2026

Myelin segmentation in cross-circular polarization birefringence microscopy

BOSTON UNIVERSITY

COLLEGE OF ENGINEERING

Thesis

MYELIN SEGMENTATION IN CROSS-CIRCULAR POLARIZATION BIREFRINGENCE MICROSCOPY

by

JIAHE NIU

B.S., Boston University, 2024

Submitted in partial fulfillment of the

requirements for the degree of

Master of Science

2026

Approved by

First Reader

Irving J. Bigio, Ph.D.
Professor of Biomedical Engineering
Professor of Electrical and Computer Engineering
Professor of Physics
Professor of Medicine

Second Reader

Jerome C. Mertz, Ph.D.
Professor of Biomedical Engineering
Professor of Electrical and Computer Engineering
Professor of Physics

Third Reader

Brian Kulis, Ph.D.
Associate Professor of Electrical and Computer Engineering
Associate Professor of Systems Engineering

Fourth Reader

Darren Roblyer, PhD.
Professor of Biomedical Engineering
Professor of Electrical and Computer Engineering

Acknowledgments

I would like to thank my thesis committee for their guidance and support throughout this work. I am deeply grateful to Professor Irving Bigio, my advisor and committee chair, for his patience, encouragement, and steady feedback over the course of this thesis. I thank Professors Brian Kulis, Jerome Mertz, and Darren Roblyer for the time they invested, and for the thoughtful questions and suggestions that strengthened this work. Additionally, I extend my thanks to Professor Lei Tian for his guidance on the computational aspects of this work, Professor Xue Han for her encouragement and advice, and Professor Ray Nagem for his continued support throughout my graduate studies.

I am grateful to the members of the Biomedical Optics Lab for creating a supportive and collaborative environment. In particular, I thank Ting Xie, Anna Novoseltseva, and Meadow Bradsby for their weekly feedback, careful reading of drafts, and detailed suggestions on figures and clarity.

I also wish to acknowledge Boston University IT and Josh Bevan for their support with GPU and cluster resources; their assistance in troubleshooting access and environment issues kept my analysis workflow running smoothly.

Finally, I thank my friends and family for their love and patience throughout this process. I am especially thankful to Tejas Singh for proofreading and thoughtful feedback on my writing. I truly could not have done it without the unwavering support of everyone around me.

MYELIN SEGMENTATION IN CROSS-CIRCULAR POLARIZATION BIREFRINGENCE MICROSCOPY

JIAHE NIU

ABSTRACT

Quantifying myelin in three dimensions is critical for studying neurodevelopment, pathology, and repair. However, dense manual annotation of high-resolution microscopy z-stacks is often impractical. This thesis develops a scalable segmentation pipeline for Cross-Circular Polarization Birefringence Microscopy (CCP-BRM) brain volumes that reduces reliance on extensive manual labels.

We generate weak supervision using two interpretable heuristics: rolling-ball background subtraction and Tenengrad (squared-gradient) focus measure. These capture complementary aspects of the myelin signal to produce pseudo-label masks for in-focus fibers. We then train a context-aware 2.5D U-Net that leverages stacked neighboring slices to predict a center-slice mask. Increasing axial context from K=1 to K=5 yields modest improvements in probabilistic segmentation quality (AUPRC) on a development crop.

We further fine-tune the selected model using a small set of densely annotated slices to better match the manual target definition. When evaluated on a physically distinct volume, fine-tuning improves AUPRC relative to both the pretrained model and the heuristic baseline. Pixel-wise overlap metrics remain moderate, reflecting the inherent challenge of segmenting sparse, thin structures in this modality. Finally, we demonstrate the downstream utility of these masks for conservative myelin volume density estimation and qualitative 3D visualization. Overall, this work establishes a practical workflow for CCP-BRM myelin fiber segmentation and highlights the data requirements for stronger generalization in future 3D myelin analysis.

Contents

List of Tables

List of Figures

List of Abbreviations

AUPRC	Area Under the Precision-Recall Curve
CCP-BRM	Cross-Circular Polarization Birefringence Microscopy
CNS	Central Nervous System
EM	Electron Microscopy
FCN	Fully Convolutional Network
GT	Ground Truth
MFD	Myelin Fiber Density
MFV	Myelin Fiber Volume
PSF	Point Spread Function
RB	Rolling Ball (background subtraction)
ROI	Region of Interest
TEN	Tenengrad (gradient energy)

Chapter 1

Introduction

1.1 The myelin connectome and the need for 3D quantification

Myelin is a critical component of the Central Nervous System (CNS), providing the electrical insulation necessary for the rapid propagation of action potentials along axons. Beyond its classical role in saltatory conduction, myelin provides essential metabolic support to neurons and is dynamically regulated by neuronal activity (Nave & Werner, 2014; Saab & Nave, 2017; Stadelmann et al., 2019).

While white matter tracts have been extensively studied, the complex organization of myelinated axons within gray matter remains less understood (Timmler & Simons, 2019; Axer, Amunts et al., 2011). In gray matter, myelin is sparse and exhibits intricate branching geometries that are critical for local circuit connectivity (Micheva et. al, 2016). Disruptions in myelin integrity are a hallmark of numerous neurological conditions, including multiple sclerosis, Alzheimer's disease, and age-related cognitive decline (Depp et al, 2023; Duncan & Radcliff, 2016).

To understand these pathologies, it is not sufficient to view myelin as a 2D texture. We must understand its three-dimensional topology, or the "myelin connectome". Accurately mapping this network in 3D is essential for quantifying myelin volume and connectivity, metrics that serve as sensitive biomarkers for neurodegeneration and repair (Wainberg, 1962; Boshkovski et al., 2021). However, capturing these structures at high resolution across significant tissue volumes presents a formidable imaging challenge.

1.2 Modality gap in myelin segmentation

Electron microscopy (EM) is widely regarded as the gold standard for imaging myelin ultrastructure because it resolves myelin boundaries with high contrast and nanoscopic structural detail (Peters, 2009; Peters & Sethares, 2002). Consequently, several deep-learning tools for myelin segmentation and 3D reconstruction, such as AxonDeepSeg and DeepACSON, were developed and validated primarily on EM datasets (Zaimi et al, 2018; Abdollahzadeh et al., 2021). However, EM is difficult to scale to large 3D volumes due to extensive sample preparation, limited field of view, and labor-intensive acquisition and reconstruction.

Confocal microscopy provides optical sectioning and high contrast when fluorescent labels are available, but it also becomes time-intensive for large fields of view and relies on specialized instrumentation and labeling protocols (Monsma & Brown, 2012). Additionally, signal degradation due to photobleaching limits the duration and spatial extent of fluorescence imaging, making large-scale 3D acquisitions challenging. As a result, scaling either EM or confocal microscopy to large-area 3D imaging remains challenging.

These constraints motivate a complementary approach: a label-free, widefield method that can image larger regions efficiently while still providing myelin-specific contrast.

1.3 Widefield Birefringence Microscopy (CCP-BRM)

These limitations motivate the use of widefield birefringence microscopy (CCP-BRM) for scalable, label-free myelin imaging (Blanke et al., 2021). Myelin is intrinsically anisotropic due to the ordered wrapping of lipid bilayers around the axon. This property allows it to be imaged label-free using polarized light. Because CCP-BRM is label-free, the same tissue section can be imaged first, then subsequently stained and co-registered with the birefringence maps to provide complementary biological context.

In this thesis, we utilize a Cross-Circular Polarizer (CCP-BRM) configuration. This

setup is cost-effective and can be implemented on standard widefield microscopes. Crucially, as a widefield technique, CCP-BRM acquires the entire field of view in a single camera exposure. In our setup, the raw images are 2960×2960 pixels with a sampling of $0.10625\,\mu$m per pixel, yielding an effective field of view of approximately $314\,\mu\text{m} \times 314\,\mu\text{m}$. This offers a speed advantage of orders of magnitude over point-scanning confocal systems and EM, making it feasible to image large regions of brain tissue in a fraction of the time (Blanke & Gray et al., 2024).

1.3.1 The "Pseudo-Sectioning" Phenomenon

A major challenge in widefield microscopy is the lack of inherent optical sectioning. Light from above and below the focal plane typically creates a prohibitive background haze. However, CCP-BRM exhibits a useful “pseudo-sectioning” behavior because its contrast is phase-dependent rather than purely intensity-based. In a cross-circular polarizer (CCP) configuration, image intensity depends on the phase retardation introduced by birefringent structures such as myelin. When a myelin fiber segment is near the focal plane, it produces a sharp, high-contrast signal. When that same structure is out of focus, the contribution does not simply appear as a uniformly blurred version of the in-focus image; instead, out-of-plane rays accumulate different phase shifts, and their contributions may partially cancel or disperse into a diffuse, low-frequency background (Novoseltseva et al., 2024).

This effect is visible in Figure 1. As the focal plane sweeps through z, the same fiber segment transitions from a broad, low-frequency haze (out of focus) to a compact, high-frequency structure (in focus), and then fades again as the focus moves past it. Different portions of the fiber come into focus at different depths due to tilt and curvature of the myelin fiber through the z-axis (Figure 1b–c). This observation motivates the core computational goal of this thesis: isolating per-pixel in-focus myelin signal from the raw z-stack can serve as a form of digital optical sectioning and enable downstream 3D reconstruction.

