COCTA: A CORTEX VESSEL SEGMENTATION BENCHMARK USING OPTICAL COHERENCE TOMOGRAPHY ANGIOGRAPHY

Jiaxiang Ren[1], Wensheng Cheng[1], Qiaochu Wang[2], Zhenghong Li[1], Rutwik Segireddy[1]
Raghuveer Tatineni[1], Hyomin Jeong[2], Yanzuo Liu[2], Ruoyu Xu, Mira Xu, Xiaoyu Hu[1], Keying Jia[1],
Congwu Du[2], Yingtian Pan[2], Haibin Ling[1*]*

[1]Department of Computer Science, [2]Department of Biomedical Engineering, Stony Brook University

ABSTRACT

Cortical vasculature plays an important role in neuroscience, and it is popularly analyzed using optical coherence tomography angiography (OCTA). Despite two decades of development, however, there remains a lack of high quality vessel segmentation benchmark for cortex OCTA data. In this paper, we introduce a novel OCTA vessel segmentation benchmark of mouse cortex, called COCTA, which largely facilitates the training and evaluation of segmentation models for brain vasculature possible. OCTA involves different types of real-world noise, including speckle noise, motion artifacts and background noise. These heterogeneous noise sources exhibit diverse characteristics and post challenges for current segmentation methods. Therefore, models trained on public retina datasets can not easily generalize to cortex OCTA. Additionally, it is hard to accurately delineate vessel boundaries using mouse so we introduce the stylus pen to ensure a refined and natural-look mask. Besides, our annotators make extensive efforts to remove artifacts and reveal underling vessels. With these corrected manual masks, our dataset is suitable for evaluation in the denoising community. Lastly, we benchmark and analyze various segmentation methods, including convolutional neural networks and transformers, providing insights for the development of new approaches in OCTA. The dataset with manual annotations are available at https://github.com/reckdk/COCTA-dataset.

Index Terms— Optical coherence tomography angiography, vessel segmentation, cortex vasculature, benchmark

1. INTRODUCTION

Vasculature is a critical component of biological systems, and effective visualization and assessment of vasculature structures are essential for medical professionals. Angiography serves as a critical imaging technique for this purpose, enabling physicians to examine blood vessels across the body. This technique aids in diagnosing and evaluating various vascular conditions, including blockages, stenosis, aneurysms, and other malformations.

Angiography techniques have evolved over time, from traditional X-ray angiography using contrast dyes to more ad-
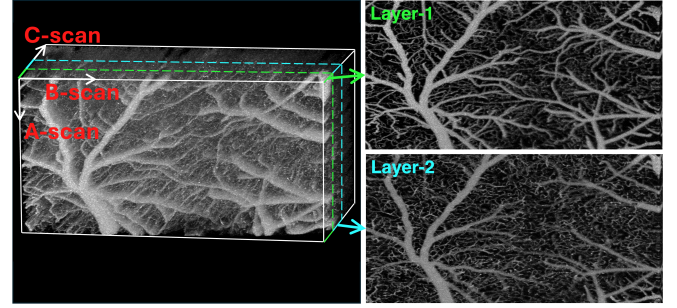

Fig. 1: 3D scan of OCTA and two annotation layers.

vanced, less invasive, higher resolution methods. New techniques include computed tomography angiography (CTA), magnetic resonance angiography (MRA), fluorescent angiography (FA), two-photon angiography, and optical coherence tomography angiography (OCTA). CTA is widely used for imaging blood vessels throughout the body, especially for cardiovascular and cerebrovascular, due to its accuracy and reliability. However, concerns remain regarding radiation exposure. MRA has been explored but, despite recent advancements, is still considered inferior in image quality compared to CTA [1]. FA can resolve the detailed retinal vasculature including capillaries, playing as the gold standard for evaluation of many retinal and choroidal vascular conditions in ophthalmology. The real-time response of FA enables the analysis of vascular flow dynamics. But FA is invasive and may cause severe side effects [2]. The recently emerged two-photon angiography method offers high-resolution 3D images of small blood vessels. Nevertheless, its depth of imaging is limited, and hence primarily employed in research settings rather than clinical practice. OCTA is a non-invasive, radiation-free, dye-free method with high spatial-temporal resolution [3], which gets increasing attention in both research and clinic.

