

Human Axon Radii Estimation at MRI Scale

Deep Learning Combined with Large-scale Light Microscopy

Laurin Mordhorst¹, Maria Morozova^{2,3}, Sebastian Papazoglou¹, Björn Fricke¹,
Jan M. Oeschger¹, Henriette Rusch³, Carsten Jäger², Markus Morawski^{2,3},
Nikolaus Weiskopf^{2,4}, Siawoosh Mohammadi^{1,2}

¹Institute of Systems Neuroscience, University Medical Center Hamburg-Eppendorf

²Department of Neurophysics, Max Planck Institute for Human Cognitive and Brain Sciences

³Paul Flechsig Institute of Brain Research, University of Leipzig

⁴Felix Bloch Institute for Solid State Physics, University of Leipzig

`l.mordhorst@uke.de`

Abstract. Non-invasive assessment of axon radii via MRI is of increasing interest in human brain research. Its validation requires representative reference data that covers the spatial extent of an MRI voxel (e.g., 1 mm^2). Due to its small field of view, the commonly used manually labeled electron microscopy (mLEM) can not representatively capture sparsely occurring, large axons, which are the main contributors to the effective mean axon radius (r_{eff}) measured with MRI. To overcome this limitation, we investigated the feasibility of generating representative reference data from large-scale light microscopy (lsLM) using automated segmentation methods including a convolutional neural network (CNN). We determined large, mis-/undetected axons as the main error source for the estimation of r_{eff} ($\approx 10\%$). Our results suggest that the proposed pipeline can be used to generate reference data for the MRI-visible r_{eff} and even bears the potential to map spatial, anatomical variation of r_{eff} .

1 Introduction

In vivo axon radius assessment in human brain using MRI-based models [1] is of neuroscientific relevance since it is, e.g., related to the neuronal conduction velocity [2]. Often, these models have been compared to manually labeled electron microscopy (mLEM) images with a very small field of view containing up to 10^3 or less axons [3,4] whereas MRI voxels of 1 mm^3 contain two to three orders of magnitude more axons. While mLEM is commonly used to report the arithmetic mean axon radii (r_{arith}) in histology studies [3,4], it can largely underestimate the axon radius indices measured with MRI such as the effective mean axon radius (r_{eff}) [1]. r_{eff} based on mLEM is often underestimated because sparsely occurring, large axons, which are the main contributors to r_{eff} , are underrepresented in the mLEM based axon distribution [5]. Here, we assessed the potential of large-scale

light microscopy (lsLM) and automated segmentation methods including a convolutional neural network (CNN) to generate representative reference data for axon radii distributions in typical MRI voxels. We evaluated the error and bias of both r_{eff} and r_{arith} on six lsLM sections across a human corpus callosum sample including a cross-microscopy comparison to their mLEM-based counterparts.

2 Materials and methods

2.1 Ensemble mean axon radii

The MRI-visible, effective mean radius [1] of an ensemble of N axons can be defined as $r_{\text{eff}} = \sqrt[4]{(\sum_{i=1}^N r_i^6)/(\sum_{i=1}^N r_i^2)}$, where r_i is the i -th radius. It deviates from the arithmetic mean radius $r_{\text{arith}} = \frac{1}{N} \sum_{i=1}^N r_i$ and is weighted towards the tail of its distribution rather than its bulk.

2.2 Normalized root-mean-square deviation and mean bias

To measure the residuals between observed (y) and estimated ($\hat{y}$) values (e.g., mean axon radii) over multiple observations, we used the normalized root-mean-square deviation $\text{NRMSD} = (\sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2}) / (\frac{1}{N} \sum_{i=1}^N y_i)$ and the relative mean bias $\bar{b}_r = (\sum_{i=1}^N \hat{y}_i - y_i) / (\sum_{i=1}^N y_i)$, where y_i and $\hat{y}_i$ are the i -th observation and estimation.

2.3 Dataset

A human optic chiasm (OC), a corticospinal tract (CST) and a corpus callosum (CC) sample were scanned at autopsy. The entire procedure of case recruitment, acquisition of the patient’s personal data, the protocols and the informed consent forms, performing the autopsy and handling the autopsy material have been approved by the responsible authorities (Approval #205/17-ek). We acquired lsLM images of semi-thin (500 nm) sections using a Zeiss AxioScan Z1 (resolution: 0.112 $\mu\text{m}/\text{px}$; resolution limit: 0.3 μm) and EM images of consecutively cut, ultra-thin (50 nm) sections using a Zeiss EM 912 Omega (resolution: 0.0043 $\mu\text{m}/\text{px}$).

