Supplementary Material

1. The Motivation in Details

Fluorescence microscopy is indispensable for volumetric cell and tissue biology, yet axial resolution is commonly 3-5 times worse than lateral resolution because of anisotropic point spread functions (PSFs), which degrades 3D analysis fidelity [8]. Optical designs such as Bessel-beam plane illumination and multiview light-sheet imaging can approach isotropy under careful alignment and calibration [3, 10]. However, engineered beams introduce side lobes and out-of-focus background that elevate photobleaching and phototoxicity risks in living specimens while increasing system complexity [4, 13]. These trade-offs have motivated deep learning reconstruction as a scalable route to isotropic 3D fluorescence imaging.

A first limitation in deep learning isotropic reconstruction is the mismatch between synthetic training pairs and the true imaging physics of fluorescence microscopes. Many pipelines build training data by convolving lateral high resolution images with a Gaussian kernel and by assuming a shift invariant point spread function, which causes models to learn the inversion of a surrogate blur rather than the physical degradation [9]. Real PSFs vary with numerical aperture, emitter orientation and aberrations, and they change with depth due to refractive index mismatch, so a single fixed kernel is insufficient [1, 7, 12]. The noise is also dominated by a mixture of Poisson shot noise and read noise, which departs from simple Gaussian assumptions and further widens the domain gap [15].

A second central limitation is the lack of explicit volumetric geometry during training and inference, which leaves cross slice continuity weak. Slice wise processing neglects depth correlations; under strong anisotropy thin or tubular structures drift in position and thickness, branches merge or break, and surfaces become stair stepped along the z direction [14]. Unpaired distribution matching can further compound the issue: matching the marginal target distribution without a volumetric anchor may hallucinate or delete features, producing volumes that appear plausible in one view but violate the true topology [2, 5]. 3D context stabilizes reconstruction and downstream quantification because features such as vessel caliber, neurite diameter, and sheet like boundaries are defined across many planes rather than a single slice.

MicroFM addresses both limitations by coupling instrument matched and spatially varying PSFs with geometry aware flow matching for isotropic reconstruction. We estimate the physical PSFs from axial images using a SFE-Net [11] and then synthesize paired training data that match the target microscope rather than an idealized Gauss kernel. A continuous implicit neural field supplies a volumetric prior that preserves morphology across slices. A flow matching network anchored to the low resolution images and guided by this prior transports the data toward a high quality endpoint with straightened trajectories and few steps. Across multiple fluorescence systems and datasets, this design yields sharper structures, more isotropic spectra and consistent gains in both full reference and no reference metrics.

2. SFE-Net for Physical PSF Prediction

SFE-Net [11] is a spatial and frequency encoding network that regresses an aberrated point spread function from a single microscopy patch by fusing a spatial branch with a Fourier branch and decoding to a compact kernel map. In our pipeline we adopt SFE-Net and build training pairs directly from the datasets' lateral XY slices. We construct semi-synthetic inputs by convolving the XY exemplars with physical PSFs generated from a Zernike expansion of the pupil function and by adding realistic photon and read noise, so the network learns a mapping from degraded content to its underlying physical blur.

To better cover higher order aberrations that arise in high numerical aperture fluorescence microscopes, we extend the set of Zernike modes used to construct the pupil in the SFE-Net based point spread function estimation stage from Wyant indices 4 through 18 to Wyant indices 4 through 25, which goes beyond the configuration used in the original SFE-Net work. As summarized in Table 1, each Zernike mode is indexed by a pair of integers (n, m), where n denotes the radial order and m denotes the azimuthal frequency. The value of m determines the number and orientation of angular lobes in the wavefront, while the value of n controls how many times the phase oscillates from the center of the pupil to its edge. This extended range still includes primary astigmatism, coma and spherical aberration and additionally introduces second and third order spherical components together with multi lobed terms such as trefoil, quadrafoil and pentafoil, which are characteristic of high numerical aperture imaging in thick or refractive index mismatched specimens. Compared with the original basis with eighteen terms, the expanded set provides a closer representation of the combined system induced and sample induced wavefront, so that the physical point spread functions used in MicroFM to synthesize training pairs reproduce the axial blur of the target microscope more faithfully. This improved agreement between simulated and real degradation helps the MicroFM reconstruction network learns 3D structure that is more realistic and more isotropic.

Table 1. Zernike modes (Wyant indices $j = 4$–25) used to construct the pupil function in the PSF synthesis stage.

