Supplementary Material: MicroDiffusion: Implicit Representation-Guided Diffusion for 3D Reconstruction from Limited 2D Microscopy Snapshots

Mude Hui[1*] Zihao Wei[2*] Hongru Zhu[3] Fei Xia[4†] Yuyin Zhou[1†]
[1]University of California, Santa Cruz [2]University of Michigan, Ann Arbor
[3]Johns Hopkins University [4]Ecole Normale Supérieure de Paris
muhui@ucsc.edu zihaowei@umich.edu {hongruzhu95, zhouyuyiner}@gmail.com fx43@cornell.edu

In this supplementary material, we detail the metrics we used in Sec. 1, including SSIM in Sec. 1.1, PSNR in Sec. 1.2, DICE in Sec. 1.3. Furthermore, we conduct an ablation study to find the best interpolation rate in Sec. 2. Finally, we visualize the MicroDiffusion reconstruction results at various step lengths on the vasculature dataset in Sec. 3.

1. Details of the metrics

In this section, we delineate the metrics employed for assessing the quality of image reconstruction. We consider a reference image x and a test image y, both being grey-level (8 bits) images of dimensions $M \times N$, drawn from respective sets X and Y. Three evaluation measures are utilized: the Structural Similarity Index Measure (SSIM), the Peak Signal-to-Noise Ratio (PSNR), and the Sørensen–Dice coefficient (DICE), each detailed below.

1.1. SSIM

The Structural Similarity Index Measure (SSIM) serves as a pivotal metric in quantifying the resemblance between two images. It evaluates three fundamental aspects: Luminance, Contrast, and Structure.

Luminance is quantified through the mean gray scale value of the pixels, encapsulated in the equation:

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \tag{1}$$

where μ_x and μ_y represent the mean luminance of images x and y, respectively. The constant C_1 prevents a zero denominator.

Contrast is gauged using the gray scale standard deviation, as:

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \tag{2}$$

where σ_x and σ_y denote the standard deviations of the images, and similarly, C_2 prevents a zero denominator.

Structure is assessed through correlation coefficients, formulated as:

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \tag{3}$$

with σ_{xy} being the covariance between x and y. The constant C_3 ensures non-zero denominators.

The aggregate SSIM value, encapsulated within the range $[0, 1]$, is derived as:

$$\text{SSIM}(x,y) = l(x,y) \cdot c(x,y) \cdot s(x,y), \tag{4}$$

offering a comprehensive measure of similarity. Notably, a SSIM score of 0 implies an absence of correlation between the images, whereas a score of 1 indicates identical images. This index is particularly adept at capturing perceptual differences, making it a robust tool in image quality assessment.

1.2. PSNR

The Peak Signal-to-Noise Ratio (PSNR) is another crucial metric, predominantly focusing on the ratio between the maximum possible power of a signal and the power of corrupting noise. It is articulated as:

$$\text{PSNR}(x,y) = 10\log_{10}\left(\frac{255^2}{\text{MSE}(x,y)}\right), \tag{5}$$

where $\text{MSE}(x,y)$, the Mean Squared Error between the two images, is computed as:

$$\text{MSE}(x,y) = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(x_{ij} - y_{ij})^2, \tag{6}$$

with x_{ij} and y_{ij} representing the pixel values at the ij^{th} position. The PSNR values range from 0 to ∞, where a higher value indicates superior image quality, reflective of lesser noise interference.

[*]Denotes equal first author
[†]Denotes equal senior author

1.3. DICE

The Sørensen–Dice coefficient (DICE), a statistical tool, quantifies the similarity between two sets. It is particularly effective in comparing the spatial arrangement of pixel values. The DICE is defined as:

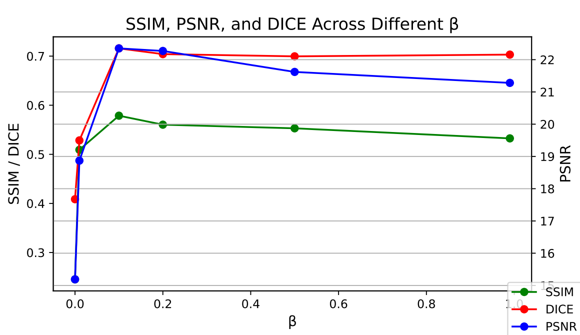
$$\text{DICE}(X, Y) = \frac{2|X \cap Y| + C}{|X| + |Y| + C}, \tag{7}$$

where $|X \cap Y|$ denotes the intersection size of sets X and Y, and $|X|$ and $|Y|$ are their respective sizes. The coefficient ranges from 0 to 1, with 1 indicating perfect agreement (complete overlap) and 0 denoting no overlap at all. This metric is particularly beneficial in scenarios where spatial correlation is a critical aspect of image similarity.

2. What is the best linear interpolation rate?

As demonstrated in the main paper, to incorporate global information and coherent 3D structures into the diffusion model, we employ a linear interpolation strategy between the Implicit Neural Representations (INR) output and the noisy image at each time step t. This approach is applied during both the training and testing phases of MicroDiffusion. Such integration of INR as prior knowledge is pivotal for guiding the diffusion process, particularly when dealing with limited 2D projection inputs.