Fig. 1 illustrates an OCTA volume of mouse cortex along with the maximum intensity projection (MIP) of the first two layers. An OCT A-scan provides a one-dimensional depth profile of tissue reflectivity and serves as the basic unit of OCT imaging. A B-scan, which is composed of multiple adjacent A-scans along the same direction, forms a two-dimensional cross-sectional image. A C-scan, in turn, is generated by stacking multiple B-scans along the direction

perpendicular to the B-scan plane. Due to shadow artifacts [4], only the MIPs are annotated in our datasets.

Currently, there are public available retina OCTA datasets that cover various subjects, modalities, conditions, field-of-view (FOV), and resolutions [5, 6, 7, 8, 9, 10, 11]. Many successful works have utilized these datasets and improve the segmentation performance [12, 13, 14, 15]. We refer to the discussion in [16, 6] for more details.

Compared with retina OCTA, mouse cortex OCTA has much deeper imaging depth and higher resolution [17, 6, 18], and thus is more susceptible to noises. The projected vasculature density also increases with imaging depth. As a result, models trained on retina datasets can not simply generalize to cortex datasets. The segmentation performance in cortex OCTA suffers from lack of high quality dataset.

In order to boost relevant research in the community, we make the following contributions in this work.

- For the first time in cortex microvasculature, we construct a mouse cortex OCTA dataset, named COCTA, with high quality manual annotations. COCTA contains 62 real high-resolution OCTA images (500×1000) with relatively large FOV ($2 \times 2.6mm^2$).
- COCTA contains multiple real-world noises, such as motion artifacts, speckle noise and background noise. Annotators create clear masks despite these challenges.
- A convenient annotation tool with stylus pens and tablets is introduced to enhance annotation of mask vesselness.
- We benchmark mainstream segmentation methods on COCTA and analyze the performance of each method, highlighting the specific challenges of COCTA.

2. CORTEX OCTA DATASET

Vasculature annotation is a labor-intensive task. Unlike typical segmentation datasets, high quality vessel annotations cannot be outlined using polygons but need marking pixel by pixel, requiring more efforts. Additionally, vasculature density and noise level can further increase the complexity of annotation. Iterative refinement is also required to maintain consistency among different annotators. Twelve annotators, mostly in computer science or biomedical engineering, underwent at least five hours of training. Senior annotators with greater expertise oversaw the refinement stage to ensure final quality. Based on our experience, initial annotation of a cortex OCTA image typically took 6–8 hours, with iterative refinements taking additional 1-3 hours, resulting in a total of approximately 9 hours per image. Such substantial time and efforts result in the scarcity of cortex OCTA segmentation datasets.

2.1. Vascular Annotation Solution

Vessels belong to curvilinear structures that need careful attention to boundary and width. Our high resolution OCTA
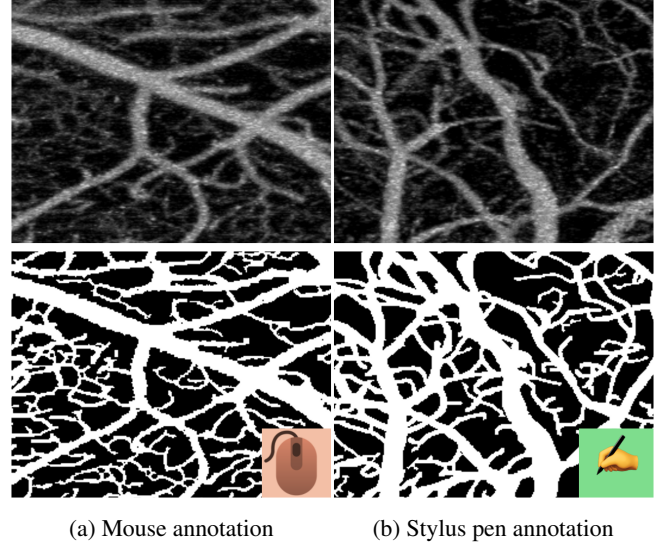

