Supplemental document for Neural-field-assisted transport-of-intensity phase microscopy: partially coherent quantitative phase imaging under unknown defocus distance

This document provides supplementary information for "Neural-field-assisted transport-of-intensity phase microscopy: partially coherent quantitative phase imaging under unknown defocus distance". We discuss in detail the principles of neural-field-assisted transport-of-intensity phase microscopy (NFTPM) and present more experimental results.

CONTENTS

1. DISCUSSION OF THE INFLUENCE OF TYPICAL HYPERPARAMETER ADJUSTING ON NFTPM

In this section, we mainly discuss the influence of typical hyperparameters of the neural network. In the main text, we have verified that changes in network scale caused by variations in C do not affect the performance of NFTPM. Hence, we performed additional simulations here to explore the impact of adjusting the hyperparameters such as the number of the network layers K, rotational intervals N_θ and frequency extensions L.

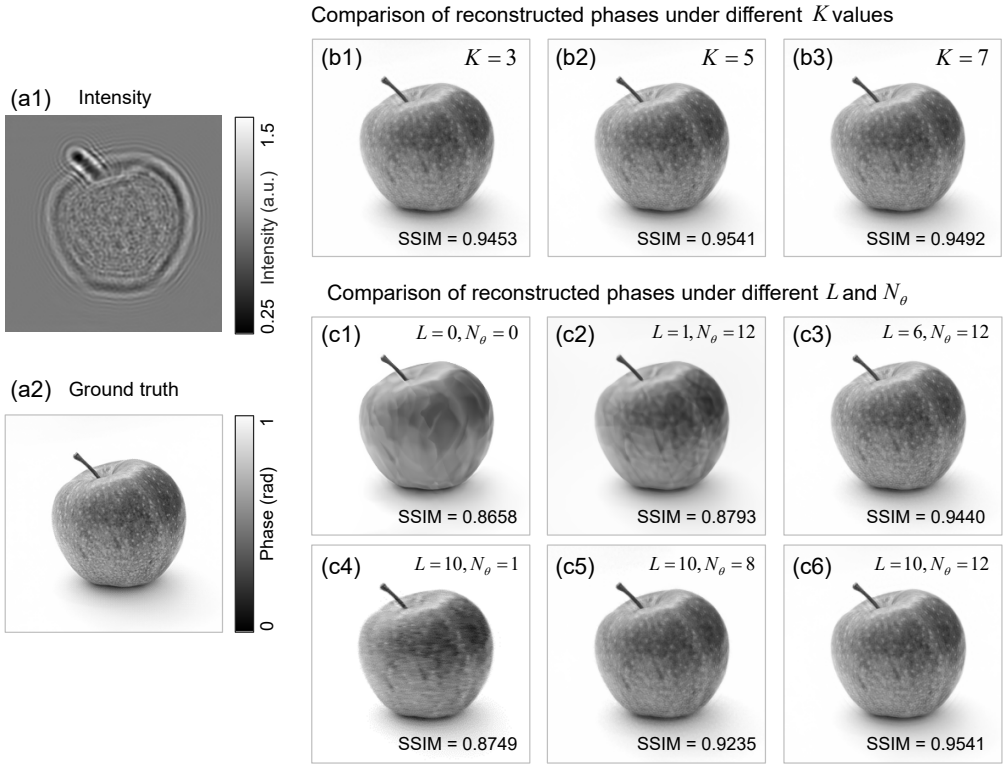

Fig. S1. Comparison of phase reconstruction results of a simulated sample using NFTPM for different hyperparameters. (a1) Intensity of the simulation object. (a2) Ground truth of the phase. (b1) - (b3) The retrieved phase under K = 3, 5 and 7. (c1) The retrieved phase under L = 0, N_θ = 0. (c2) The retrieved phase under L = 1, N_θ = 12. (c3) The retrieved phase under L = 6, N_θ = 12. (c4) The retrieved phase under L = 10, N_θ = 1. (c5) The retrieved phase under L = 10, N_θ = 8. (c6) The retrieved phase under L = 10, N_θ = 12.

For the hyperparameter K, excessive reduction (K = 1) degrades the neural network training to logistic regression, while over-increasing this parameter ($K \gg 5$) defeats the original purpose of forming a lightweight MLP. Thus, the hyperparameter K is selected as a smaller value of K = 5 in the main text. In fact, the phase retrieval quality remains stable when K is slightly increased or decreased. The retrieved phase results for different number of network layers K are shown in Figs. S1 (b1) - (b3). It can be seen that the SSIM is stable around 0.95 for K = 3, 5 and 7, which means that changes in the number of network layers K do not have much effect on NFTPM recovered phase.

To explore the effect of the hyperparameters N_θ and L, we carried out a set of comparative simulation as illustrated in Fig. S1 (c). In Figs. S1 (c2), (c3), (c6), we show the reconstruction phase results for different L, and Figs. S1 (c4) - (c6) are the retrieved results for different N_θ. It can be seen that when only L is adjusted, SSIM stabilizes around 0.95 (with accurate high-frequency reconstruction) for L of 6 and 10, but decreases to around 0.88 when L is 1. Besides, as can be seen, the impact of N_θ on NFTPM is consistent with the case of adjusting L. When both hyperparameters L and N_θ turn zero [Fig. S1 (c1)], the SSIM reaches its lowest value, and it is clearly found that only low-frequency information is present in the recovered phase with a lack

of high-frequency details. In short, the two parameters of N_θ and L do not significantly affect the imaging accuracy in NFTPM when they are both large value. However, when N_θ and L become quite small (equivalent to no radial encoding), it is difficult for the encoding module to map the spatial coordinates to the high-dimensional space, which hinders the MLP from learning to represent high-frequency information.

In addition, the speed of the NFTPM is mainly determined by the operating speed of forward propagation of MLP, the computation process of physical model, and the backpropagation. Therefore, the variation of the network parameter count (associated with K and C) affects the floating point operations (FLOPs) of network, and hence changes the iteration speed. However, modifying N_θ and L has no effect on the iteration speed of NFTPM, because N_θ and L are not involved in the parameter optimization of NFTPM, but only concern changing the size of the input feature of MLP by preprocessing the 2D coordinates.

2. ROBUSTNESS OF NFTPM TO DEFOCUS DISTANCE INITIALIZATION

For the defocus distance prediction process of NFTPM, there is no need for explicit initialization of the defocus parameter. Instead, the initial value of defocus distance z is randomly set within the range of 0 to 20 μm and no physical prior on the defocus distance has been introduced in our proposed NFTPM method.

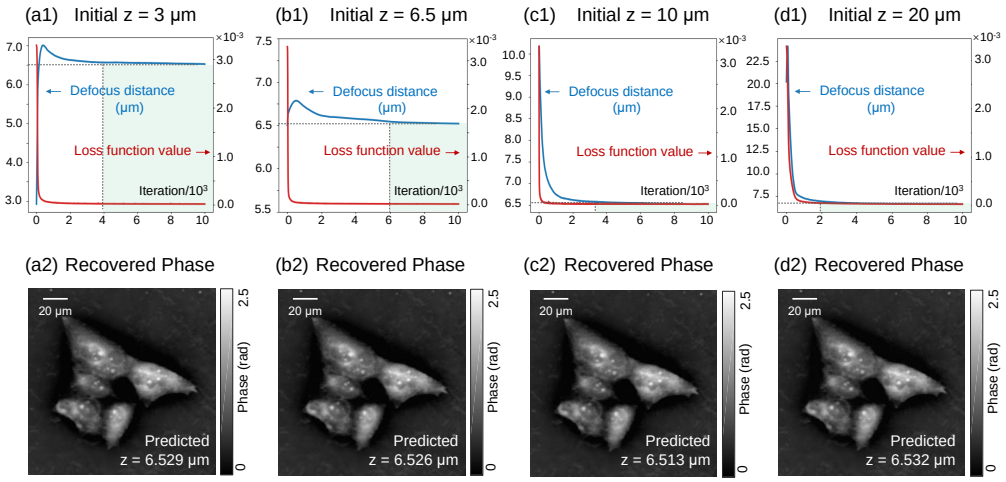

Fig. S2. Recovered phase by NFTPM for different initial defocus distance values. (a1) - (d1) Defocus distance prediction curves and loss function values under different initial defocus distance z. (a2) - (d2) NFTPM-retrieved phase corresponding to (a1) - (d1).