Wyant index j	(n, m)	**Classical aberration**	**Physical interpretation**
4	$(2, +2)$	primary astigmatism	Elliptical blur along horizontal and vertical axes.
5	$(2, -2)$	primary astigmatism	Elliptical blur rotated by 45°.
6	$(3, +1)$	primary coma	Comet-shaped PSF with a tail along the x direction.
7	$(3, -1)$	primary coma	Comet-shaped PSF with a tail along the y direction.
8	$(4, 0)$	primary spherical	Axially elongated PSF with symmetric side lobes.
9	$(3, +3)$	trefoil	Three-lobed PSF with 120° symmetry.
10	$(3, -3)$	trefoil	Three-lobed PSF rotated with respect to the axes.
11	$(4, +2)$	secondary astigmatism	Astigmatism whose ellipticity varies with pupil radius.
12	$(4, -2)$	secondary astigmatism	Rotated secondary astigmatism with diagonal axes.
13	$(5, +1)$	secondary coma	Higher-order coma with a longer, curved tail along x.
14	$(5, -1)$	secondary coma	Higher-order coma with a longer, curved tail along y.
15	$(6, 0)$	secondary spherical	Higher-order spherical aberration with strong axial broadening.
16	$(4, +4)$	quadrafoil	Four-lobed aberration with lobes along the axes.
17	$(4, -4)$	quadrafoil	Four-lobed aberration rotated by 45°.
18	$(5, +3)$	higher-order trefoil	Finer three-lobed structure near the pupil edge.
19	$(5, -3)$	higher-order trefoil	Rotated version of the higher-order trefoil.
20	$(6, +2)$	tertiary astigmatism	Third-order astigmatism with complex radial dependence.
21	$(6, -2)$	tertiary astigmatism	Rotated tertiary astigmatism with diagonal axes.
22	$(7, +1)$	tertiary coma	Third-order coma with a long, curved tail along x.
23	$(7, -1)$	tertiary coma	Third-order coma with a long, curved tail along y.
24	$(8, 0)$	tertiary spherical	Third-order spherical aberration with strong axial broadening.
25	$(5, +5)$	pentafoil	Five-lobed aberration with a petal-like PSF pattern.

After training, we run SFE-Net on axial XZ and YZ slices and infer per-slice instrument-matched PSFs on a fixed 21×21 grid; these kernels act as the physics operator for MicroFM data synthesis and reconstruction, narrowing the gap between synthetic and real optical formation.

Let the physical PSF be defined by a Zernike pupil model with Fourier propagation,

$$h_{\mathrm{phys}}(x, y) = \left| \mathcal{F}\big\{ A(\rho)\, \exp\big(\mathrm{i} \textstyle\sum_n a_n Z_n(\rho, \theta)\big) \big\} \right|^2, \quad (1)$$

and let a semi-synthetic observation follow the standard fluorescence formation model

$$I \sim \mathrm{Poisson}\big(S * h_{\mathrm{phys}}\big) + \mathcal{N}(0, \sigma^2), \quad (2)$$

with S the lateral exemplar and $*$ convolution. SFE-Net is supervised to predict $\hat{h}$ on a 21×21 grid. The loss is the mean squared error between the predicted kernel and the physical target,

$$\mathcal{L}_{\mathrm{PSF}} = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} \big(\hat{h}_{ij} - h_{ij}^{\mathrm{phys}}\big)^2, \qquad H = W = 21, \quad (3)$$

which encourages accurate recovery of instrument-matched blur that is subsequently reused for physics-guided degradation and MicroFM reconstruction.

3. Additional Experiment

3.1. Detail of Experiment Setting

Following Volume Tells [6], we construct a fair and reproducible baseline where lateral XY slices provide full reference supervision and axial XZ or YZ slices are evaluated with no reference metrics. Concretely, for the full reference track we create controlled low-quality inputs from the native XY images by first estimating instrument-matched physical point spread functions from the axial views with our physics-guided PSF module, then convolving each XY slice with the predicted PSFs and finally decimating in plane by an integer factor that matches the native anisotropy. The convolved and downsampled XY slices are used as inputs, and the original XY slices serve as reference targets to compute PSNR, SSIM, LPIPS and VIF, consistent with the protocol of Volume Tells. For the axial track, we feed the raw XZ or YZ slices without extra synthesis and report NIQE, PIQE and NRQM. All competing unsupervised baselines are evaluated on exactly the same inputs and test splits to ensure comparability. As indicated by the voxel sizes in Table 2, we downsample the XY plane by a factor of five for all datasets except Mouse Liver, which is downsampled by a factor of three.