The residual background in CCP-BRM is fundamentally different from the additive

out-of-focus blur encountered in fluorescence microscopy. In fluorescence imaging, out-of-focus light contributes additively as an intensity signal, which is well modeled by linear convolution with a point spread function (PSF) (Sibarita et al., 2005; Wallace et al., 2001). In CCP-BRM, the measured signal arises from interference of polarized light; out-of-plane structures can perturb the measured phase in a non-additive manner, producing structured "phase noise" rather than a simple blurred replica of the in-focus content (Novoseltseva et al., 2024). As a result, standard PSF-based deconvolution methods are poorly matched to this setting. Instead, separating true in-focus myelin from this complex background requires a non-linear, semantic approach.

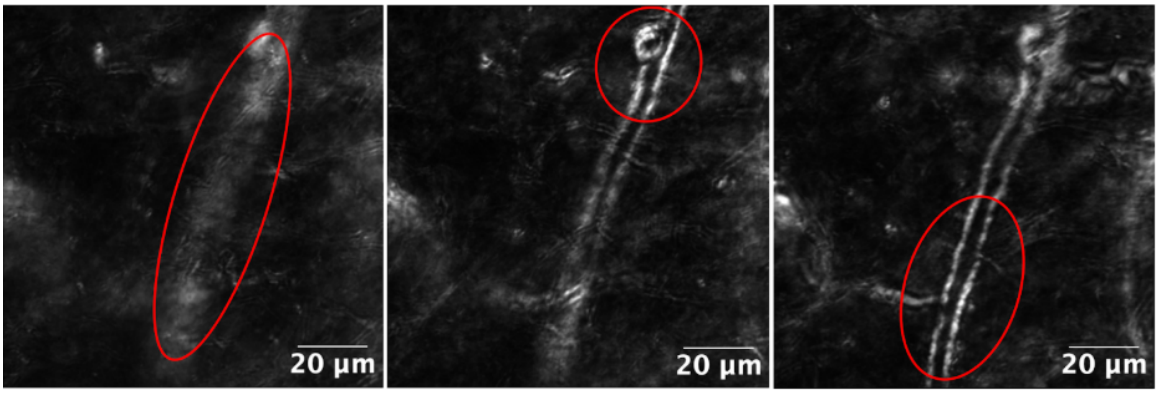

Figure 1: CCP-BRM z-stack illustrating pseudo-sectioning of myelinated fibers. The red-circled fiber segment appears diffuse and low-contrast when out of focus (a, z=10), becomes sharply localized with higher contrast near its focal depth (b, z=20), and then fades as focus moves past it while a different red-circled segment becomes visible at a deeper plane (c, z=30), consistent with fiber tilt/curvature through z.

1.4 The Computational Challenge: Scaling via Weak Supervision

Extracting the in-focus signal is non-trivial due to the phase noise and low-intensity in-focus fibers described above. Simple thresholding based on pixel intensity is insufficient, as it cannot reliably distinguish between dim in-focus fibers and background noise. To address this challenge, we introduce a weakly supervised pipeline that eliminates the need

for dense manual labels. We leverage two interpretable heuristics that are well-suited to the morphological diversity of myelin:

- **Tenengrad (Sobel Gradient):** Computes the gradient magnitude in x and y directions, making it highly sensitive to the sharp edges of transverse myelin rings (Yeo et al., 1993).

$$G_x = I * S_x, \quad G_y = I * S_y$$

$$\text{TEN}_{\text{response}} = G_x^2 + G_y^2$$

- **Rolling Ball (background subtraction):** Estimates a smooth, low-frequency background per slice using the rolling-ball (ImageJ/Fiji 'Subtract Background') algorithm and subtracts it from the original image (with negative values clipped), which enhances fiber-like signal especially longitudinal segments that may not exhibit strong edge responses (Sternberg, 1983).

By taking the union of the binary masks produced by these two heuristics, we create a high-recall training target that exposes the deep learning models to the full range of myelin orientations.

Chapter 2

Background and Related Work

2.1 Existing Approaches to Myelin Quantification

Manual and Semi-Automated Stereology. Historically, myelin quantification relied on manual stereology or semi-automated 2D thresholding (Arous et al., 2011; Gonsalvez et al., 2019). While accurate for small regions, these methods are labor-intensive and subject to inter-rater variability. Furthermore, they restrict analysis to disjointed planes that cannot fully capture the complex 3D topology described in Section 1.1. Simple automated alternatives, such as global intensity thresholding, are similarly ineffective for CCP-BRM. The wide dynamic range of in-focus myelin results in dim fibers that overlap with the background noise floor, making it impossible for a single threshold to reliably distinguish signal from haze.

Limitations of Existing Deep Learning Tools. In recent years, deep learning toolboxes such as AxonDeepSeg (Zaimi et al., 2018) and DeepACSON (Abdollahzadeh et al., 2021) have automated myelin segmentation EM. However, these models are trained on images where myelin consistently appears as a distinct, high-contrast dark ring against a light background. They do not generalize to CCP-BRM data, where myelin appears as bright, birefringent structures with variable intensity. Unlike the uniform cross-sections in EM, CCP-BRM signals manifest as complex geometries ranging from long, linear fiber segments to circular cross-sections, depending on the fiber's orientation relative to the imaging plane.

2.2 Deep Learning in Biomedical Segmentation

The U-Net standard. Fully convolutional networks (FCNs), especially the U-Net architecture (Ronneberger et al., 2015), are widely used in biomedical segmentation. The encoder-decoder design with skip connections preserves fine spatial detail, enabling precise localization of thin structures in tasks such as cell segmentation and neurite tracing. 2D vs. 3D strategies. Applying deep learning to volumetric microscopy data typically involves choosing between 2D, 3D, or "2.5D" modeling (Zhang et al. 2020).

3D U-Net. In principle, 3D U-Nets can exploit full volumetric context. In practice, they are often computationally expensive and can overfit when dense 3D labels are limited. They can also be awkward for volumes with strongly anisotropic sampling, where the voxel spacing in z is much larger than the in-plane sampling (Wang et al. 2025; Valanarasu et al. 2020). In CCP-BRM, we sample at approximately 0.106 μm per pixel in x-y but 1.0 μm between slices. This anisotropy directly affects what a 3D network learns, because standard 3D kernels operate on the discrete voxel grid rather than physical distance. For example, under our sampling a $3 \times 3 \times 3$ kernel spans roughly $0.32 \times 0.32 \times 3.0$ μm in physical space, so it mixes information over a much larger extent in z than in x–y. This mismatch can make it harder to learn consistent features and may encourage overly smooth behavior along z.

2.5D U-Net. A common compromise is 2.5D segmentation, where the network takes a small stack of adjacent slices as input but predicts the mask on the center slice (Siddique et al., 2020; Kumar et al. 2024). This approach adds axial context to resolve ambiguities (e.g., whether a faint structure is part of a continuous fiber) while keeping the efficiency and strong in-plane modeling of 2D convolutions. Because the z-slices are provided as separate input channels (rather than processed with 3D convolutions), the network can learn how much to trust each z-offset differently. This design allows for the use of axial context in an anisotropically sampled volume without imposing an isotropic 3D receptive field.

2.3 Weak supervision and pseudo-labeling

Label scarcity bottleneck. Applying U-Net-style models to new microscopy modalities is often limited by the need for dense manual labels, which are time-consuming to produce and require expert annotation (Siddique et al., 2020).

Focus-based weak supervision. Li et al. (2023) showed that classical focus operators (e.g., local variance-style measures) can generate dense pseudo-labels in widefield z-stacks, enabling networks to learn to predict focus-derived targets without manual annotation. In this framing, each voxel receives a continuous focus score, and the network is trained to reproduce this score field as a proxy for segmenting in-focus structure.

Relevance to CCP-BRM myelin. While using focus-based targets is directly relevant, CCP-BRM myelin introduces additional difficulty: fibers are thin, low-contrast, and orientation-dependent, with "tram-track" sheaths on the order of 0.5 to 2 μm. These properties can challenge generic focus metrics and motivate adapting the pseudo-label design to myelin-specific cues.

2.4 Summary and motivation for this thesis

This review identifies three intersecting barriers to scalable myelin quantification: the domain gap preventing the use of existing EM tools (Section 2.1), the geometric limitations of full 3D networks on anisotropic data (Section 2.2), and the prohibitive cost of dense manual annotation (Section 2.3). Collectively, these constraints create a bottleneck for scalable analysis. In this thesis, we address this by combining interpretable focus and contrast heuristics (Tenengrad and rolling-ball background subtraction) to generate pseudo-labels. We then train a 2.5D U-Net to denoise and regularize these targets using limited axial context, utilizing small-scale manual fine-tuning for final alignment and evaluation.

