PGDM: Physics-Guided Noise-Free Diffusion Model Based on Point Spread Function for Light-Scattering Removal in Unpaired Biomedical Images

Supplementary Material

1. Introduction of Datasets

NIR-II Wide-Field Images of Mice (NIR-II WFM) Dataset: Unpaired in vivo wide-field NIR-IIa and NIR-IIb images of mice are used for model training and testing. The training and testing sets follow a 5:1 image ratio. NIR-IIa images depict various anatomical features, such as blood vessels, labeled with biocompatible probes. NIR-IIb images represent the same structures but are labeled with higher-performance NIR-IIb probes. Each image has a resolution of 640×512 pixels. [4]

NIR-IIa In-House (In-House) Dataset: For our In-House NIR-IIa Dataset, the imaging conditions (contrast agent and equipment) are different from those of the public NIR-II WFM Dataset. We used indocyanine green (ICG), a probe approved by the U.S. Food and Drug Administration (FDA) for clinical use. ICG has a fluorescence tail that extends into the NIR-IIa window, so detectable signals can still be captured in this range. Based on this property, we applied a 1000 nm long-pass filter to obtain the NIR-IIa fluorescence signal. Each mouse received an ICG dose of 2 mg/kg. The images were acquired using an InGaAs detector with 808 nm excitation, and the final resolution was 640×512 pixels. Our animal experiments were conducted in accordance with the principles of the Declaration of Helsinki and were approved by the institutional ethics committee.

Please note that our purpose in collecting an in-house dataset is to assess whether our model can maintain high performance across different imaging devices and contrast agents, thereby demonstrating stronger cross-device stability and generalization ability.

NIR-II Light-Sheet Microscopy (NIR-II LSM) Dataset: Unlike wide-field in vivo images, ex vivo LSM images capture the cerebral vasculature of mice. Images are acquired in the NIR-IIa window using the p-FE nanoparticle probe and in the NIR-IIb window using the PbS/ CdS core-shell quantum dot (CSQD) probe. These images are used as an unpaired dataset for model training and evaluation. Each image has a resolution of 320×256 pixels. [4]

Structured Illumination Microscopy Light-Scattering (SIM-LS) Dataset: We also evaluate our model on paired super-resolution microscopy images of living mammalian cells reconstructed by different methods. The Wiener-reconstructed images containing light scattering artifacts serve as input, while the BF-reconstructed images suppressing light scattering act as ground truth. The dataset includes three subcellular structures from two cell types: mitochondria and endoplasmic reticulum in COS-7 cells (monkey kidney) and actin filaments in U2OS cells (human osteosarcoma). All images were acquired using structured illumination microscopy (Mo et al. 2023). [5]

Lattice Light-Sheet Microscope (LLSM) Dataset: To further assess the versatility of our model across different tasks, we evaluate it on a paired dataset of out-of-focus and in-focus images. This task is relevant because the image degradation caused by instrumental misalignment resembles light scattering. The reference images are maximum intensity projections (MIPs) of fixed HeLa cells with Alexa Fluor 488-labeled actin networks, acquired under optimal alignment conditions on a lattice light-sheet microscope (LLSM). The corresponding offset images are generated through a controlled 1 μm axial misalignment between the excitation light sheet and the detection focal plane. [1]

2. Details of Adaptive Fussion in Dual Sample

As described in our paper, we obtain two sampling results based on residual- and convolution-based processes, respectively denoted as $\hat{x}_0^{res}$ and $\hat{x}_0^{cov}$. Fusion weights are then adaptively computed based on GM [7] and LBP [3] to combine these results to produce the final high-quality results without light scattering. The procedure for calculating the fusion weights is detailed below.

GM primarily reflects the structural information of an image. It is derived by calculating the partial derivatives of each pixel along two orthogonal directions, as shown in Eq.1.

$$G(i,j) = \sqrt{G_X^2(i,j) + G_Y^2(i,j)} \tag{1}$$

LBP features capture the local information, effectively preserving details of weak texture structures. Since LBP extraction is based on computing the relative intensity of local neighborhoods, it can indirectly reflect information such as the true scattering-free background. The formula is presented in Eq.2.

$$LBP_{P,R}(i,j) = \sum_{n=0}^{P-1} s(I_n - I_{(i,j)}) \cdot 2^n \tag{2}$$

where $R = 1.0$ and $P = 8$ represent the radius and the number of neighboring points on the circumference centered at pixel $I_{(i,j)}$, respectively. The gray values of non-integer pixel locations are calculated using bilinear interpolation.

