Diffusion-driven lensless fiber endomicroscopic quantitative phase imaging towards digital pathology

This document provides supplementary information to "Diffusion-driven lensless fiber endomicroscopic quantitative phase imaging towards digital pathology". Included are supplementary figures for phase reconstruction on ImageNet, resolution test chart, human tissue images and the IoU distribution in the cell segmentation task.

1. EXPERIMENTAL SETUP FOR DATA ACQUISITION

The optical setup for training data acquisition is demonstrated in Fig. S1. The system employs a modulated laser beam (532nm ,Verdi, Coherent Inc.), which is expanded using a telescope system (L1-L2) to ensure uniform illumination on the SLM. The SLM displays the phase image, thereby modulating the phase of the laser beam. To remove unwanted diffraction orders, a spatial filtering system (L3-L4, ID) is utilized. The phase-modulated beam is then projected onto the proximal facet of the MCF (FIGH-350S; Fujikura) at the measurement side, using a microscope system (L5, MO1; 10× plan achromat objective, 0.25 NA, Olympus). A part of the incident beam is reflected by the MCF facet at the measurement side and directed onto the alignment camera (CAM1; uEye LE, IDS), facilitating the precise alignment between the holographic display plane and the fiber facet. Far-field speckle images are captured on the detection camera (CAM2; uEye CP, IDS) through an additional microscope system (MO2, L6, LP2; 10× plan fluo objective, 0.3 NA, Nikon).

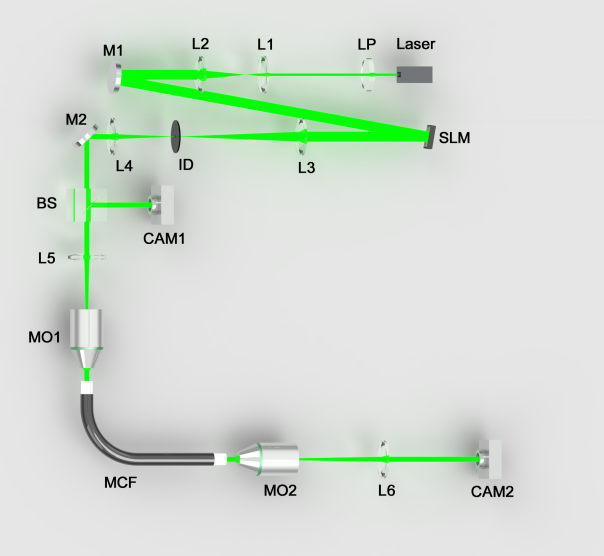

Fig. S1. Scheme of the optical system for training dataset acquisition. The real phase images are displayed on the SLM and then projected on the MCF facet at the measurement side. The SLM and the detection camera (CAM2) are synchronously triggered to record the corresponding paired phase and speckle images. LP1 and LP2, linear polarizer; L1-L6, achromatic lenses; M1 and M2, mirrors; ID, iris diaphragm; BS, beam spliter; CAM1, alignment camera; MO1 and MO2, microscope objectives.

2. PHASE RECONSTRUCTION WITH DIFFERENT INITIALIZATION

In our experiment, the phase reconstruction results of SpecDiffusion demonstrate robustness to random initialization. As illustrated in Fig. S2, under three different initializations, SpecDiffusion

Table S1. Averaged evaluation metrics of U-Net and SpecDiffusion on ImageNet

Method	MAE (rad)	PSNR	SSIM	2D Correlation
U-Net	0.1143	19.63	0.5410	0.8394
SpecDiffusion	0.0879	21.72	0.6707	0.8989

successfully reconstructs the phase image in each case, and the reconstructed images are highly similar to each other. This indicates that SpecDiffusion effectively extracts image information from MCF speckles to adapt the reconstruction process for each random initialization, robustly transforming the initial image into the target phase.

