Supplemental Document

Optics Letters

Lensless masked imaging with self-calibrated phase retrieval: supplement

SHENGHAO ZHENG,[1] ZHIHUI DING,[1] RUI JIANG,[1] AND CHENG GUO[1,2,*]

[1] *Center of Ultra-precision Optoelectronic Instrument Engineering, Harbin Institute of Technology, Harbin 150080, China*
[2] *Faculty of Computing, Harbin Institute of Technology, Harbin 150001, China*
[*] *guocheng_27@163.com*

Supplement DOI: https://doi.org/10.6084/m9.figshare.23189936

Parent Article DOI: https://doi.org/10.1364/OL.492476

Lensless masked imaging with self-calibrated phase retrieval: supplemental document

The supplemental document is organized as follows:

(1) The detailed derivation of binary-coding phase retrieval is added;
(2) The numerical simulation results of binary-coding phase retrieval are added, where the imaging performances on the selection of denoiser, pixel size, occlusion ratio, and diffractive distance are discussed;
(3) The auto-focusing curve of Z_2^n estimation is given;
(4) The auto-focusing curve of Z_1 estimation is given;
(5) The workflow of the multi-SPICA algorithm is given;
(6) The single-shot reconstructed results of our method and SPICA algorithm are added;
(7) The comparison of our method and multi-distance phase retrieval is supplemented;
(8) The comparison of our method with different initial guess is added;
(9) Ablation studies of our method are supplemented;
(10) The imaging performances on CNN denoisers are given;
(11) The parameter tuning of SPICA and Multi-SPICA is discussed;
(12) The selection of binary mask is discussed;
(13) The reconstructed results of H&E-stained human tongue fungiform papillae, H&E-stained cow lung tissue, and pure phase target are added;
(14) The reason why the mismatch error exists is given.

1. Derivation of binary-coding phase retrieval

In this section, we provide a detailed derivation of binary-coding phase retrieval. In our method, the diffraction propagation calculation is defined by angular spectrum model. Thus, the forward A_Z and backward A_Z^{-1} propagation operators can be expressed as follows

$$\begin{cases} A_Z(\cdot) = \mathcal{F}^{-1}\left[\mathcal{F}(\cdot) \odot H_Z\right] \\ A_Z^{-1}(\cdot) = \mathcal{F}^{-1}\left[\mathcal{F}(\cdot) \odot H_Z^*\right] \end{cases} \tag{S1}$$

$$H_Z(f_x, f_y) = \begin{cases} \exp\left(\frac{2\pi \mathrm{i} Z}{\lambda}\sqrt{1-(\lambda f_x)^2-(\lambda f_y)^2}\right), (\lambda f_x)^2+(\lambda f_y)^2 \le 1 \\ 0, \qquad\qquad (\lambda f_x)^2+(\lambda f_y)^2 > 1 \end{cases} \tag{S2}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote Fourier transform and inverse form. (f_x, f_y) is frequency coordinate, and λ represents the wavelength of the incident light, the superscript '*' denotes a complex conjugate operation. Binary-coding phase retrieval is constructed as a minimization problem as

$$O_{k+1}^M = \arg\min_{O_k^M} \frac{1}{N}\sum_{n=1}^{N}\left\| \left| A_{Z_2^n}\left(M \odot O_k^M\right)\right|^2 - I_n^w \right\|_2^2, \; n \in [1, N] \tag{S3}$$

to decouple the incident wavefield by using a given mask function (M) and intensity diffraction patterns (I_n^w). Here we formulate two sub-problems to solve the Eq. (S3) as follows

$$T_k^M = \frac{1}{N}\sum_{n=1}^{N} A_{Z_2^n}^{-1}\left[\frac{A_{Z_2^n}\left(M \odot O_k^M\right)}{\left|A_{Z_2^n}\left(M \odot O_k^M\right)\right|}\sqrt{I_n^w}\right], \tag{S4}$$

$$O_{k+1}^{M}=\arg\min_{O_k^M}\frac{1}{2}\left\|M\odot O_k^M-T_k^M\right\|_2^2. \tag{S5}$$

The Eq. (S4) is a classical phase retrieval solver for multi-distance intensity datasets [1-3], which can be conducted by alternative propagation projection and amplitude replacement. After the transmitted wavefield of the mask T_k^M is obtained, the minimization of Eq. (S5) can be used to decode the incident wavefield. Since the diffraction field is represented by a complex amplitude function, the Eq. (S5) should be decomposed into real and imaginary parts to be iteratively solved as

$$X_{j+1}^{\mathrm{Re}}=\arg\min_{X_j^{\mathrm{Re}}}\frac{1}{2}\left\|M\odot X_j^{\mathrm{Re}}-\mathrm{Re}\left(T_k^M\right)\right\|_2^2,\ X_1^{\mathrm{Re}}=\mathrm{Re}\left(T_k^M\right), \tag{S6}$$

$$X_{j+1}^{\mathrm{Im}}=\arg\min_{X_j^{\mathrm{Im}}}\frac{1}{2}\left\|M\odot X_j^{\mathrm{Im}}-\mathrm{Im}\left(T_k^M\right)\right\|_2^2,\ X_1^{\mathrm{Im}}=\mathrm{Im}\left(T_k^M\right), \tag{S7}$$

where j is the index of iterations (j =1, 2, …, J), $\odot$ is an element-wise product. With the use of the plug-and-play alternating direction method of multipliers (PnP-ADMM), the Eq. (S6) is expanded as

$$X_{j+1}^{\mathrm{Re}}=\arg\min_{X_j^{\mathrm{Re}}}\frac{1}{2}\left\|M\odot X_j^{\mathrm{Re}}-\mathrm{Re}\left(T_k^M\right)\right\|_2^2+\frac{\rho}{2}\left\|X_j^{\mathrm{Re}}-\left(V_j^{\mathrm{Re}}-U_j^{\mathrm{Re}}\right)\right\|_2^2, \tag{S8}$$

$$V_{j+1}^{\mathrm{Re}}=\mathcal{D}_\sigma\left(X_{j+1}^{\mathrm{Re}}+U_j^{\mathrm{Re}}\right), \tag{S9}$$

$$U_{j+1}^{\mathrm{Re}}=U_j^{\mathrm{Re}}+X_{j+1}^{\mathrm{Re}}-V_{j+1}^{\mathrm{Re}}. \tag{S10}$$

where $\mathcal{D}_\sigma$ denotes a denoising filter with a noise variance of σ, ρ is a regularization parameter. The minimization of Eq. (S8) can be solved as

$$\frac{\partial G\left(X_j^{\mathrm{Re}}\right)}{\partial X_j^{\mathrm{Re}}}=\frac{\partial\left\{\frac{1}{2}\left\|M\odot X_j^{\mathrm{Re}}-\mathrm{Re}\left(T_k^M\right)\right\|_2^2+\frac{\rho}{2}\left\|X_j^{\mathrm{Re}}-\left(V_j^{\mathrm{Re}}-U_j^{\mathrm{Re}}\right)\right\|_2^2\right\}}{\partial X_j^{\mathrm{Re}}}=0. \tag{S11}$$

Accordingly, the closed form of the above partial derivative is expressed as

$$\begin{aligned}\frac{\partial G\left(X_j^{\mathrm{Re}}\right)}{\partial X_j^{\mathrm{Re}}}&=M^T\odot\left(M\odot X_j^{\mathrm{Re}}-\mathrm{Re}\left(T_k^M\right)\right)+\rho\left[X_j^{\mathrm{Re}}-\left(V_j^{\mathrm{Re}}-U_j^{\mathrm{Re}}\right)\right]\\&=\left(M^T\odot M+\rho E\right)\odot X_j^{\mathrm{Re}}-\left[M^T\odot\mathrm{Re}\left(T_k^M\right)+\rho\left(V_j^{\mathrm{Re}}-U_j^{\mathrm{Re}}\right)\right].\end{aligned} \tag{S12}$$

With the use of Eq. (S11) and Eq. (S12), the closed form of Eq. (S8) is arranged as

$$X_{j+1}^{\mathrm{Re}}=\frac{\left[M^T\odot\mathrm{Re}\left(T_k^M\right)+\rho\left(V_j^{\mathrm{Re}}-U_j^{\mathrm{Re}}\right)\right]}{M^T\odot M+\rho E}. \tag{S13}$$

where the division is an element-wise operation. Based on the above derivation, the solution of Eq. (S6) is given as

$$X_{j+1}^{\mathrm{Re}}=\frac{\left[M^T\odot\mathrm{Re}\left(T_k^M\right)+\rho\left(V_j^{\mathrm{Re}}-U_j^{\mathrm{Re}}\right)\right]}{M^T\odot M+\rho E}, \tag{S14}$$

$$V_{j+1}^{\mathrm{Re}}=\mathcal{D}_\sigma\left(X_{j+1}^{\mathrm{Re}}+U_j^{\mathrm{Re}}\right), \tag{S15}$$

$$U_{j+1}^{\mathrm{Re}}=U_j^{\mathrm{Re}}+X_{j+1}^{\mathrm{Re}}-V_{j+1}^{\mathrm{Re}}. \tag{S16}$$

Similarly, the solution of the Eq. (S7) is given as:

$$X_{j+1}^{\mathrm{Im}}=\frac{\left[M^T\odot\mathrm{Im}\left(T_k^M\right)+\rho\left(V_j^{\mathrm{Im}}-U_j^{\mathrm{Im}}\right)\right]}{M^T\odot M+\rho E}, \tag{S17}$$

$$V_{j+1}^{\mathrm{Im}} = \mathcal{D}_{\sigma}\left(X_{j+1}^{\mathrm{Im}} + U_{j}^{\mathrm{Im}}\right), \tag{S18}$$

$$U_{j+1}^{\mathrm{Im}} = U_{j}^{\mathrm{Im}} + X_{j+1}^{\mathrm{Im}} - V_{j+1}^{\mathrm{Im}}. \tag{S19}$$

2. Numerical simulation of binary-coding phase retrieval

2.1 Pixel size and occlusion ratio of mask

In this section, we conduct the numerical simulation to get the optimal parameters of binary mask for the lensless masked imaging system. The grey-scaled biological tissue images captured from a commercial microscope (Olympus, IX71) are used for ground truth. The simulated parameters are listed as follows: (1) the amplitude image of the object is a grey-scale image and its phase is set as a constant value; (2) the sampling size of the diffraction field is 512×512; (3) the wavelength is 532nm; (4) the sample-to-mask distance is 3mm; (5) the initial value and interval of the mask-to-sensor distances are set as 3mm and 0.1mm; (6) the number of multi-distance intensity patterns is 11; (7) iterative number is 30. In the simulation, the object is propagated to the mask plane and sieved by a binary amplitude mask. The sieved diffraction field is propagated with different mask-to-sensor distances to generate a set of intensity patterns. Binary-coding phase retrieval is used to reconstruct the complex-valued image of an object. Peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) are used to assess the reconstructed accuracy. The recursive filter is selected as the denoiser in the calculation.