From the total of 20 lsLM sections (18 CC, 1 CST, 1 OC), we annotated 17 subsections (11 CC, 3 CST, 3 OC; size: $\sim 100 \times 100 \mu\text{m}^2$) as background, axon or myelin for training and validation. For testing, we reserved six regions (no overlap with training/validation regions) of the CC sample (Fig. 1) and annotated EM sections and subsections of their corresponding lsLM sections at different granularity: EM subsections (one per region; size: $\sim 50 \times 50 \mu\text{m}^2$) and small lsLM subsections were annotated entirely (five per region; size: $\sim 28 \times 28 \mu\text{m}^2$), whereas larger subsections (two of size $\sim 110 \times 110 \mu\text{m}^2$ per region to evaluate the segmentation performance; one per region of size $\sim 550 \times 550 \mu\text{m}^2$ to evaluate the estimation of r_{eff}) were annotated selectively for MRI-relevant, large axons with $r \geq 1.6 \mu\text{m}$. The threshold was determined so that r_{eff} was halved when axons with radii above this threshold were removed.

2.4 Axon radius estimation pipeline

Axon radius estimation was split up into three parts (Fig. 2): pixel-wise classifications were obtained using a U-Net [6] (Fig. 2a; details below) in a sliding window manner (patch size 512×512 px). After identification of axon instances (Fig. 2b) through connected-component labeling, axon radii were extracted (Fig. 2c) as circular area equivalent radii from the segmented instance areas.

Network Architecture We used a U-Net [6] implemented in PyTorch [7] and employed transfer learning, i.e., we used pretrained (on the ImageNet dataset) ResNet18 [8] encoders and adapted all weights during training.

Training and Validation The training process was split into pseudo-epochs of 75 random patches, matching the number of available patches. Augmentations were employed on-the-fly: geometric, brightness and saturation transformations, Gaussian blurring and stain augmentation [9]. Patches were standardized with respect to the ImageNet dataset. We used Adam [10] with a learning rate of 10^{-3} and a learning rate decay of $\gamma = 0.75$ every 25 epochs to minimize a cross-entropy loss with classes weighted inversely proportional to their occurrences. Hyperparameter optimization was carried out in a 4-fold cross-validation split at the level of entire lsLM subsections. The model with the best hyperparameters was determined in terms of the averaged dice score for axon and myelin and trained for 250 epochs.

2.5 Experiment 1: Segmentation performance

We assessed the general segmentation quality on small, entirely labeled subsections (Sec. 2.3) as well as the MRI-relevant, large-axon specific segmentation

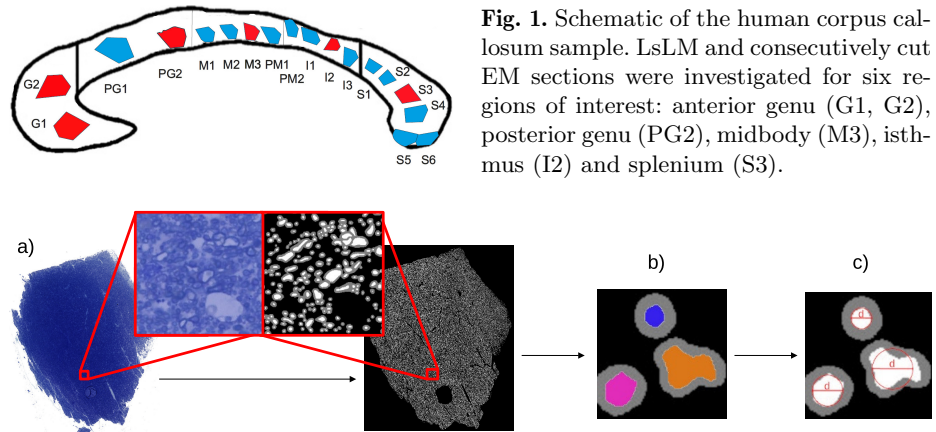


Fig. 1. Schematic of the human corpus callosum sample. lsLM and consecutively cut EM sections were investigated for six regions of interest: anterior genu (G1, G2), posterior genu (PG2), midbody (M3), isthmus (I2) and splenium (S3).

Fig. 2. The axon radius estimation pipeline. Radius estimation was performed in three steps: semantic segmentation (a), instance segmentation (b) and radius extraction (c).

quality on large, selectively labeled subsections (Sec. 2.3) in terms of accuracy, dice, precision and recall for the binary classification task of discriminating between axon and background.

