Single-shot lensless masked imaging with enhanced self-calibrated phase retrieval: supplemental document

The supplemental document is organized as follows:

(1) The detailed derivation of sparse regularized phase retrieval (SrPR) is added;
(2) The detailed derivation of denoising regularized phase retrieval (DrPR) is added;
(3) The ablation study and parameter tuning of SrPR algorithm are given;
(4) The ablation study and parameter tuning of DrPR algorithm are given;
(5) The derivation and sharpness metric selection of mask-guided auto-focusing are added;
(6) The discussion of joint recovery using the mask of 2.5μm pixel size is given;
(7) The performance comparison of SPICA, SCPR, and our method on different mask's pixel sizes is given;
(8) The extra reconstructed results of Rongchi grating, stained tissue, phase mask, and SLM target are given;
(9) The reconstructed results under low signal-to-noise ratios are given.

1. Derivation of sparse regularized phase retrieval (SrPR)

We provide a detailed derivation of sparse regularized phase retrieval (SrPR). In this paper, the diffraction propagation calculation is defined by angular spectrum model. Thus, the forward A_Z and backward A_Z^{-1} propagation operators with a diffractive distance of Z are expressed as

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

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

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. $\odot$ denotes an operation of Hadamard product. With a given intensity image (I_M), we formulate SrPR algorithm to iteratively solve a minimization problem as follows

$$M^{k_1+1} = \arg\min \left\| \left| A_{Z_2} M^{k_1} \right|^2 - I_M \right\|_2^2 + D_\sigma\left(M^{k_1}\right) + \delta \left\| M^{k_1} \right\|_1 + t\nabla \left\| M^{k_1} \right\|, \; k_1 \in [1, \mathrm{K}_1], \tag{S3}$$

to reconstruct the mask distribution function. k_1 is the index of the iterative number and K_1 is the total iterations. $\|\cdot\|_1$ denotes L1 norm of a matrix. $\nabla\|\cdot\|$ denotes a total variation (TV) norm and ∇ is a gradient operator. D_σ represents a denoiser with a noise variance of σ. With the use of PnP-FISTA solver, Eq. (S3) can be decomposed into as follows:

$$S^{k_1} = \arg\min_{M^{k_1}} \left\| \left| A_{Z_2} M^{k_1} \right|^2 - I_M \right\|_2^2 + \delta \left\| M^{k_1} \right\|_1 + t\nabla \left\| M^{k_1} \right\|, \tag{S4}$$

$$\tilde{M}^{k_1+1} = \arg\min_{\tilde{M}^{k_1}} \left\{ \left\| S^{k_1} - \tilde{M}^{k_1} \right\|_2^2 + D_\sigma\left(\tilde{M}^{k_1}\right) \right\}, \tag{S5}$$

$$Q_{k_1+1}=\frac{1}{2}\left(1+\sqrt{1+4Q_{k_1}^2}\right), \tag{S6}$$

$$M^{k_1+1}=\tilde{M}^{k_1}+\frac{Q_{k_1}-1}{Q_{k_1+1}}\left(\tilde{M}^{k_1+1}-\tilde{M}^{k_1}\right). \tag{S7}$$

In Eq. (S4), it is difficult to acquire a closed-form derivative of L1 and TV terms. Thus, we convert Eq. (S4) into two alternating sub-problems as follows

$$S^{k_1}=\begin{cases}\arg\min \mathcal{L}_1=\arg\min\limits_{M^{k_1}}\left\{\left\|\left|A_{Z_2}M^{k_1}\right|^2-I_M\right\|_2^2+\delta\left\|M^{k_1}\right\|_1\right\}, \mathrm{mod}\left(k_1,5\right)=0,\\ \arg\min \mathcal{L}_2=\arg\min\limits_{M^{k_1}}\left\{\left\|\left|A_{Z_2}M^{k_1}\right|^2-I_M\right\|_2^2+t\nabla\left\|M^{k_1}\right\|\right\}, \textit{others},\end{cases} \tag{S8}$$

where $\mathrm{mod}\left(k_1,5\right)$ denotes a remainder after dividing by 5. According to alternative projection and sparse regularization [1], the minimization of $\mathcal{L}_1$ is derived as

$$\begin{cases}g^{k_1}=A_{Z_2}^{-1}\left[\sqrt{I_M}\dfrac{A_{Z_2}\left(M^{k_1}\right)}{\left|A_{Z_2}\left(M^{k_1}\right)\right|}\right],\\ S^{k_1}=\max\left\{\left|g^{k_1}\right|-\delta,0\right\}\odot\dfrac{g^{k_1}}{\left|g^{k_1}\right|}.\end{cases} \tag{S9}$$