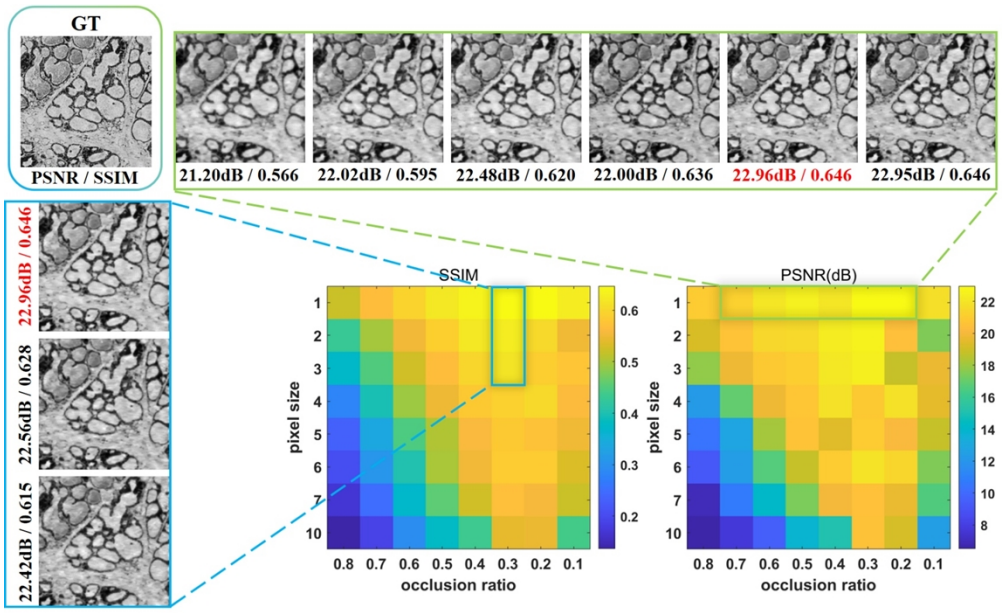

Fig. S1. Numerical reconstructed results with different pixel sizes and occlusion ratios for binary amplitude mask.

Figure S1 is given to discuss the influence of the mask's pixel size and occlusion ratio on the reconstructed accuracy. For the binary amplitude mask, the pixels with '0' and '1' could be used to switch on and off the incident light, where the incident wave is blocked at '0' and passed at '1'. Thus, we introduce an occlusion ratio to describe how much the incident wave is blocked, and the definition of occlusion ratio is presented as

$$occlusion\ ratio = \frac{S_{block}}{S_{total}}, \tag{S20}$$

where S_{block} denotes the pixel number of '0', and S_{total} is the total pixel number of the mask. As the pixel size of the mask is set from 1μm to 10μm and the occlusion ratio is set from 0.8 to 0.1, SSIM and PSNR maps obtained from 64 retrieved images are pictured in Fig. S1. When the occlusion ratio is 0.4, the retrieved results with pixel sizes of 1μm, 2μm, and 3μm are shown in the blue box. When the pixel size is set as 1μm, the retrieved results with different occlusion ratios are displayed in the green box. It can be observed from Fig. S1 that the optimal value of the occlusion ratio is ~0.3. Also, the reconstructed accuracy will be enhanced if the pixel size of mask is smaller. Based on this rule, we fabricate two masks with pixel sizes of 2μm and 4μm to verify the imaging performance. Unfortunately, the mask-to-sensor distances cannot be acquired by an auto-focusing algorithm for the mask (pixel size is 2μm). This poor performance is attributed that a smaller pixel size introduces a stronger scattering. The strong scattering scrambles diffraction field, and thus back-propagated Z-stack data is wrapped with a strong speckle pattern. In this situation, it is difficult to get the sharpest image from the Z-stack data. Therefore, the parameters of the mask are specified as: pixel size is 4μm, and occlusion ratio is 0.3.

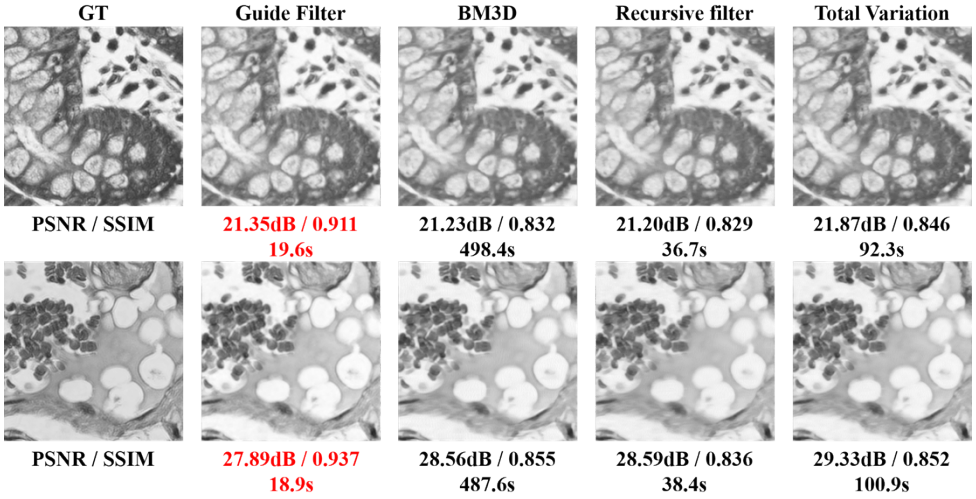

Fig. S2. Numerical reconstructed results with different denoisers.

2.2 Denoiser selection

As used in Eq. (S15) and Eq. (S18), an image filter needs to be embedded as a denoising prior. Here we consider BM3D [4], guide filter [5], recursive filter [6], and total variation [7] to find an optimal denoiser. Two gray-scale natural images are chosen to form object images, and the corresponding reconstructed images using different denoisers are shown in Fig. S2, where PSNR and SSIM values are labelled. It is noted that the guide filter gets the highest metric values for the two object images. In addition to reconstructed accuracy, running time should be considered for each filter. As the four filters are embedded into binary-coding phase retrieval, the total running time related to the four filters are listed at the bottom of each picture. Based on the above analysis, the guide filter not only has the highest reconstructed accuracy but also owns the shortest running time. Therefore, we use the guide filter as a denoising prior.

2.3 Convergence analysis

To analyze the convergence of binary-coding phase retrieval, we use a biological tissue image to test the performance of reconstruction, and the retrieved results are presented in Fig. S3. Fig. S3(a) is chosen as the ground truth. The simulated parameters are same as the above setting. The binary amplitude mask with an occlusion ratio of 0.3 is shown in Fig. S3(b). After diffraction propagation and binary selection, the intensity pattern at the initial plane is given in Fig. S3(c). After 50 iterations, the convergence curve of binary-coding phase retrieval is plotted

in Fig. S3(d). It is noted that the curve becomes flat while the iterative number is larger than 10. Therefore, we set the iterative number as 10, i.e., K=10.

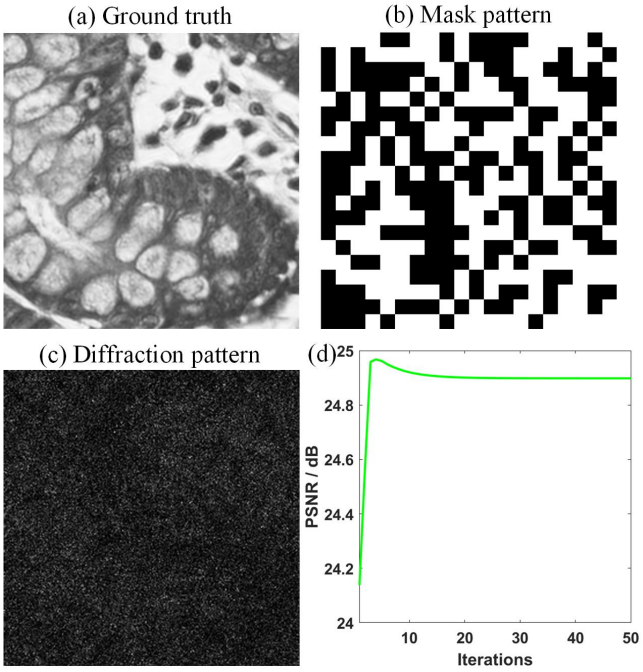

Fig. S3. Simulation results and convergence curve.

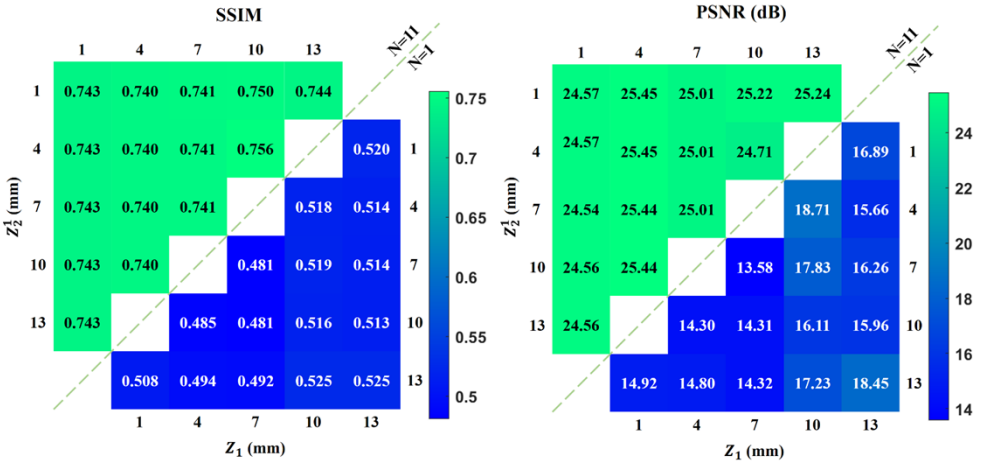

Fig. S4. Reconstructed accuracy comparison with different sample-to-mask distances and initial distances under N=11 and N=1.