Chapter 3

Methods

3.1 Dataset

This study utilizes high-resolution volumetric microscopy data obtained via Cross-Circular Polarization Birefringence Microscopy (CCP-BRM). Unlike standard brightfield or fluorescence imaging, CCP-BRM exploits the intrinsic optical anisotropy (birefringence) of the myelin sheath. When imaged under cross-circular polarization, myelin fibers appear as high-contrast, bright structures against a darker background. The data are z-stack images of translucent neural tissue with densely packed myelin fibers. Despite the high contrast provided by the birefringence signal, segmentation remains challenging due to the out-of-focus light and complex 3D fiber crossing events where multiple fibers overlap at different depths.

3.1.1 Image Acquisition and Dimensions

The raw data comprises 3D volumes with lateral dimensions of 2960×2960 pixels and an axial depth of 30 slices.

Voxel Anisotropy. A defining characteristic of this dataset is its high degree of spatial anisotropy, resulting from the acquisition settings:

- **Lateral Sampling (xy)**: 0.10625 μm per pixel
- **Axial Sampling (z)**: 1.0 μm per slice

This geometry results in an aspect ratio of approximately 1:9.4. While the lateral resolution

is sufficient to resolve individual myelin sheaths, the relatively large physical gap between z-slices necessitates segmentation approaches that can infer structural continuity across disjointed focal planes.

3.1.2 Dataset Organization

To ensure rigorous evaluation of model generalization, we organized the data into two physically distinct volumes:

- **Training Volume (Stack A):** The primary $2960 \times 2960 \times 30$ pixel volume used for heuristic pseudo-label generation and model pretraining. A subset of this volume ("Dev-Crop", $512 \times 512 \times 30$ pixels) was manually annotated to tune heuristic parameters and fine-tune the deep learning model.
- **Generalization Volume (Stack B):** A separate, held-out $2960 \times 2960 \times 30$ pixel volume acquired from a distinct tissue section. This volume was reserved exclusively for final testing. No data from Stack B was used during hyperparameter tuning or training. A "Test-Crop" ($512 \times 512 \times 30$ pixels) was extracted from this volume and densely annotated (central slices 11-17) to serve as the test set for assessing the pipeline's performance on unseen data.

3.2 Weakly Supervised Pseudo-Labeling

To enable large-volume pretraining without manual 3D labels, we develop a weakly supervised pseudo-labeling pipeline that produces slice-wise binary masks using two interpretable heuristics: rolling-ball background subtraction (RB) and Tenengrad gradient energy (TEN). These heuristics capture complementary appearance cues, specifically background-corrected intensity for RB and local high-frequency structure for TEN. Their union provides a high-recall approximation of in-focus myelin fibers that is suitable for supervising a learning-based model.

All pseudo-label parameters, including thresholds, pooling window size, and morphology settings, were selected only on the manually annotated development crop and then frozen for full-stack inference. This ensures a strict separation between pseudo-label design and downstream evaluation.

3.2.1 Motivation and Task Definition

The goal of pseudo-labeling is to identify the in-focus myelin fibers that appear as sharp and localized birefringence signals near their focal depth. This task differs from classical autofocus methods that select a single focal plane. Within a single slice, some fibers may be sharply resolved, others may appear blurred, and some may be entirely absent due to their orientation, curvature, or depth within the tissue volume.

Because manual labeling of in-focus fibers across full 3D stacks is prohibitively slow, we seek a scalable approach that produces high-recall but imperfect labels suitable for pretraining a segmentation network. Deep learning can later refine these labels using spatial context, axial context, and regularization. The resulting problem is therefore to generate binary masks that identify candidate in-focus myelin pixels using only slice-wise information, with the expectation that later stages of the pipeline will improve precision using learned representations.

3.2.2 Candidate Pseudo-Label Generators and Screening Procedure

We evaluated types of heuristic generators:

1. dense focus measures that produce a per-pixel focus score based on local texture or gradient information, and

2. a background-subtraction approach implemented using a rolling-ball filter, which enhances fiber-like structures by removing smooth low-frequency background.

Focus-measure screening. Inspired by Li et al. (2023), we evaluated six classical autofocus operators spanning derivative- and statistic-based families. These included the Brenner and Tenengrad gradient measures, two Laplacian-based variants, the Vollath autocorrelation metric, and a variance-based operator (GLVN). Each operator was applied using 3×3 derivative kernels followed by local mean pooling with window sizes $w \in \{3,5,7,9,11,15,21\}$. For every operator and pooling size, we swept 64 quantile-based thresholds to binarize the continuous response map. Thresholds were selected by maximizing Dice on training slices within the dev-crop, and performance was subsequently evaluated on four held-out slices to reduce overfitting.

Rolling-ball background subtraction. RB estimates a smooth background for each slice using a morphological rolling-ball filter. Subtracting this background highlights bright, structured signals that correspond to myelin. After threshold optimization, RB consistently produced spatially coherent masks and did not require post-processing. This screening process established the performance range and limitations of classical operators. Gradient-based operators often fragmented dim structures, while RB occasionally included bright non-fiber background regions. These observations guided the selection of two complementary methods.

3.2.3 Selected Generators and Complementarity

Two generators were retained for large-scale pseudo-labeling:

1. **Rolling-ball background subtraction (RB).** This method achieved the strongest single-generator performance with Dice = 0.476, precision = 0.509, and recall = 0.446. RB recovered many faint longitudinal fibers that were missed by gradient-based methods.

2. **Tenengrad gradient energy (TEN).** TEN was robust across window sizes and effectively detected sharp boundaries of transverse myelin. It achieved Dice = 0.459,

precision = 0.473, and recall = 0.445.

RB and TEN emphasize different aspects of myelin structure. Because the rolling-ball background subtraction uses a radius of 6 pixels, it sometimes suppresses small transverse myelin profiles whose bright rings fall within the size of the background-estimation kernel. As a result, these circular cross-sections can be partially or fully removed. In contrast, RB performs well on longitudinal fibers, which extend over many pixels and therefore survive the background subtraction (red circles in Figure 2). TEN provides the opposite behavior, giving strong responses on sharp transverse boundaries but fragmenting dim longitudinal fibers (blue circles in Figure 2). Their complementary behavior motivates combining them.

- The intersection has higher precision (0.682) at the cost of lower recall (0.335).
- The union has higher recall (0.556) with moderate precision (0.420).

We use the union mask (RB $\cup$ TEN) for pretraining because, at this stage, higher recall and broader coverage of in-focus fibers are prioritized over precision. A learning-based model can later suppress false positives using spatial and axial context. Figure 2 provides an example of RB, TEN, and the union mask.

3.2.4 Final Thresholding, Post-Processing, and Frozen Parameters

For each slice, each generator produces a continuous response map. This map is then binarized using a global threshold selected on the development crop. The final hyperparameters are:

- **TEN:** Window size $w = 5$, threshold $\tau_{TEN} = 0.00644355$, followed by a 3×3 opening and a 3×3 closing (one iteration each).
- **RB:** Rolling-ball radius $r = 6$, threshold $\tau_{RB} = 0.0265$, with no post-processing.

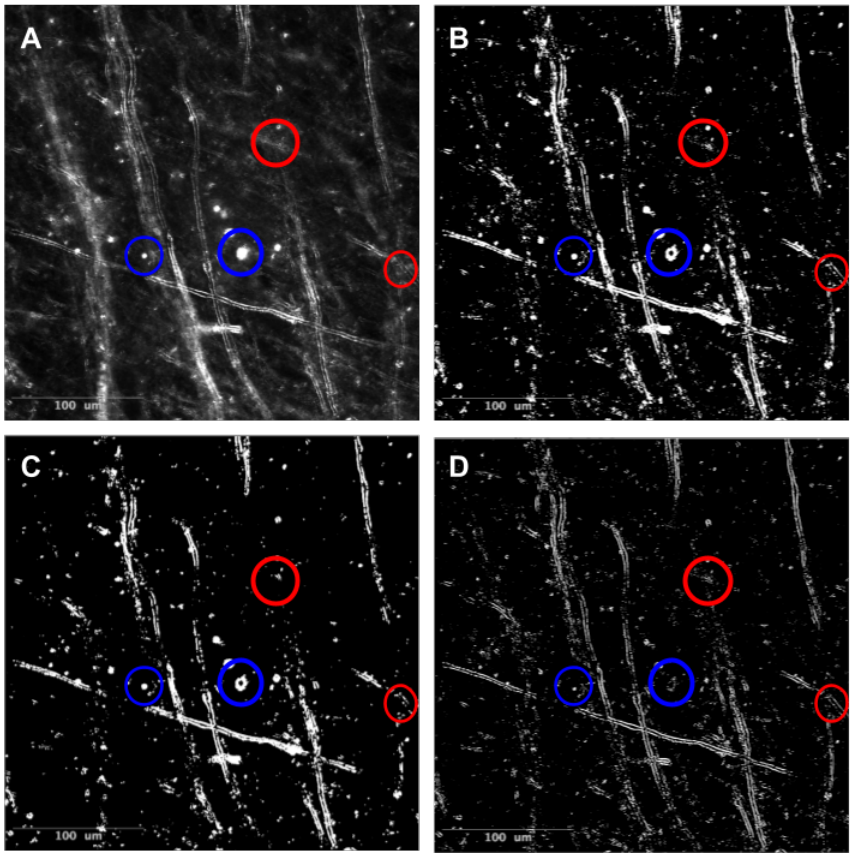

Figure 2: Example pseudo-label outputs on at z-plane 18 (1-indexed). Pseudo-labels were generated on the full 2960×2960×30 stack using thresholds tuned on the manually annotated Dev-Crop. Top left: original CCP-BRM slice. Top right: union pseudo-labels formed by taking the pixelwise union of the Rolling Ball and Tenengrad masks. Bottom left: Tenengrad pseudo-label. Bottom right: Rolling Ball pseudo-label. Red circles indicate regions where RB captures longitudinal fibers that are missed by TEN, while blue circles highlight sharp transverse fiber boundaries that are detected by TEN but attenuated in the RB mask.

