

Blood Flow Speed Estimation with Optical Coherence Tomography Angiography Images —Supplementary Materials—

1. Dataset Properties

The OCTA and ODT images in the dataset have a width of 1,000 pixels, while the height varies between 500 and 625 pixels, depending on the hardware system settings. Each OCTA image is pixel-wise aligned with its corresponding ODT data to form a paired dataset, ensuring that both images have the same spatial dimensions. OCTA images are 8-bit grayscale, with theoretical integer values ranging from 0 to 255. In contrast, ODT data is 16-bit, with theoretical integer values ranging from 0 to 65,535. This substantial range in blood flow data values adds to the complexity of the task. The properties of Anesthetized Dataset and Awake Dataset are shown in Table 1.

Table 1. Properties of Anesthetized Dataset and Awake Dataset.

Property	Anesthetized	Awake
Sample count	66	40
Motion artifacts	no/slight	obvious
Blood flow speed	medium	high

2. Evaluation Metric Details

In our experiments, the relative absolute error (Abs Rel) and root mean squared error (RMSE) are used as evaluation metrics. Below, we provide their formal definitions.

The Abs Rel is defined as:

$$\text{Abs Rel} = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \bar{y}_i|}{\bar{y}_i}, \quad (1)$$

where N is the number of pixels in an image. y_i is the model’s blood flow speed estimation for the i -th pixel, and \bar{y}_i is the corresponding flow speed measurement from ODT.

Abs Rel emphasizes the relative magnitude of errors, ensuring that both small and large values contribute equitably to the overall evaluation. This property makes it particularly suitable for assessing blood flow speed data that spans varying scales.

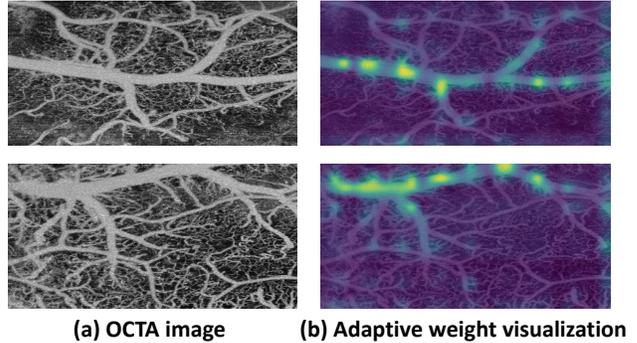


Figure 1. Adaptive weight visualization results. (a) shows the OCTA input image, and (b) overlays it with the adaptive weight, where brighter regions represent higher weights. Key regions like main vessels, branches, and junctions critical for blood flow estimation are effectively highlighted. This demonstrates the Adaptive Window Fusion module’s ability to capture vessels of varying sizes and complexities, adaptively focusing on important features for blood flow speed estimation.

RMSE is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y}_i)^2}. \quad (2)$$

RMSE emphasizes large errors due to its quadratic nature, making it sensitive to outliers or significant deviations between predictions and labels. This sensitivity is crucial for blood flow speed estimation, where avoiding substantial errors is important.

By combining Abs Rel and RMSE, we achieve a balanced evaluation framework: Abs Rel ensures robustness across varying scales, while RMSE highlights the impact of large errors. Together, they provide a comprehensive assessment of the model’s performance.

3. Adaptive Weight Visualization

Fig. 1 intuitively demonstrates the effectiveness of the proposed Adaptive Window Fusion module. For the dynamic weight $\mathcal{G} \in \mathbb{R}^{H_3 \times W_3 \times 3}$, we first compute its absolute val-