2.4 Diffractive distances and number of multi-distance measurement

In our system, diffractive distances are composed of two parts: sample-to-mask distance (Z_1) and mask-to-sensor distances ($Z_2^n, n \in [1, N]$). To simplify the distance setting, the mask-to-sensor distances are represented by initial distance (Z_2^1) and interval (d) as follows:

$$Z_2^n = Z_2^1 + (n-1)d, \quad n \in [1, N]. \tag{S21}$$

As Z_1 and Z_2^1 are both set from 1mm to 13mm, and the interval is fixed at 0.1mm, the SSIM and PSNR maps of different retrieved images under N=1 and N=11 are pictured in Fig. S4(a)

and Fig. S4(b), respectively. The two metric values are displayed with color bars, in which the metric value gradually decreases from green to blue. It can be observed that the SSIM and PSNR values of N=11 hold at ~0.74 and ~25dB for different distances, which indicates that our method is not sensitive to the selection of Z_1 and Z_2^1. Compared to the single-shot results (N=1), the results of N=11 get higher recovery accuracy. This remarkable enhancement accords with the experimental results (Figs. 3(a1-a3) versus Figs. 3(b1-b3)), which proves that multi-distance measurement brings a stable convergence for binary-coding phase retrieval. As Z_1 and Z_2^1 are both fixed at 3mm, and the interval ranges from 0.1mm to 4.1mm, the SSIM and PSNR maps of different retrieved images under N=11 is pictured in Fig. S5(a) and Fig. S5(b), respectively. Similar to the results of Fig. S4, the metric values of our method still keep a stable level.

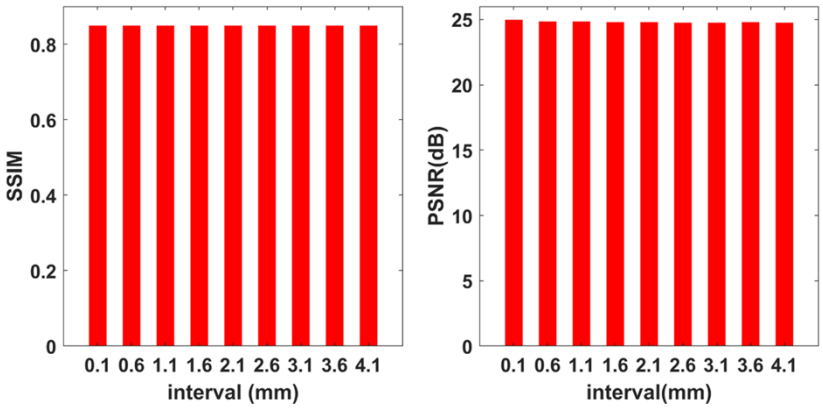

Fig. S5. Reconstructed accuracy comparison with different intervals under N=11

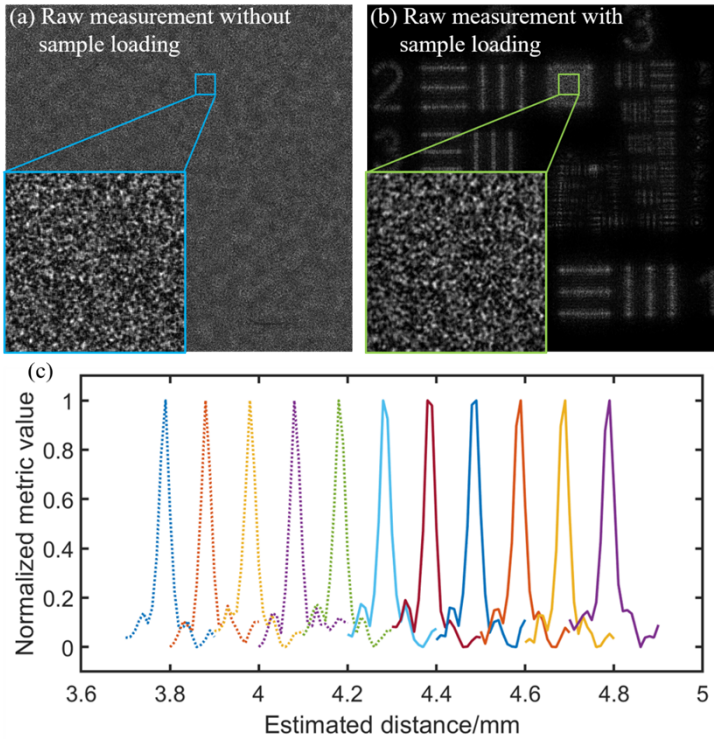

Fig. S6. The estimation of mask-to-sensor distances for USAF resolution target.

3. Distance estimation

3.1 mask-to-sensor distances

Distance estimation is regarded as a process of auto-focusing, where a real distance is acquired by the back-propagation of diffraction patterns. Our previous work [8] is used to accomplish this task. As an example, the auto-focusing results of USAF resolution chart are given to exhibit the workflow of distance estimation. Fig. S6(a) is the recorded intensity pattern of the initial plane without sample loading, and Fig. S6(b) corresponds to the pattern with sample loading. In our system, the mask only needs to be calibrated for once, thus the intensity patterns without sample are used to find the true mask-to-sensor distances. The corresponding auto-focusing workflow is generalized as: (1) the intensity patterns without sample $I_n^{w/o}$ are inversely propagated with a set of diffractive distances

$$S_n^r = \mathcal{F}^{-1}\left[\mathcal{F}\left(\sqrt{I_n^{w/o}}\right) \odot H_{Z_r}^*\right],\ r = 1, 2, ..., R, \tag{S22}$$

to generate a Z-stack data (S_n^r); (2) nuclear norm of gradient (NoG) metric is used to assess the Z-stack data:

$$\mathrm{NoG}_n\left(r\right) = \left\| \left|\nabla_x\left(S_n^r\right)\right|^2 + \left|\nabla_y\left(S_n^r\right)\right|^2 \right\|_*, \tag{S23}$$

where $\|X\|_*$ denotes a nuclear norm of X, which can be calculated by summing up all singular values of X; (3) the peak value of $\mathrm{NoG}_n\left(r\right)$ is related to the real distance; (4) repeating the step (1) to step (3) so that all mask-to-sensor distances are obtained. For the accuracy of the distance estimation, the step size of distance searching is 0.01mm for pixel-limited scene and 0.001mm for pixel super-resolved scene, which can be found in Ref. [8] and Ref. [13]. We do not introduce a pixel super-resolved scanning module in the present system, thus the step size of auto-focusing is 0.01mm for all experiments. We utilize the NoG metric to plot an auto-focusing curve in Fig. S6(c), where the initial distance and interval of the mask-to-sensor distances are specified as 3.79mm and 0.10mm.

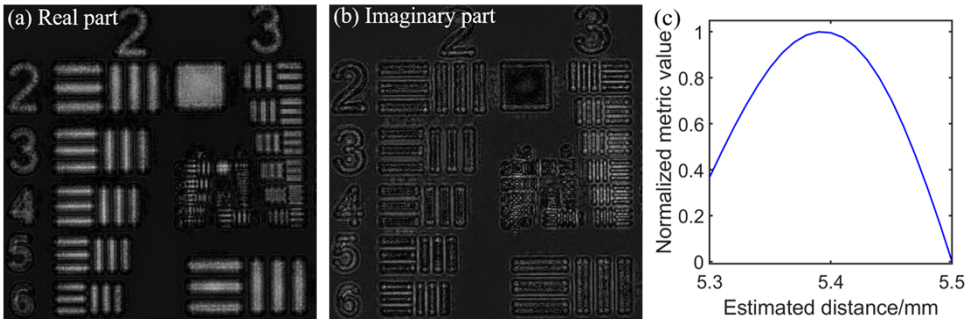

Fig. S7. The estimation of sample-to-mask distance for USAF resolution target. (a) and (b) are real and imaginary parts of O_K^M . (c) is the auto-focusing curve for Z_1 estimation.

3.2 sample-to-mask distance

As shown in Fig. 1(c) in the manuscript, the incident wavefield of the mask plane O_K^M can be reconstructed by binary-coding phase retrieval. The real and imaginary parts of O_K^M are displayed in Fig. S7(a) and Fig. S7(b). Thus, we still use the NoG metric to acquire the real sample-to-mask distance, which is expressed as follows

$$S^r = \mathcal{F}^{-1}\left[\mathcal{F}\left(O_K^M\right) \odot H_{Z_r}^*\right],\ r = 1, 2, ..., R, \tag{S24}$$

$$\mathrm{NoG}(r)=\left\|\left|\nabla_x\left(S^r\right)\right|^2+\left|\nabla_y\left(S^r\right)\right|^2\right\|_*. \tag{S25}$$

Using the Eq. (S24) and Eq. (S25), the auto-focusing curve of sample-to-mask distance is plotted in Fig. S7(c), where the real distance points at 5.43mm.

4. Multi-SPICA algorithm

SPICA algorithm combines compressive Fresnel holography with support-based phase retrieval to reconstruct complete wavefield from a single-shot coded pattern. Considering that our method needs multi-distance patterns, we update SPICA with the multi-distance dataset to form a multi-SPICA algorithm for a fair comparison. The flowchart of multi-SPICA is generalized as: (1) the original single-shot alternative projection is corrected as the multi-plane projection, which is expressed as

$$T_{k+1}^M=\frac{1}{N}\sum_{n=1}^{N}A_{Z_2^n}^{-1}\left[\frac{A_{Z_2^n}\left(M\odot O_k^M\right)}{\left|A_{Z_2^n}\left(M\odot O_k^M\right)\right|}\sqrt{I_n^w}\right],\ k=1,2,...,K, \tag{S26}$$

(2) after the loop of Eq. (S26) is converged, the output is written as T_K^M; (3) a two-step iterative shrinkage/thresholding (TwIST) algorithm [9] is constructed as

$$\hat{O}^S=\arg\min_{O^S}\frac{1}{2}\left\|M\odot A_{Z_1}\left(O^S\right)-T_K^M\right\|_2^2+\rho\left|\nabla O^S\right|, \tag{S27}$$

to acquire a final estimation of the sample.