(a) Mouse annotation   (b) Stylus pen annotation

Fig. 2: OCTA images (top) and corresponding masks (bottom) annotated with mouse and stylus pen, respectively.

images cover the vessels of different scales, including arteries/veins, arterioles/venules, and capillaries. Delineating dense vasculature using mouse is laborious and can lead to waist strain due to overuse. Drawing pad for graphic designers ensures the quality of mask but slow interaction response and software adaption restrict its usage in our cases. Benefiting from the evolution of tablet computers and compatible annotation toolkits, we introduce an intuitive annotation solution, consisting of a stylus pen, an iPad (or Android tablets), and the toolkit called webKnossos [19]. Specifically, we visualize and annotate an image on the same screen with a stylus pen. We use the brush tool in webKnossos with one pixel width to delineate the thinnest capillaries. Zoom-in and out are supported via finger gestures, especially useful for annotating capillaries. Compared to traditional mouse-based solution, the intuitive interaction enhances vesselness of mask and efficiency, making it more suitable for extended work sessions.

Fig. 2 illustrates two OCTA images annotated with mouse and stylus pen. Benefiting from direct feedback, stylus-based masks tend have smoother boundaries than mouse-based ones. Additionally, the curves and diameters of capillaries appear more natural and closer than those in Fig. 2b. In some neuron studies, capillaries play more important roles than large vessels so we should treat them more carefully. Furthermore, small holes often appear sporadically in mouse-based masks, particularly at junctions or along thin vessels. These issues are minimized with stylus-based annotations.

2.2. Multi-layer Annotation

The original cortex OCTA is a 3D scan, as illustrated in Fig. 1, and provides significantly deeper FOV than retinal OCTA. When projecting as the MIP image, vessels at different depths
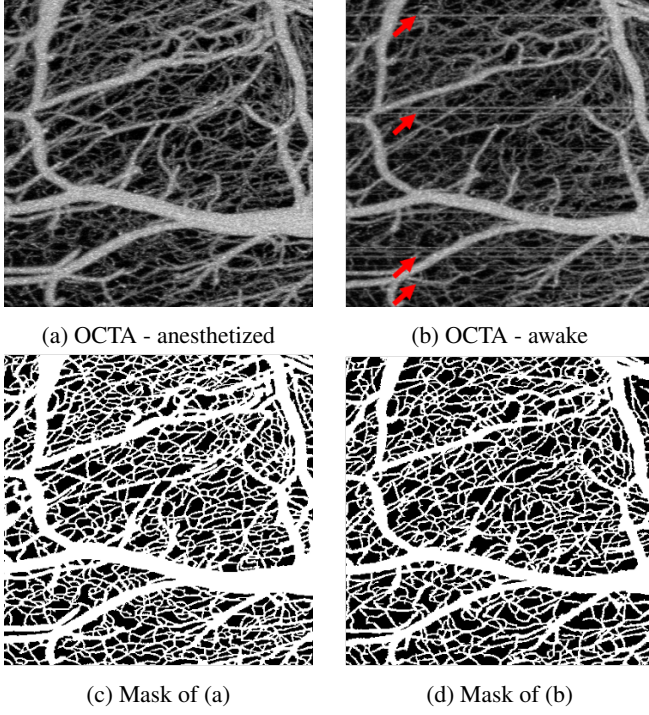
(a) OCTA - anesthetized   (b) OCTA - awake