In Fig. 4 of the main text, the defocus distance of NFTPM was set with 1 μm as an initial value, and the final predicted defocus distance is around 6.52 μm. To further illustrate the robustness of NFTPM to different initializations of defocus distance, we performed phase retrieval in different cases with initial defocus values z that are smaller, nearer, larger, or overly large, respectively (3 μm, 6.5 μm, 10 μm, 20 μm). From Figs. S2 (a1) - (d1), it can be seen that all the different initial values of the defocus distance converge to the correct result as the loss function converges. Finally, NFTPM can achieve high-quality phase retrieval in all situations, as shown in Figs. S2 (a2) - (d2). It is obvious that different defocus distance initialization has almost no effect on the quality and efficiency of phase retrieval by NFTPM. Thus, our proposed NFTPM method can achieve accurate partially coherent quantitative phase imaging (QPI) under unknown defocus distance with random initialization z.

3. ADAPTABILITY OF NFTPM TO ARBITRARY SOURCE DISTRIBUTION

The forward image formation model of NFTPM is the Abbe imaging model, according to which the generated intensity is equivalent to a linear superposition of intensity images from coherent imaging of each point source in a partially coherent light source. There is no restriction on the distribution of the light source in Abbe theory, and it can be applied to arbitrary illumination

patterns. Theoretically, the illumination pattern in the forward image formation model can be modified by changing the distribution function $S_{pc}(\mathbf{u})$ in Eq. (9) of the main text.

In optical microscopic imaging, circular and annular illuminations are widely used as partially coherent light sources and the corresponding simulations have been shown in our manuscript. In fact, NFTPM is also applicable to other illumination patterns when their distributions are known. For example, as shown in Figs. S3 (a2) - (a4), we also simulated other special light sources (asymmetric semicircular illumination, Gaussian-distributed illumination), and as in the case of circular illumination [Fig. S3 (c1)], the NFTPM recovers the high-quality phase in all cases [as shown in Figs. S3 (c2) - (c4)].

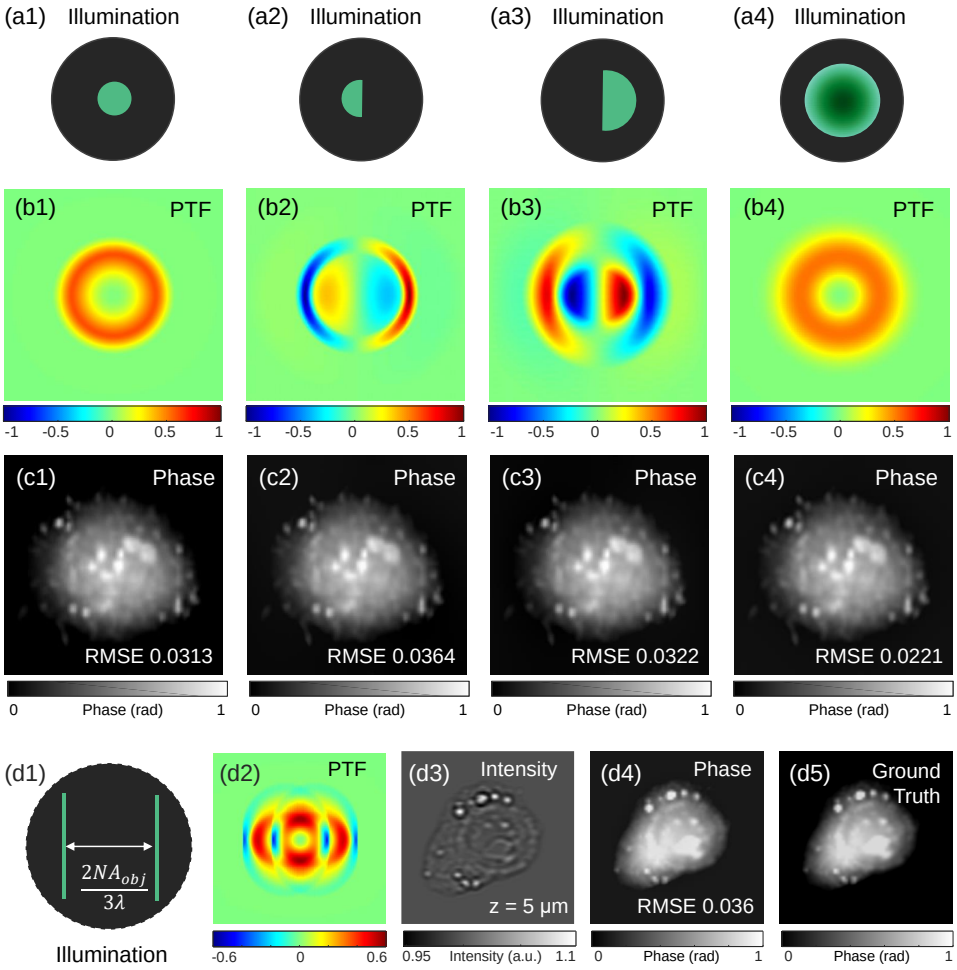

Fig. S3. NFTPM recovered phase based on different partially coherent illumination patterns. (a1) - (a4) Source patterns. (b1) - (b4) Phase transfer function (PTF) corresponding to (a1) - (a4). (c1) - (c4) Retrieved phase via NFTPM under different illuminations in (a1) - (a4). (d1) Double-slit illumination. (d2) PTF corresponding to (d1). (d3) Defocused intensity. (d4) Recovered phase by NFTPM under illumination of (d1). (d5) Ground truth corresponding to (d4).

In addition, Figs. S3 (d1) - (d4) show the special illumination distribution of a double slit and corresponding phase transfer function (PTF). Compared to the ground truth displayed in Fig. S3 (d5), our method can also achieve high-precision phase recovery in double slit illumination. In summary, as long as the illumination distribution is known and the illumination together with the defocus distance can provide sufficient phase contrast, NFTPM can achieve high-quality phase recovery for pure phase samples under arbitrary illumination distribution.

4. ANALYSIS OF THE IMPACT OF SOURCE DISTRIBUTION, DEFOCUS DISTANCE AND NOISE ON NFTPM

The previous Section 3 demonstrated the phase recovery capability of NFTPM under special known illumination distributions. Indeed, in addition to the shape characteristics such as circular and annular, the numerical value of the coherence parameter S also affects the phase information contained in the measured intensity images in partially coherent illumination distributions. Hence, in this section, we based on the transfer function theory to analyze the impact of the coherence parameter S, thereby providing a theoretical basis for the selection strategy of coherence parameter S. As shown in Figs. S4 (a) - (c), for circular illumination, as the coherence parameter S increases from 0.5 to 1.0, the cut-off frequency corresponding to PTF, which is the sum of the numerical aperture (NA) of the illumination and the objective lens, increases from 1.5NA to 2NA. It is evident that an increase in the coherence parameter S (from 0 to 1) corresponds to a higher frequency information contained in the intensity image of the sample, thereby enhancing the theoretical resolution from coherent to incoherent diffraction limit.