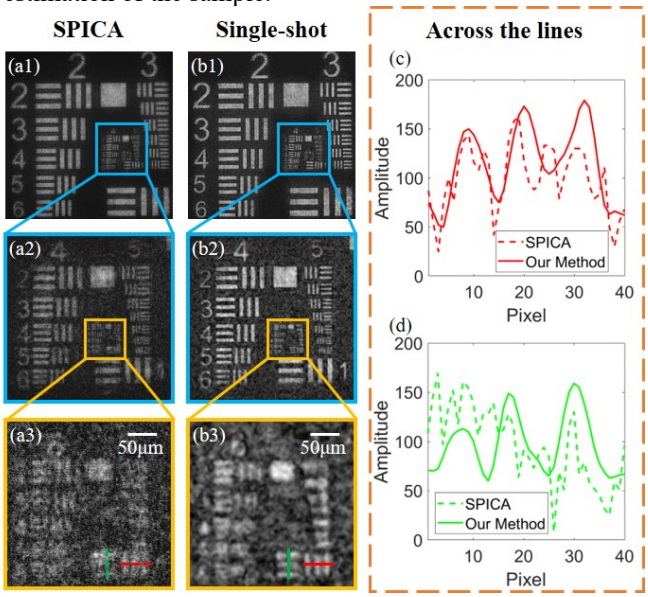

Fig. S8. Single-shot reconstructed results of SPICA and binary-coding phase retrieval. (a1-a3) and (b1-b3) are reconstructed images of the SPICA algorithm and our method by using a single-frame intensity pattern. (c1-c2) are plotlines captured from (a3) and (b3).

5. Single-shot reconstructed results of SPICA and binary-coding phase retrieval

We correct our binary-coding phase retrieval to a single-shot mode by replacing the multi-plane projection (Eq. 3) with a single-shot projection, which is operated as follows

$$T_k^M=A_{Z_2^1}^{-1}\left[\frac{A_{Z_2^1}\left(M\odot O_k^M\right)}{\left|A_{Z_2^1}\left(M\odot O_k^M\right)\right|}\sqrt{I_1^w}\right]. \tag{S28}$$

With the use of single-frame intensity data (Fig. S6(b)), the retrieved amplitude images of SPICA and binary-coding phase retrieval are shown in Figs. S8(a1-a3) and Figs. S8(b1-b3). To further show the difference, the plotlines along green and red dash lines captured from Fig. S8(a3) and Fig. S8(b3) are pictured in Fig. S8(c1) and Fig. S8(c2). As shown in Fig. S8, even using single-frame intensity data, our method still outperforms the SPICA algorithm with a higher resolution and better imaging contrast. However, compared to multi-image results (Figs. 3(c1-c3) in the manuscript), single-shot reconstructed results in Fig. S8(b1-b3) show a poor imaging resolution, which accords with the simulated results, i.e., the use of multi-distance dataset enables a higher reconstructed accuracy.

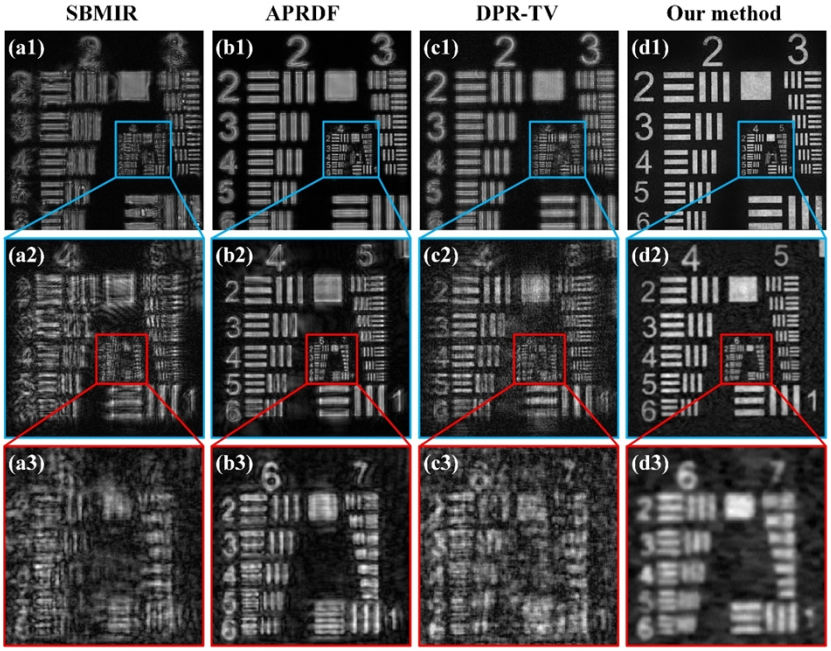

Fig. S9. Reconstructed resolution target of binary-coding phase retrieval and MDPR methods.

6. Comparison of binary-coding phase retrieval and multi-distance phase retrieval

We compare our method with multi-distance phase retrieval (MDPR) methods in Fig. S9. For a fair comparison, our method and MDPR share a common experimental configuration except that the binary amplitude mask is removed for the MDPR system. Regular MDPR methods, including SBMIR [1], APRDF [2], and DPR-TV [3], are used to form the comparison group. After 50 iterations, the retrieved amplitude images of SMBIR, APRDF, DPR-TV, and our method are displayed in Figs. S9(a1-a3), Figs. S9(b1-b3), Figs. S9(c1-c3), and Figs. S9(d1-d3). It is observed that the retrieved image of our method shows a good reconstruction quality for both imaging contrast and resolution. The results of MDPR methods are constrained by noisy backgrounds and blurred edges. This bad performance of MDPR is attributed that the used lensless system is not an on-chip setup but adopts a spherical wave illumination. Under the spherical illumination, defocused intensity patterns scale up with the diffractive distance increasing, which leads to a scale mismatch for the multi-distance dataset. However, the MDPR methods utilize angular spectrum diffraction model for calculation, and thus the intrinsic property of angular spectrum confines the MDPR methods under plane wave illumination. In contrast, our method imposes a binary mask to modulate the scale-up diffraction field of sample, which contributes a strong physical constraint for phase retrieval. Thus, the encode and decode process of the mask is beneficial to good convergence for image reconstruction.

7. Reconstructed results with different initial values

As shown in Fig. 1 (c), our binary-coding phase retrieval algorithm starts with an initial guess of incident wavefield of the mask plane, i.e., $O_1^M = U^w$. Here we discuss how this initial guess influences the performance and convergence rate of our method. As the initial guess is set as a random image or $O_1^M = U^w$, the corresponding retrieved results are shown in Fig. S10. We can see that when the initial guess is set as $O_1^M = U^w$, our method has a faster convergence than using a random initial guess.

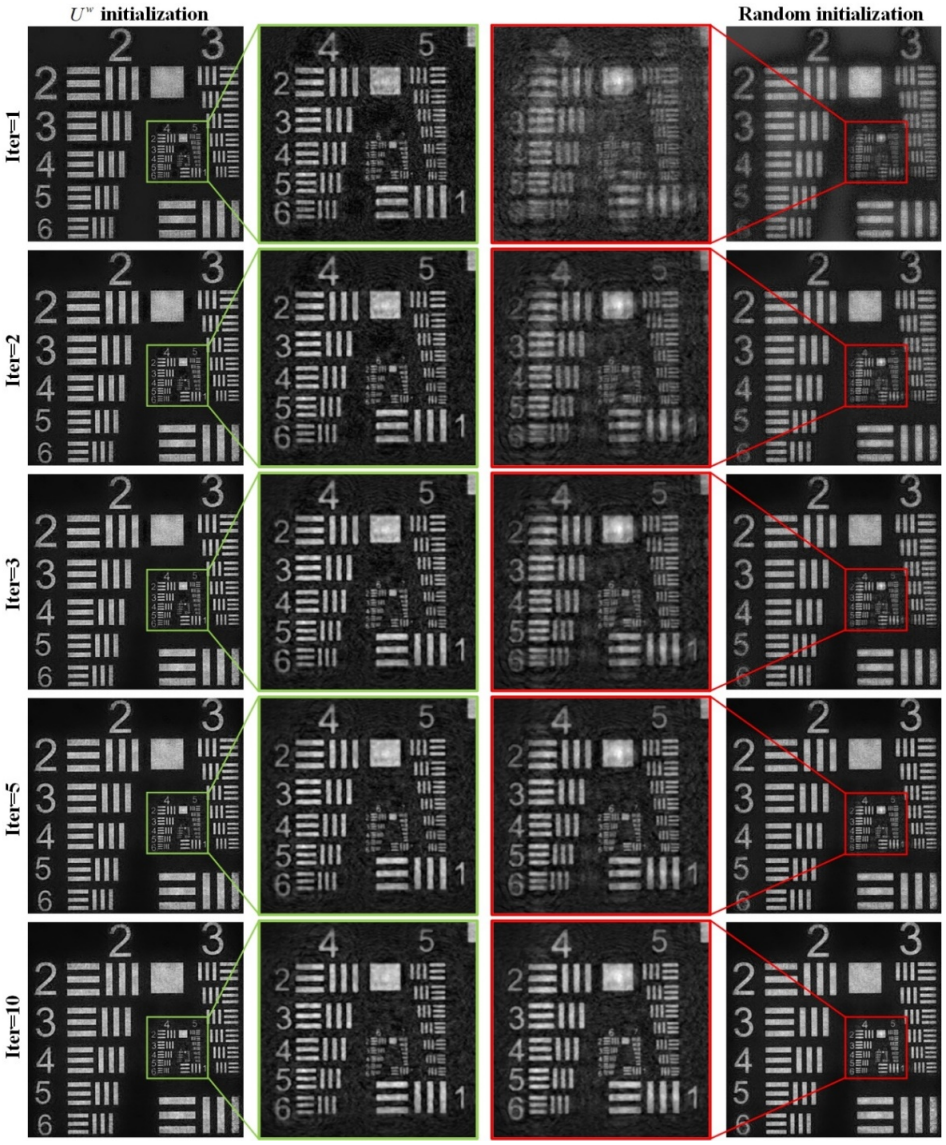

Fig. S10. Reconstruction results with different initial guess values

8. Ablation studies of our method

We add ablation studies to analyze how the parameters in our method influence the performance of our method so that the optimal parameters can be obtained.

