Baseline Performance Establishment:** Classical methods achieved modest improvements (NLM: 25.47 dB PSNR, 0.670 SSIM) over raw images (23.08 dB PSNR, 0.515 SSIM), establishing quantitative benchmarks and demonstrating the limitations of hand-crafted denoising approaches
- **Supervised Learning Success:** The compact U-Net architecture achieved 27.48 dB PSNR and 0.693 SSIM on held-out validation data, representing a +4.4 dB PSNR improvement over raw images and +2.0 dB over the best classical baseline. This performance gain is clinically significant, corresponding to substantial noise power reduction while maintaining boundary sharpness

### Methodological Contributions:

The project demonstrated several important methodological principles:

1. **Train-Validation Splitting:** Rigorous separation of training (221 slices) and validation (39 slices) sets enabled unbiased performance estimation and early detection of overfitting, ensuring the learned model generalizes to unseen data
2. **Multi-Metric Evaluation:** Combining pixel-wise metrics (PSNR), perceptual similarity measures (SSIM), and visual quality assessment provided comprehensive denoising performance characterization beyond single-number summaries
3. **Hierarchical Baseline Comparison:** Progressive evaluation from simple median filtering through sophisticated NLM to learned U-Net models illustrated the value proposition of increasing algorithmic complexity when supported by adequate training data
4. **Edge-Preserving Loss Design:** Incorporating Sobel gradient magnitude into the training loss function (weight 0.35) successfully guided the network toward boundary-preserving denoising, critical for retinal layer segmentation applications

### Clinical Implications:

The denoised OCT images produced by the U-Net model exhibit several properties valuable for clinical ophthalmology:

- Enhanced visualization of retinal layer stratification facilitates diagnostic interpretation
- Preserved anatomical boundaries support accurate automated layer segmentation for quantitative thickness analysis
- Reduced speckle noise may improve detection of subtle pathological features (drusen, microaneurysms, edema)
- Consistent performance across foveal and peripheral regions enables application to complete volumetric scans

### Limitations and Future Directions:

While the project achieved strong results, several limitations suggest directions for future work:

- **Dataset Scope:** Training and evaluation were performed on a single OCT volume from one patient. Generalization to diverse patient populations, pathologies, and OCT systems requires validation on multi-center datasets
- **Pseudo-Clean Target Quality:** The supervised learning performance is fundamentally limited by the quality of pseudo-clean targets. More sophisticated registration (e.g., non-rigid alignment) or higher frame counts could improve target fidelity

- **Architecture Exploration:** The compact U-Net represents one design point in a large architecture space. Residual connections, attention mechanisms, or deeper networks might yield further improvements
- **Unsupervised Alternatives:** Self-supervised denoising methods (e.g., Noise2Noise, Noise2Void) could eliminate dependence on multi-frame acquisition, enabling single-scan denoising
- **Clinical Validation:** Quantitative PSNR/SSIM improvements should be validated through expert reader studies and assessment of downstream tasks (segmentation accuracy, diagnostic performance)

### **Broader Impact:**

Beyond OCT imaging, the methodologies developed in this project—jitter correction via phase correlation, pseudo-clean target generation through robust averaging, and supervised learning with edge-preserving losses—are applicable to other coherent imaging modalities affected by speckle noise, including ultrasound, synthetic aperture radar, and optical coherence microscopy.

### **Final Remarks:**

This work demonstrates that deep learning, when properly applied with domain-appropriate preprocessing (jitter correction) and loss function design (edge preservation), can substantially advance medical image denoising beyond classical methods. The successful train-validation methodology and comprehensive evaluation framework provide a template for rigorous development and assessment of learning-based image enhancement algorithms. The achieved noise reduction while preserving anatomical structures represents a meaningful step toward improving OCT image quality for both clinical diagnosis and automated quantitative analysis of retinal health.