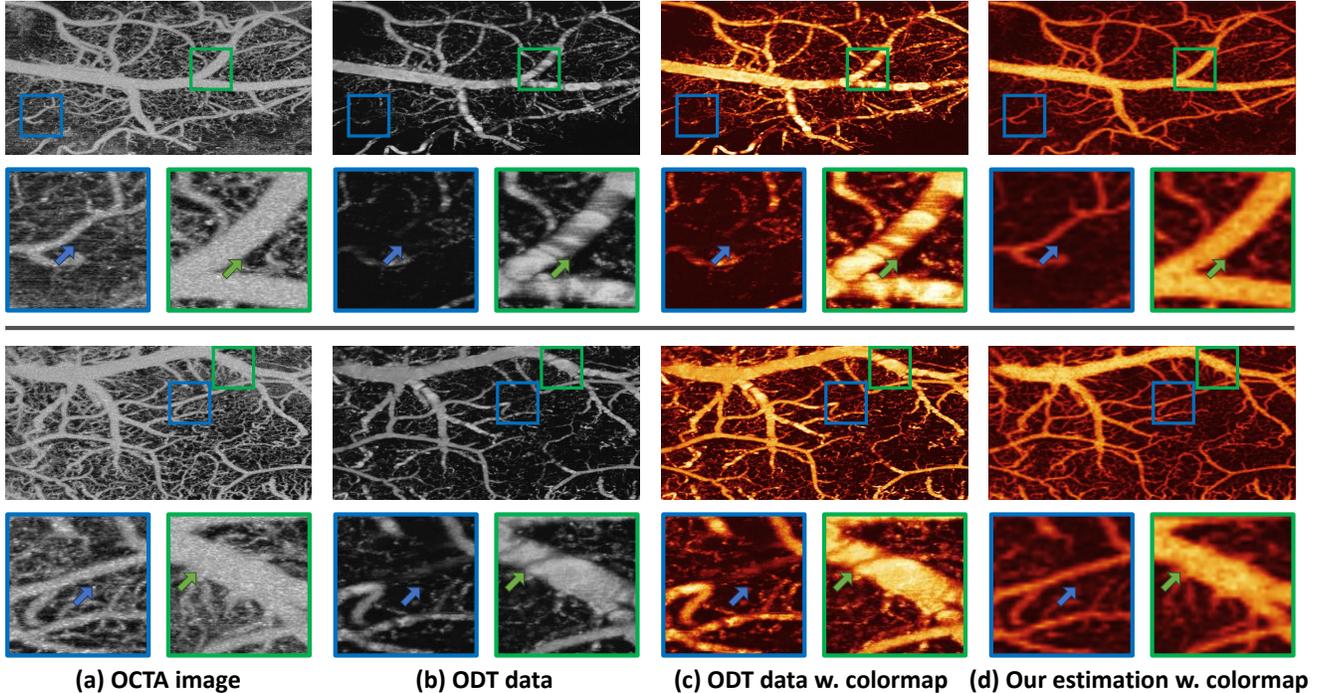


Figure 2. Additional qualitative results on the Anesthetized Dataset. Zoomed-in regions with arrows highlight details. ODT data is severely influenced by measurement artifacts, causing inconsistent results (marked by green arrows), which may even totally break the blood flow (marked by blue arrows). In contrast, our method mitigates the measurement artifacts, producing smoother and more continuous blood flow estimations.

ues. Then, for each spatial location, we extract the maximum value across its three channels, producing a spatial map. This map is subsequently upsampled to the original image size using bilinear interpolation, and overlaid onto the OCTA image.

The visualization results demonstrate that critical regions for blood flow speed estimation, such as main vessels, branches, and junctions, are emphasized by the adaptive weights. This validates the capability of the Adaptive Window Fusion (AWF) module to effectively capture vessels of varying sizes and complex structures, which significantly impact blood flow speed estimation. Moreover, the module adaptively adjusts its focus based on the input vascular image, ensuring alignment with diverse vascular patterns. This adaptiveness enhances the model’s ability to prioritize essential features, improving its performance in blood flow speed estimation.