However, as can be seen from the PTF profile, as S increases, the amplitude of the PTF response gradually decreases, leading to a reduction in the phase contrast conveyed to the intensity image (more susceptible to noise). Especially when $S = 1$ (completely incoherent illumination), the phase component of the sample can hardly be coupled in the intensity due to the zero value of frequency response [Fig. S4 (c)], where NFTPM cannot recover valid information of phase from a single-shot intensity. Thus, the object phase cannot be retrieved under completely incoherent illumination. In comparison, the annular illumination can solve the contradictory problem between accuracy and phase contrast in circular illumination.

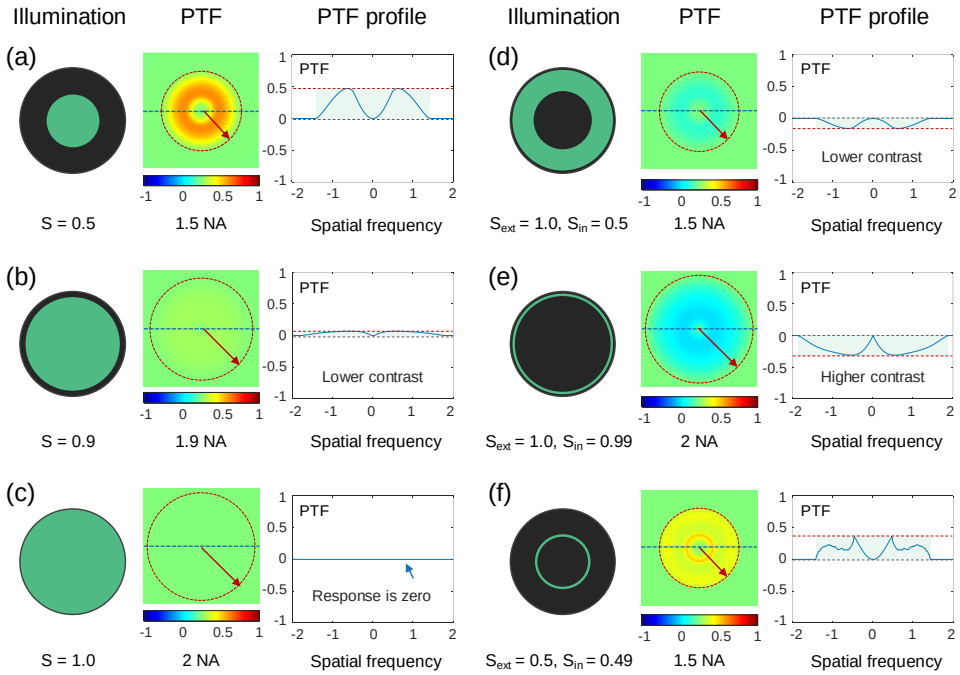

Fig. S4. PTF for different coherence parameters S under (a) - (c) circular illumination and (d) - (f) annular illumination.

From Figs. S4 (d) - (f), it can be found that the PTF of annular illumination generally has a uniformly high response. In particular, when the outer diameter circle of the ring is 1 and the inner diameter circle is 0.99 [Fig. S4 (e)], the distribution of the PTF is optimal due to that it reaches the maximum cut-off frequency and the response is sufficiently homogeneous to ensure high-accuracy and high-contrast phase retrieval. Although our method can achieve high-quality phase retrieval under known illumination, considering factors such as phase contrast and signal-to-noise ratio (SNR), NFTPM prefers utilizing NA-matched annular illumination to provide measurements necessary for high-resolution phase recovery.

In Section 2, we discussed that the initialization of the defocus distance has a minimal impact on the algorithm. As a means of obtaining phase contrast, the choice of defocus distance will affect the quality of phase recovery, because the information contained in the captured intensity determines whether the sample to be measured can be reconstructed with high precision. Besides, the convergence of NFTPM can be facilitated by placing the sample within the appropriate range of defocus distances. Therefore, here we have simply compared the impact of defocus distance on phase contrast under circular illumination (S = 0.1) and NA-matched annular illumination through the transfer function theory. As shown in Figs. S5 (a1) - (b1), a smaller defocus distance allows for a strong high-frequency response of phase transfer function, thereby ensuring the acquisition of the high-contrast defocused intensity image. However, at smaller defocus distances, there is the problem of lower response to the low frequency domain. The larger the defocus distance, the better the response amplitude of the low frequency domain, but as shown in Figs. S5 (a2) - (b2), the non-uniform response of the passband presents a weak tolerance to noise, resulting in low SNR intensity measurements at large defocus distances.

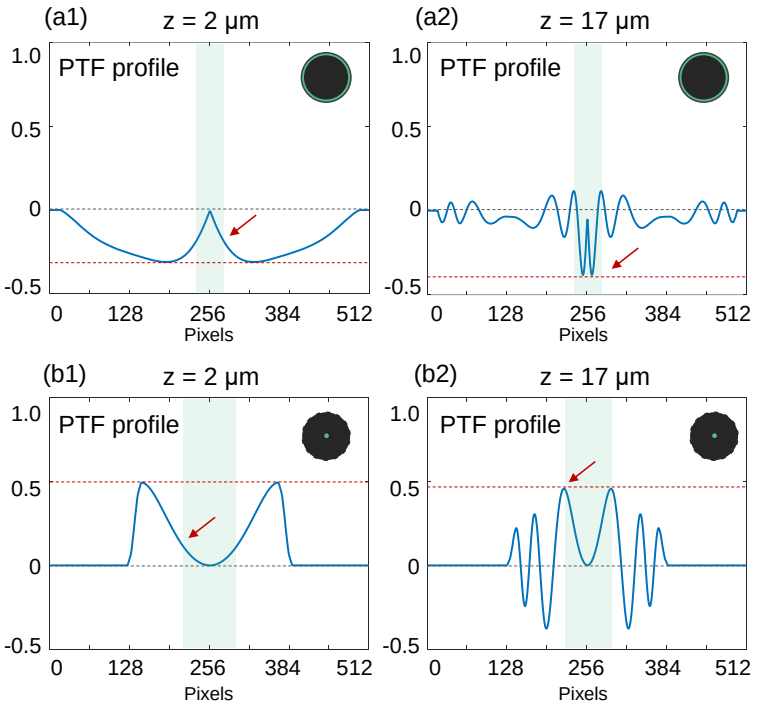

Fig. S5. PTF for different defocus distances z under (a1), (a2) NA-matched annular illumination and (b1), (b2) circular illumination (S = 0.1).

Nonetheless, in the absence of noise, NFTPM accurately predicts the corresponding defocus distance even at a large defocus distance. To illustrate this point, we simulated the intensity image under a NA-matched annular illumination at a defocus distance of 17 μm [see Fig. S6 (a1)], and Fig. S6 (b1) shows that the high-quality phase can be recovered using the NFTPM method. However, the impact of noise on the phase retrieval capability of NFTPM cannot be ignored in practice. When considering Gaussian noise with a standard deviation of 0.005, the intensity obtained at z = 17 μm [Fig. S6 (a2)] has lower contrast compared to the one obtained at an appropriate defocus distance of 7 μm [Fig. S6 (c2)], displaying that the noise has a greater contamination of the phase contrast at a larger defocus distance. Besides, the recovered phase at the z = 17 μm [Fig. S6 (b2)] is not as accurate as those at the one in z = 7 μm [Fig. S6 (d2)] (RMSE 0.069 > 0.045). In addition, as displayed in Figs. S6 (a3) and (c3), when the noise increases (Gaussian noise with a standard deviation of 0.05) at z = 17 μm and 7 μm, the high-frequency information of the object phase can no longer be recovered [see Figs. S6 (b3) and (d3)] due to the lower SNR.