(c) Mask of (a)   (d) Mask of (b)

Fig. 3: OCTA images and masks for mice under anesthetized and awake statuses, respectively. Red arrow: motion artifacts.

interleave with each other. In this dataset, we provide masks in three subsets of different depths, which archives optimal trade-off between annotation abundance and confidence. To be specific, the first layer MIP is projected from the superficial 80 voxels; the second layer is projected from the 70th to 150th voxels, with 12.5% overlap with the first layer for consistency; the third subset is the fusion of the first and second layers. The first layer has more large vessels while the second layer has more mid-to-small vessels. Directly annotating the fused images may lose details due to the relatively low signal-to-noise ratio (SNR) and denser vasculature. Therefore, we annotate the first two layers separately and then fuse the masks. As a result, the mask in the fusion subset are of highest density among the three subsets.

2.3. Heterogeneous Noise Sources

Our dataset involves multiple real-world noise sources: (1) motion artifacts caused by heartbeat, respiration, and mechanical jittering, severely degrade the image quality; (2) speckle noise due to the imaging principle; and (3) background noise with the compromised SNR due to the trade-off for improving temporal resolution. For noise distribution, motion artifacts appear in both the first and second layers. The first layer has less background noise than the second layer. Annotators work to differentiate and label structures despite these noises, producing a clear segmentation mask.
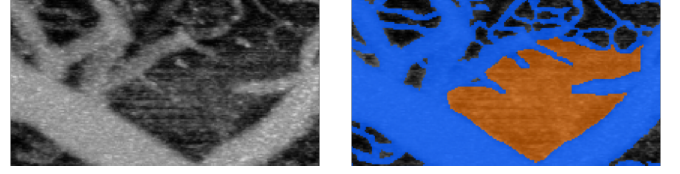

Fig. 4: Left: OCTA image with extreme-noise; right: corresponding mask. Blue: vessel; orange: ignored area.

Motion artifacts. Fig. 3 illustrates two OCTA images with masks for mice under anesthetized and awake statuses, respectively. Motion artifacts are more prevalent in awake animals, as indicated by the red arrows in Fig. 3b. We remove the motion artifacts from masks and make our best effort to reveal the underlying vessels. We provide paired clear and motion-affected images, which can be used to evaluate denoising methods.

Extreme-noise Handling. Although the vasculature is discernible in most regions, some areas remain extremely noisy and complex, make them too indistinct to annotate. Therefore, we create the ignore-mask to exclude these indiscernible areas from both training and evaluation. For example, Fig. 4 illustrates an OCTA image with extreme-noise and corresponding mask. The central area is too blurry so we only keep the discernible vessels while ignoring the rest. Most samples in our dataset have an ignore-mask covering less than 10% of the area. In rare cases, the ignore-mask covers up to 30% of the area.

Iterative Refinement. Different annotators show different preferences. The annotation process follow the same standard and includes two stages for consistency. In the first stage, each mask is iteratively refined by the initial annotator for one to three rounds. In the second stage, the senior annotators check the results and make necessary corrections if needed, and then add the ignore-mask for indiscernible regions.

3. EXPERIMENT

Our mouse cortex OCTA dataset includes three subsets at different depths: Layer1, Layer2, and Layer12, containing 32, 30, and 30 images, respectively. The resolutions along the C-, B-, and A-scans range in $(500-650)\times 1000\times 1024$, with the FOV spanning in $(3.0-4.0)\mu m\times(1.9-2.6)\mu m\times 1.0\mu m$. All OCTA images are collected from a customized ultra-high resolution optical coherence angiography system (μOCA) with a center wavelength of $1310nm$. Additionally, we separate all images based on subject statuses into *Anesthetized* and *Awake*. The Anesthetized subset contains 23 Layer1 and 23 Layer2 samples. The Awake subset contains 9 Layer1 and 7 Layer2 samples. Manual annotations cover all images.