3.2.5 Crop-to-Full Consistency Check

To confirm that crop-level parameter tuning transfers cleanly to full-volume inference, we compared TEN masks generated in two ways: (1) by running TEN on the full normalized volume and then cropping to the Dev-Crop region, and (2) by running TEN only on the Dev-Crop. The two masks were nearly identical, with Dice = 0.9984 and disagreement fraction of 5.210^{-5}. Minor differences occurred only near the boundaries due to padding effects. This confirms that the pseudo-label generator behaves consistently across both cropped and full-stack settings and is not sensitive to the region of application.

3.3 Pretraining Dataset Construction

3.3.1 Full-Stack Pseudo-Label Inference

Using the fixed thresholds and parameters defined in Section 3.2, we generated pseudo-labels for the entire $2960 \times 2960 \times 30$ training volume. For each slice, we computed the RB and TEN masks and took their union to obtain a high-recall foreground estimate. This produced a complete binary pseudo-label map for the full volume.

Across the stack, the fraction of pixels labeled as myelin is summarized in Table 1. RB and TEN capture different subsets of the fiber signal, and their union provides the highest overall foreground coverage, which is desirable for weak supervision. The intersection is included only to characterize complementarity between the two heuristics and is not used in model training.

Table 1: Myelin pixel percentages for all pseudo-label methods.

Method	Myelin Pixel Percent
RB	2.16
TEN	2.71
Intersection	1.26
Union	3.60

3.3.2 Patch Extraction and Leakage Avoidance

To construct a large and diverse training set, we extracted overlapping patches of size $256 \times 256 \times 30$ from the full-volume pseudo-labeled stack. A stride of 128 pixels was used along both x and y, resulting in substantial overlap that increases dataset size and enables effective data augmentation.

Because the development crop contains manual annotations used later for fine-tuning and evaluation, it is critical to prevent any form of leakage into the pretraining stage. We therefore excluded all patches whose spatial footprint overlapped the development region, expanded by a 64-pixel margin in all directions. This ensures that the pretraining dataset contains no pixels that are ever manually evaluated, even indirectly.

After applying this exclusion rule, the final pretraining dataset contained 442 patches.

3.3.3 Dataset Split and Effective Sample Count

The patch dataset was divided into 354 training patches and 88 validation patches (an 80/20 split). Each 30-slice patch contributes multiple training examples, depending on the axial context size K used in the segmentation models. Only slices with complete context can serve as valid center slices, which results in 30 valid centers for $K = 1$, 28 for $K = 3$, and 26 for $K = 5$. Table 2 summarizes the dataset size for each context value K.

Table 2: Valid center slices and total examples per K.

K	Valid Centers	Training Patches	Validation Patches
1	30	10,620	2,640
3	28	9,912	2,464
5	26	9,204	2,288

3.4 Segmentation Models

We trained a family of U-Net based models to segment in-focus myelin fibers from CCP-BRM slices. The primary goal is to learn a mapping that corrects errors in the heuris-

tic pseudo-labels and improves generalization to unseen tissue. We evaluated both a 2D baseline that uses a single slice as input and context-aware 2.5D models that incorporate neighboring slices to resolve depth-dependent ambiguity. All models were trained on the pseudo-labeled patches described in Section 3.3 and were evaluated using manual annotations on the development crop for model selection.

3.4.1 2D U-Net Baseline (K = 1)

The baseline model is a standard 2D U-Net with an encoder-decoder structure and skip connections between corresponding layers. The encoder contains four downsampling stages. Each stage consists of two 3×3 convolutional layers with ReLU activation and Group Normalization, followed by 2×2 max pooling. Feature channels double at each stage, reaching 1024 channels in the bottleneck. The decoder mirrors the encoder through 2×2 transposed convolutions for upsampling, concatenation with encoder features, and two 3×3 convolutional layers to refine the output. A spatial dropout layer is applied in the bottleneck to improve robustness to noise in the pseudo-labels. The input to this model is a single axial slice and the output is a single-channel probability map of the same spatial resolution.

3.4.2 Context-Aware 2.5D U-Net (K = 3 and K = 5)

To incorporate information from neighboring slices, we extended the baseline to a 2.5D model. The network architecture is identical, but the input consists of K consecutive slices stacked as channels, centered on the target slice. The model predicts the mask for the center slice only. The overall architecture is illustrated in Figure 3.

This design provides axial context while avoiding the drawbacks of full 3D convolutions, which are less appropriate for anisotropic CCP-BRM data. The slice spacing in z is approximately ten times larger than the in-plane pixel size, so applying isotropic 3D kernels would combine information across mismatched physical scales. The 2.5D formulation allows the model to assign different weights to different z-offsets and to learn how faint or

ambiguous structures evolve across neighboring slices.

We evaluated $K = 3$ and $K = 5$ to study how increasing the amount of context influences segmentation performance.

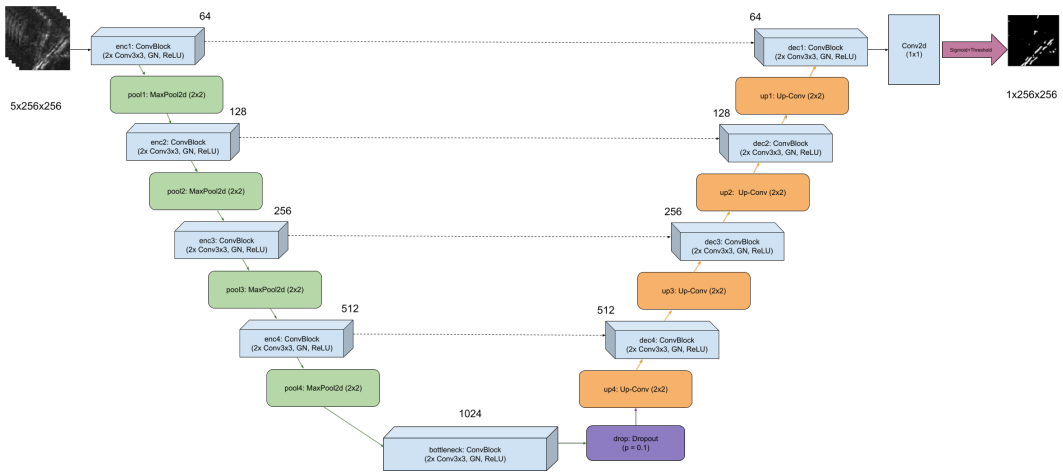

Figure 3: The 2.5D U-Net Architecture. The model takes 5 depth slices as a multi-channel 2D input ($5 \times 256 \times 256$). It uses a standard U-Net encoder-decoder structure with Group Normalization and skip connections. The network outputs a single-channel prediction that is processed via Sigmoid activation and thresholding to create a binary segmentation mask for the center slice.

3.4.3 Optimization and Loss Function

All models were trained using a weighted combination of binary cross-entropy with logits and soft Dice loss. The total loss L is:

$$L = 0.5 \times L_{\text{Dice}} + 1.0 \times L_{\text{BCE}}$$

The Dice component L_{Dice} encourages spatial overlap with the target mask, while the cross-entropy term L_{BCE} stabilizes training under class imbalance. Models were optimized using AdamW with a learning rate of 3×10^{-4} and weight decay of 1×10^{-4}. We applied

gradient clipping with a maximum norm of 1.0 to stabilize training under noisy pseudo-labels. Training proceeded for 10 epochs for all K values.

3.4.4 Data Augmentation

To improve generalization and reduce overfitting, we applied on-the-fly data augmentation that preserved the physical properties of the imaging process. Specifically, we used random 90 degree rotations and horizontal and vertical flips. These augmentations were applied consistently to the input slices and the corresponding target mask.

3.4.5 Training Protocol

For each K value, models were trained on the pseudo-labeled patches from Section 3.3. The same preprocessing, normalization, and augmentation pipeline was used across all experiments.

3.5 Model Selection and Fine-Tuning Protocol

3.5.1 Development Set and Evaluation Procedure

Model selection was performed using the manually annotated development crop described in Section 3.1.2. This crop was used only for selecting hyperparameters and evaluating intermediate models, and not for pseudo-label generation or pretraining.

Given that myelin fibers occupy only a small fraction of the total voxels in CCP-BRM volumes, we utilized the Area Under the Precision-Recall Curve AUPRC as our main evaluation metric. This metric summarizes the trade-off between precision and recall across all possible decision thresholds, providing a robust measure of segmentation quality for imbalanced data.By integrating performance over the full range of probabilities, AUPRC ensures that the evaluation focuses on the accurate detection of the fiber signal rather than the correct classification of the abundant background.