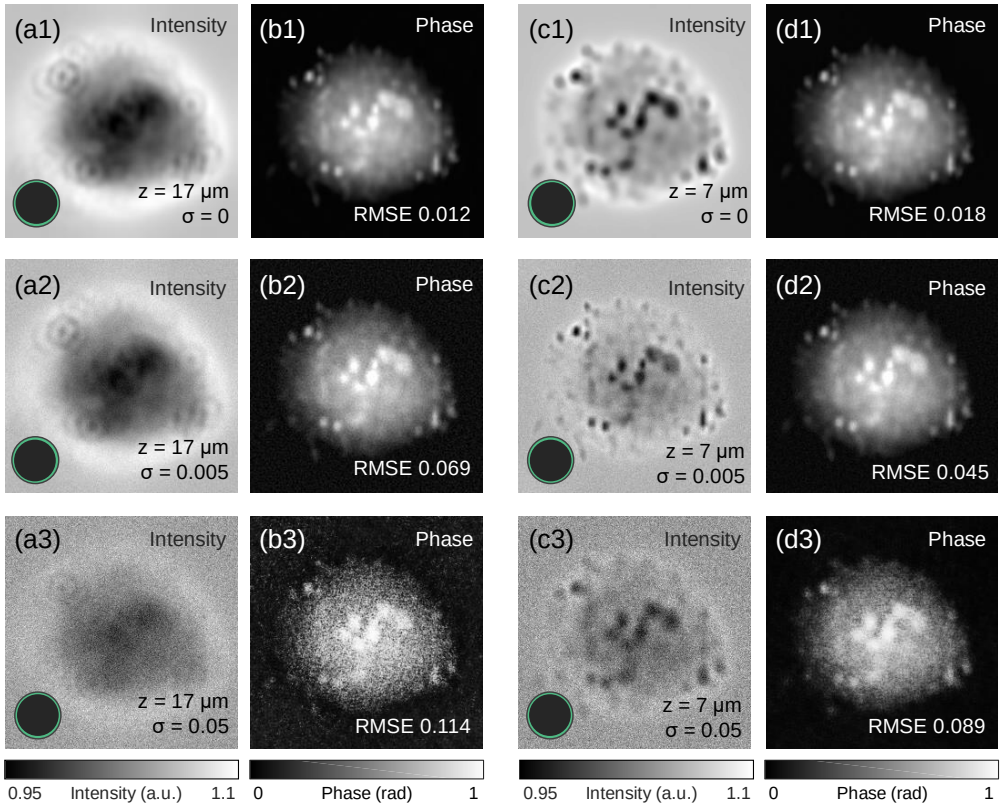

Fig. S6. (a) Defocused intensity images under annular illumination for different noise cases. (b) Corresponding recovered phase by NFTPM of (a).

5. PIXEL SUPER-RESOLUTION QPI VIA NFTPM

Deep image prior (DIP) [1] shows that the network itself can be used as an image prior for tasks such as denoising and pixel super-resolution. The image prior is implicitly captured by the choice of a particular network structure, before any of its parameters are learned [1]. In normal NFTPM, we use MLP as the image prior and Abbe imaging model as the physics prior. To implement pixel super-resolution phase retrieval, here we are inspired to expand the application of NFTPM to the pixel aliasing condition by additionally introducing pixel binning as a prior, which essentially can be considered as part of the physical model. As shown in Fig. S7 (a), the main modification in NFTPM for 2-fold pixel super-resolution (corresponding to incoherent diffraction limit) is the linear interpolation of the input coordinates and the 2-fold pixel binning of the generated intensity.

Given a pixel-aliased intensity I, the spatial coordinates $\mathbf{R} = \{\mathbf{r}_i\}_{i=0}^{M\times N-1}$ need to be linearly interpolated as $\mathbf{R}_{int} = \{\mathbf{r}_i\}_{i=0}^{2M\times 2N-1}$ and are then fed into $\Phi_{\mathbf{W}}$ to obtain unaliased phase $\Phi_{\mathbf{W}}(\mathbf{r})$. Next, we process $\Phi_{\mathbf{W}}(\mathbf{r})$ through the physical model $H\{\phi(\mathbf{r}), z\}$ to generate intensity $\tilde{I}$, and perform pixel binning $f_b(\cdot)$ on $\tilde{I}$ to get $f_b(\tilde{I})$ for comparison with I using the mean square error (MSE) loss function. Similar to Eq. (10), the above operations can be formulated as

$$\mathbf{W}^\dagger, z^\dagger = \arg\min_{\mathbf{W},z} \|f_b(H\{\Phi_{\mathbf{W}}(\mathbf{R}_{int}), z\}) - I(\mathbf{R})\|_2^2, \tag{S1}$$

where $\Phi_{\mathbf{W}^\dagger}(\mathbf{r})$ is the pixel-super-resolved phase and $z^\dagger$ is the predicted defocus distance.

To verify that NFTPM has the capability of pixel super-resolution phase imaging, in the presence of pixel aliasing [the objective NA is 0.4 (10 ×), the pixelsize is 6.5 μm, and the partially coherent light source with the wavelength of 520 nm provides an annular NA-matched illumination], we simulated a 360 × 360 pixels intensity image [Fig. S7 (b2)] at a defocus distance of 30 μm by using USAF resolution test target as the ground truth phase [Fig. S7 (b3)]. As shown in Fig. S7 (b2), the intensity has mosaic effect due to the pixel aliasing of the simulated system. Figure S7 (b1) shows