2.6 Experiment 2: Accuracy of mean axon radii estimation

Cross-microscopy comparison with mLEM The estimation of r_{arith} from lsLM sections was compared to its counterpart obtained from consecutively cut mLEM subsections with ($r \geq 0.3 \mu\text{m}$) and without ($r \geq 0.0 \mu\text{m}$) rejection of small axons irresolvable due to the lsLM resolution limit (Sec. 2.3).

Spatial variance of r_{arith}/r_{eff} and their correlation with lsLM image intensity To assess the spatial variance, r_{arith} and r_{eff} were computed for randomly positioned subsections (size: 0.12 mm^2) of each lsLM sample and the average of overlapping subsections was computed locally, resulting in spatially smoothed axon radii maps. To investigate the influence of lsLM image intensity on the spatial variance of the axon radii maps, we first estimated the r_{arith} and r_{eff} maps on an equally spaced grid (pixel area: 0.12 mm^2) and resampled the corresponding original lsLM images to the same size. Then, we estimated the correlation between image intensity and mapped axon radii across all pixels in the grid and all sections with similar axon radii distributions (i.e., G1, G2, PG2, M3 and I2).

Error propagation of small axons into r_{eff} To measure the error in the axon radii range irresolvable in lsLM ($r < 0.3 \mu\text{m}$), we compared r_{eff} obtained from lsLM with ($r \geq 0.3 \mu\text{m}$) and without ($r \geq 0.0 \mu\text{m}$) rejection of small axons to a mLEM-informed reference. To generate the mLEM-informed reference, we added small axon radii drawn from the mLEM distribution to its lsLM equivalent, so that the ratio between small ($r < 0.3 \mu\text{m}$) and remaining ($r \geq 0.3 \mu\text{m}$) axons equaled the ratio observed in mLEM.

Error propagation of large, false negative/positive axons into r_{eff} To assess the error introduced by large, false negative (fn) and false positive (fp) axons, we compared the estimated r_{eff} to a manually corrected r_{eff} . The latter was generated by manually correcting large axons in the aforementioned subsections (Sec. 2.3). As false negative, we determined a matching predicted axon for each labeled axon in terms of the highest pairwise dice score and defined labeled axons with no matching predicted axon or a dice score < 0.75 . As false positive, we determined predicted axons with no matching labeled axon and $r \geq 1.6 \mu\text{m}$. Small axons ($r < 0.3 \mu\text{m}$) irresolvable in lsLM were rejected for both estimated and manually corrected r_{eff} .

Table 1. Mean axon radii estimation results. The errors of r_{arith} and r_{eff} (Sec. 2.6) are given as normalized root-mean-square deviations and relative mean biases $\bar{b}_r$ (Sec. 2.2) over six corpus callosum regions with ($r \geq 0.3 \mu\text{m}$) and without ($r \geq 0.0 \mu\text{m}$) rejection of small axons irresolvable in lsLM.

Measure Experiment		$r \geq 0.0 \mu\text{m}$		$r \geq 0.3 \mu\text{m}$	
		NRMSD [%]	$\bar{b}_r$ [%]	NRMSD [%]	$\bar{b}_r$ [%]
r_{arith}	Cross-microscopy comparison	10.1	4.5	8.4	6.7
r_{eff}	Error prop. of small axons	0.3	0.2	1.7	1.6
	Error prop. of large fn/fp axons			9.6	-6.1

3 Results

3.1 Experiment 1: Segmentation performance

The segmentation metrics are given as (mean for all axons/mean for large axons): accuracy was (0.92/0.98), axon dice was (0.74/0.84), axon precision was (0.77/0.83) and axon recall was (0.72/0.86).

3.2 Experiment 2: Accuracy of mean axon radii estimation

Cross-microscopy comparison with mLEM r_{arith} was overestimated by 5 % to 7 %. The rejection of small axons decreased the NRMSD from 10 % to 8 %.

Spatial variance of $r_{\text{arith}}/r_{\text{eff}}$ and their correlation with lsLM image intensity r_{eff} , estimated with or without inclusion of small axons, had a high local heterogeneity (Fig. 3a, bottom) and no correlation with the image intensity (Fig. 3c, bottom). The smooth spatial pattern across the sample visible in r_{arith} (Fig. 3a, top) roughly coincided with the lsLM image intensity (Fig. 3b). The visual correspondence was supported by the significant correlation between r_{arith} and the image intensity (Fig. 3c). Removing smaller axons, reduced the correlation coefficient for r_{arith} (without small axons: $\rho = 0.67$, $p < 10^{-5}$; with small axons: $\rho = 0.83$, $p < 10^{-5}$).