All models were evaluated using AUPRC on the same set of center slices to ensure a fair comparison across different context sizes. For completeness, we also report the Dice score, precision, and recall at a fixed operating threshold.

3.5.2 Model Selection: Choice of Axial Context K

We compared three context configurations: $K = 1$ (2D baseline), $K = 3$, and $K = 5$ (2.5D models). All models used the same U-Net backbone and training protocol; the only difference was the number of input slices provided to the network. Increasing K consistently improved AUPRC on the development crop. The $K = 5$ model achieved the highest ranking performance and produced more stable probability maps, particularly for dim and partially visible fibers. Dice values changed only slightly across K, which is expected since Dice is sensitive to threshold choice and boundary alignment. Based on the overall improvements in AUPRC and qualitative stability, the $K = 5$ model was selected for fine-tuning and downstream evaluation.

3.5.3 Fine-Tuning on Manual Annotations

After pretraining on pseudo-labels, the selected $K = 5$ model was fine-tuned using manual annotations from the development crop. Fine-tuning was conducted to reduce systematic noise in the pseudo-labels and improve consistency with the manually annotated examples.

To avoid spatial leakage within the crop, the 512×512 region was divided into four non-overlapping 256×256 quadrants. Three quadrants were used for fine-tuning and one quadrant was held out as a small validation region for model selection. Fine-tuning used the same preprocessing and data augmentation as pretraining. The learning rate was reduced to 1×10^{-4}, and gradient clipping was applied with a maximum norm of 1.0. The checkpoint with the highest validation AUPRC was selected for evaluation on the held-out Stack B volume.

3.6 Experimental Setup for Generalization (Stack B)

To evaluate how well the model generalizes to new tissue, we used a second CCP-BRM volume, referred to as Stack B. This stack has the same spatial dimensions as the training volume ($2960 \times 2960 \times 30$), but it was acquired from a different tissue section. As a result, it contains independent variations in illumination, background structure, and fiber organization that were not seen during pretraining or fine-tuning.

From Stack B, we extracted a $512 \times 512 \times 20$ sub-volume (Test-Crop) for quantitative evaluation. Although the raw stack contains 30 slices, we limited analysis to a contiguous region where tissue is present and fibers are in focus. The remaining slices primarily contained background and were excluded.

Manual annotations were produced for the central slices of Test-Crop ($z = 11$ to 17, seven slices in total). These slices were used only for evaluation. Predictions from the context-aware model were generated using the same preprocessing and normalization applied during training, with Stack B normalized using its own global intensity statistics.

For 2.5D models, neighboring slices were included as input to provide axial context, but quantitative metrics were computed exclusively on the manually annotated center slices. Probabilities were optionally converted into binary masks using a fixed threshold of 0.9 for volumetric density analyses, which prioritizes high-confidence fiber predictions and suppresses low-level background activations.

This setup simulates a realistic deployment scenario in which the global intensity distribution of a new scan is known, but no manual labels are available for adaptation. It also provides a controlled way to assess the model's ability to transfer to physically distinct tissue.

3.7 3D Reconstruction and Volumetric Quantification

3.7.1 Myelin Fiber Density and Volume Calculation

We quantified myelin directly from thresholded binary segmentation masks on the native (non-interpolated) CCP-BRM stack. Each z-slice is a 2D image, and consecutive slices are sampled at 1.0 μm spacing. Myelin fiber density (MFD) was computed as the fraction of pixels labeled as myelin within a region of interest (ROI), aggregated across all slices:

$$\text{MFD} = \frac{\text{number of myelin pixels in ROI across the stack}}{\text{total number of pixels in ROI across the z-stack}}$$

To convert this dimensionless density into an absolute physical volume, we multiplied by the physical volume of the ROI. With in-plane sampling of 0.10625 μm per pixel in x and y and axial slice spacing of 1.0 μm in z, the ROI volume is:

$$\text{ROI volume} = (N_x \times 0.10625) \times (N_y \times 0.10625) \times (N_z \times 1.0)\ \mu m^3$$

where N_x and N_y are the ROI dimensions in pixels and N_z is the number of included slices. The total myelin fiber volume (MFV) is then:

$$\text{MFV} = \text{MFD} \times (\text{ROI physical volume})$$

All quantitative measurements were computed on the native anisotropic grid. We avoid z-interpolation for quantification because interpolation can introduce artificial structure between slices and inflate volume estimates by creating fiber-like signal that is not directly supported by the acquired data.

3.7.2 3D Visualization and Interpolation

CCP-BRM stacks are highly anisotropic (dz = 1.0 μm compared to 0.10625 μm in-plane), which produces stair-step artifacts in 3D renderings. For qualitative visualization only, we

thresholded predicted probability maps at $\tau = 0.9$ and upsampled only along the z-axis to an isotropic 0.10625 μm grid using linear interpolation in Fiji/ImageJ. The interpolated volumes were then rendered in MATLAB. Interpolation was used only for visualization, and all density and volume calculations used the native acquisition grid.

Chapter 4

Results

4.1 Effect of axial context on Dev-Crop segmentation performance

On the Dev-Crop, we compared axial context sizes $K = 1, 3$, and 5 on the same set of center slices ($z = 2$ to 27). Because myelin pixels are sparse, we used AUPRC as the primary model-selection metric.

Increasing axial context yielded a modest but consistent improvement in probabilistic segmentation quality. AUPRC increased from 0.3824 ($K = 1$) to 0.4017 ($K = 3$) and 0.4233 ($K = 5$) (Table 3). In contrast, the best Dice score over a threshold sweep changed only slightly (0.4785 to 0.4840), which is expected because Dice is computed after binarization and is sensitive to small boundary shifts.

At the best Dice operating point (threshold = 0.8 for all K), larger context shifted the precision-recall tradeoff toward higher precision and lower recall. Precision increased from 0.4419 ($K = 1$) to 0.4816 ($K = 5$), while recall decreased from 0.5217 to 0.4863. The predicted foreground fraction also decreased with K and became better calibrated to the ground truth (0.0214 to 0.0183, compared to GT 0.0181). Based on the highest AUPRC and improved foreground calibration, we selected $K = 5$ for subsequent experiments and fine-tuning.

Table 3: Dev-Crop ($z = 2$ to 27) comparison of axial context size K. AUPRC is computed over continuous prediction scores across all pixels. Best Dice is the maximum Dice achieved over a fixed threshold sweep of the predicted probability maps (best threshold = 0.8 for all K).

K	AUPRC	Best Dice	Precision	Recall	Pred FG frac
1	0.3824	0.4785	0.4419	0.5217	0.0214
3	0.4017	0.4814	0.4555	0.5104	0.0203
5	0.4233	0.4840	0.4816	0.4863	0.0183

4.2 Generalization to Held-out Volume (Stack B)

To evaluate generalization to unseen tissue, we tested the best-performing 2.5D U-Net ($K = 5$) on a physically distinct CCP-BRM volume (Stack B) under two training conditions: (i) trained on union pseudo-labels only, and (ii) initialized from that model and fine-tuned on Dev-Crop manual annotations. Stack B was normalized using its own global intensity statistics. We extracted a $512 \times 512 \times 20$ evaluation crop and computed metrics on the manually annotated central region (7 slices).

Qualitative comparison. Figure 4 shows representative inference results on Stack B (slice $z = 13$). Both models detect the dominant in-focus fiber bundle. Compared to the pretrained-only model, the fine-tuned model produces more localized and thinner predictions, and it more consistently preserves the paired bright boundaries of the myelin sheath rather than merging them into a thick band. The fine-tuned probability map also shows more continuous activation along faint fibers without a large increase in diffuse background activation.

Quantitative performance. Table 4 summarizes segmentation performance on the 7 manually annotated Stack B slices. AUPRC is computed from the continuous prediction scores, while Dice, precision, and recall are computed after binarization. For the U-Net models, probability maps were thresholded at 0.8. For the heuristic baseline, binary masks were generated using thresholds frozen from the Dev-Crop.