With the use of alternative projection and gradient descent [2], the minimization of $\mathcal{L}_2$ is derived as

$$\begin{cases}g^{k_1}=A_{Z_2}^{-1}\left[\sqrt{I_M}\dfrac{A_{Z_2}\left(M^{k_1}\right)}{\left|A_{Z_2}\left(M^{k_1}\right)\right|}\right],\\ S^{k_1}=\dfrac{1}{2}\left[g^{k_1}+\dfrac{t}{2}\nabla\left(\dfrac{\nabla g^{k_1}}{\left|\nabla g^{k_1}\right|}\right)\right].\end{cases} \tag{S10}$$

The sub-problem of Eq. (S5) can be viewed as a denoising process and its closed-form solution is expressed as follows

$$\tilde{M}^{k_1+1}=D_\sigma\left[\mathrm{Re}\left(S^{k_1}\right)\right]+\mathrm{j}\cdot D_\sigma\left[\mathrm{Im}\left(S^{k_1}\right)\right], \tag{S11}$$

where 'Re' and 'Im' denote the real and imaginary parts of S^{k_1}. A learning-based denoiser, TNRD filter, is selected for this denoising task. Here we do not train TNRD filter with a new dataset but adopt the defaulted version, in which only 400 natural images are used for network training. Therefore, the entire iterative process of SrPR algorithm is generalized into Eqs. (S12-S15) as follows

$$g^{k_1}=A_{Z_2}^{-1}\left[\sqrt{I_M}\frac{A_{Z_2}\left(M^{k_1}\right)}{\left|A_{Z_2}\left(M^{k_1}\right)\right|}\right], \tag{S12}$$

$$S^{k_1}=\begin{cases}\max\left\{\left|g^{k_1}\right|-\delta,0\right\}\odot\dfrac{g^{k_1}}{\left|g^{k_1}\right|}, & \mathrm{mod}\left(k_1,5\right)=0\\ \dfrac{1}{2}\left[g^{k_1}+\dfrac{t}{2}\nabla\left(\dfrac{\nabla g^{k_1}}{\left|\nabla g^{k_1}\right|}\right)\right], & \textit{others}\end{cases} \tag{S13}$$

$$\tilde{M}^{k_1+1} = D_\sigma\left[\mathrm{Re}\left(S^{k_1}\right)\right] + \mathrm{j}\cdot D_\sigma\left[\mathrm{Im}\left(S^{k_1}\right)\right], \tag{S14}$$

$$M^{k_1+1} = \tilde{M}^{k_1} + \frac{2\left(Q_{k_1}-1\right)\left(\tilde{M}^{k_1+1}-\tilde{M}^{k_1}\right)}{1+\sqrt{1+4Q_{k_1}^2}}. \tag{S15}$$

2. Derivation of denoising regularized phase retrieval (DrPR)

In this section, we provide a detailed derivation of denoising regularized phase retrieval (DrPR). With the given mask function (M) and intensity image (I_O), the iterative optimization process of DrPR algorithm is constructed as

$$O^{k_3+1} = \arg\min \left\| \left| A_{Z_2}\left[M \odot \left(A_{Z_1} O^{k_3} \right) \right] \right|^2 - I_O \right\|_2^2 + D_\sigma\left(O^{k_3}\right). \tag{S16}$$

k_3 is the index of iterative number for DrPR. With the use of PnP-FISTA solver, Eq. (S16) can be decomposed into as follows:

$$S^{k_3} = \arg\min_{O^{k_3}} \left\| \left| A_{Z_2}\left[M \odot \left(A_{Z_1} O^{k_3} \right) \right] \right|^2 - I_O \right\|_2^2, \tag{S17}$$

$$\tilde{O}^{k_3+1} = \arg\min_{\tilde{O}^{k_3}} \left\{ \left\| S^{k_3} - \tilde{O}^{k_3} \right\|_2^2 + D_\sigma\left(\tilde{O}^{k_3}\right) \right\}, \tag{S18}$$

$$Q_{k_3+1} = \frac{1}{2}\left(1+\sqrt{1+4Q_{k_3}^2}\right), \tag{S19}$$

$$O^{k_3+1} = \tilde{O}^{k_3} + \frac{Q_{k_3}-1}{Q_{k_3+1}}\left(\tilde{O}^{k_3+1}-\tilde{O}^{k_3}\right). \tag{S20}$$