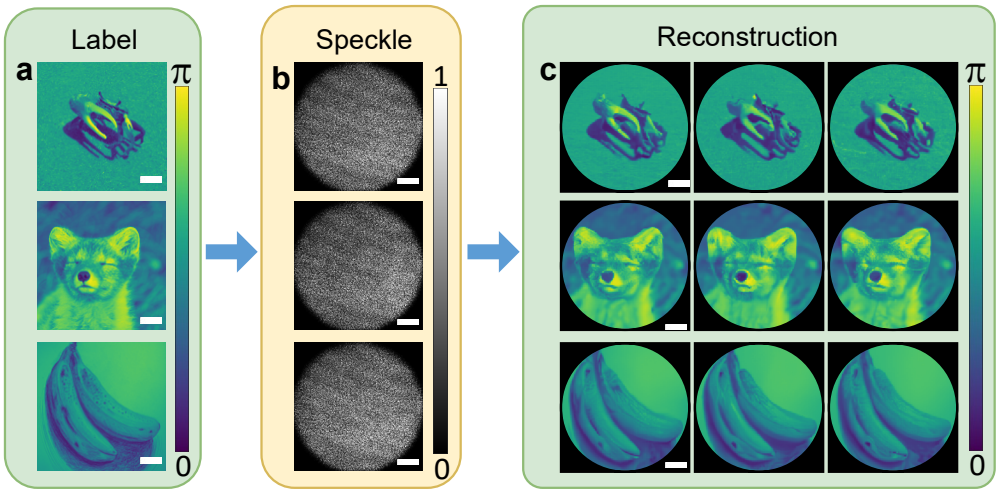

Fig. S2. Phase reconstruction of complex images through the MCF with different initialization. **a** Ground truth phase images. **b** Speckle patterns from MCF. **c** Reconstructed phase images by SpecDiffusion under different initializations. Scale bars 50μm.

3. PHASE RECONSTRUCTION ON COMPLEX IMAGES FROM IMAGENET

To more comprehensively demonstrate the superior reconstruction results achieved by SpecDiffusion, we present an expanded collection of reconstructed images along with the distribution of their performance metrics across the test set, as illustrated in Fig. S3. The averaged evaluation metrics are summarized in Tab. S1. Compared to U-Net [1], SpecDiffusion enhances all performance metrics on ImageNet [2]. These improvements confirm SpecDiffusion's capacity to accurately reconstruct phase image from the speckle, indicating its effectiveness in MCF endoscope.

4. RESOLUTION CHART RECONSTRUCTION

In our experiment, U-Net's resolution capacity on the test chart is limited to separating lines in Group 3, Element 3, as depicted in Fig. S4.c. Nevertheless, the reconstructed lines suffer from severe blurring and content missing. In comparison, the reconstruction result of SpecDiffusion, as depicted in Fig. S4.d, demonstrates a clearer and more detailed recovery of the lines. The resolution capacity difference is further verified in Fig.S4.e, where SpecDiffusion achieves a higher contrast between peaks and valleys in the reconstructed image. SpecDiffusion's superior resolution capacity enables it to reconstruct detailed regions more effectively than U-Net, thus providing more information in practical application.

5. TISSUE RECONSTRUCTION

In the experiment setup, we simulate the phase retrieval process of tissues using a SLM. The human tissue images are sourced from Stanford Tissue Microarray Database [3]. Throughout the