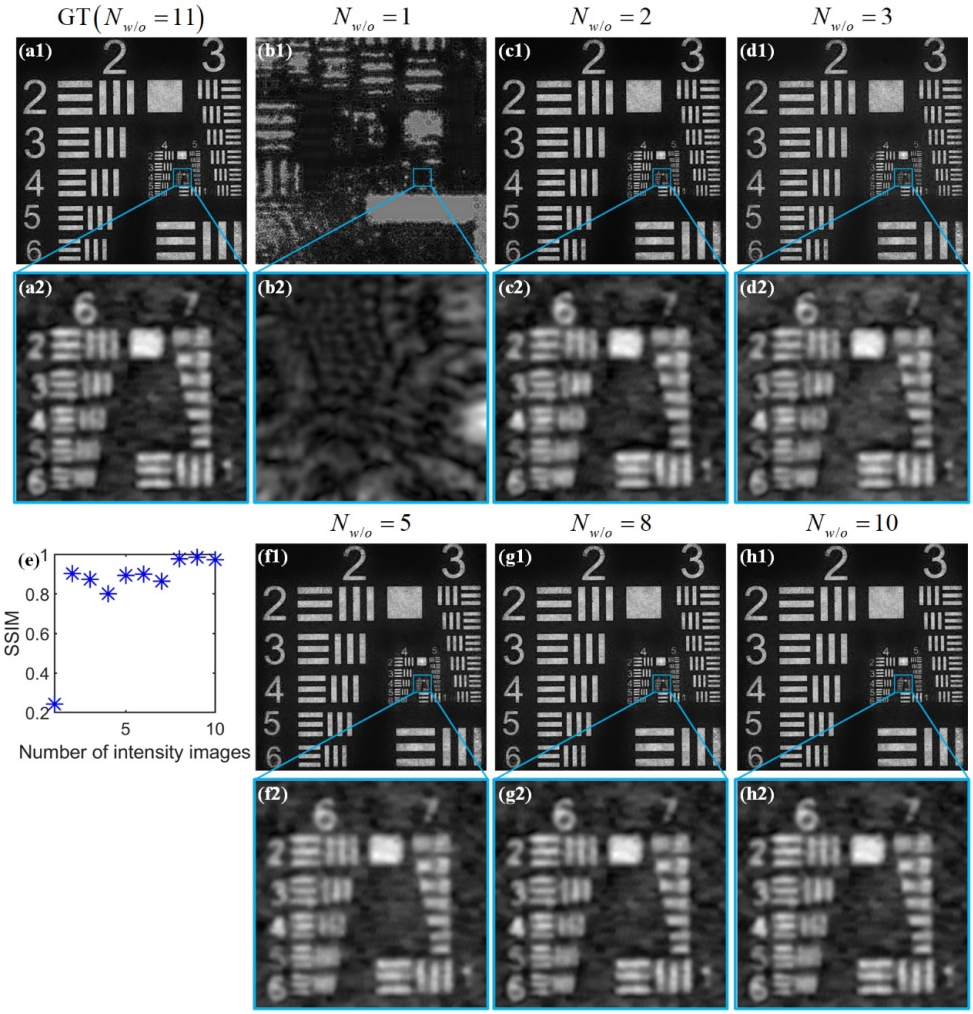

Fig. S11. Reconstruction results with the change of $N_{w/o}$ when N_w is held at 11.

8.1 the number of intensity patterns

We analyze how the number of intensity images influence the performance of our method. Considering that two groups of multi-distance intensity images are required in our method, we vary the number of intensity images of one group and fix that of another to analyze the influence.

For simple description, we use N_w and $N_{w/o}$ to represent the number of multi-distance intensity images with/without sample loading. As N_w is held at 11, the retrieved images using different $N_{w/o}$ are shown in Fig. S11. It is noted that image recovery cannot be accomplished only in $N_{w/o}$ =1 . To quantitatively show this, the retrieved result with $N_{w/o}$ =11 is assigned as a ground truth image (GT), and the SSIM values between the reconstructed result and GT are plotted in Fig. S11(e). The distribution of Fig. S11(e) accords with the visual judgment. We can conclude from Fig. S11(e) that the fewest number of intensity images is 2 for mask calibration process.

Then, we set $N_{w/o}$ as 11 and vary N_w from 1 to 11. The corresponding reconstructed results are shown in Fig. S12. Similarly, the result of N_w =11 is set as a GT image, and the SSIM

values related to different N_w are plotted in Fig. S12(e), where the SSIM value rises with the increase of the number of intensity images. It can be observed that the resolved resolution under N_w =1 is lower than under other conditions. Therefore, the fewest number of intensity images is 2 for the process of image recovery, in which the imaging resolution can be protected from damage.

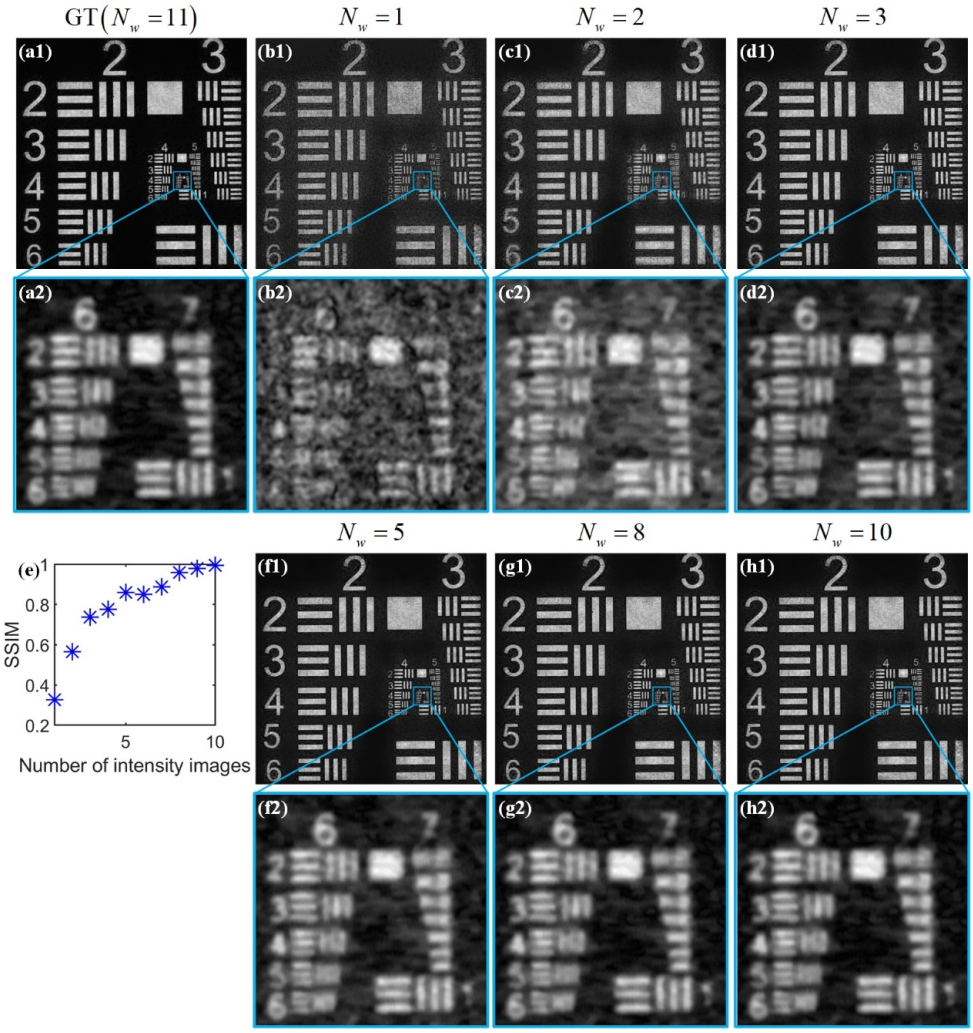

Fig. S12. Reconstruction results with the change of N_w when $N_{w/o}$ is held at 11

8.2 the regularized parameter

We analyze how the regularized parameter ρ influences the performance of our method. The reconstruction results with different regularized parameters are shown in Fig. S13. It is noted that when ρ changes from 0.004 to 0.4, the reconstructed quality varies slightly, which proves that our method is not sensitive to the regularized parameter selection. In all experiments, we set the regularized parameter as 0.1.

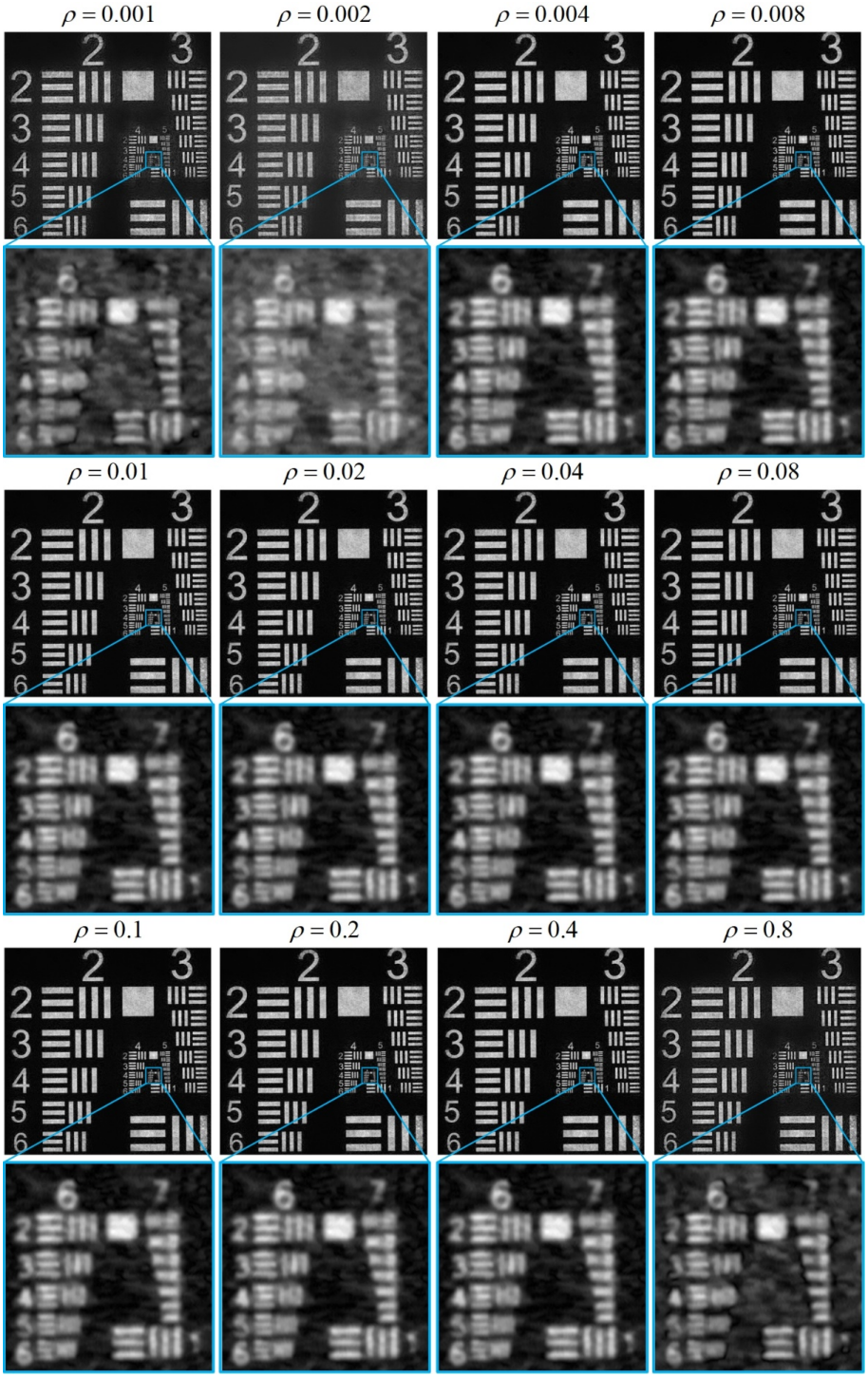

Fig. S13. Reconstruction results with different regularized parameters

8.3 the binarized threshold of the mask

We analyze how the binarized threshold of the mask influences the performance of our method. We set the threshold from 0.3 to 0.65, and provide the reconstruction results in Fig. S14. We can find that our method can achieve the highest imaging contrast when the threshold is 0.4, 0.45, or 0.5. Therefore, we set the threshold as 0.45 for all experiments.