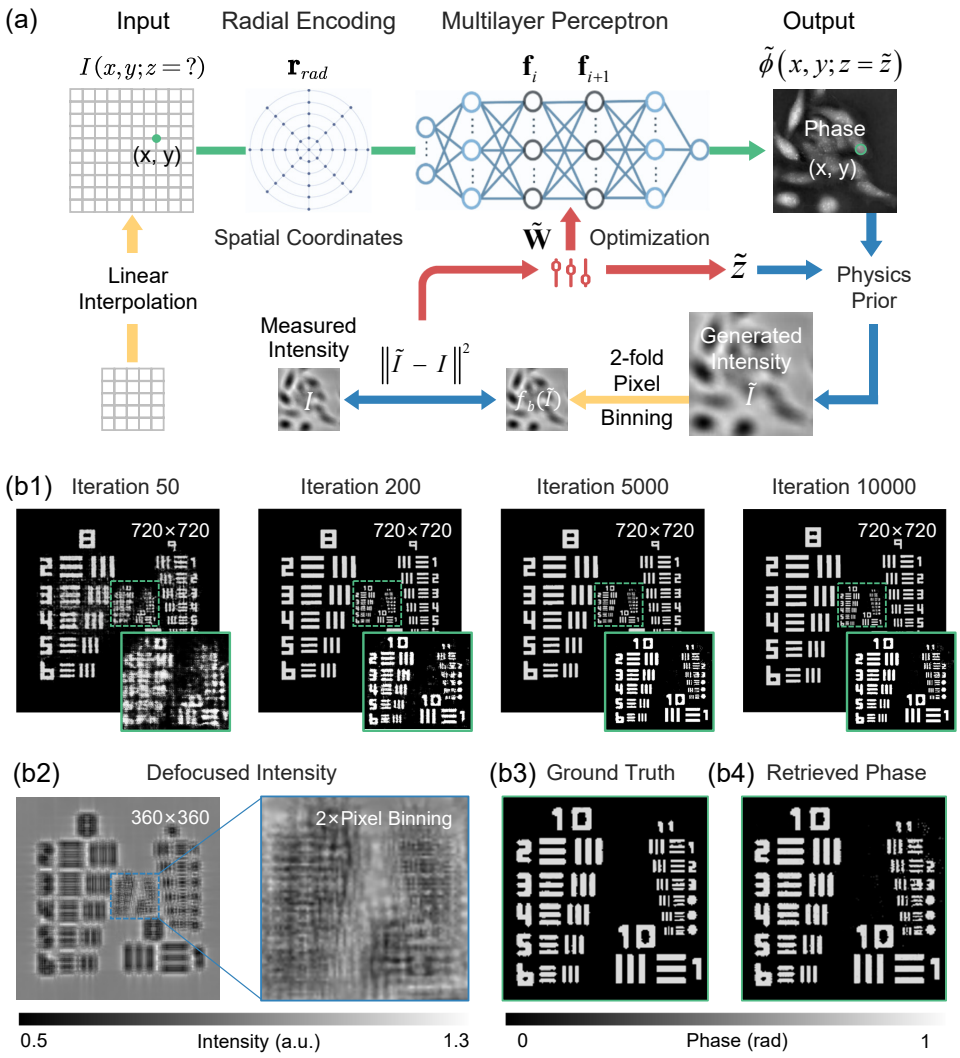

Fig. S7. (a) The schematic diagram of NFTPM for pixel super-resolution QPI. (b1) Recovered phase for different number of iterations. (b2) The simulated pixel-aliased intensity (spectrum aliasing). (b3) The USAF resolution test target used as the ground truth phase (central region). (c3) The recovered phase by NFTPM (central region).

the iterative process of pixel super-resolution reconstruction results (720 × 720 pixels) via NFTPM, and as the number of iterations increases, the high-frequency information of the recovered phase gradually becomes clear. The recovered phase illustrated in Fig. S7 (b4) is USAF group 10 and 11, where the highest distinguishable bar is group 11 elements 2. It is highly consistent with the ground truth [Fig. S7 (b3)] and overcomes the mosaic effect to obtain the incoherent diffraction limit resolution, clearly exhibiting the pixel-super-resolution capability of the NFTPM.

The reason why NFTPM can be used for pixel unaliasing by a single-frame defocused intensity is that instead of recovering the phase through the deconvolution of measurements via the PTF, NFTPM is essentially a DIP-based process that allows the MLP network (updated with a gradient-based optimization algorithm) to "guess" the phase, and forward feed the phase into the complex transfer function of the imaging system, ultimately meeting that the generated intensity image agrees with the actual measurement.

6. FORWARD MODEL OF PARTIALLY COHERENT IMAGING

The NFTPM method can achieve high-precision partially coherent illumination phase retrieval under unknown defocus distances. Compared with existing untrained methods, one of the

innovations of our algorithm is the incorporation of the image formation model under partially coherent illumination into the neural network framework. Therefore, this section mainly introduces the principles of connecting the physical model with the neural network based on matrix language.

Firstly, to facilitate understanding, let's briefly clarify the basic definitions of the characters. The matrix $\mathbf{I}^{PC}$ denotes the intensity image under partially coherent illumination, while matrix $\mathbf{I}_C^{(i)}$ denoting the intensity image under coherent illumination of the i-th point source. With $\mathbf{\Psi}$ denoting the phase represented by the neural field, $\exp(j\mathbf{\Psi})$ represents the distribution of the complex transmittance $t(\mathbf{r})$ of the sample. $\mathbf{X}_i$ denotes the distribution of $\mathbf{u}_i\mathbf{r}$ in the space domain. $\mathbf{P}$ is the complex matrix of the transfer function of the imaging system. $\mathbf{T}_i$ denotes the light field in the object plane after the illumination of the object by the i-th point source, and $\tilde{\mathbf{T}}_i$ denotes its corresponding spectrum in Fourier domain. $\mathbf{U}_C^{(i)}$, $\tilde{\mathbf{U}}_C^{(i)}$ are the spatial and spectral distribution of the light field in the imaging plane of the sample illuminated by the i-th point source, respectively. In addition, the implementation of the 2D Fourier transform is based on the 1D Fourier transform

$$\mathbf{D}_N = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & W_N^1 & W_N^2 & \cdots & W_N^{N-1} \\ 1 & W_N^2 & W_N^4 & \cdots & W_N^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & W_N^{N-1} & W_N^{2(N-1)} & \cdots & W_N^{(N-1)^2} \end{bmatrix}, \tag{S2}$$

where $W_N = e^{-j2\pi/N}$.

Firstly, the plane wave from the i-th point source passes through the sample and forms a light field

$$[\mathbf{T}_i]_{M\times N} = [\exp(j(\mathbf{\Psi} + 2\pi\mathbf{X}_i))]_{M\times N}, \tag{S3}$$

and the corresponding spectrum in the Fourier domain is

$$\left[\tilde{\mathbf{T}}_i\right]_{M\times N} = [\mathbf{D}_M]_{M\times M}[\mathbf{T}_i]_{M\times N}[\mathbf{D}_N]_{N\times N}. \tag{S4}$$

After passing through the aperture of the objective lens, the spectrum of the light field in the image plane can be expressed as

$$\left[\tilde{\mathbf{U}}_C^{(i)}\right]_{M\times N} = \left[\tilde{\mathbf{T}}_i\right]_{M\times N} \odot [\mathbf{P}]_{M\times N}, \tag{S5}$$

and can be transformed by the 2D Fourier inverse transform as

$$\left[\mathbf{U}_C^{(i)}\right]_{M\times N} = \frac{1}{MN}[\mathbf{D}_M^*]_{M\times M}\left[\tilde{\mathbf{U}}_C^{(i)}\right]_{M\times N}[\mathbf{D}_N^*]_{N\times N}, \tag{S6}$$

where $\odot$ is Hadamard product (element-wise multiplication on two matrices), and $\mathbf{D}_N^*$ is the conjugate matrix of $\mathbf{D}_N$. Eventually, the coherent light field originating from the i-th point source is captured by the camera in the form of intensity image

$$\left[\mathbf{I}_C^{(i)}\right]_{M\times N} = \left|\left[\mathbf{U}_C^{(i)}\right]_{M\times N}\right|^2 = \left(\mathrm{Re}\left\{\left[\mathbf{U}_C^{(i)}\right]_{M\times N}\right\}\right)^2 + \left(\mathrm{Im}\left\{\left[\mathbf{U}_C^{(i)}\right]_{M\times N}\right\}\right)^2. \tag{S7}$$

According to Abbe's theory of imaging, the intensity on the image plane under a partially coherent illumination can be represented as

$$\left[\mathbf{I}^{PC}\right]_{M\times N} = \sum_i \left[\mathbf{I}_C^{(i)}\right]_{M\times N}. \tag{S8}$$

7. OPTIMIZATION PROCESS OF NFTPM

A. Backpropagation in partially coherent imaging model

The optimization problem is defined as the Eq. (10), according to which the loss function can be expressed as

$$\mathcal{L}\left(\Phi_{\mathbf{W}}(\mathbf{r}), I(\mathbf{r})\right) = \frac{1}{MN}\sum_{\mathbf{r}\in\mathbf{R}} \|H(\Phi_{\mathbf{W}}(\mathbf{r})) - I(\mathbf{r})\|_2^2 = \frac{1}{MN}\sum_m\sum_n \left(\mathbf{I}_{m,n}^{PC} - \mathbf{I}_{m,n}\right)^2. \tag{S9}$$