Experimental Settings. We benchmark state-of-the-art segmentation models using 62 samples from the Layer1 and Layer2 subsets. The training set and validation set consist of 50 and 12 samples, respectively. The validation set has 8

Table 1: Segmentation IoU (%) and Dice (%) of different methods on COCTA. The backbone of each method is listed in the second row.

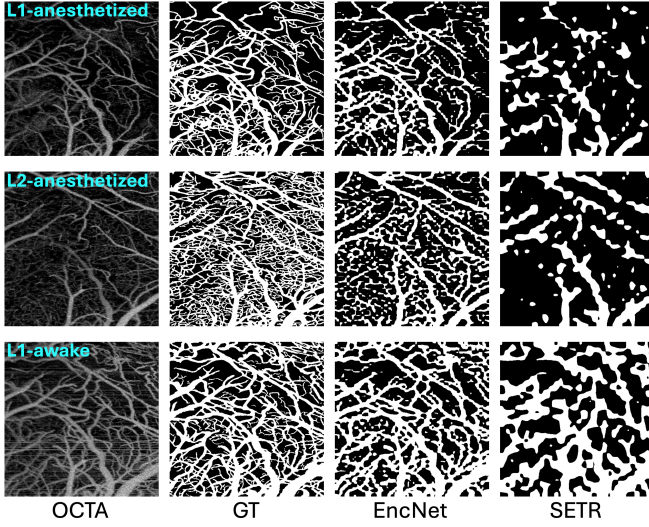|		FCN ResNet50	EncNet ResNet50	PSANet ResNet50	Segmenter ViT-Base	SETR ViT-Large
Layer1	IoU	60.16	60.65	58.27	47.72	48.09
	Dice	74.94	75.33	73.39	64.32	64.63
Layer2	IoU	52.79	52.93	52.16	42.02	41.78
	Dice	68.87	68.97	68.23	58.86	58.58
Total	IoU	56.47	56.79	55.21	44.87	44.93
	Dice	71.90	72.15	70.81	61.59	61.60

Fig. 5: Segmentation results of different methods. From left to right: OCTA, manual annotations, results by EncNet [21] and SETR [24], respectively. From top to bottom: anesthetized layer1, layer2, and awake layer1, respectively.

anesthetized samples and 4 awake samples. Dice coefficient and Intersection over Union (IOU) are used as metrics.

All baselines are implemented with PyTorch and trained on 4 NVIDIA RTX A6000 GPUs. All convolutional neural network (CNN) methods (FCN [20], EncNet [21], PSANet [22]) use the same training setting with the SGD optimizer, where initial learning rate is 0.01, momentum is 0.9, weight decay is 0.0005, and poly learning rate schedule with power as 0.9 in the whole training process. Segmenter [23] uses initial learning rate as 0.001 without weight decay. SETR [24] uses initial learning rate as 0.01 without weight decay. All models are trained with 1000 iterations. Other settings follow the default one of each method. For data preprocessing, each image with size of H×W (H<W) is cropped to two square images with size of H×H, covering the original whole image. The right side of one image is overlapped with the left side of the other image. This leads to 100 training images and 24 validation images. For data augmentation, random resize, random crop, and random flip are applied to all methods.

Tab. 1 shows the results of different methods on the proposed COCTA. All methods exhibit degraded performance on the mouse cortex compared to common medical segmentation benchmarks. A primary factor is the complexity and density of the cortical vasculature. Additionally, CNN-based methods outperform transformer-based approaches, such as Segmenter [23] and SETR [24]. This could be partially due to the modest dataset size, which is insufficient to fully train transformers, as they typically require more data than CNNs. Data augmentation could help mitigate this issue by expanding the training set. Furthermore, all methods perform better on Layer1 compared to Layer2. The vessels in Layer2 are generally smaller and more dense than Layer1's, as illustrated in Fig. 5, making segmentation more challenging. Fig. 5 also provides a comparison of the same subject under both anesthetized and awake conditions. In the awake images, motion artifacts are observed as prominent horizontal stripes. Both EncNet [21] and SETR [24] exhibit the capability to mitigate motion artifacts through training. However, these methods demonstrate relatively low sensitivity to small vessels, indicating that they have not yet achieved optimal performance in cortical vessel segmentation. Future advancements will be driven by the development of more effective methods and the incorporation of larger datasets.