3.2. Additional Ablation Experiment

We further study the influence of the number of 3D prior slices used to construct the volumetric geometry prior in MicroFM. Increasing the number of prior slices from four

Table 2. Detailed information on the microscopy datasets used in MicroFM

Dataset	Sample	Source	System	Pixel number $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$	Pixel size $(\mathbf{d_x}, \mathbf{d_y}, \mathbf{d_z})$ μm
Dense neuron cluster	Dense neuron clusters (C57BL/6J, S1 AAV-YFP)	Self-Net [8]	CS-fMOST	$(893, 1153, 194)$	$(0.20, 0.20, 1.00)$
Cleared Thy1-GFP	Cleared Thy1-GFP mouse brain	Self-Net [8]	Confocal	$(1024, 1024, 90)$	$(0.21, 0.21, 1.00)$
Mouse Kidney	Cleared mTmG mouse kidney	Self-Net [8]	Two-photon	$(1024, 1024, 150)$	$(0.21, 0.21, 1.00)$
Mouse Liver	Cleared mTmG mouse liver	Self-Net [8]	Wide-field	$(1800, 1700, 246)$	$(0.32, 0.32, 1.00)$

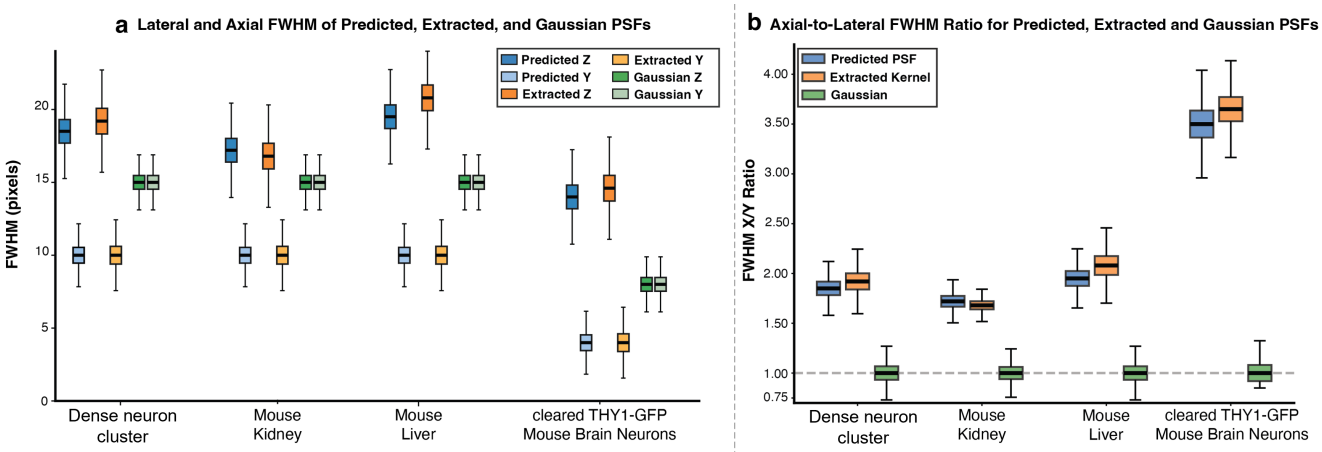

Figure 1. Spatial characterization of PSF anisotropy. (a) Box plots of lateral and axial FWHM for predicted, extracted and Gaussian PSFs across the four datasets, measured on axial YZ sections. Predicted PSFs reproduce the axial elongation of the extracted kernels, whereas Gaussian PSFs remain nearly isotropic. (b) Axial-to-lateral FWHM ratio, $\mathrm{FWHM_{ax}}/\mathrm{FWHM_{lat}}$. Values near one correspond to isotropic blur, as in the Gaussian baseline, while ratios above one indicate axial elongation. Across all datasets the anisotropy of the predicted PSFs closely matches that of the extracted axial kernels and differs clearly from the Gaussian PSFs.

Table 3. Ablation on the number of 3D prior slices n. Best result is shown in **bold**.