The fine-tuned 2.5D U-Net ($K = 5$) achieved the best overall performance on Stack

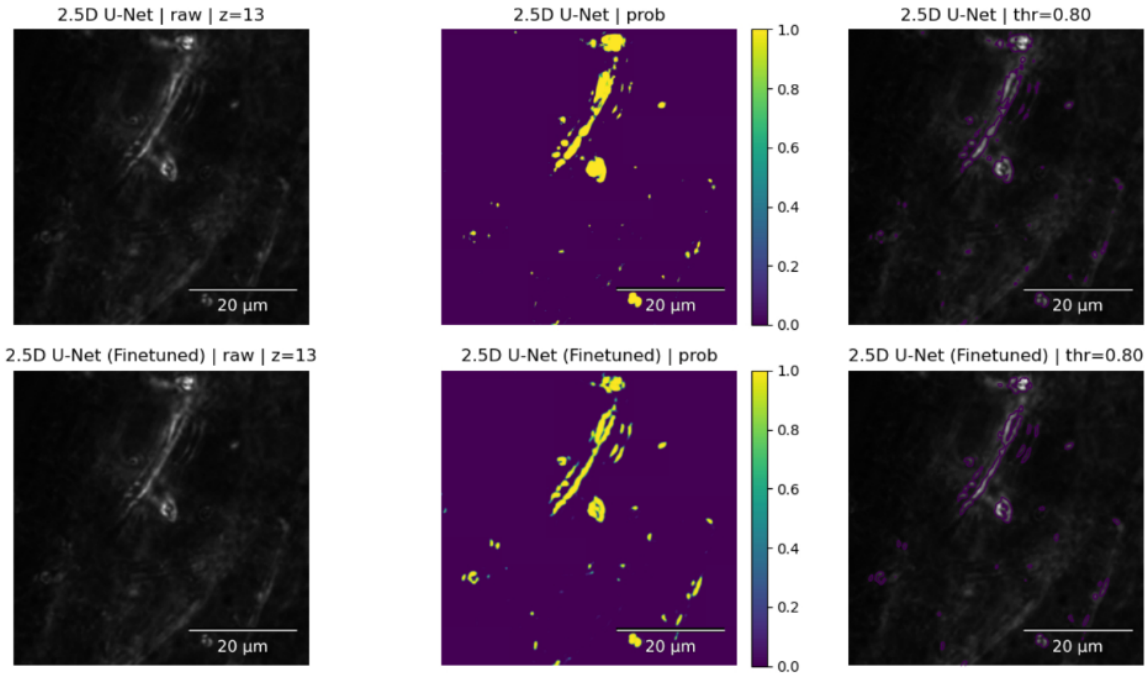

Figure 4: Qualitative comparison on Stack B (z = 13) between the pre-trained 2.5D U-Net ($K = 5$) and the fine-tuned 2.5D U-Net ($K = 5$). Left column: raw CCP-BRM slice. Middle columm: predicted probability map. Right column: thresholded mask ($thr = 0.80$) overlaid on the raw image.

B (AUPRC = 0.427; Dice = 0.466), improving over the pretrained-only model (AUPRC = 0.332; Dice = 0.406). This corresponds to a +0.095 absolute AUPRC gain (+28.6% relative). At the operating threshold of 0.8, fine-tuning improved both precision and recall (0.449 and 0.484 versus 0.397 and 0.414), indicating improved recovery of true fibers without a large increase in false positives.

For additional context, the heuristic union baseline achieved AUPRC = 0.382 and Dice = 0.409 on Stack B. The fine-tuned model exceeded this baseline by +0.045 AUPRC (+11.8% relative) and also achieved higher Dice, indicating improved agreement with manual annotations on unseen tissue.

Table 4: Stack B segmentation performance on 7 manually annotated slices. AUPRC is computed from continuous scores over all pixels. Dice, precision, and recall are computed after binarization. U-Net predictions are binarized with a fixed threshold of 0.8. The heuristic baseline uses frozen Dev-Crop thresholds defined in Section 3.2.4.

Method	AUPRC	Dice	Precision	Recall
Heuristic Baseline (Union)	0.382	0.409	0.362	0.470
2.5D U-Net (K = 5)	0.332	0.406	0.397	0.414
2.5D U-Net (K = 5) Fine-tuned	0.427	0.466	0.449	0.484

4.3 3D Volumetric Density Quantification

For volumetric quantification, we thresholded model probability maps at $thr = 0.9$ to prioritize precision and obtain a conservative, high-confidence estimate of myelin fiber density and volume.

Density agreement on annotated slices. Across the 7 manually annotated slices (slices 10˜16), the predicted myelin fiber density was 0.0195 compared to the ground truth 0.0201 (absolute difference 0.0006). Slice–wise densities followed similar trends, with the largest deviations occurring in the lowest–density slices. Pearson correlation across slices was $r = 0.69$ ($n = 7$). Bland–Altman analysis showed negligible systematic bias (-0.0006)

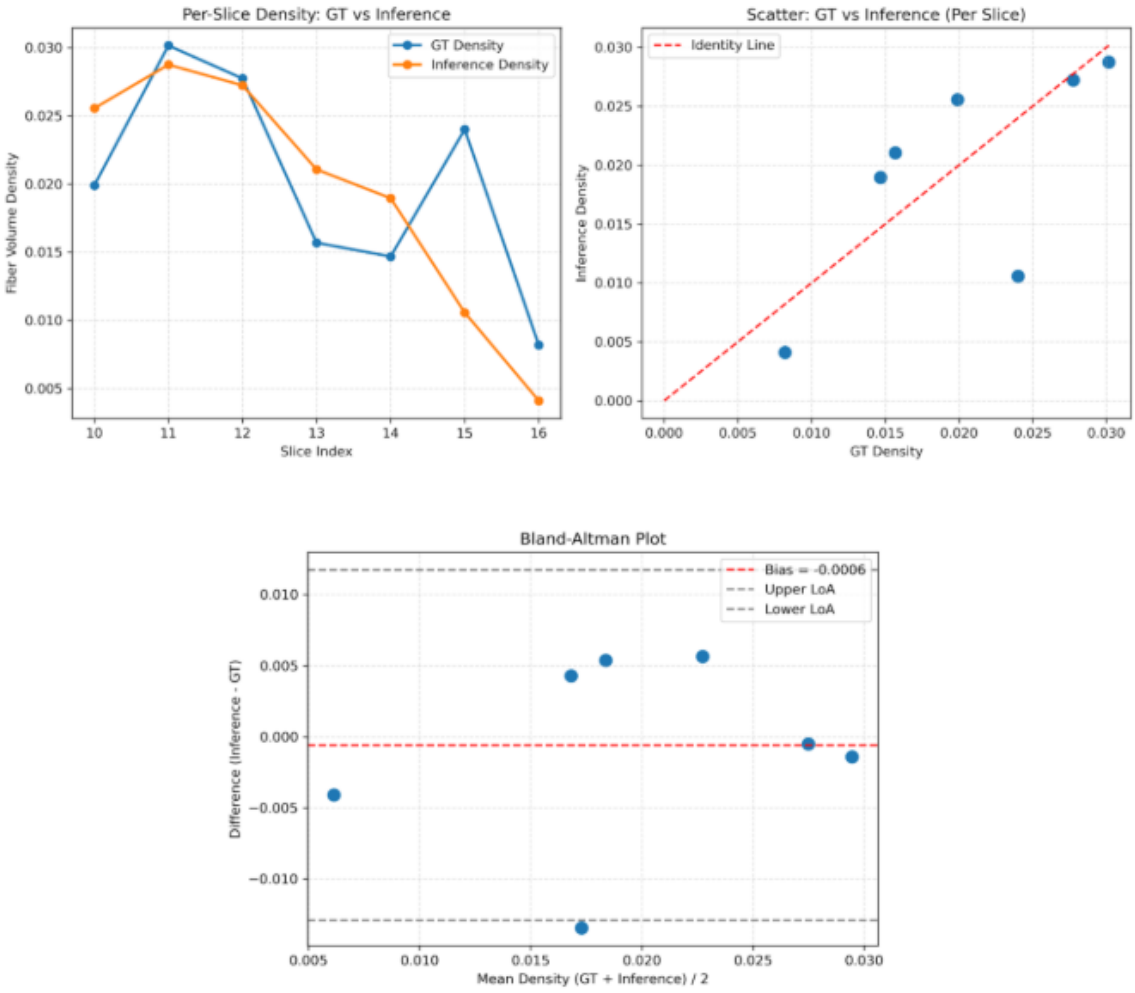

Figure 5: Per–slice myelin fiber density agreement between manual ground truth (GT) and model inference on Stack B (annotated slices 10–16). Top left: slice–wise myelin fiber density (fraction of fiber pixels) for GT and the thresholded prediction mask plotted across slice index. Top Right: inferred density versus GT density per slice with the identity line (y = x). Bottom: Bland–Altman plot showing the slice–wise difference (Inference – GT) versus the mean density; the red dashed line indicates the mean bias and gray dashed lines indicate the 95% limits of agreement.

with limits of agreement $[-0.0129, 0.0117]$, suggesting that density estimates are broadly stable across slices in this region.

Total fiber volume (annotated region). We computed myelin fiber volume from the native anisotropic grid as described in Section 3.7.1. For the annotated $512 \times 512 \times 7$ region ($54.4 \times 54.4 \times 7.0$ μm), the predicted myelin fiber density was 0.0195, yielding a total predicted myelin fiber volume of 403.8 μm^3. This value represents the volume of high-confidence segmented myelin within the annotated portion of the Stack B crop.

4.4 3D Rendering

We visualized Stack B predictions in 3D using volume rendering of the predicted segmentation volume (Figure X). Although CCP-BRM is acquired anisotropically (dz = 1.0 μm), the predicted masks show sufficient slice-to-slice consistency to support z-axis interpolation for visualization (Section 3.7.2). The resulting 3D rendering highlights extended longitudinal fibers and branching patterns that are difficult to interpret from individual 2D slices alone. This visualization is intended for qualitative assessment of spatial continuity and for downstream morphology exploration, not for quantitative measurements.