Error propagation of small axons into r_{eff} The rejection of small axons ($r < 0.3 \mu\text{m}$) increased the bias from 0.2 % to 1.6 % and the error from 0.3 % to 1.7 % (Tab. 1).

Error propagation of large false negative/positive axons into r_{eff} The comparison between manually corrected r_{eff} and predicted r_{eff} (Tab. 1) revealed an underestimation of 6 % and an error of 10 %.

4 Discussion

We demonstrated that our pipeline employing a CNN on large-scale light microscopy (lsLM) data is more suitable to estimate the MRI-visible effective mean radius r_{eff} than the arithmetic mean radius r_{arith} . Albeit r_{arith} had comparable bias (r_{eff} : -6.1 % and r_{arith} : 4.5 %) and error (r_{eff} : 9.6 % and r_{arith} : 10.1 %), the significant correlation with the image/staining intensity proved our pipeline incapable to capture the spatial, anatomical variation. The correlation was particularly high when small axons (i.e., axons below the lsLM resolution limit: $r < 0.3 \mu\text{m}$) were included, which was reflected in consistently lower segmentation performance for small axons. For r_{eff} , the error and bias was dominated by false negative and positive predicted large axons ($r \geq 1.6 \mu\text{m}$). Consequently, the presented pipeline is potentially suited to map anatomical, spatial variation at MRI voxel resolution, but requires further quantification of errors in the large axon range, e.g., systematic under-/overestimation to qualify it as a robust reference for the validation of MRI-based axon diameter models.

References

1. Veraart J, Nunes D, Rudrapatna U, et al. Noninvasive Quantification of Axon Radii Using Diffusion MRI. *elife*. 2020;9:e49855.
2. Schmidt H, Knösche TR. Action potential propagation and synchronisation in myelinated axons. *PLoS Comput Biol*. 2019;15:e1007004.
3. Aboitiz F, Scheibel AB, Fisher RS, et al. Fiber Composition of the Human Corpus Callosum. *Brain Res*. 1992;598(1):143–153.

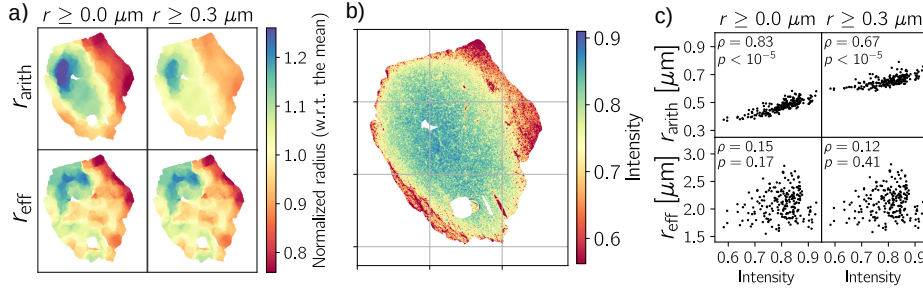


Fig. 3. Spatial variance of $r_{\text{arith}}/r_{\text{eff}}$ and their correlation with lsLM image intensity. Depicted are: (a) spatially smoothed r_{arith} and r_{eff} maps, (b) lsLM image of section PG2 (a) adjusted to illustrate the correlation with r_{arith} map in (a), and scatter plots between $r_{\text{arith}}/r_{\text{eff}}$ maps and lsLM image intensities (c). The r_{arith} and r_{eff} maps in (a) and (c) were generated with ($r \geq 0.3 \mu\text{m}$) and without ($r \geq 0.0 \mu\text{m}$) rejection of small axons. The correlation plots in (c) pools across 5 sections, its p -value has been multiplied by the number of regions, to correct for multiple comparisons (ρ is the correlation coefficient). The grid overlaid in (b) illustrates a typical MRI voxel (1 mm² resolution).

4. Caminiti R, Ghaziri H, Galuske R, et al. Evolution Amplified Processing with Temporally Dispersed Slow Neuronal Connectivity in Primates. *PNAS USA*. 2009;106(46):19551–19556.
5. Mordhorst L, Morozova M, Papazoglou S, et al. Towards a Representative Estimation of the MRI-Visible Axon-Radius Distribution Using Large-Scale Light Microscopy Images. *ESMRMB*. 2020;.
6. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *MICCAI*. 2015; p. 234–241.
7. Paszke A, Gross S, Massa F, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv:1912.01703*. 2019;.
8. He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*. 2016; p. 770–778.
9. Macenko M, Niethammer M, Marron JS, et al. A Method for Normalizing Histology Slides for Quantitative Analysis. *ISBI*. 2009; p. 1107–1110.
10. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980*. 2017;.