No. of 3D prior slices n	Full Reference Metrics			
	PSNR↑	SSIM↑	VIF↑	LPIPS↓
4	35.048	0.942	0.365	0.085
6	**40.186**	**0.964**	**0.378**	**0.075**
8	35.666	0.944	0.359	0.082
10	34.289	0.940	0.357	0.088
12	34.953	0.943	0.352	0.089

to 6 leads to a clear improvement across all full reference metrics, indicating that a moderate axial neighborhood provides richer structural cues and allows the flow matching reconstruction network to better preserve fine morphology while enhancing contrast. When we use more than 6 slices, the performance consistently degrades: PSNR and SSIM decrease, VIF drops, and LPIPS increases. This trend suggests that including too many neighboring slices causes the prior to over smooth along the axial direction and blend incompatible structures from distant planes, which weakens the effective guidance of the prior and partially cancels the benefit of anchoring the probability flow to the observed low quality slice. In contrast, 6 slices provide a compact yet informative axial context that is sufficient to propagate volumetric geometry without washing out local details. These results confirm that MicroFM benefits most from a balanced volumetric prior that couples a limited axial support with observation anchored flow matching, and they justify the choice of 6 prior slices in all reported experiments.

3.3. Predicted PSFs Analysis

We estimate reference axial point spread functions directly from acquired YZ slices using an image based edge spread procedure. Each YZ slice is normalised in intensity and filtered with Sobel operators to compute gradients and detect strong edges. For each selected edge we determine the local normal direction, sample a one dimensional intensity profile across the edge, align the profiles by the position of maximum gradient, and average them to obtain an edge spread function. A finite difference of this profile yields a line spread function. We compute separate line spread functions for mainly lateral and mainly axial edge orientations as well as an additional profile from oblique edges. These one dimensional line spread functions are lifted to the YZ plane and combined into a single anisotropic PSF. When too few reliable edges are present we instead estimate a Gaus-