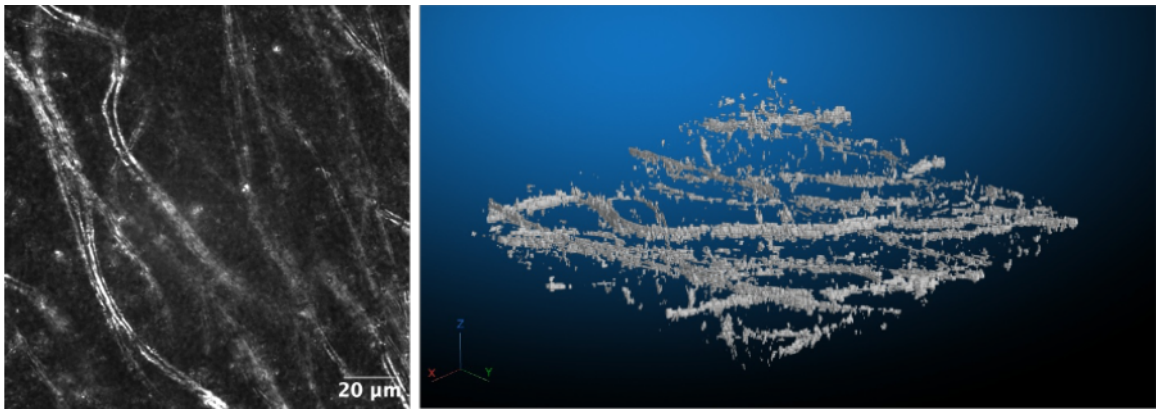

Figure 6: Stack B qualitative visualization of myelin fibers in 2D and 3D. Left: raw CCP-BRM image from the central 1024×1024 px region at $z = 15$. Right: 3D volume rendering of the corresponding region generated from the model's predicted segmentation mask after z-interpolation for visualization.

Chapter 5

Discussion

This thesis developed a weakly supervised pipeline for segmenting in-focus myelin fibers in CCP-BRM z-stacks, motivated by the high cost and ambiguity of dense manual annotation for thin structures. The approach combines heuristic pseudo-labels (RB and TEN) with a context-aware 2.5D U-Net, followed by fine-tuning on a small set of manual annotations. Overall, the results suggest that weak supervision can provide a useful initialization for learning in this modality, but reliable agreement with manual annotations and robust transfer to new volumes still depends strongly on additional labeled data and broader validation.

Weak supervision: coverage versus correctness. RB and TEN capture different image cues, and their union increases coverage of candidate fiber signal. However, union labels also introduce systematic noise, including background speckle and thickened or merged structures. On Stack B, the model trained only on pseudo-labels did not consistently outperform the heuristic baseline, indicating that pseudo-labels can scale training signal but do not by themselves guarantee alignment with the manual annotation target.

Axial context helped modestly, while fine-tuning mattered most. Increasing axial context from $K = 1$ to $K = 5$ improved AUPRC on the Dev-Crop, but the effect size was limited and Dice changed only slightly. The largest improvement came from fine-tuning on manual annotations. On Stack B, fine-tuning increased AUPRC relative to the pretrained-only model and improved overlap at the fixed operating threshold. Qualitatively, fine-tuning reduced over-segmentation and produced more localized predictions, including

clearer separation of paired sheath boundaries in representative examples. These qualitative gains should be interpreted cautiously, since they may reflect calibration to the limited labeled region rather than fully transferable morphology.

Generalization remains the primary open question. Although Stack B is a physically distinct tissue section, evaluation used only seven annotated slices within one crop. This provides evidence of transfer, but it is not sufficient to characterize robustness across different tissue regions, imaging conditions, or depth-dependent variation. Because the task involves sparse, thin structures and anisotropic sampling, performance estimates from small labeled subsets may have high variance.

What would strengthen the conclusions. The clearest next step is expanding manual annotations in a way that directly targets generalization. This includes adding additional held-out stacks from different tissue sections, annotating multiple crops per stack at different depths, and reporting uncertainty using confidence intervals or bootstrap resampling. Multi-annotator agreement would also help quantify label ambiguity for thin structures. In addition, the weak supervision scheme could be made less brittle by treating RB and TEN outputs as noisy constraints, for example by using the intersection as high-confidence positives and treating union-only regions as uncertain rather than forcing a single hard binary target (Fan et al. 2022).

Model and optimization improvements. We used a standard U-Net with minimal tuning to keep the study focused on supervision and axial context. A small hyperparameter sweep over learning rate, batch size, weight decay, and threshold selection could also improve calibration and generalization. Potential future work includes evaluating U-Net variants that are known to perform well on thin structures, such as ASFU-Net, KiU-Net, and attention/shape-aware U-Nets (Yue et al., 2024; Valanarasu et al. 2020; Hussain & Shouno, 2024; Cao & Cheng, 2025). In addition, different losses that better target sparse boundaries could be used including Boundary Loss (Kervadec et al., 2018; Zheng et al.,

2025). Further validation on diverse datasets and with multiple annotators is necessary to establish broader applicability (Griebel et al., 2023). Comparing model predictions to expert annotations from multiple raters would help quantify the inherent ambiguity in myelin fiber labeling. Aligning EM-derived measurements on the same samples would additionally establish the reliability of downstream quantitative estimates.

Toward orientation fields and connectome-relevant representations. Connectomics typically benefits more from vector-field or skeleton-based representations than from dense masks alone (Axer, Gräßel et al., 2011; Schiavi et al., 2022). A natural extension of this work is to use the in-focus probability map as a confidence weight for estimating local fiber orientation and reconstructing a 3D orientation vector field. In this framing, segmentation is an intermediate step toward physically meaningful fields and downstream metrics, such as orientation dispersion. Future work should therefore focus on joint modeling of (i) in-focus probability, (ii) orientation, and (iii) confidence, and evaluate success by the stability and anatomical plausibility of the reconstructed 3D myelin fiber organization across stacks.

Chapter 6

Conclusion

This thesis developed a practical segmentation-to-quantification pipeline for in-focus myelin fiber analysis in CCP-BRM z-stacks under limited manual annotation. Using interpretable heuristic pseudo-labels (RB and Tenengrad) as scalable supervision, we trained a context-aware 2.5D U-Net and found that fine-tuning on a small set of manual annotations improves performance and transfer to a physically distinct held-out volume. Although pixel-level overlap remains moderate, which is consistent with the difficulty of segmenting thin, sparse structures and the ambiguity of manual labeling, the model produces consistent probability maps that enable downstream analysis. Specifically, the pipeline supports qualitative 3D visualization via z-axis interpolation for rendering and conservative volumetric quantification of myelin density and volume computed directly on the native acquisition grid. Overall, this work provides a reproducible foundation for scalable 3D myelin analysis and motivates future validation with additional annotated regions and independent volumes to better characterize robustness across tissue and acquisition variability.

References

Abdollahzadeh, A., Belevich, I., Jokitalo, E., Sierra, A., & Tohka, J. (2021). DeepACSON automated segmentation of white matter in 3D electron microscopy. Communications biology, 4(1), 179.

Arous, J., Binding, J., Léger, J., Léger, J., Léger, J., Casado, M., Casado, M., Casado, M., Topilko, P., Topilko, P., Topilko, P., Gigan, S., Boccara, A., Bourdieu, L., Bourdieu, L., & Bourdieu, L. (2011). Single myelin fiber imaging in living rodents without labeling by deep optical coherence microscopy. Journal of biomedical optics, 16 11, 116012 . https://doi.org/10.1117/1.3650770.

Avesta, A., Hossain, S., Lin, M., Aboian, M., Krumholz, H. M., & Aneja, S. (2023). Comparing 3D, 2.5 D, and 2D approaches to brain image auto-segmentation. Bioengineering, 10(2), 181.

Axer, M., Amunts, K., Grässel, D., Palm, C., Dammers, J., Axer, H., Pietrzyk, U., & Zilles, K. (2011). A novel approach to the human connectome: Ultra-high resolution mapping of fiber tracts in the brain. NeuroImage, 54(2), 1091–1101. https://doi.org/10.1016/j.neuroimage.2010.08.075

Axer, M., Gräßel, D., Kleiner, M., Dammers, J., Dickscheid, T., Reckfort, J., Hütz, T., Eiben, B., Pietrzyk, U., Zilles, K., & Amunts, K. (2011). High-Resolution Fiber Tract Reconstruction in the Human Brain by Means of Three-Dimensional Polarized Light Imaging. Frontiers in Neuroinformatics, 5.

Blanke, N., Go, V., Rosene, D. L., & Bigio, I. J. (2021). Quantitative birefrin-

gence microscopy for imaging the structural integrity of CNS myelin following circumscribed cortical injury in the rhesus monkey. Neurophotonics, 8(1), 1–15. https://doi.org/10.1117/1.NPh.8.1.015010

Blanke, N., Gray, A. J., Robinson, R. E., Novoseltseva, A., Rosene, D. L., & Bigio, I. J. (2024). Practical considerations for birefringence microscopy of myelin structure: Microscope design and tissue processing for effective imaging. Imaging Neuroscience, 2, 1-22.