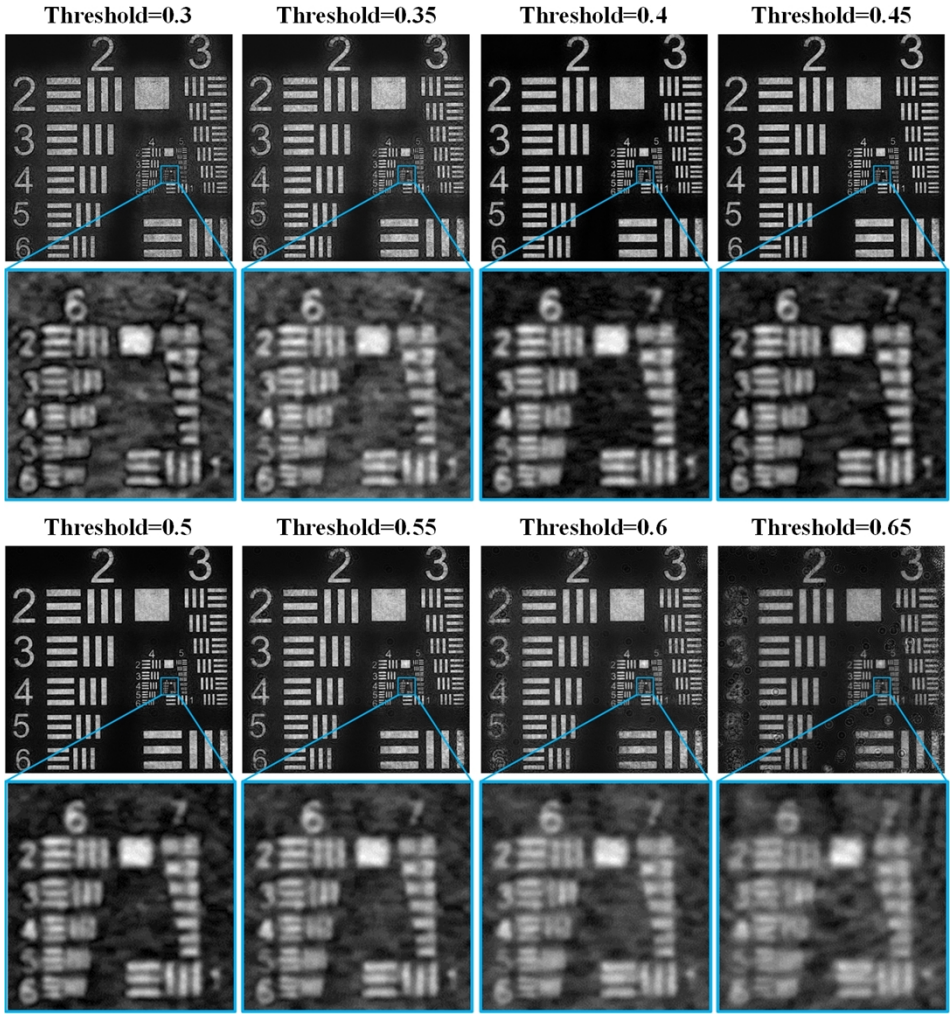

Fig. S14. Reconstruction results with different thresholds for binarization of the mask

9. Parameter tuning of SPICA and Multi-SPICA

As described Eq. (S26) and Eq. (S27), SPICA and Multi-SPICA algorithms have a regularized parameter ρ at their optimization functions. To get an optimal regularized parameter, we set regularized parameter ρ as 0.005, 0.01, 0.02, 0.04, 0.08, 0.1, 0.2, 0.4 for image recovery. With the above parameters, the retrieved results of SPICA and Multi-SPICA are given in Fig. S15 and Fig. S16. As shown in Fig. S15 and Fig. S16, we can find that $> \rho 0.04$ removes the detailed information and outputs an oversmoothed image. Therefore, we select the regularized parameter as 0.01 for SPICA and Multi-SPICA algorithms.

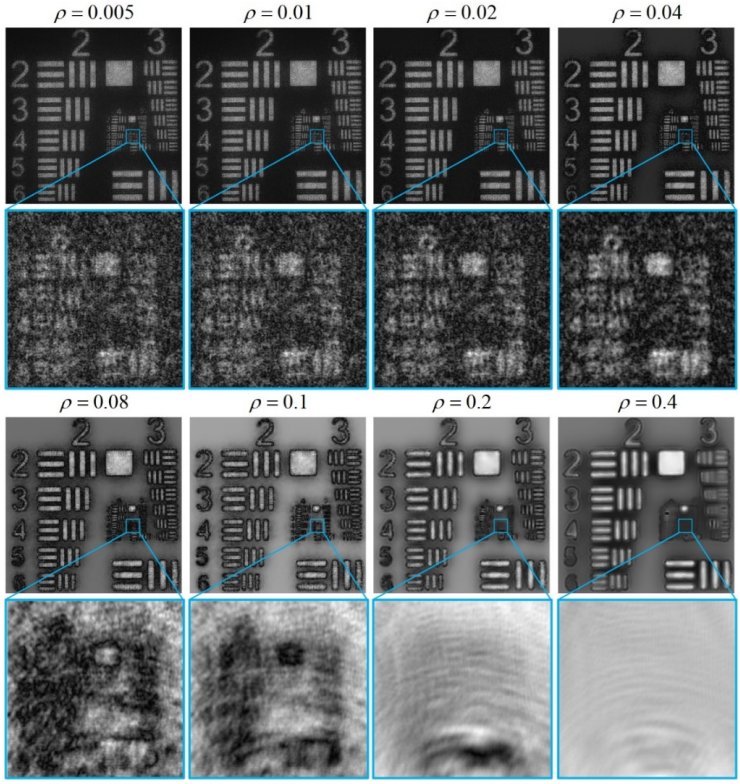

Fig. S15. Reconstruction results of SPICA with different regularized parameters

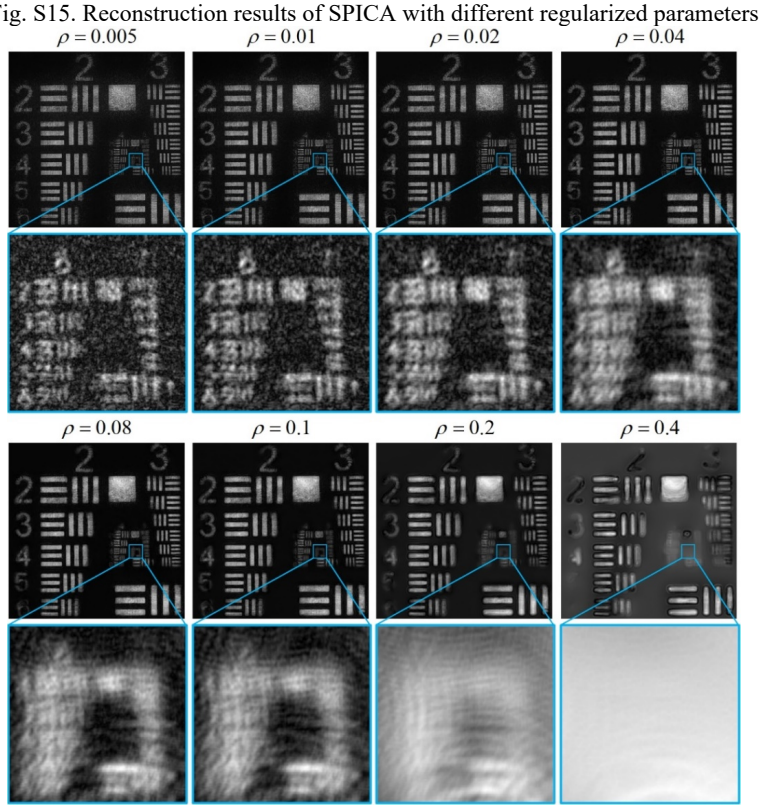

Fig. S16. Reconstruction results of Multi-SPICA with different regularized parameters

10. Reconstructed results with CNN-based denoisers

We analyze the performance of our method when combined with CNN-based denoisers. The fast processing speed of CNN-based denoisers is attributed to the GPU acceleration. However, for a large-scaled image task, the running time of the CNN-based denoisers is limited by GPU's memory capacity. For example, the GPU's memory capacity of our laptop is 8GB (NVIDIA GeForce RTX3070). If we run a TNRD filter (trained by diffusion model) [10] for image denoising, the acceptable maximum size of input image is approximately specified as 400×400 pixels. In our system, the size of intensity pattern is 3400×3400 pixels, which obviously exceeds memory limit of GPU. If this large-scaled wavefield data is input to the CNN-based denoiser, we must crop the data into a series of overlapped sub-images and then serially feed them into the CNN-based denoiser. After wavefield denoising, we have to stitch the denoised sub-images to form a full-size output image. Unfortunately, this serial processing remarkably increases the running time of wavefield denoising. To prove this drawback, we select two typical CNN-based denoisers, TNRD [10] and DnCNN [11] into our method for image recovery. When guide filter, DnCNN and TNRD are embedded into our method and the total iterative number is set as 10, the corresponding retrieved images are displayed in Fig. S17. The running time related to guide filter, DnCNN and TNRD is 189.5s, 515.5s and 6263.3s. It can be observed from Fig. S17 that the results of DnCNN and TNRD are undermined with remarkable resolution loss. This bad performance of the two learning-base denoiser is attributed to that the CNN-based denoisers are trained in Gaussian or Poisson noise scenes, but our application scene is more complicated. Thus, in our system, a feasible CNN-denoiser should be trained by enforcing a large amount experimental dataset. Considering the computational efficiency and robustness, we finally choose the guide filter as a denoiser for wavefield denoising.