4. Window Attention Blocks of AWF

We perform the ablation study on the number of window attention blocks in AWF. Tab. 2 shows that three blocks are sufficient to capture the relations, and additional blocks with increasing window sizes do not bring extra benefit. Here 6 window attention blocks in Tab. 2 means using 6 blocks with window sizes of 3×3 , 5×5 , 7×7 , 9×9 , 11×11 ,

Table 2. Effect of the number of window attention blocks in AWF. Results in bold are the best.

Metric	3 (ours)	4	5	6
Abs Rel ↓	0.328	0.330	0.331	0.330
RMSE ↓	6.661	6.712	6.706	6.684

and 13×13 , respectively. Hence we use window attention with three blocks in AWF.

5. Evaluation on Simulation Data

We create a simulation ODT dataset based on real data properties, largely removing the ODT artifacts (inconsistent/alternating flow), shown in Fig. 3. We then evaluate all methods on this simulation dataset. The results in Tab. 3 further validate the superiority of our method.

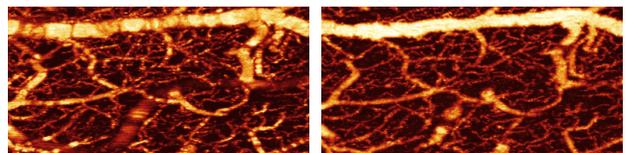


Figure 3. Raw ODT (left) and simulated ODT (right) images.

Table 3. Evaluation results on the simulation dataset. Results in bold are the best.

Method	Abs Rel ↓	RMSE ↓
BTS [2]	0.292	5.699
IEBins [4]	0.316	8.521
NeuWin [5]	0.283	5.102
Ord Ent [6]	0.276	5.310
Diff Depth [1]	0.378	6.100
ECoDepth [3]	0.403	7.289
OCTA-Flow (ours)	0.267	4.988

6. Additional Qualitative Results

Fig. 2 presents additional qualitative results from the Anesthetized Dataset. The ODT data shown in columns (b) and (c) exhibit significant measurement artifacts, primarily caused by the angle between the incident light and the vasculature. These artifacts appear as alternating bright and dim regions, resulting in inconsistent and erroneous measurements, as highlighted by the green rectangular region. Moreover, these artifacts can even break the blood flow in the vessels, as evident in the blue rectangular regions when comparing columns (a) with (b) and (c).

In contrast, our method (column (d)) generates consistent and smooth blood flow estimation results while preserving the continuity and integrity of the blood flow. This demonstrates the robustness of our approach in mitigating system-level artifacts in blood flow speed measurements, underscoring its superiority over laboratory-based ODT measurements in this aspect.

References

- [1] Yiquan Duan, Xianda Guo, and Zheng Zhu. Diffusiondepth: Diffusion denoising approach for monocular depth estimation. In *European Conference on Computer Vision*, pages 432–449. Springer, 2024. 3
- [2] Jin Han Lee, Myung-Kyu Han, Dong Wook Ko, and Il Hong Suh. From big to small: Multi-scale local planar guidance for monocular depth estimation. *arXiv preprint arXiv:1907.10326*, 2019. 3
- [3] Suraj Patni, Aradhya Agarwal, and Chetan Arora. Ecodepth: Effective conditioning of diffusion models for monocular depth estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 28285–28295, 2024. 3
- [4] Shuwei Shao, Zhongcai Pei, Xingming Wu, Zhong Liu, Weihai Chen, and Zhengguo Li. Iebins: Iterative elastic bins for monocular depth estimation. *Advances in Neural Information Processing Systems*, 36, 2023. 3
- [5] Weihao Yuan, Xiaodong Gu, Zuozhuo Dai, Siyu Zhu, and Ping Tan. Neural window fully-connected crfs for monocular depth estimation. In *Proceedings of the IEEE/CVF confer-*

ence on computer vision and pattern recognition, pages 3916–3925, 2022. 3

- [6] Shihao Zhang, Linlin Yang, Michael Bi Mi, Xiaoxu Zheng, and Angela Yao. Improving deep regression with ordinal entropy. *arXiv preprint arXiv:2301.08915*, 2023. 3