And the function $s(\cdot)$ is shown in Eq. 3.

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{3}$$

For the computation of GM similarity, given two images with GM values GM_1 and GM_2, the similarity at each pixel is calculated in Eq.4.

$$S^{GM}(i,j) = \frac{2GM_1(i,j) \cdot GM_2(i,j) + C_{GM}}{GM_1^2(i,j) + GM_2^2(i,j) + C_{GM}} \tag{4}$$

where we set $C_{GM} = 160$ as suggested in [7]. And LBP similarity $S^{LBP}(i,j)$ can be calculated in a similar manner, as shown in Eq. 5. And we set $C_{LBP} = 100$.

$$S^{LBP}(i,j) = \frac{2LBP_1(i,j) \cdot LBP_2(i,j) + C_{LBP}}{LBP_1^2(i,j) + LBP_2^2(i,j) + C_{LBP}} \tag{5}$$

The overall fusion similarity is defined in Eq.6. And $\alpha = 0.7$ is used to adjust the weight between GM and LBP similarity.

$$S^{GL}(i,j) = \alpha \cdot S^{GM}(i,j) + (1-\alpha) \cdot S^{LBP}(i,j) \tag{6}$$

After obtaining the similarity $S_{res}^{GL}(i,j)$ between the residual-based de-scattered image $\hat{x}_0^{res}$ and the original scattering image y_0, as well as the similarity $S_{cov}^{GL}(i,j)$ between the convolution-based de-scattered image $\hat{x}_0^{cov}$ and y_0, the fusion weights are determined accordingly based on these similarities, as shown in Eq.7.

$$\begin{aligned} w_{res}(i,j) &= \frac{e^{S_{res}^{GL}(i,j)}}{e^{S_{res}^{GL}(i,j)} + e^{S_{cov}^{GL}(i,j)}} \\ w_{cov}(i,j) &= \frac{e^{S_{cov}^{GL}(i,j)}}{e^{S_{res}^{GL}(i,j)} + e^{S_{cov}^{GL}(i,j)}} \end{aligned} \tag{7}$$

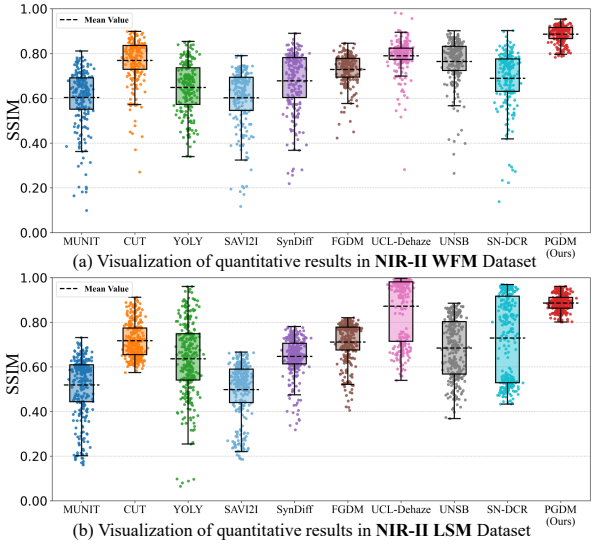
The final fused image is then obtained as shown in Eq.8, where $\odot$ denotes element-wise (pixel-wise) multiplication.

$$\hat{x}_0^{fused} = w_{res} \odot \hat{x}_0^{res} + w_{cov} \odot \hat{x}_0^{cov} \tag{8}$$

3. Stability Analysis of Model Performance on the Training Datasets

To evaluate the stability of different models on the training datasets, we analyze the distribution of SSIM values across all samples in both the NIR-II WFM and NIR-II LSM datasets, as shown in Fig. 1. Models that exhibit narrow interquartile ranges (IQRs) and the density of high-quality samples are considered more stable and reliable.