4. CONCLUSION

Cortical vasculature plays a crucial role in neuron studies. In this paper, we introduce an OCTA dataset of mouse cortex to mitigate the data insufficiency for brain vessel segmentation. Our OCTA images are projected from deeper volumes than retina OCTA and contains multiple real-world noise types like speckle noise, motion artifacts, and background noise. Annotators create clear masks despite these challenging heterogeneous noises. Instead of interacting with mouse, we introduce the stylus pen to ensure a mask of good vesselness. Our solution takes advantage of the seamless integration of a stylus pen, a smart tablet, and webKnossos to enable intuitive and accurate annotation. Last but not least, we provide a benchmark of segmentation with mainstream methods. We discuss the performance of each approach, which can be instructive to the development of new approaches in OCTA.

We plan to collect more images, annotate deeper layers with severe noises, and design methods to effectively address complex noise challenges.

5. REFERENCES

[1] M. P. Chae, D. J. Hunter-Smith, and W. M. Rozen, "Comparative analysis of fluorescent angiography, computed tomographic angiography and magnetic resonance angiography for planning autologous breast reconstruction," *Gland surgery*, vol. 4, no. 2, pp. 164, 2015. 1

[2] I. S. Kornblau and J. F. El-Annan, "Adverse reactions to fluorescein angiography: A comprehensive review of the literature," *Surv. Ophthalmol.*, vol. 64, no. 5, pp. 679–93, 2019. 1

[3] R. F. Spaide, J. M. Klancnik, and M. J. Cooney, "Retinal vascular layers imaged by fluorescein angiography and optical coherence tomography angiography," *JAMA ophthalmology*, vol. 133, no. 1, pp. 45–50, 2015. 1

[4] P. Anvari, M. Ashrafkhorasani, A. Habibi, and K. G. Falavarjani, "Artifacts in optical coherence tomography angiography," *JOVR*, vol. 16, no. 2, pp. 271, 2021. 2

[5] Y. Ma, H. Hao, J. Xie, H. Fu, J. Zhang, J. Yang, Z. Wang, J. Liu, Y. Zheng, and Y. Zhao, "Rose: a retinal oct-angiography vessel segmentation dataset and new model," *TMI*, vol. 40, no. 3, pp. 928–39, 2020. 2

[6] M. Li, K. Huang, Q. Xu, J. Yang, Y. Zhang, Z. Ji, K. Xie, S. Yuan, Q. Liu, and Q. Chen, "Octa-500: a retinal dataset for optical coherence tomography angiography study," *MedIA*, vol. 93, pp. 103092, 2024. 2

[7] Y. Wang, Y. Shen, M. Yuan, J. Xu, B. Yang, C. Liu, W. Cai, W. Cheng, and W. Wang, "A deep learning-based quality assessment and segmentation system with a large-scale benchmark dataset for optical coherence tomographic angiography image," *arXiv preprint arXiv:2107.10476*, 2021. 2

[8] J. Kuhlmann, K. Rothaus, X. Jiang, H. Faatz, D. Pauleikhoff, and M. Gutfleisch, "3d retinal vessel segmentation in octa volumes: Annotated dataset more3d and hybrid u-net with flattening transformation," in *DAGM GCPR*. Springer, 2023. 2

[9] H. Ning, C. Wang, X. Chen, and S. Li, "An accurate and efficient neural network for octa vessel segmentation and a new dataset," in *ICASSP*, 2024. 2