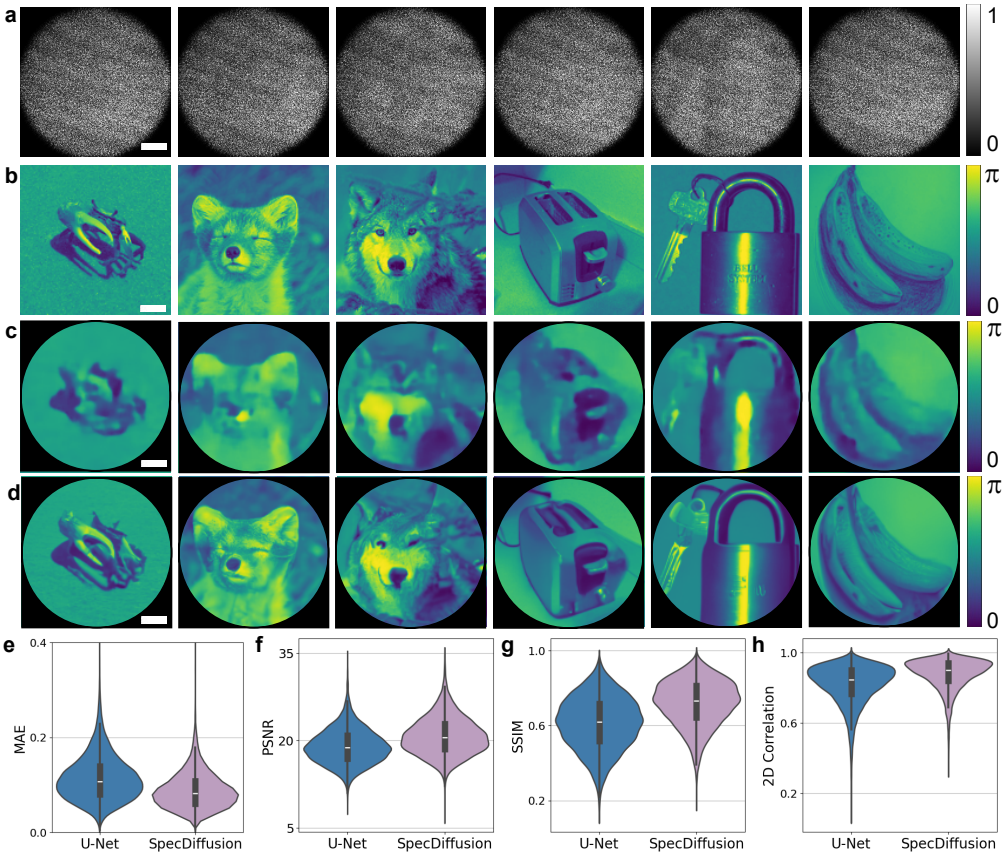

Fig. S3. Phase reconstruction of ImageNet images through the MCF. **a** Speckle patterns from MCF. **b** Ground truth phase images. **c** and **d** Reconstructed phase images by U-Net and SpecDiffusion. **e**-**g** PSNR, SSIM and 2D correletion coefficient distribution of U-Net and SpecDiffusion on the test set. Scale bars $50\mu m$.

experiment process, the tissue images are converted into computer-generated holograms and subsequently projected onto the SLM. The resulting speckles are captured from the detection side of MCF. For the transfer learning process, both SpecDiffusion and U-Net are initially pretrained on the ImageNet dataset, and then undergo transfer learning on a limited set of tissue images. Their performance is evaluated on a separate set of $1,000$ images.

To better illustrate our reconstruction results on human tissue dataset [3], we display more reconstructed images and their corresponding residual maps in Fig. S5. The distinction is particularly evident when examining the residual maps of reconstructed phase images from both models, as illustrated in Fig. S5.e. The residual map for the U-Net reconstructed phase image exhibits numerous details of the tissue, indicating a considerable deviation from the original image in terms of both structural detail and texture. Conversely, the residual map of the SpecDiffusion reconstructed phase image demonstrates a much lower overall amplitude, suggesting a higher similarity to the original image. Notably, the residuals in SpecDiffusion's map primarily concentrate around the cell outlines. This observation is crucial because the outlines of cells are where information is most susceptible to being lost, especially considering the discrete distribution of fiber cores in a Multi-Core Fiber (MCF). To illustrate the phase reconstruction process of SpecDiffusion, we record a video in Visualization 1. The SpecDiffusion model is pretrained on ImageNet and then undergoes a transfer learning with a tissue dataset of 150 images. It takes 100 denoising steps to finish the phase reconstruction process.