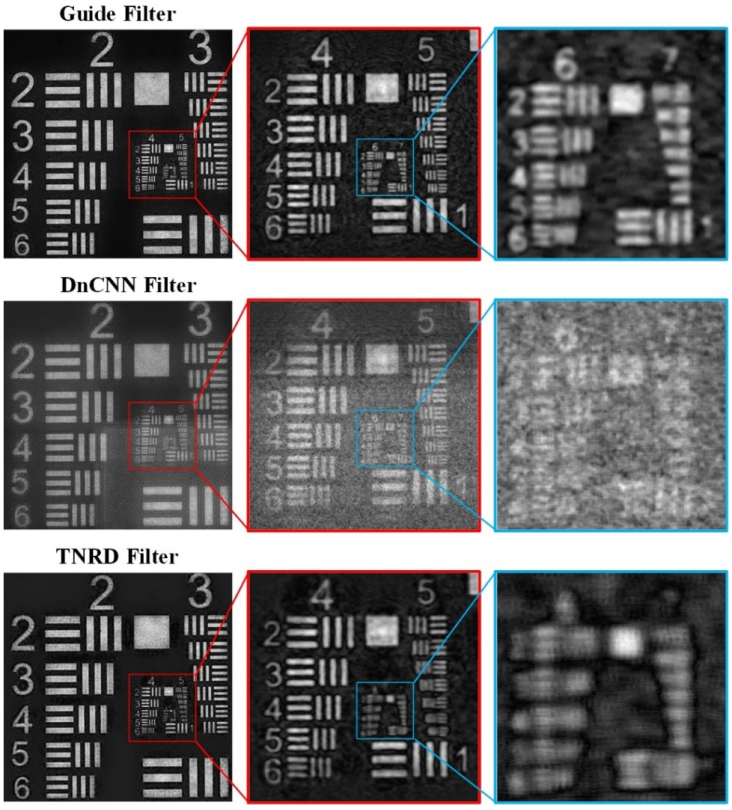

Fig. S17. Reconstruction results of our method with different denoisers

11. Resolution limitation analysis

As described in [12], the lateral resolution R_L of a lensless imaging system can be calculated as follows:

$$R_L = \frac{0.5\lambda}{NA} = \frac{0.5\lambda}{n_r \sin\left[\arctan\left(\frac{L_0}{2Z_0}\right)\right]} \tag{S29}$$

where NA is the equivalent numerical aperture of a leneless system, λ is the wavelength of incident light, n_r is the refractive index of the medium between the sample and the imaging sensor, L_0 is the width of observation region for the sensor, Z_0 is the diffraction distance between the sample and the imaging sensor. In our system, the above parameters are listed as follows: L_0 is 4.556mm (3400×1.34μm), λ is 532nm, n_r is 1, the sample-to-sensor distance is approximately 10mm. Based on Eq. (S29), the theoretical lateral resolution of our system is 1.20μm. Considering the pixelation of imaging sensor, the theoretical lateral resolution of our system is 1.34μm.

In practice, the smallest resolved part of our method is element 6 of Group 6 for USAF target, in which the real lateral resolution is 4.38μm. The resolution loss is caused by wavefield denoising process. Because the guide filter is a low-pass denoiser, which easily removes high-frequency information. Unfortunately, for the decoupling of mask function and transmitted wavefield, high-frequency information is hidden in noisy speckle background and the guide filter could remove the two components. Thus, the imaging resolution has to be impaired by the wavefield denoising.

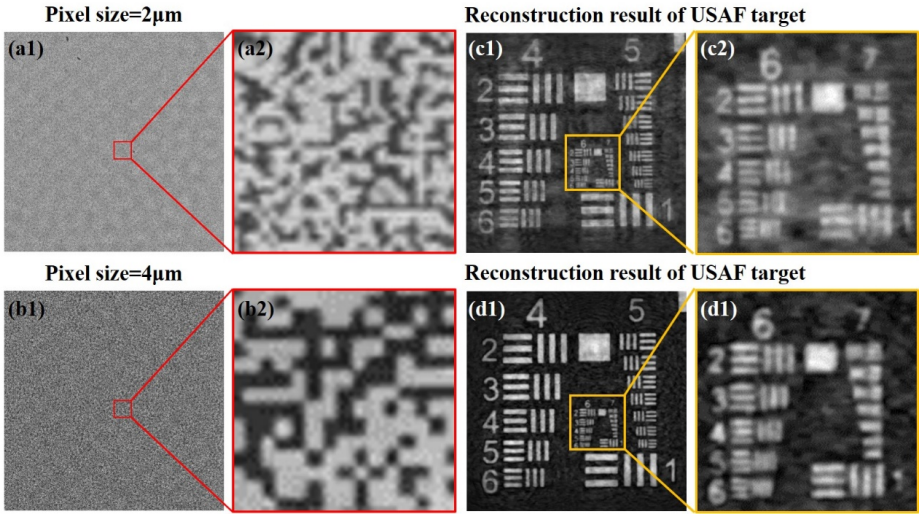

Fig. S18. Reconstructed results for different masks (pixel size: 2μm and 4μm)

12. Mask parameter selection

As discussed in Fig. S1, we conclude that the optimal value of the occlusion ratio is ~0.3. Also, the reconstructed accuracy will be enhanced if the pixel size of mask is smaller. Based on this rule, we fabricate two masks with pixel sizes of 2μm and 4μm to further verify the imaging performance. The mask functions with the pixel size of 2μm and 4μm are reconstructed and presented in Fig. S18(a) and Fig. S18(b). As the pixel size of the used sensor is 1.34μm, the pixel size of 4μm is nearly three multiples of the sensor's pixel, while the pixel size of 2μm cannot reach the multiple ratios. With the above mask functions, the amplitude images of resolution target are retrieved in Fig. S18(c) and Fig. S18(d). We can see that the recovered

image using the mask with pixel sizes of 4μm has a higher imaging contrast than that using the mask with pixel sizes of 2μm. This imaging performance comparison indicates that the pixel ratio between mask and sensor should be an integer for better reconstruction quality. Therefore, we choose the mask with pixel size of 4μm for other experiments.

13. Extra experimental results

Additionally, we rebuild the experimental configuration to further show the robustness of the auto-focusing algorithm on lensless masked imaging. Here three samples, including H&E-stained human tongue fungiform papillae, H&E-stained cow lung tissue, and pure phase target, are employed for test. The pure phase target is produced by a spatial light modulator (Holoeye GAEA-2-VIS-036, 4160×2464, pixel pitch: 3.74μm). Since the spatial light modulator (SLM) is a reflective configuration, we load a natural image to the SLM and utilize a beam splitter to record the reflective diffraction patterns. Using our auto-focusing algorithm, the mask-to-sensor distances are acquired and plotted in Fig. S19(a), where the initial distance and interval are specified as 3.19mm and 0.10mm. For the three samples, their auto-focusing curves are plotted in Figs. S19 (b-d), where the sample-to-mask distances of human tongue fungiform papillae, cow lung tissue, and pure phase target are specified as 7.77mm, 8.73mm, and 58.29mm. The biggest obstacle of auto-focusing is multi-peak crosstalk that originated from the twin-image artifact. As shown in Figs. S19 (b-d), we can find that our auto-focusing curves are unimodal over long distance range, which proves the robustness of our method. This good unimodality is because that our method utilizes multi-distance dataset to effectively remove the twin-image artifact and thus the crosstalk disturbance from twin-image can be mitigated.

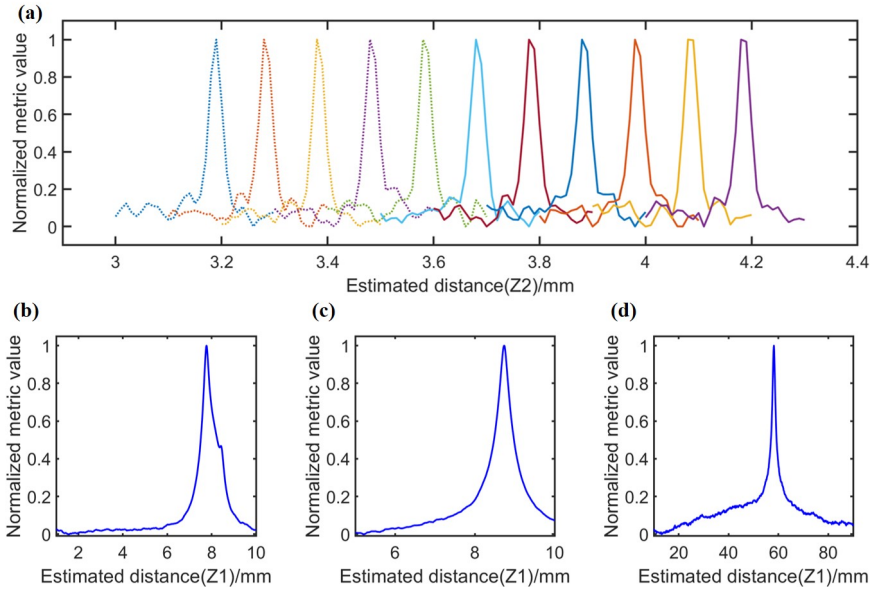

Fig. S19. The auto-focusing curves for additional experiments. (a) mask-to-sensor distance estimation. (b-d) are sample-to-mask distance estimation curves for human tongue fungiform papillae, cow lung tissue, and pure phase target.