Our proposed PGDM demonstrates the most stable performance in both datasets. The IQR of PGDM is significantly narrower, and the overall SSIM values cluster closely around the mean with minimal low-end outliers. This indicates that the PSF-based diffusion mechanism not only enhances reconstruction quality but also ensures consistent behavior across diverse samples.



(a) Visualization of quantitative results in **NIR-II WFM** Dataset

(b) Visualization of quantitative results in **NIR-II LSM** Dataset

Figure 1. Visualizations of quantitative results in NIR-II WFM and NIR-II LSM Datasets.

Overall, these results verify that PGDM achieves superior stability compared to all competing methods, reflecting both strong training convergence and high intra-dataset generalization ability.

4. More Qualitative Results on Different Datasets

NIR-II WFM Dataset: Fig. 2 presents additional results of different methods on the NIR-II WFM dataset. With the integration of the PSF into the diffusion process, the forward process in PGDM can more accurately reflect the physical degradation caused by light scattering. And the integration helps preserve structural details and enhances the contrast between blood vessels and the background. In contrast, other methods show varying degrees of real structural detail loss, leading to suboptimal image quality in their generated results.

NIR-II LSM Dataset: More comparative results of various methods on the NIR-II LSM dataset are shown in Fig. 5. Unlike the macroscopic mouse fluorescence images in the NIR-II WFM and In-House datasets, this dataset corresponds to fluorescence light-sheet microscopy data. The original image is shown in yellow pseudo-color, while the outputs of different methods are displayed in blue pseudo-color. Although most methods partially enhance vascular structures, many fail to preserve details adequately. In contrast, our proposed PGDM (in green pseudo-color) clearly delineates the vascular structures, demonstrating its strong cross-modal capabilities in fluorescence microscopy. **SIM-LS Dataset:** Fig. 6 shows visual comparisons of different methods on the SIM-LS dataset, including actin filaments