Our another experiment focuses on the impact of varying the number of denoising steps on the reconstruction quality of SpecDiffusion, as depicted in Fig. S6. As the denoising steps increase, all evaluated metrics get improved, indicating higher reconstruction quality. The reconstruction quality reaches its optimum when the denoising steps are increased to 100. At this stage, at this

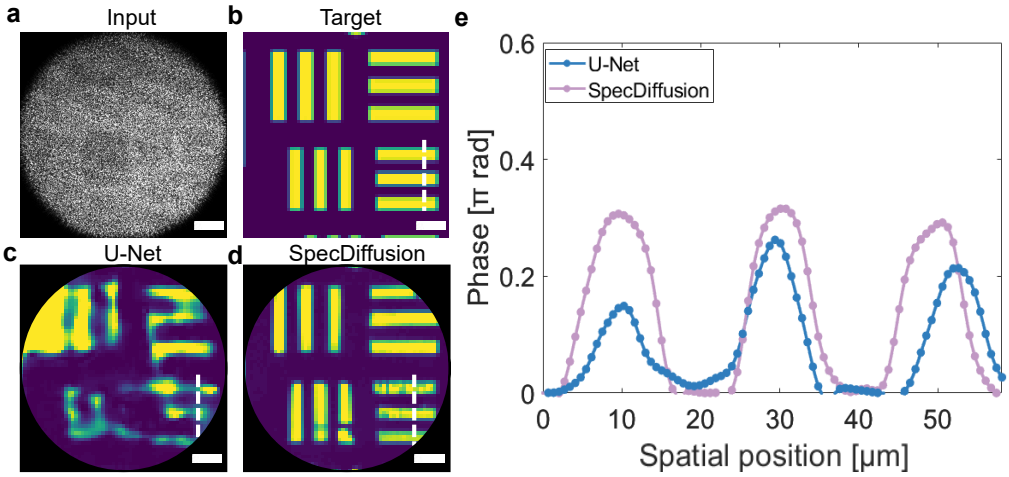

Fig. S4. Phase reconstruction of USAF resolution chart. **a** Speckle pattern from MCF. **b** Ground truth. **c** U-Net reconstructed phase image. **d** SpecDiffusion reconstructed phase image. *e* Phase reconstruction constrast between U-Net and SpecDiffusion. Scale bars $50\mu m$

time, the reconstruction speed achieves 1.77 FPS, which is acceptable for medical scenarios where time sensitivity is less critical.

6. CELL SEGMENTATION

In cell segmentation task, we adopt IoU to quantitatively assess the segmentation results of reconstructed images. IoU measures the agreement between segmentation results derived from reconstructed images and their ground truth counterparts. The IoU distributions of U-Net and SpecDiffusion are illustrated in Fig. S7. It reveals a distinctly superior IoU distribution for SpecDiffusion compared to U-Net, with a significant portion of SpecDiffusion's results clustering within the higher IoU range. The superiority of SpecDiffusion's segmentation results not only highlights its advanced reconstruction capabilities, but also signifies its potential to significantly maintain the performance of downstream tasks.

7. EVALUATION METRICS

In our experiment, we employ Mean Absolute Error (MAE), Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), 2D Correlation coefficient as the evaluation metrics for reconstructed phase images. These metrics provide a quantitative measure of the reconstruction fidelity.

The mean absolute error (MAE) is a critical metric for assessing the accuracy of reconstructed images, particularly in the context of phase reconstruction. It measures the average of the absolute differences between the original phase image and the reconstructed phase image, across all pixels. A lower MAE value indicates a closer match between the reconstructed and original phases, signifying higher fidelity in the reconstruction process. This high fidelity is crucial for applications that rely on precise phase information, such as in the calculation of biophysical properties including refractive index, cell volume and dry mass. For original image I and reconstructed image I', the MAE is calculated as:

$$\mathrm{MAE}(I, I') = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |I(i,j) - I'(i,j)|. \tag{S1}$$

The peak signal-to-noise ratio (PSNR) is a widely utilized metric in the field of image processing to evaluate the quality of reconstructed images relative to their originals. It quantifies the ratio of the maximum possible power of the original image to the power of corrupting noise that affects the fidelity of the reconstructed image. Essentially, PSNR measures the level of distortion introduced into the original image upon reconstruction, which also reflects the amount of information derived from the speckle. A higher PSNR value typically indicates more information derived