With above parameters, the retrieved images of three samples are given in Fig. S20, Fig. S21, and Fig. S22. Here the reconstructed results using 2 and 11 intensity patterns are both displayed. Fig. S20 shows the results of human tongue fungiform papillae reconstructed by our method and Multi-SPICA. The regions labelled by blue and red boxes are cropped and zoomed for further comparison. It is noted that when recovering dense biological tissue, our method shows a better imaging quality, while the results of Multi-SPICA are contaminated by noisy backgrounds. As a sparse sample, the retrieved results of lung tissue are shown in Fig. S21, where our method can reconstruct the lung tissue under dual-plane and multi-plane where our method can reconstruct the lung tissue under dual-plane and

multi-plane measurement, but Multi-SPICA only works well with multi-plane measurement. Fig. S22 shows the reconstructed result of phase-only target that loaded in the SLM. It is observed that both methods can reconstruct the pure phase image but our method shows a better imaging contrast.

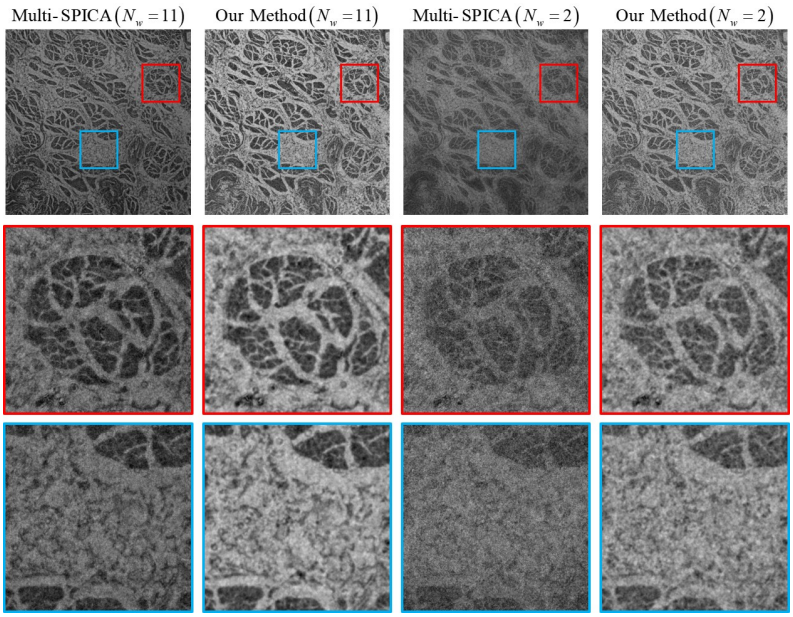

Fig. S20. Reconstructed results of human tongue fungiform papillae for dual-plane and multi-plane measurement.

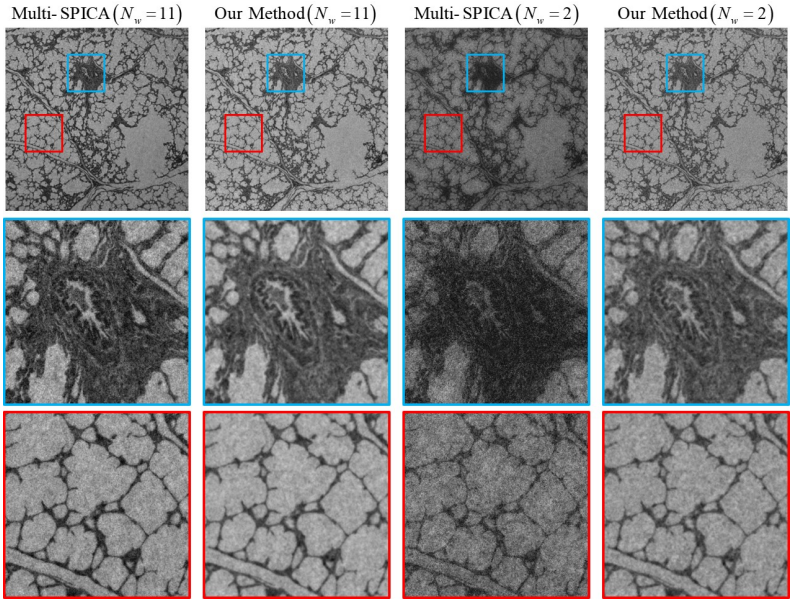

Fig. S21. Reconstructed results of cow lung tissue for dual-plane and multi-plane measurement.

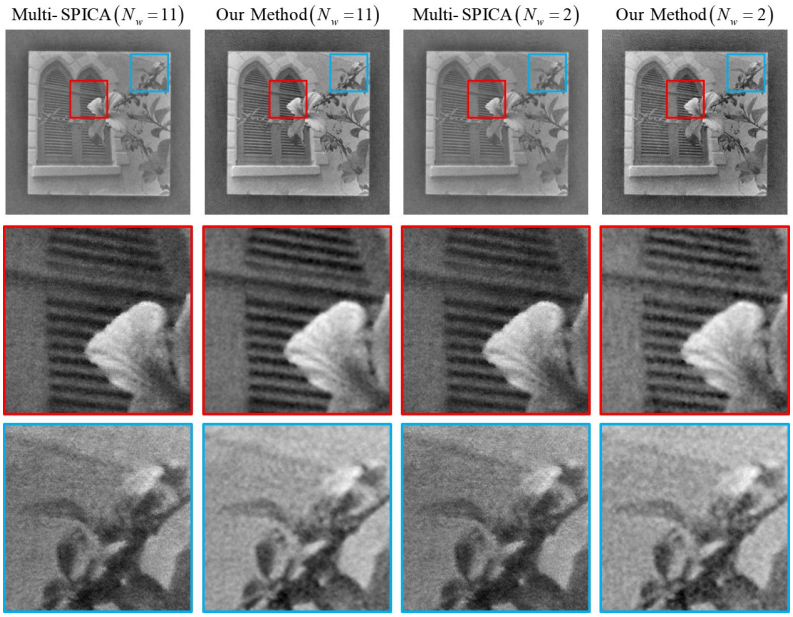

Fig. S22. Reconstructed results of pure phase target for dual-plane and multi-plane measurement.

14. The reason why the mismatch error exists

In this section, the reason why the mismatch error exists is given. In a real system, incident light usually illuminates sample with a tilt angle. As sample is fixed or fabricated on a glass slide (or substrate), the thickness of glass (or substrate) could introduce an extra optical path length. Under tilt illumination, the incident light will be refracted at the interface between air and glass, which gives rise to a spatial position shift for transmitted light. The corresponding phenomenon has been illustrated in Fig. S23. As shown in Fig. S23(a), a sample is not loaded and the incident light illuminates the mask with a tilt angle. In Fig. S23(b), a sample is loaded and the refraction effect between air and glass causes a lateral shift Δx. The above position shift leads to a mismatch error for two groups of intensity patterns under sample loading and sample removing. According to the above analysis, the two groups of intensity datasets should be aligned before being fed into the binary-coding phase retrieval algorithm. As for the binary amplitude mask, once the mask is installed, the relative position between the mask and the imaging sensor is fixed and will not change even if different samples are loaded. Therefore, in our system, the mask only needs to be calibrated for once, where the mask-to-sensor distances and mask function are acquired. When different samples are loaded, we just need to align the mask to the new sample instead of calibrating the mask again.

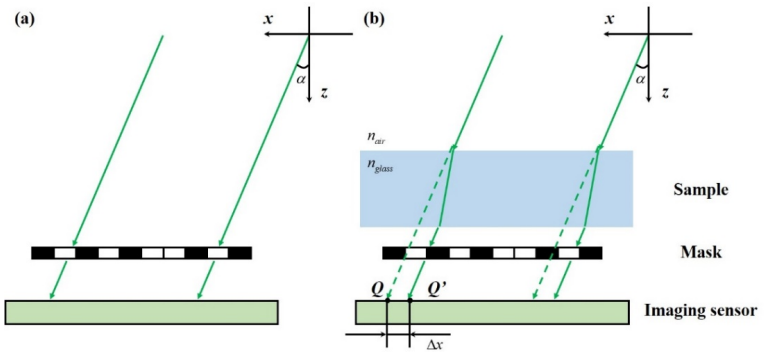

Fig. S23. Diagram of lensless masked imaging under tilt illumination. (a) Sample is not loaded. (b) Sample is loaded.

References

1. G. Pedrini, W. Osten, and Y. Zhang, “Wave-front reconstruction from a sequence of interferograms recorded at different planes,” Opt. Lett. 30(8), 833-835 (2005).
2. C. Guo, C. Shen, Q. Li, J. Tan, S. Liu, X. Kan, and Z. Liu, “A fast-converging iterative method based on weighted feedback for multi-distance phase retrieval,” Sci. Rep. 8(1), 6436 (2018).
3. C. Guo, X. Liu, X. Kan, F. Zhang, J. Tan, S. Liu, and Z. Liu, “Lensfree on-chip microscopy based on dual-plane phase retrieval,” Opt. Express 27(24), 35216–35229 (2019).
4. K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3-d transform domain collaborative filtering,” IEEE Trans. Image Process. **16**, 2080-2095 (2007).
5. K. He, J. Sun, and X. Tang, “Guided image filtering,” IEEE Trans. Pattern Anal. **35**, 1397-1409 (2012).
6. Y. Chi and S. H. Chan, “Fast and Robust Recursive Filter for Image Denoising,” in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (IEEE, 2018), pp. 1708-1712.
7. S. H. Chan, R. Khoshabeh, K. B. Gibson, P. E. Gill, and T. Q. Nguyen, “An augmented lagrangian method for total variation video restoration,” IEEE Trans. Image Process. **20**, 3097-3111 (2011).
8. C. Guo, Feilong Zhang, Xianming Liu, Qiang Li, Shenghao Zheng, Jiubin Tan, Zhengjun Liu, Weibo Wang, “Lensfree auto-focusing imaging using nuclear norm of gradient,” Opt. Lasers Eng. 156, 107076 (2022).
9. J. M. Bioucas-Dias, and M. A. T. Figueiredo, “A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration,” IEEE Trans. Image Process. 16(12), 2992-3004 (2007).
10. Y Chen, T Pock, “Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration,” IEEE Transactions on pattern analysis and machine intelligence 39(6), 1256-1272 (2016).
11. K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, “Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising,” IEEE Trans. Image Process. 26(7), 3142-3155 (2017).
12. A. Ozcan, E. McLeod, “Lensless Imaging and Sensing,” Annu. Rev. Biomed. Eng. 18,77–102 Annu. Rev. Biomed. Eng. 18, 77-102 (2016).
13. Y. Zhang, H. Wang, Y. Wu, M. Tamamitsu, and A. Ozcan, “Edge sparsity criterion for robust holographic autofocusing,” Opt. Lett. 42(19), 3824–3827 (2017).