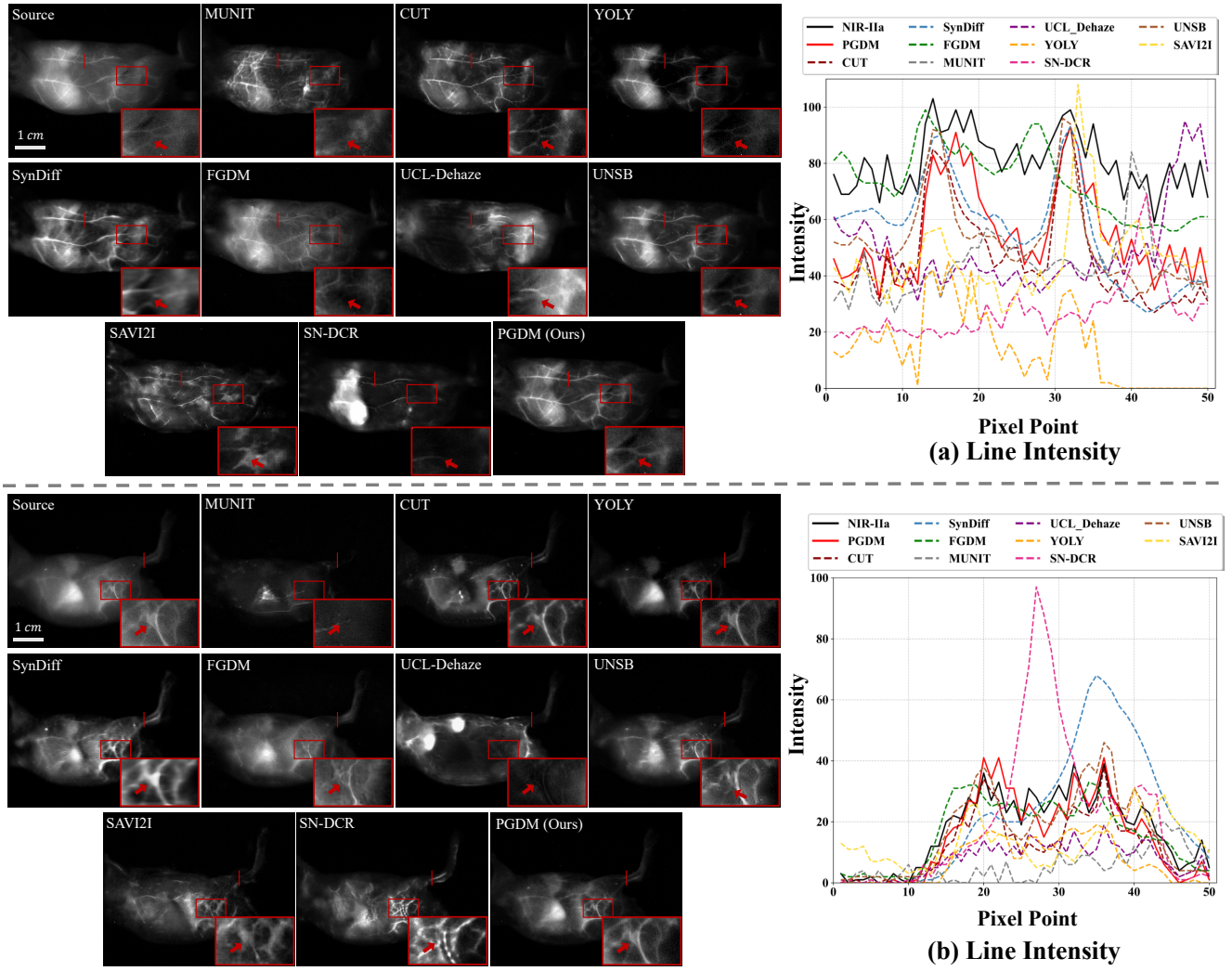


Figure 2. Visual comparisons of different methods on the NIR-II WFM dataset, with zoomed-in regions highlighting structural details, scale bar: 1 cm. The intensity profiles along the red line in the mouse fluorescence images are shown in the left of the figure.

in U2OS cells and mitochondria in COS-7 cells. Zoomed-in regions and error maps (brighter regions indicate larger discrepancies) highlight structural details. Most competing methods often fail to preserve fine details, especially across different structures and cell types. In contrast, our proposed PGDM clearly recovers structural details and maintains close correspondence with the ground truth,

Fig. 7 shows direct visual comparisons between PGDM outputs, source images, and target images. Panels (a) and (b) depict actin filaments in U2OS cells and endoplasmic reticulum in COS-7 cells, respectively. PGDM effectively suppresses light scattering preserving anatomical structures, demonstrating its strong generalization across diverse subcellular structures and cell types.

LLSM Dataset: Figure 4 presents visual comparisons on the defocused dataset, where images were intentionally offset from the focal plane to simulate defocus-induced degradation. PGDM effectively reconstructs actin networks and preserves structural details under defocus. These results demonstrate that PGDM not only mitigates light scattering but also generalizes robustly to other imaging challenges, such as defocus, highlighting its strong cross-task versatility.

In-House Dataset: Fig. 3 presents visual comparisons of different methods on the In-House dataset, which was acquired under imaging conditions distinct from the public NIR-II WFM, including a different contrast agent (ICG) and detection system. Despite these variations, PGDM consistently reconstructs vascular structures with high fidelity, preserving fine details and vessel morphology that are partially lost in outputs from other competing methods. Results demonstrate PGDM's robust cross-device performance and

generalization across different contrast agents in NIR-IIa fluorescence imaging.

5. Code Repository

To ensure reproducibility, we also provide detailed descriptions of the network configuration, hyperparameter settings, and training strategies in the **Experimental Details** section. The complete code will be made publicly available on GitHub upon acceptance.

6. Experimental Details

All experiments are conducted on an NVIDIA RTX 4090 GPU. The code for all competitions was obtained from their official implementations to ensure consistency and reproducibility. Models are trained in a Python-3.9 Conda environment using PyTorch 1.8.0 + CUDA 11.1.