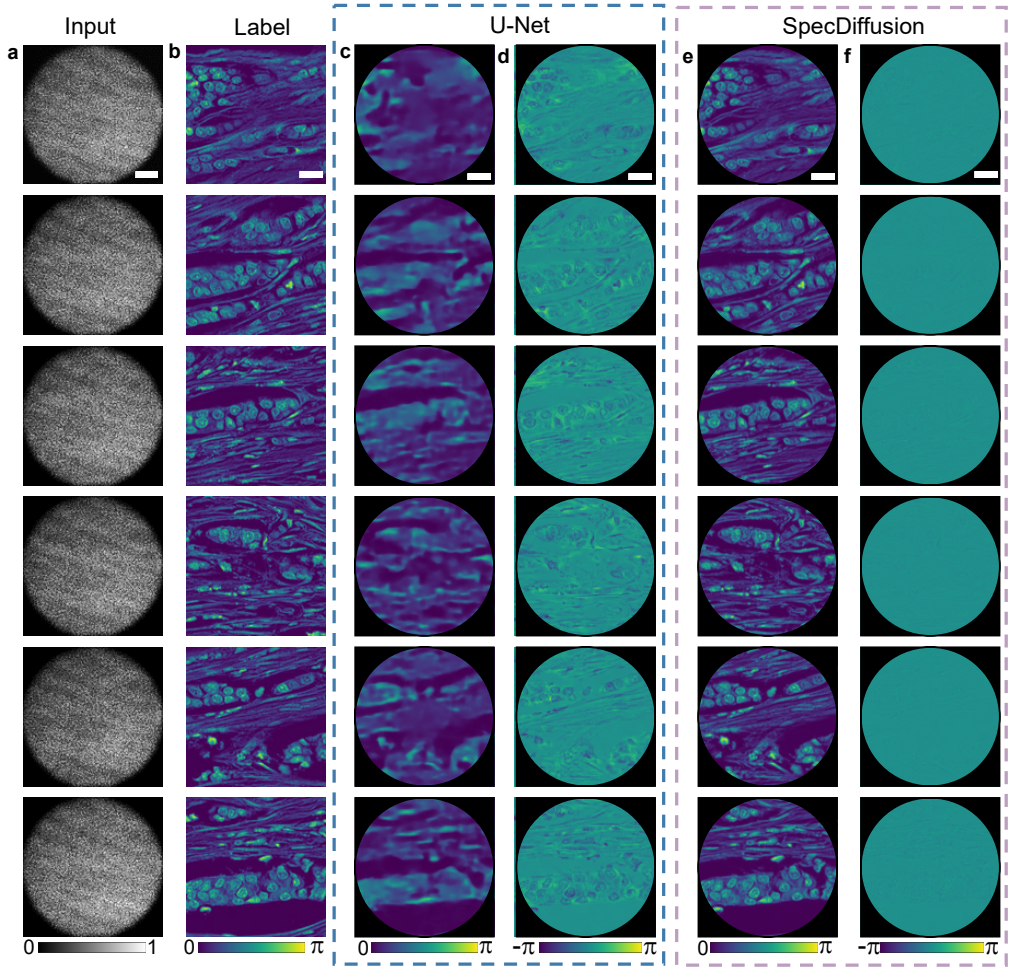

Fig. S5. Tissue reconstruction results by SpecDiffusion and U-Net. *a* Speckle patterns from MCF. **b** Ground truth phase images. **c** and **e** Reconstructed phase images by U-Net and SpecDiffusion. **d** and **f** Residual maps of the reconstructed tissue images by U-Net and SpecDiffusion. Scale bars $50\mu m$.