[10] B. Qian, H. Chen, X. Wang, Z. Guan, T. Li, Y. Jin, Y. Wu, Y. Wen, H. Che, G. Kwon, et al., "Drac 2022: A public benchmark for diabetic retinopathy analysis on ultra-wide optical coherence tomography angiography images," *Patterns*, vol. 5, no. 3, 2024. 2

[11] J. Xue, Z. Feng, L. Zeng, S. Wang, X. Zhou, J. Xia, and A. Deng, "Soul: An octa dataset based on human machine collaborative annotation framework," *Scientific Data*, vol. 11, no. 1, pp. 838, 2024. 2

[12] Y. Liu, A. Carass, L. Zuo, Y. He, S. Han, L. Gregori, S. Murray, R. Mishra, J. Lei, P. A. Calabresi, et al., "Disentangled representation learning for octa vessel segmentation with limited training data," *TMI*, vol. 41, no. 12, pp. 3686–98, 2022. 2

[13] X. Tan, X. Chen, Q. Meng, F. Shi, D. Xiang, Z. Chen, L. Pan, and W. Zhu, "Oct2former: A retinal oct-angiography vessel segmentation transformer," *Comput Methods Progs Biomed*, vol. 233, pp. 107454, 2023. 2

[14] L. Kreitner, J. C. Paetzold, N. Rauch, C. Chen, A. M. Hagag, A. E. Fayed, S. Sivaprasad, S. Rausch, J. Weichsel, B. H. Menze, et al., "Synthetic optical coherence tomography angiographs for detailed retinal vessel segmentation without human annotations," *TMI*, 2024. 2

[15] B. Wittmann, L. Glandorf, J. C. Paetzold, T. Amiranashvili, T. Wälchli, D. Razansky, and B. Menze, "Simulation-based segmentation of blood vessels in cerebral 3d octa images," in *MICCAI*, 2024. 2

[16] Y. Giarratano, E. Bianchi, C. Gray, A. Morris, T. MacGillivray, B. Dhillon, and M. O. Bernabeu, "Automated segmentation of optical coherence tomography angiography images: benchmark data and clinically relevant metrics," *TVST*, vol. 9, no. 13, pp. 5, 2020. 2

[17] B. Alamouti and J. Funk, "Retinal thickness decreases with age: an oct study," *BJO*, vol. 87, no. 7, pp. 899–901, 2003. 2

[18] Y. Pan, K. Park, J. Ren, N. D. Volkow, H. Ling, A. P. Koretsky, and C. Du, "Dynamic 3d imaging of cerebral blood flow in awake mice using self-supervised-learning-enhanced optical coherence doppler tomography," *Commun. Biol.*, vol. 6, no. 1, pp. 298, 2023. 2

[19] K. M. Boergens, M. Berning, T. Bocklisch, D. Bräunlein, F. Drawitsch, et al., "webknossos: efficient online 3d data annotation for connectomics," *Nature Methods*, vol. 14, no. 7, pp. 691–4, 2017. 2

[20] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *CVPR*, 2015. 4

[21] H. Zhang, K. Dana, J. Shi, Z. Zhang, X. Wang, A. Tyagi, and A. Agrawal, "Context encoding for semantic segmentation," in *CVPR*, 2018. 4

[22] H. Zhao, Y. Zhang, S. Liu, J. Shi, C. C. Loy, D. Lin, and J. Jia, "Psanet: Point-wise spatial attention network for scene parsing," in *ECCV*, 2018. 4

[23] R. Strudel, R. Garcia, I. Laptev, and C. Schmid, "Segmenter: Transformer for semantic segmentation," in *ICCV*, 2021. 4

[24] S. Zheng, J. Lu, H. Zhao, X. Zhu, Z. Luo, Y. Wang, Y. Fu, J. Feng, et al., "Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers," in *CVPR*, 2021. 4