Based on Eq. (S9), the gradient of the loss function with respect to the matrix $\mathbf{I}^{PC}$ is

$$\left[\frac{\partial \mathcal{L}}{\partial \mathbf{I}^{PC}}\right]_{M\times N} = \frac{2}{MN}\left(\left[\mathbf{I}^{PC}\right]_{M\times N} - [\mathbf{I}]_{M\times N}\right). \tag{S10}$$

According to the law of chains and Eq. (S8), the gradient of the loss function with respect to the matrix $\mathbf{I}_C^{(i)}$ can be jointly represented by $\frac{\partial \mathcal{L}}{\partial \mathbf{I}^{PC}}$ and the gradient of $\mathbf{I}^{PC}$ with respect to each coherent mode $\mathbf{I}_C^{(i)}$

$$\left[\frac{\partial \mathcal{L}}{\partial \mathbf{I}_C^{(i)}}\right]_{M\times N} = \left[\frac{\partial \mathcal{L}}{\partial \mathbf{I}^{PC}}\right]_{M\times N} \odot \left[\frac{\partial \mathbf{I}^{PC}}{\partial \mathbf{I}_C^{(i)}}\right]_{M\times N} = \left[\frac{\partial \mathcal{L}}{\partial \mathbf{I}^{PC}}\right]_{M\times N} \odot \mathbf{1} = \left[\frac{\partial \mathcal{L}}{\partial \mathbf{I}^{PC}}\right]_{M\times N}, \tag{S11}$$

where $\mathbf{1}$ is a matrix with all elements of 1 in the same shape as $\mathbf{I}^{PC}$. Similarly, the gradients of the loss function regarding the real and imaginary parts of $\mathbf{U}_C^{(i)}$ can be obtained according to equation (S7)

$$\left[\frac{\partial \mathcal{L}}{\partial \mathrm{Re}\left\{\mathbf{U}_C^{(i)}\right\}}\right]_{M\times N} = \left[\frac{\partial \mathcal{L}}{\partial \mathbf{I}_C^{(i)}}\right]_{M\times N} \odot \left[\frac{\partial \mathbf{I}_C^{(i)}}{\partial \mathrm{Re}\left\{\mathbf{U}_C^{(i)}\right\}}\right]_{M\times N} = \left[2\frac{\partial \mathcal{L}}{\partial \mathbf{I}_C^{(i)}} \odot \mathrm{Re}\left\{\mathbf{U}_C^{(i)}\right\}\right]_{M\times N}, \tag{S12}$$

$$\left[\frac{\partial \mathcal{L}}{\partial \mathrm{Im}\left\{\mathbf{U}_C^{(i)}\right\}}\right]_{M\times N} = \left[\frac{\partial \mathcal{L}}{\partial \mathbf{I}_C^{(i)}}\right]_{M\times N} \odot \left[\frac{\partial \mathbf{I}_C^{(i)}}{\partial \mathrm{Im}\left\{\mathbf{U}_C^{(i)}\right\}}\right]_{M\times N} = \left[2\frac{\partial \mathcal{L}}{\partial \mathbf{I}_C^{(i)}} \odot \mathrm{Im}\left\{\mathbf{U}_C^{(i)}\right\}\right]_{M\times N}. \tag{S13}$$

According to the 2D Fourier inverse transform, the gradients of the loss function with respect to the real and imaginary parts of $\tilde{\mathbf{U}}_C^{(i)}$ are obtained

$$\begin{aligned}\left[\frac{\partial \mathcal{L}}{\partial \mathrm{Re}\left\{\tilde{\mathbf{U}}_C^{(i)}\right\}}\right]_{M\times N} = \frac{1}{MN}\Bigg(&\left[\mathrm{Re}\left\{\mathbf{D}_M^*\right\}\frac{\partial \mathcal{L}}{\partial \mathrm{Re}\left\{\mathbf{U}_C^{(i)}\right\}}\mathrm{Re}\left\{\mathbf{D}_N^*\right\}\right]_{M\times N} - \left[\mathrm{Im}\left\{\mathbf{D}_M^*\right\}\frac{\partial \mathcal{L}}{\partial \mathrm{Re}\left\{\mathbf{U}_C^{(i)}\right\}}\mathrm{Im}\left\{\mathbf{D}_N^*\right\}\right]_{M\times N} \\ &+\left[\mathrm{Im}\left\{\mathbf{D}_M^*\right\}\frac{\partial \mathcal{L}}{\partial \mathrm{Im}\left\{\mathbf{U}_C^{(i)}\right\}}\mathrm{Re}\left\{\mathbf{D}_N^*\right\}\right]_{M\times N} + \left[\mathrm{Re}\left\{\mathbf{D}_M^*\right\}\frac{\partial \mathcal{L}}{\partial \mathrm{Im}\left\{\mathbf{U}_C^{(i)}\right\}}\mathrm{Im}\left\{\mathbf{D}_N^*\right\}\right]_{M\times N}\Bigg),\end{aligned} \tag{S14}$$

$$\begin{aligned}\left[\frac{\partial \mathcal{L}}{\partial \mathrm{Im}\left\{\tilde{\mathbf{U}}_C^{(i)}\right\}}\right]_{M\times N} = -\frac{1}{MN}\Bigg(&\left[\mathrm{Im}\left\{\mathbf{D}_M^*\right\}\frac{\partial \mathcal{L}}{\partial \mathrm{Re}\left\{\mathbf{U}_C^{(i)}\right\}}\mathrm{Re}\left\{\mathbf{D}_N^*\right\}\right]_{M\times N} + \left[\mathrm{Re}\left\{\mathbf{D}_M^*\right\}\frac{\partial \mathcal{L}}{\partial \mathrm{Re}\left\{\mathbf{U}_C^{(i)}\right\}}\mathrm{Im}\left\{\mathbf{D}_N^*\right\}\right]_{M\times N} \\ &-\left[\mathrm{Re}\left\{\mathbf{D}_M^*\right\}\frac{\partial \mathcal{L}}{\partial \mathrm{Im}\left\{\mathbf{U}_C^{(i)}\right\}}\mathrm{Re}\left\{\mathbf{D}_N^*\right\}\right]_{M\times N} + \left[\mathrm{Im}\left\{\mathbf{D}_M^*\right\}\frac{\partial \mathcal{L}}{\partial \mathrm{Im}\left\{\mathbf{U}_C^{(i)}\right\}}\mathrm{Im}\left\{\mathbf{D}_N^*\right\}\right]_{M\times N}\Bigg).\end{aligned} \tag{S15}$$