Boshkovski, T., Kocarev, L., Cohen-Adad, J., Mišić, B., Lehéricy, S., Stikov, N., & Mancini, M. (2021). The R1-weighted connectome: complementing brain networks with a myelin-sensitive measure. Network Neuroscience, 5(2), 358-372.

Cao, Y., & Cheng, Y. (2025). SACU-Net: Shape-Aware U-Net for Biomedical Image Segmentation With Attention Mechanism and Context Extraction. IEEE Access, 13, 5719-5730. https://doi.org/10.1109/access.2025.3526602.

Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. Proceedings of the 23rd International Conference on Machine Learning, 233–240.

Depp, C., Sun, T., Sasmita, A., Spieth, L., Berghoff, S., Nazarenko, T., Overhoff, K., Steixner-Kumar, A., Subramanian, S., Arinrad, S., Ruhwedel, T., Möbius, W., Göbbels, S., Saher, G., Werner, H., Damkou, A., Zampar, S., Wirths, O., Thalmann, M., Simons, M., Saito, T., Saido, T., Krueger-Burg, D., Kawaguchi, R., Willem, M., Haass, C., Geschwind, D., Ehrenreich, H., Stassart, R., & Nave, K. (2023). Myelin dysfunction drives amyloid-β deposition in models of Alzheimer's disease. Nature, 618, 349 - 357. https://doi.org/10.1038/s41586-023-06120-6.

Duncan, I. D., & Radcliff, A. B. (2016). Inherited and acquired disorders of myelin: the

underlying myelin pathology. Experimental neurology, 283, 452-475.

Fan, S., Zhu, F., Feng, Z., Lv, Y., Song, M., & Wang, F. (2022). Conservative-Progressive Collaborative Learning for Semi-Supervised Semantic Segmentation. IEEE Transactions on Image Processing, 32, 6183-6194. https://doi.org/10.1109/tip.2023.3242819.

Gonsalvez, D., Yoo, S., Fletcher, J., Wood, R., Craig, G., Murray, S., & Xiao, J. (2019). Imaging and Quantification of Myelin Integrity After Injury With Spectral Confocal Reflectance Microscopy. Frontiers in Molecular Neuroscience, 12. https://doi.org/10.3389/fnmol.2019.00275.

Griebel, M., Segebarth, D., Stein, N., Schukraft, N., Tovote, P., Blum, R., & Flath, C. (2023). Deep learning-enabled segmentation of ambiguous bioimages with deepflash2. Nature Communications, 14. https://doi.org/10.1038/s41467-023-36960-9.

Hussain, T., & Shouno, H. (2024). MAGRes-UNet: Improved Medical Image Segmentation Through a Deep Learning Paradigm of Multi-Attention Gated Residual U-Net. IEEE Access, 12, 40290-40310. https://doi.org/10.1109/access.2024.3374108.

Kervadec, H., Bouchtiba, J., Desrosiers, C., Granger, E., Dolz, J., & Ayed, I. (2018). Boundary loss for highly unbalanced segmentation. Medical image analysis, 67, 101851. https://doi.org/10.1016/j.media.2020.101851.

Kumar, A., Jiang, H., Imran, M., Valdes, C., Leon, G., Kang, D., Nataraj, P., Zhou, Y., Weiss, M., & Shao, W. (2024). A Flexible 2.5D Medical Image Segmentation Approach with In-Slice and Cross-Slice Attention. Computers in biology and medicine, 182, 109173 . https://doi.org/10.48550/arxiv.2405.00130.

Li, R., Kudryashev, M., & Yakimovich, A. (2023). A weak-labelling and deep learning approach for in-focus object segmentation in 3D widefield microscopy. Scientific reports, 13(1), 12275. https://doi.org/10.1038/s41598-023-38490-2.

Micheva, K., Wolman, D., Mensh, B., Pax, E., Buchanan, J., Smith, S., & Bock, D. (2016). A large fraction of neocortical myelin ensheathes axons of local inhibitory neurons. eLife, 5. https://doi.org/10.7554/elife.15784.

Monsma, P. C., & Brown, A. (2012). FluoroMyelinTM Red is a bright, photostable and non-toxic fluorescent stain for live imaging of myelin. Journal of Neuroscience Methods, 209(2), 344–350. https://doi.org/10.1016/j.jneumeth.2012.06.015.

Nave, K. A., & Werner, H. B. (2014). Myelination of the nervous system: Mechanisms and functions. Annual Review of Cell and Developmental Biology, 30, 503–533. https://doi.org/10.1146/annurev-cellbio-100913-013101.

Novoseltseva, A., Xie, T., & Bigio, I. J. (2024, April). Extracting 3D Information from Widefield Birefringence Microscopy Images of Sparse Myelinated Fibers in Human Brain Gray Matter. In Microscopy Histopathology and Analytics (pp. JS4A-41). Optica Publishing Group.

Peters, A. (2009). The effects of normal aging on myelinated nerve fibers in monkey central nervous system. Frontiers in Neuroanatomy, 3, 1–10.

Peters, A., & Sethares, C. (2002). Aging and the myelinated fibers in prefrontal cortex and corpus callosum of the monkey. Journal of Comparative Neurology, 442(3), 277–291. https://doi.org/10.1002/cne.10099

Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Cham: Springer international publishing.

Saab, A. S., & Nave, K. A. (2017). Myelin dynamics: Protecting and shaping neuronal functions. Current Opinion in Neurobiology, 47, 104–112.

Saito, T., & Rehmsmeier, M. (2015). The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. PLOS ONE, 10(3), .

Schiavi, S., Lu, P., Weigel, M., Lutti, A., Jones, D., Kappos, L., Granziera, C., & Daducci, A. (2022). Bundle myelin fraction (BMF) mapping of different white matter connections using microstructure informed tractography. NeuroImage, 249, 118922 - 118922. https://doi.org/10.1016/j.neuroimage.2022.118922.

Sibarita, J.-B. (2005). Deconvolution microscopy. Advances in Biochemical Engineering/Biotechnology, 95, 201–243.

Siddique, N., Paheding, S., Elkin, C., & Devabhaktuni, V. (2020). U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications. IEEE Access, 9, 82031-82057. https://doi.org/10.1109/access.2021.3086020.

Stadelmann, C., Timmler, S., Barrantes-Freer, A., & Simons, M. (2019). Myelin in the central nervous system: Structure, function, and pathology. Physiological Reviews, 99(3), 1381–1431. https://doi.org/10.1152/physrev.00031.2018

Sternberg, S. R. (1983). Biomedical image processing. Computer, 16(1), 22–34. https://doi.org/10.1109/MC.1983.1654163

Timmler, S., & Simons, M. (2019). Grey matter myelination. Glia, 67, 2063 - 2070. https://doi.org/10.1002/glia.23614.

Valanarasu, J., Sindagi, V., Hacihaliloglu, I., & Patel, V. (2020). KiU-Net: Overcomplete Convolutional Architectures for Biomedical Image and Volumetric Segmentation. IEEE Transactions on Medical Imaging, 41, 965-976.

Wainberg, M., Forde, N. J., Mansour, S., Kerrebijn, I., Medland, S. E., Hawco, C., &

Tripathy, S. J. (2024). Genetic architecture of the structural connectome. Nature Communications, 15(1), 1962.

Wallace, W., Schaefer, L. H., & Swedlow, J. R. (2001). A Workingperson's Guide to Deconvolution in Light Microscopy. BioTechniques, 31(5), 1076–1097.

Wang, H., Deng, Z., Yang, Z., Su, Y., Liu, J., Zhang, S., Qiao, Y., Gu, L., & He, J. (2025). Revisiting model scaling with a U-net benchmark for 3D medical image segmentation. Scientific Reports, 15. https://doi.org/10.1038/s41598-025-15617-1.

Yeo, T.T., Ong, S.H., Jayasooriah, & Sinniah, R. (1993). Autofocusing for tissue microscopy. Image and Vision Computing, 11(10), 629–639. https://doi.org/10.1016/0262-8856(93)90059-P.

Yue, J., Jin, S., Wang, S., Zeng, J., Shan, S., Liu, B., Jiang, N., & Zhou, F. (2024). A shape-supervised feature fusion U-Net for tubular structure segmentation. Computers and Electrical Engineering, 118(Part A), 109522. https://doi.org/10.1016/j.compeleceng.2024.109522.

Zaimi, A., Wabartha, M., Herman, V., Antonsanti, P. L., Perone, C. S., & Cohen-Adad, J. (2018). AxonDeepSeg: automatic axon and myelin segmentation from microscopy data using convolutional neural networks. Scientific reports, 8(1), 3816.

Zhang, Y., Liao, Q., Ding, L., & Zhang, J. (2020). Bridging 2D and 3D segmentation networks for computation-efficient volumetric medical image segmentation: An empirical study of 2.5D solutions. Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society, 99, 102088 . https://doi.org/10.1016/j.compmedimag.2022.102088.

Zheng, Y., Tian, B., Yu, S., Yang, X., Yu, Q., Zhou, J., Jiang, G., Zheng, Q., Pu, J., & Wang, L. (2025). Adaptive boundary-enhanced Dice loss for image segmentation. Biomedical

signal processing and control, 106, 107741. https://doi.org/10.1016/j.bspc.2025.107741.

Curriculum Vitae