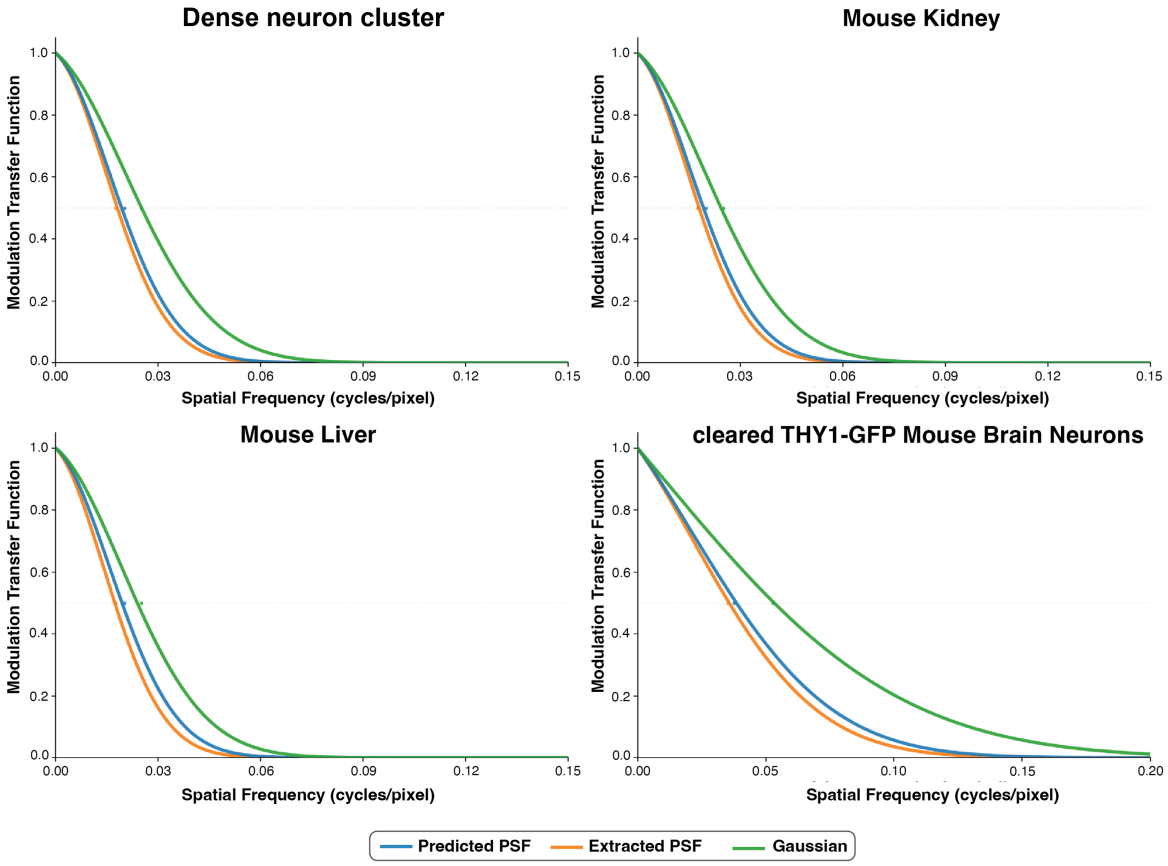

Figure 2. Axial modulation transfer functions of predicted, extracted and Gaussian PSFs. For each dataset, we apply the predicted PSFs, the PSFs extracted from axial YZ slices, and matched Gaussian kernels to high resolution lateral images and compute the resulting axial modulation transfer function as a function of spatial frequency. The Predicted PSFs closely follow the MTF of the extracted axial kernels, while Gaussian kernels with the same half power frequency show a different high frequency falloff, which indicates that their blur shape does not match the true axial degradation.

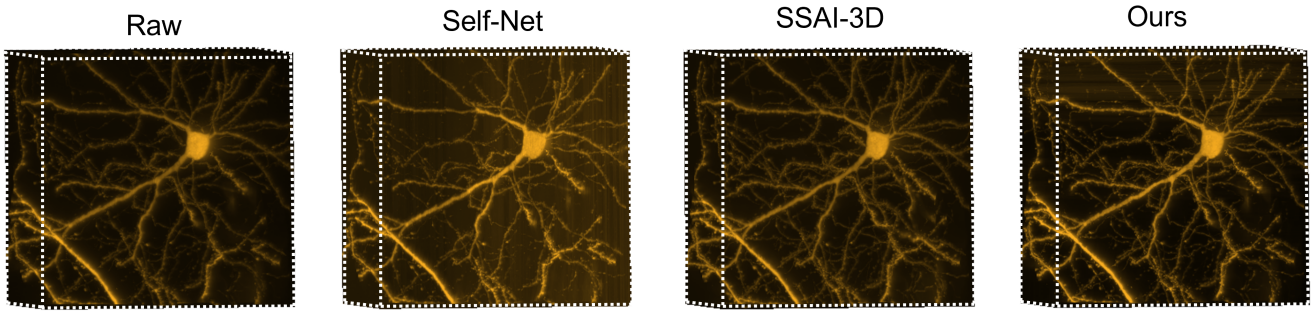

Figure 3. 3D comparison of isotropic reconstruction on the Dense neuron cluster dataset acquired with CS-fMOST. Volume renderings of the anisotropic input stack, Self-Net, SSAI-3D, and MicroFM are shown from left to right.

sian kernel by fitting the decay of the radial power spectrum at high spatial frequencies. The extracted PSFs serve as a data driven axial reference that we compare with the PSFs predicted by our model from the same YZ slices and with matched Gaussian kernels.

We first analyse the spatial structure of these kernels. As shown in Fig. 1, for each dataset we measure lateral and axial full width at half maximum on axial YZ sections and plot the distributions for predicted, extracted and Gaussian PSFs, together with the axial to lateral FWHM ratio

$\mathrm{FWHM_{ax}}/\mathrm{FWHM_{lat}}$. The predicted PSFs reproduce the axial elongation and moderate asymmetry of the extracted kernels across all datasets, while Gaussian PSFs remain nearly isotropic with ratios close to one.

We then examine the frequency behaviour of the three kernels. We apply each PSF to the same high resolution XY images and compute the corresponding axial modulation transfer functions as a function of spatial frequency, as shown in Fig. 2. The curves obtained with the predicted PSFs almost overlap with those from the extracted axial kernels, indicating that our model reproduces the characteristic attenuation of mid and high spatial frequencies. In contrast, the matched Gaussian kernels show a different decay pattern even though their half power frequency is the same.

Because the learned PSFs closely match the anisotropic axial blur of the microscope in both spatial and frequency domains, using them to degrade lateral images produces synthetic low quality views that are well aligned with real axial observations. Training on these PSF based degradations therefore enables the network to learn how to reconstruct high resolution structures along the axial direction. For the same reason, our held out twenty percent of lateral images, degraded with the predicted PSFs, form a realistic test set that approximates unseen axial blur and provides a meaningful evaluation of generalisation.