We train our models by the Adam optimizer and use the cosine learning rate decay for training both Fre-DCP based network G_θ and discriminator D_ϕ. We also apply an exponential moving average (EMA) to G_θ. And the time steps of forward diffusion process in PGDM is 4. We sumarrize the optimization hyper-parameters in Table. 1.

Table 1. Optimization hyper-parameters.

	Hyper-parameters
Initial learning rate for G_θ	1.5×10^{-4}
Initial learning rate for D_ϕ	10^{-4}
Adam optimizer β_1	0.5
Adam optimizer β_2	0.9
EMA decay	0.999
Batch size	1
Epochs	160

Our proposed Fre-DCP based network G_θ integrates the Fre-DCP module for domain-consistent guidance in the reverse diffusion process, built upon a U-Net structure [6] with ResNet and attention blocks. Detailed key hyper-parameters are provided in Table. 2.

Table 2. Key hyper-parameters for G_θ.

	Hyper-parameters
Number of diffusion steps	4
ResNet blocks per scale	2
Number of resolutions	4
Initial feature channels	64
Channel multipliers per scale	[1,2,2,2]
Activation function	SiLU
Time embedding type	Positional
Fir kernel	[1,3,3,1]

For the time-dependent discriminator D_ϕ, its role is to determine whether a real or generated $\hat{x}_{t-1}$ is correctly paired with the corresponding real x_t at each time step t, as described in our main paper. In addition, we append a minibatch standard deviation layer [2] after all ResNet-based downsampling blocks to enhance discrimination stability. Further architectural details are provided in Table. 3.

Table 3. The hyper-parameters for D_ϕ.

	Hyper-parameters
Number of downsampling blocks	6
Timestep embedding dimension	128
Initial feature channels	64
Start-conv output channels	$2 \times 64 = 128$
Activation function	LeakyReLU(0.2)
Dimension change in final linear layer	512→1

References

[1] Chad M Hobson, Min Guo, Harshad D Vishwasrao, Yicong Wu, Hari Shroff, and Teng-Leong Chew. Practical considerations for quantitative light sheet fluorescence microscopy. *Nature Methods*, 19(12):1538–1549, 2022. 1

[2] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020. 4

[3] Song Ke-Chen, YAN Yun-Hui, CHEN Wen-Hui, et al. Research and perspective on local binary pattern. *Acta Automatica Sinica*, 39(6):730–744, 2013. 1

[4] Zhuoran Ma, Feifei Wang, Weizhi Wang, Yeteng Zhong, and Hongjie Dai. Deep learning for in vivo near-infrared imaging. *Proceedings of the National Academy of Sciences*, 118 (1):e2021446118, 2021. 1

[5] Yanquan Mo, Kunhao Wang, Liuju Li, Shijia Xing, Shouhua Ye, Jiayuan Wen, Xinxin Duan, Ziying Luo, Wen Gou, Tongsheng Chen, et al. Quantitative structured illumination microscopy via a physical model-based background filtering algorithm reveals actin dynamics. *Nature Communications*, 14 (1):3089, 2023. 1

[6] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *International Conference on Learning Representations*, 2021c. 4

[7] Wufeng Xue, Lei Zhang, Xuanqin Mou, and Alan C Bovik. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE transactions on image processing*, 23(2):684–695, 2013. 1, 2