According to Eq. (S5), the gradients of the loss function regarding the real and imaginary parts of $\tilde{\mathbf{T}}_i$ are

$$\begin{aligned}\left[\frac{\partial \mathcal{L}}{\partial \mathrm{Re}\left\{\tilde{\mathbf{T}}_i\right\}}\right]_{M\times N} &= \left[\frac{\partial \mathcal{L}}{\partial \mathrm{Re}\left\{\tilde{\mathbf{U}}_C^{(i)}\right\}} \odot \frac{\partial \mathrm{Re}\left\{\tilde{\mathbf{U}}_C^{(i)}\right\}}{\partial \mathrm{Re}\left\{\tilde{\mathbf{T}}_i\right\}}\right]_{M\times N} + \left[\frac{\partial \mathcal{L}}{\partial \mathrm{Im}\left\{\tilde{\mathbf{U}}_C^{(i)}\right\}} \odot \frac{\partial \mathrm{Im}\left\{\tilde{\mathbf{U}}_C^{(i)}\right\}}{\partial \mathrm{Re}\left\{\tilde{\mathbf{T}}_i\right\}}\right]_{M\times N} \\ &= \left[\frac{\partial \mathcal{L}}{\partial \mathrm{Re}\left\{\tilde{\mathbf{U}}_C^{(i)}\right\}} \odot \mathrm{Re}\left\{\mathbf{P}\right\}\right]_{M\times N} + \left[\frac{\partial \mathcal{L}}{\partial \mathrm{Im}\left\{\tilde{\mathbf{U}}_C^{(i)}\right\}} \odot \mathrm{Im}\left\{\mathbf{P}\right\}\right]_{M\times N},\end{aligned} \tag{S16}$$

$$\begin{aligned}\left[\frac{\partial \mathcal{L}}{\partial \mathrm{Im}\left\{\tilde{\mathbf{T}}_i\right\}}\right]_{M\times N} &= \left[\frac{\partial \mathcal{L}}{\partial \mathrm{Re}\left\{\tilde{\mathbf{U}}_C^{(i)}\right\}} \odot \frac{\partial \mathrm{Re}\left\{\tilde{\mathbf{U}}_C^{(i)}\right\}}{\partial \mathrm{Im}\left\{\tilde{\mathbf{T}}_i\right\}}\right]_{M\times N} + \left[\frac{\partial \mathcal{L}}{\partial \mathrm{Im}\left\{\tilde{\mathbf{U}}_C^{(i)}\right\}} \odot \frac{\partial \mathrm{Im}\left\{\tilde{\mathbf{U}}_C^{(i)}\right\}}{\partial \mathrm{Im}\left\{\tilde{\mathbf{T}}_i\right\}}\right]_{M\times N} \\ &= -\left[\frac{\partial \mathcal{L}}{\partial \mathrm{Re}\left\{\tilde{\mathbf{U}}_C^{(i)}\right\}} \odot \mathrm{Im}\left\{\mathbf{P}\right\}\right]_{M\times N} + \left[\frac{\partial \mathcal{L}}{\partial \mathrm{Im}\left\{\tilde{\mathbf{U}}_C^{(i)}\right\}} \odot \mathrm{Re}\left\{\mathbf{P}\right\}\right]_{M\times N}.\end{aligned} \tag{S17}$$

From the 2D Fourier transform, the gradients of the loss function with respect to the real and imaginary parts of $\mathbf{T}_i$ are given as

$$\begin{aligned}\left[\frac{\partial \mathcal{L}}{\partial \mathrm{Re}\{\mathbf{T}_i\}}\right]_{M\times N} = &\left[\mathrm{Re}\left\{\mathbf{D}_M\right\}\frac{\partial \mathcal{L}}{\partial \mathrm{Re}\{\tilde{\mathbf{T}}_i\}}\mathrm{Re}\left\{\mathbf{D}_N\right\}\right]_{M\times N} - \left[\mathrm{Im}\left\{\mathbf{D}_M\right\}\frac{\partial \mathcal{L}}{\partial \mathrm{Re}\{\tilde{\mathbf{T}}_i\}}\mathrm{Im}\left\{\mathbf{D}_N\right\}\right]_{M\times N} \\ &+\left[\mathrm{Im}\left\{\mathbf{D}_M\right\}\frac{\partial \mathcal{L}}{\partial \mathrm{Im}\{\tilde{\mathbf{T}}_i\}}\mathrm{Re}\left\{\mathbf{D}_N\right\}\right]_{M\times N} + \left[\mathrm{Re}\left\{\mathbf{D}_M\right\}\frac{\partial \mathcal{L}}{\partial \mathrm{Im}\{\tilde{\mathbf{T}}_i\}}\mathrm{Im}\left\{\mathbf{D}_N\right\}\right]_{M\times N},\end{aligned} \tag{S18}$$

$$\left[\frac{\partial \mathcal{L}}{\partial \mathrm{Im}\{\mathbf{T}_i\}}\right]_{M\times N} = -\left[\mathrm{Im}\{\mathbf{D}_M\}\frac{\partial \mathcal{L}}{\partial \mathrm{Re}\{\tilde{\mathbf{T}}_i\}}\mathrm{Re}\{\mathbf{D}_N\}\right]_{M\times N} - \left[\mathrm{Re}\{\mathbf{D}_M\}\frac{\partial \mathcal{L}}{\partial \mathrm{Re}\{\tilde{\mathbf{T}}_i\}}\mathrm{Im}\{\mathbf{D}_N\}\right]_{M\times N} + \left[\mathrm{Re}\{\mathbf{D}_M\}\frac{\partial \mathcal{L}}{\partial \mathrm{Im}\{\tilde{\mathbf{T}}_i\}}\mathrm{Re}\{\mathbf{D}_N\}\right]_{M\times N} - \left[\mathrm{Im}\{\mathbf{D}_M\}\frac{\partial \mathcal{L}}{\partial \mathrm{Im}\{\tilde{\mathbf{T}}_i\}}\mathrm{Im}\{\mathbf{D}_N\}\right]_{M\times N}. \tag{S19}$$

Finally, the gradient of the loss function with respect to the phase matrix $\mathbf{\Psi}$ can be obtained according to Eq. (S3)

$$\begin{aligned}\left[\frac{\partial \mathcal{L}}{\partial \mathbf{\Psi}}\right]_{M\times N} &= \sum_i \left[\frac{\partial \mathcal{L}}{\partial \mathrm{Re}\{\mathbf{T}_i\}} \odot \frac{\partial \mathrm{Re}\{\mathbf{T}_i\}}{\partial \mathbf{\Psi}}\right]_{M\times N} + \sum_i \left[\frac{\partial \mathcal{L}}{\partial \mathrm{Im}\{\mathbf{T}_i\}} \odot \frac{\partial \mathrm{Im}\{\mathbf{T}_i\}}{\partial \mathbf{\Psi}}\right]_{M\times N} \\ &= -\sum_i \left[\frac{\partial \mathcal{L}}{\partial \mathrm{Re}\{\mathbf{T}_i\}} \odot \sin(\mathbf{\Psi} + 2\pi\mathbf{X}_i)\right]_{M\times N} + \sum_i \left[\frac{\partial \mathcal{L}}{\partial \mathrm{Im}\{\mathbf{T}_i\}} \odot \cos(\mathbf{\Psi} + 2\pi\mathbf{X}_i)\right]_{M\times N}.\end{aligned} \tag{S20}$$

B. Backpropagation in MLP

The Eq. (7) describes the forward propagation of MLP. The gradient of the loss function with respect to the feature $\mathbf{f}_i$ is expressed as

$$\left[\frac{\partial \mathcal{L}}{\partial \mathbf{f}_i}\right]_{MN\times C_i} = \left[\frac{\partial \mathcal{L}}{\partial \mathbf{f}_{i+1}} \odot \sigma'_i(\mathbf{f}_i\mathbf{W}_i)\right]_{MN\times C_{i+1}} \left[\mathbf{W}_i^{\mathrm{T}}\right]_{C_{i+1}\times C_i}. \tag{S21}$$