An ablation study focusing on the interpolation rate γ was conducted, with the results summarized in Figure 1. Our findings indicate that the effectiveness of γ plateaus beyond a threshold of 0.1. Further increments in γ yield minimal improvements, as evidenced by a marginal decrease in both Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) metrics. This trend suggests that, while the incorporation of INR prior is beneficial, an excessive reliance on it, particularly in the absence of Gaussian noise, can compromise the model's generalization capabilities.



(a) SSIM, PSNR and DICE

Figure 1. Performance metrics across different linear interpolation rates.

3. Visualization of MicroDiffusion reconstruction of vasculature at various Step Lengths

We visualize the reconstructed images in Figure 2, which clearly demonstrates that the difficulty of reconstruction escalates with increasing step length, leading to a noticeable decline in model performance. This trend is quantitatively supported by the rapid decrease in metrics such as SSIM, PSNR, and DICE. Despite this challenge, it is noteworthy that satisfactory reconstruction quality is still achievable at step lengths of approximately 6 to 8. This finding is significant as it implies the potential to increase the speed of volumetric imaging by a factor of 6 to 8, enhancing imaging efficiency substantially. Looking ahead, our research aims to further improve model performance at even higher step lengths, pushing the boundaries of efficient and high-quality imaging in MicroDiffusion processes.

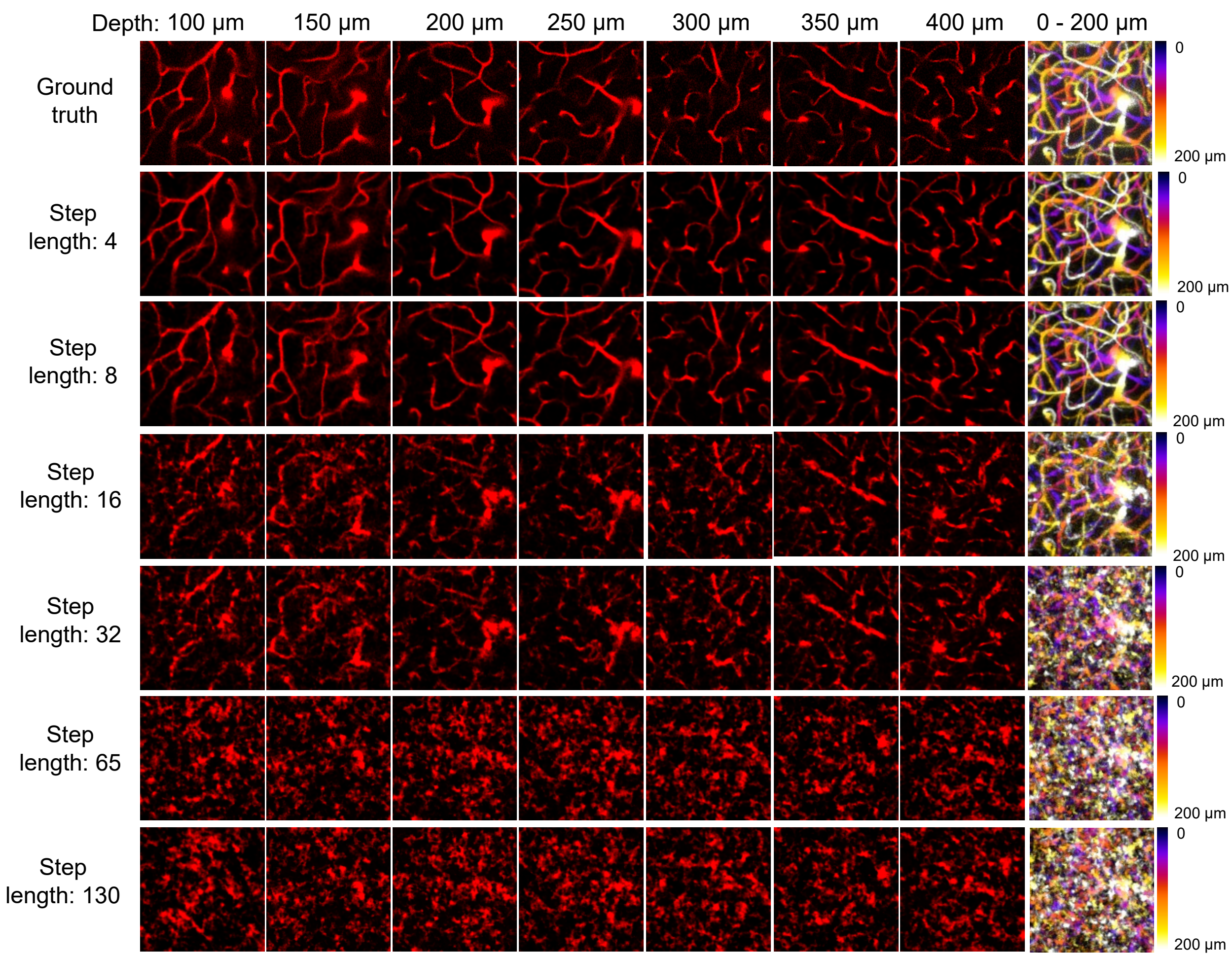


Figure 2. Reconstruction of the depth-resolved vasculature images and depth-resolved volumetric projections with different step lengths.