from the speckle, implying a higher quality of the reconstructed image. For original image I and reconstructed image I', the PSNR is calculated as:

$$\mathrm{PSNR}(I, I') = 10 \cdot \log_{10}[\frac{\mathrm{MAX}_I^2}{\mathrm{MSE}(I, I')}], \tag{S2}$$

where MAX_I is the max pixel value of image I and I', MSE is the mean square error between image I and I', which is calculated as

$$\mathrm{MSE}(I, I') = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - I'(i,j)]^2. \tag{S3}$$

The structural similarity of index measure (SSIM) [4] is an advanced metric designed to assess the perceptual quality of images. Unlike traditional metrics that primarily focus on pixel-level differences, SSIM evaluates the visual impact of three key components: luminance, contrast, and structure, which correspond to the perception of brightness, contrast, and patterns or textures in human vision, respectively. For original image I and reconstructed image I', the SSIM is calculated as

$$\mathrm{SSIM}(I, I') = \frac{(2\mu_I\mu_{I'} + c_1)(2\sigma_{II'} + c_2)}{(\mu_I^2 + \mu_{I'}^2 + c_1)(\sigma_I^2 + \sigma_{I'}^2 + c_2)}, \tag{S4}$$

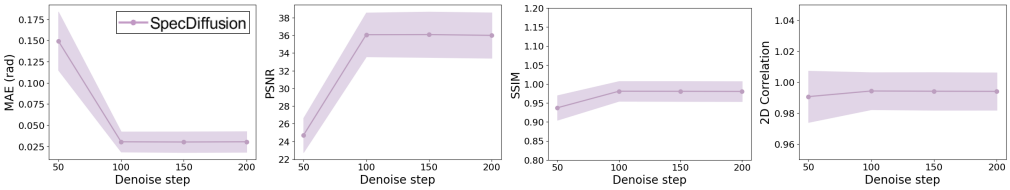

Fig. S6. The relationship between denoise step and reconstruction quality of SpecDiffusion. *a-d* MAE, PSNR, SSIM and 2D correlation coefficient evaluated by U-Net and SpecDiffusion with varying denoise step.

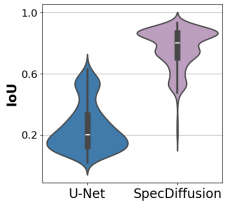

Fig. S7. IoU distribution of U-Net and SpecDiffusion in cell segmentation task.

where μ indicates pixel sample mean, σ indicates the variance, c_1 and c_2 are two variables to stabilize the division to avoid zero denominator.

The 2D correlation coefficient is a statistical measure that quantifies the degree of linear correlation between two images. It serves as an indicator of the overall fidelity of the reconstructed image by assessing how well the pixel intensity variations in the reconstructed image match those in the original image. For original image I and reconstructed image I', the 2D correlation coefficient is calculated as

$$\rho_{II'} = \frac{\mathrm{Cov}(I, I')}{\sigma_I \sigma_{I'}}, \tag{S5}$$

where $\mathrm{Cov}(I, I')$ is the convariance between I and I', σ_I and $\sigma_{I'}$ are the standard variance of I and I'.

For assessing the accuracy of cell segmentation on the reconstructed images, we adopted the intersection over union (IoU) metric. IoU quantifies the overlap between two segmentation maps, thereby providing an assessment of the similarity between the predicted segmentation result and the ground truth. This metric is particularly effective for comparing the precision of segmentation boundaries and the overall segmentation quality. For segmentation map M of ground truth and M' of reconstructed image, IoU is computed as

$$\mathrm{IoU}(M, M') = \frac{|M \cap M'|}{|M \cup M'|}, \tag{S6}$$

where $|M \cap M'|$ indicates the overlap area between M and M', $|M \cap M'|$ indicates the union area between M and M'.

REFERENCES

1. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, (Springer, 2015), pp. 234–241.
2. J. Deng, W. Dong, R. Socher, *et al.*, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, (Ieee, 2009), pp. 248–255.
3. R. J. Marinelli, K. Montgomery, C. L. Liu, *et al.*, "The stanford tissue microarray database," Nucleic acids research **36**, D871–D877 (2007).
4. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE transactions on image processing **13**, 600–612 (2004).