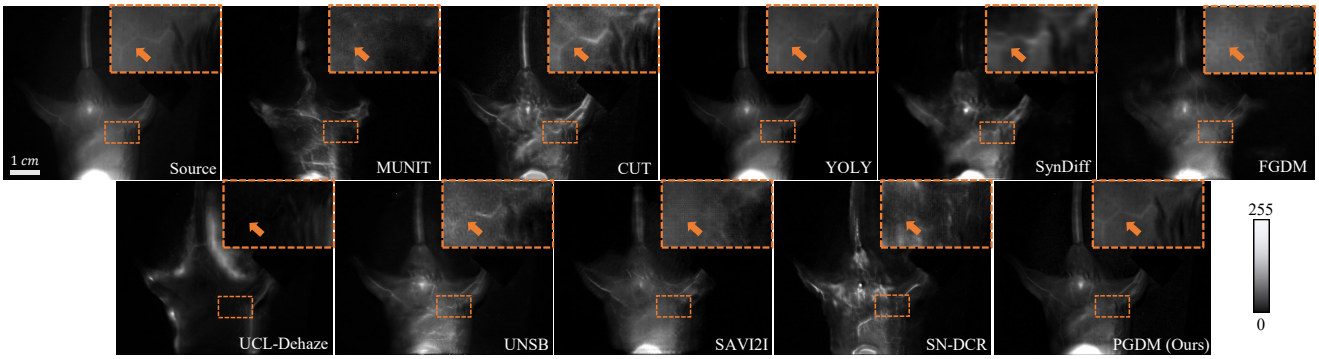


Figure 3. Visual comparisons of different methods on In-House Dataset, with zoomed-in regions highlighting structural details, scale bar: 1 cm.

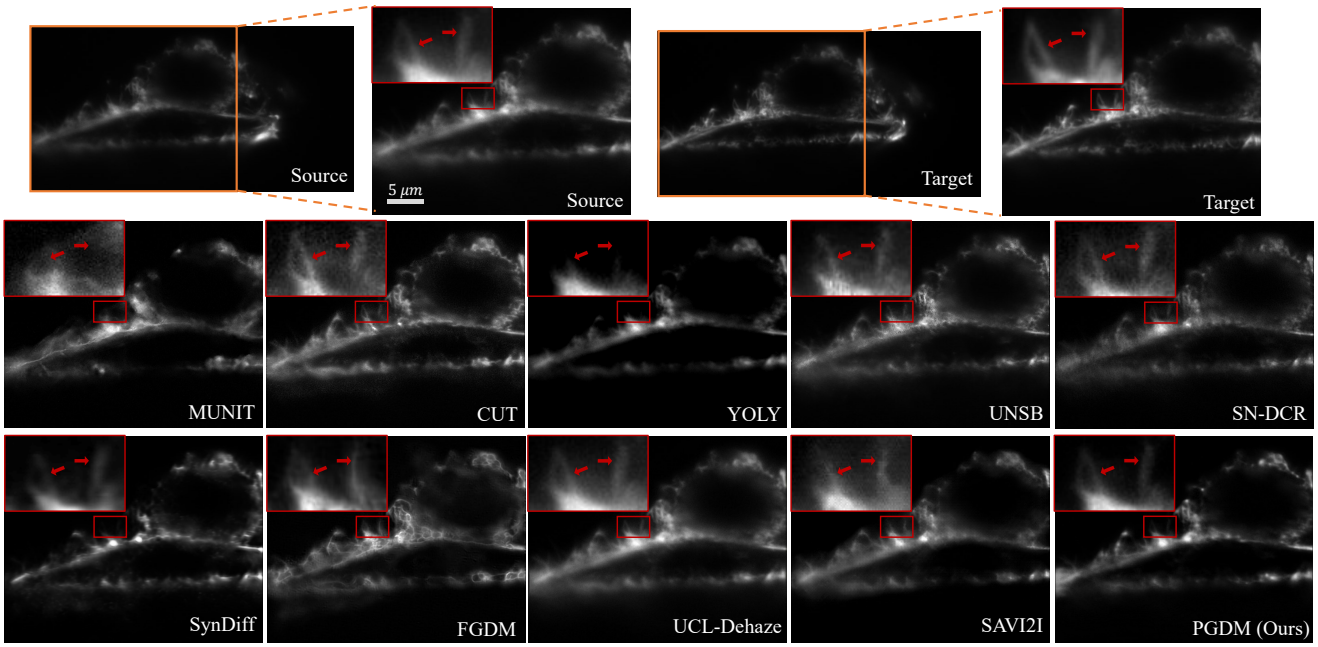


Figure 4. Visual comparisons of different methods on LLSM Dataset, with zoomed-in regions highlighting structural details, scale bar: 5 μm.