Similarly, the gradient of the loss function with respect to the network parameters $\mathbf{W}_i$ can be formulated as

$$\left[\frac{\partial \mathcal{L}}{\partial \mathbf{W}_i}\right]_{C_i\times C_{i+1}} = \left[\mathbf{f}_i^{\mathrm{T}}\right]_{C_i\times MN} \left[\frac{\partial \mathcal{L}}{\partial \mathbf{f}_{i+1}} \odot \sigma'_i(\mathbf{f}_i\mathbf{W}_i)\right]_{MN\times C_{i+1}}. \tag{S22}$$

When the activation function σ_i in the MLP is a Sigmoid function, the first order derivative is

$$\sigma'_i(\mathbf{f}_i\mathbf{W}_i) = \sigma_i(\mathbf{f}_i\mathbf{W}_i)(1 - \sigma_i(\mathbf{f}_i\mathbf{W}_i)). \tag{S23}$$

When σ_i is a LeakyReLU function (with negative slope s) there are

$$\sigma'_i(\mathbf{f}_i\mathbf{W}_i) = \max((\mathbf{f}_i\mathbf{W}_i), s\sigma_i(\mathbf{f}_i\mathbf{W}_i))/\sigma_i(\mathbf{f}_i\mathbf{W}_i). \tag{S24}$$

C. Phase retrieval of NFTPM

In order to make the partially coherent imaging model interface with backpropagation in MLP, it is necessary to reshape $\left[\frac{\partial \mathcal{L}}{\partial \mathbf{\Psi}}\right]_{M\times N}$ [Eq. (S20)] to flatten as a vector $\left[\frac{\partial \mathcal{L}}{\partial \mathbf{f}_K}\right]_{MN\times 1}$, thus obtaining the gradient of the loss function with respect to the output feature $\mathbf{f}_K$ of the MLP.

Then the gradient $\frac{\partial \mathcal{L}}{\partial \mathbf{f}_i}$ of the features at any layer of the MLP can be computed by using Eq. (S21) and $\frac{\partial \mathcal{L}}{\partial \mathbf{f}_K}$. Similarly, the gradient $\frac{\partial \mathcal{L}}{\partial \mathbf{W}_i}$ of the parameters of the network at any layer can be computed according to Eq. (S22), and the parameters can then be updated at each iteration by an optimization algorithm such as gradient descent

$$\mathbf{W}_i \leftarrow \mathbf{W}_i - \alpha \cdot \frac{\partial \mathcal{L}}{\partial \mathbf{W}_i}, \tag{S25}$$

where α is the learning rate. After a number of iterations of gradient backpropagation and network parameter updating, the optimization of the MLP network can finally be completed to generate the phase matrix $\mathbf{\Psi}$.

D. Defocus distance prediction of NFTPM

According to Eq. (S5), Eq. (S14) and Eq. (S15), the gradient of the loss function with respect to the defocus distance z is further derived in this section.

Firstly, matrix $\mathbf{V}$ is used to represent the distribution of $|\mathbf{u}|^2$ in the frequency domain, with the pupil function $|\mathbf{P}|$ as a circular low-pass filter with a response amplitude of 1, k as the number of waves, and z as the defocus distance. Since the defocus parameter z is univariate, it is necessary to repeat its value through the broadcast mechanism to yield a matrix $\mathbf{Z} = z\mathbf{1}$, which matches the shape of $|\mathbf{P}|$. Therefore, the transfer function matrix corresponding to the point spread function $p(\mathbf{r})$ can be expressed as

$$\mathbf{P} = |\mathbf{P}| \odot e^{jkz\sqrt{1-\lambda\mathbf{V}}} = |\mathbf{P}| \odot e^{jk\mathbf{Z}\odot\sqrt{1-\lambda\mathbf{V}}}. \tag{S26}$$

According to Eq. (S5), with respect to the real and imaginary parts of the transfer function matrix $\mathbf{P}$, the gradients of the loss function are

$$\left[\frac{\partial \mathcal{L}}{\partial \mathrm{Re}\{\mathbf{P}\}}\right]_{M\times N} = \sum_i \left[\frac{\partial \mathcal{L}}{\partial \mathrm{Re}\left\{\tilde{\mathbf{U}}_C^{(i)}\right\}} \odot \mathrm{Re}\left\{\tilde{\mathbf{T}}_i\right\}\right]_{M\times N} + \sum_i \left[\frac{\partial \mathcal{L}}{\partial \mathrm{Im}\left\{\tilde{\mathbf{U}}_C^{(i)}\right\}} \odot \mathrm{Im}\left\{\tilde{\mathbf{T}}_i\right\}\right]_{M\times N}, \tag{S27}$$

$$\left[\frac{\partial \mathcal{L}}{\partial \mathrm{Im}\{\mathbf{P}\}}\right]_{M\times N} = \sum_i \left[-\frac{\partial \mathcal{L}}{\partial \mathrm{Re}\left\{\tilde{\mathbf{U}}_C^{(i)}\right\}} \odot \mathrm{Im}\left\{\tilde{\mathbf{T}}_i\right\}\right]_{M\times N} + \sum_i \left[\frac{\partial \mathcal{L}}{\partial \mathrm{Im}\left\{\tilde{\mathbf{U}}_C^{(i)}\right\}} \odot \mathrm{Re}\left\{\tilde{\mathbf{T}}_i\right\}\right]_{M\times N}. \tag{S28}$$

The gradient of the loss function concerning $\mathbf{Z}$ is

$$\left[\frac{\partial \mathcal{L}}{\partial \mathbf{Z}}\right]_{M\times N} = \left[\frac{\partial \mathcal{L}}{\partial \mathrm{Re}\{\mathbf{P}\}}\right]_{M\times N} \odot \left[\frac{\partial \mathrm{Re}\{\mathbf{P}\}}{\partial \mathbf{Z}}\right]_{M\times N} + \left[\frac{\partial \mathcal{L}}{\partial \mathrm{Im}\{\mathbf{P}\}}\right]_{M\times N} \odot \left[\frac{\partial \mathrm{Im}\{\mathbf{P}\}}{\partial \mathbf{Z}}\right]_{M\times N}, \tag{S29}$$

where the gradients of the real and imaginary parts of $\mathbf{P}$ with respect to $\mathbf{Z}$ are as flows

$$\frac{\partial \mathrm{Re}\{\mathbf{P}\}}{\partial \mathbf{Z}} = -|\mathbf{P}| \odot k\sqrt{1-\lambda\mathbf{V}} \odot \sin\left(k\mathbf{Z} \odot \sqrt{1-\lambda\mathbf{V}}\right), \tag{S30}$$

$$\frac{\partial \mathrm{Im}\{\mathbf{P}\}}{\partial \mathbf{Z}} = |\mathbf{P}| \odot k\sqrt{1-\lambda\mathbf{V}} \odot \cos\left(k\mathbf{Z} \odot \sqrt{1-\lambda\mathbf{V}}\right). \tag{S31}$$

The gradient with respect to the matrix $\mathbf{Z}$ is then converted to a gradient with respect to z by summation

$$\frac{\partial \mathcal{L}}{\partial z} = \sum_{m=1}^{M}\sum_{n=1}^{N} \frac{\partial \mathcal{L}}{\partial \mathbf{Z}_{m,n}}. \tag{S32}$$

Finally, the parameters are updated at each iteration based on the calculated gradient

$$z \leftarrow z - \alpha_z \cdot \frac{\partial \mathcal{L}}{\partial z}, \tag{S33}$$

where α_z is the learning rate. According to Eq. (S25) and Eq. (S33), after synchronously updating the network parameters and the defocus distance parameter, phase recovery and defocus distance prediction can eventually be achieved.

REFERENCES

1. D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2018), pp. 9446–9454.