3.4. The 3D Quantitative Comparison

To provide a more intuitive comparison, we visualize three dimensional renderings of the dense neuron cluster volume, as shown in Fig. 3. In this dense neuron cluster volume, MicroFM produces the most faithful three dimensional morphology among all methods. It reconstructs a compact soma, smoothly tapering dendrites, and thin terminal branches that remain continuous across depth. The reconstructed volume also shows a more homogeneous background and fewer spurious structures, which demonstrates that MicroFM preserves fine neuronal topology while effectively suppressing artifacts in dense tissue.

References

[1] Muthuvel Arigovindan, John W Sedat, and David A Agard. Effect of depth dependent spherical aberrations in 3d structured illumination microscopy. *Optics Express*, 20(6):6527–6541, 2012. 1

[2] Joseph Paul Cohen, Margaux Luck, and Sina Honari. Distribution matching losses can hallucinate features in medical image translation. In *International conference on medical image computing and computer-assisted intervention*, pages 529–536. Springer, 2018. 1

[3] Uros Krzic, Stefan Gunther, Timothy E Saunders, Sebastian J Streichan, and Lars Hufnagel. Multiview light-sheet microscope for rapid in toto imaging. *Nature Methods*, 9(7):730–733, 2012. 1

[4] P Philippe Laissue, Rana A Alghamdi, Pavel Tomancak, Emmanuel G Reynaud, and Hari Shroff. Assessing phototoxicity in live fluorescence imaging. *Nature Methods*, 14(7):657–661, 2017. 1

[5] Xinyang Li, Guoxun Zhang, Hui Qiao, Feng Bao, Yue Deng, Jiamin Wu, Yangfan He, Jingping Yun, Xing Lin, Hao Xie, et al. Unsupervised content-preserving transformation for optical microscopy. *Light: Science & Applications*, 10(1):44, 2021. 1

[6] Zelin Li, Chenwei Wang, Zhaoke Huang, Yiming Ma, Cunming Zhao, Zhongying Zhao, and Hong Yan. Volume tells: Dual cycle-consistent diffusion for 3d fluorescence microscopy de-noising and super-resolution. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 16091–16100, 2025. 2

[7] Ryan McGorty, Joerg Schnitzbauer, Wei Zhang, and Bo Huang. Correction of depth-dependent aberrations in 3d single-molecule localization and super-resolution microscopy. *Optics letters*, 39(2):275–278, 2014. 1

[8] Kefu Ning, Bolin Lu, Xiaojun Wang, Xiaoyu Zhang, Shuo Nie, Tao Jiang, Anan Li, Guoqing Fan, Xiaofeng Wang, Qingming Luo, et al. Deep self-learning enables fast, high-fidelity isotropic resolution restoration for volumetric fluorescence microscopy. *Light: Science & Applications*, 12(1):204, 2023. 1, 3

[9] Hyoungjun Park, Myeongsu Na, Bumju Kim, Soohyun Park, Ki Hean Kim, Sunghoe Chang, and Jong Chul Ye. Deep learning enables reference-free isotropic super-resolution for volumetric fluorescence microscopy. *Nature Communications*, 13(1):3297, 2022. 1

[10] Thomas A Planchon, Liang Gao, Daniel E Milkie, Michael W Davidson, James A Galbraith, Catherine G Galbraith, and Eric Betzig. Rapid three-dimensional isotropic imaging of living cells using bessel beam plane illumination. *Nature Methods*, 8(5):417–423, 2011. 1

[11] Chang Qiao, Haoyu Chen, Run Wang, Tao Jiang, Yuwang Wang, and Dong Li. Deep learning-based optical aberration estimation enables offline digital adaptive optics and super-resolution imaging. *Photonics Research*, 12(3):474–484, 2024. 1

[12] Sjoerd Stallinga and Bernd Rieger. Accuracy of the gaussian point spread function model in 2d localization microscopy. *Optics Express*, 18(24):24461–24476, 2010. 1

[13] Sota Takanezawa, Takashi Saitou, and Takeshi Imamura. Wide field light-sheet microscopy with lens-axicon controlled two-photon bessel beam illumination. *Nature Communications*, 12(1):2979, 2021. 1

[14] Martin Weigert, Loic Royer, Florian Jug, and Gene Myers. Isotropic reconstruction of 3d fluorescence microscopy images using convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 126–134. Springer, 2017. 1

[15] Yide Zhang, Yinhao Zhu, Evan Nichols, Qingfei Wang, Siyuan Zhang, Cody Smith, and Scott Howard. A poisson-gaussian denoising dataset with real fluorescence microscopy images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11710–11718, 2019. 1