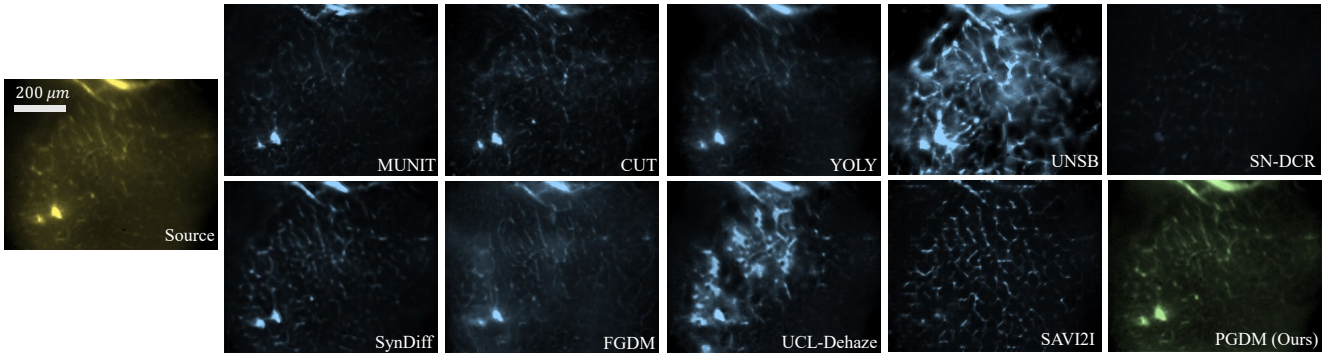
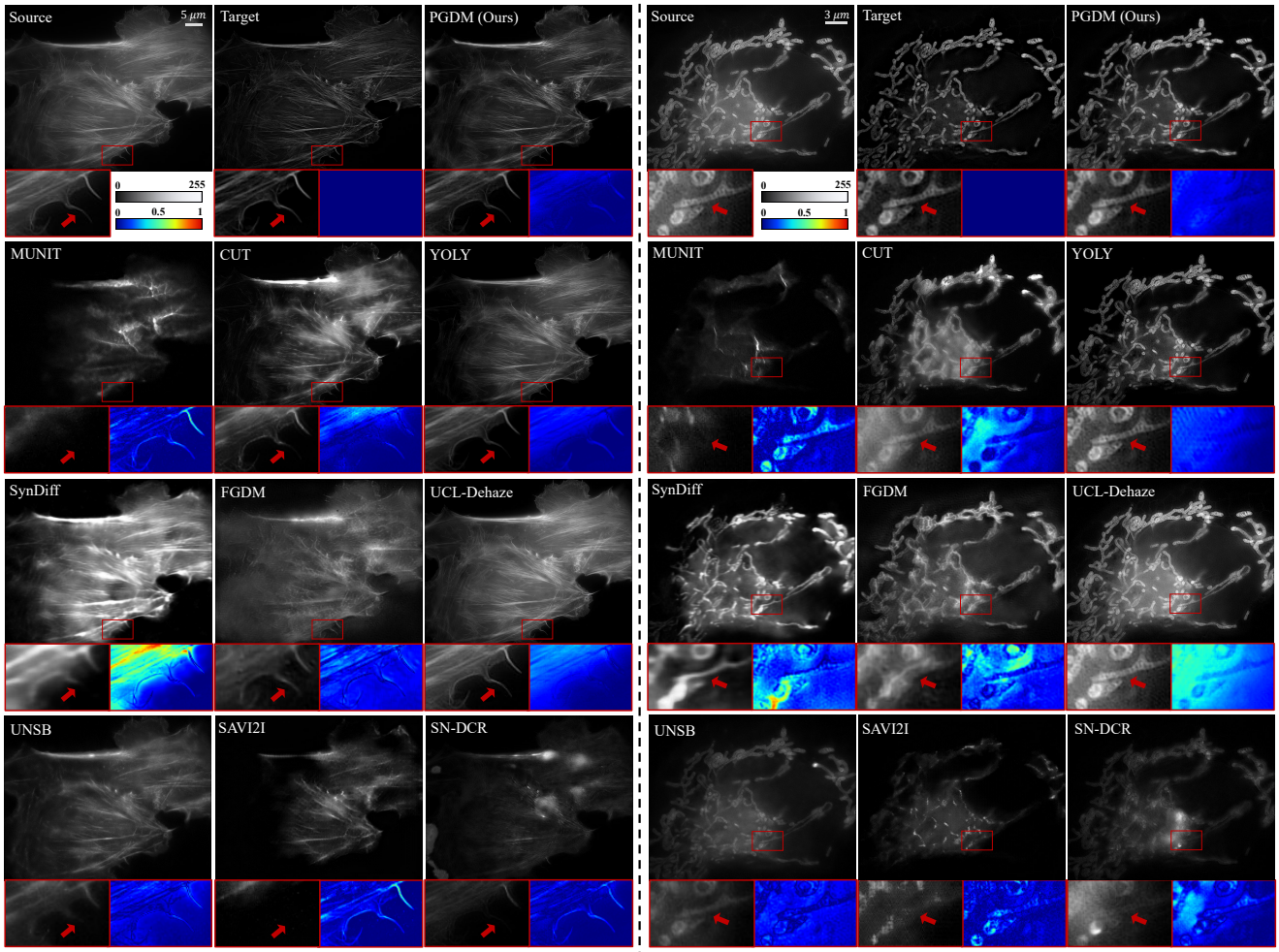
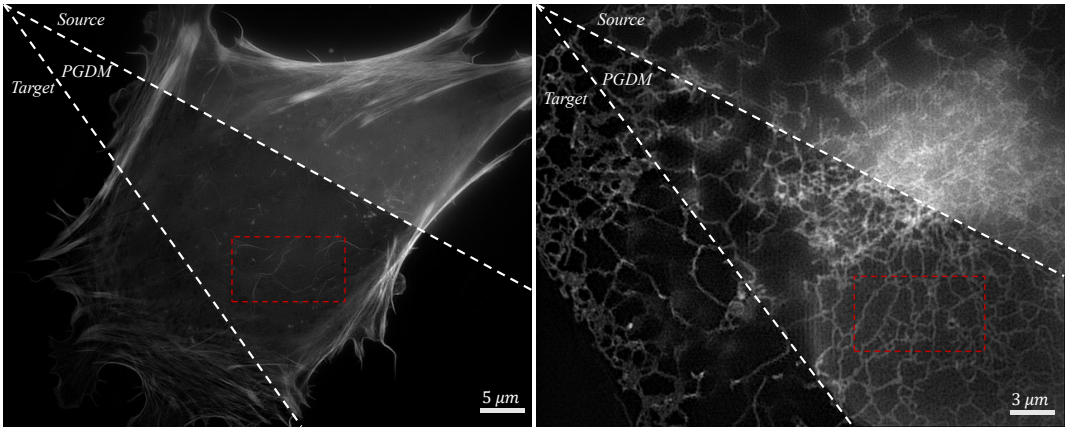


Figure 5. Visual comparisons of different methods on NIR-II LSM Dataset, scale bar: 200 μm.The source image is shown in yellow pseudo-color and the outputs of different methods are displayed in blue pseudo-color. The result of our proposed PGDM is in green pseudo-color.



(a) Image of actin filaments labeled with LifeAct-EGFP in a live U2OS cell. ***(b) Images of mitochondria in COS-7 cells stained with MitoTracker Green.***

Figure 6. Visual comparisons of different methods on SIM-LS Dataset. Zoomed-in regions of details and their corresponding error maps are shown in red boxes. In the error maps, brighter textures indicate a larger discrepancy from the target image. (a) Images of actin filaments in U2OS cells, scale bar: 5 μm. (b) Images of mitochondria in COS-7 cells, scale bar: 3 μm.



(a) The actin filaments (LifeAct-EGFP) of a living U2OS cell

(b) The endoplasmic reticulum labeled with mCherry-Cytb5ER in living COS-7 cells

Figure 7. Visual comparisons of PGDM on SIM-LS Dataset. (a) The actin filaments (LifeAct-EGFP) of a living U2OS cell, scale bar 5 μm. (b) The endoplasmic reticulum labeled with mCherry-Cytb5ER in living COS-7 cells, scale bar, 3 μm.