The sub-problem of Eq. (S17) can be divided into alternative projection and wavefield decoupling, which is sequentially unfolded as

$$T^{k_3} = A_{Z_2}^{-1}\left\{ \frac{A_{Z_2}\left[M \odot \left(A_{Z_1} O^{k_3} \right) \right]}{\left| A_{Z_2}\left[M \odot \left(A_{Z_1} O^{k_3} \right) \right] \right|} \sqrt{I_O} \right\}, \tag{S21}$$

$$B^{k_3+1} = \arg\min_{B^{k_3}} \left\| B^{k_3} \odot M - T^{k_3} \right\|_2^2, s.t.\ B^{k_3} = A_{Z_1} O^{k_3}, \tag{S22}$$

$$S^{k_3} = A_{Z_1}^{-1} B^{k_3+1}. \tag{S23}$$

where T^{k_3} and B^{k_3} are the transmitted and incident wavefield of the mask plane, respectively. The solution of Eq. (S22) is to decouple the incident wavefield from the transmitted wavefield. In our previous work [3], we regard this wavefield decoupling operation as a complex-valued image inpainting, and then construct an ADMM solver to accomplish this task. However, this complex-valued image inpainting strategy imposes an inner loop in the phase retrieval framework, which increases the running time. In addition, this inpainting-based phase retrieval method assumes that the mask is an amplitude-based matrix (only includes '1'and '0' pixels). But the plane wave illumination also imposes a weak phase variation for the mask plane, and thus the mask function should be assumed as a complex-valued matrix for wavefield decoupling. To address this problem, we follow the separation modality of probe and wavefront with ptychographic iterative engine [4], and utilize a gradient descent method to decouple the incident wavefield as follows

$$B^{k_3+1} = B^{k_3} + \frac{M^*\left[T^{k_3} - M \odot B^{k_3} \right]}{\left| M \right|_{\max}^2}, s.t.\ B^{k_3} = A_{Z_1} O^{k_3}. \tag{S24}$$

Similar to SrPR algorithm, the sub-problem of Eq. (S18) can be solved by complex-valued denoising, thus Eq. (S18) is expressed by TNRD filter as

$$\tilde{O}^{k_3+1} = D_\sigma\left[\mathrm{Re}\left(S^{k_3}\right)\right] + \mathrm{j}\cdot D_\sigma\left[\mathrm{Im}\left(S^{k_3}\right)\right]. \tag{S25}$$

Based on the above derivation, the entire iterative process of DrPR algorithm can be generalized as follows

$$\begin{cases} T^{k_3} = A_{Z_2}^{-1}\left\{\dfrac{A_{Z_2}\left[M\odot\left(A_{Z_1}O^{k_3}\right)\right]}{\left|A_{Z_2}\left[M\odot\left(A_{Z_1}O^{k_3}\right)\right]\right|}\sqrt{I_O}\right\}, \\ S^{k_3} = A_{Z_1}^{-1}\left\{A_{Z_1}O^{k_3} + \dfrac{M^*\odot\left[T^{k_3} - M\odot\left(A_{Z_1}O^{k_3}\right)\right]}{\left|M\right|_{\max}^2}\right\}, \\ \tilde{O}^{k_3+1} = D_\sigma\left[\mathrm{Re}\left(S^{k_3}\right)\right] + \mathrm{j}\cdot D_\sigma\left[\mathrm{Im}\left(S^{k_3}\right)\right], \\ O^{k_3+1} = \tilde{O}^{k_3} + \dfrac{2\left(Q_{k_3}-1\right)\left(\tilde{O}^{k_3+1} - \tilde{O}^{k_3}\right)}{1+\sqrt{1+4Q_{k_3}^2}}. \end{cases} \tag{S26}$$

3. Derivation of mask-guided auto-focusing method and sharpness metric selection

In this section, the derivation process of mask-guided auto-focusing method is given. The mask-guided auto-focusing method decouples the incident wavefield of the mask plane and then performs an auto-focusing sharpness metric to obtain Z_1. The corresponding optimization can be expressed as

$$C^{k_2+1} = \arg\min\left\|\left|A_{Z_2}\left(C^{k_2}\odot M\right)\right|^2 - I_O\right\|_2^2 + D_\sigma\left(C^{k_2}\right), \tag{S27}$$

$$Z_1 \leftarrow \underset{e_1\le e_n\le e_N}{\arg\max}\left\{\mathrm{G}\left(\nabla\left|A_{e_n}^{-1}C^{\mathrm{K}_2}\right|\right)\right\}, n\in[1,N], \tag{S28}$$

where $\{e_1, e_2 \ldots e_N\}$ are the refocused distances for inverse diffraction propagation. C^{k_2} is the incident wavefield at the mask plane. The function 'G' denotes an auto-focusing sharpness quantitative metric. Similar to DrPR algorithm, the optimization problem of Eq. (S27) is iteratively solved as follows

$$\begin{cases} T^{k_2} = A_{Z_2}^{-1}\left\{\dfrac{A_{Z_2}\left(M\odot C^{k_2}\right)}{\left|A_{Z_2}\left(M\odot C^{k_2}\right)\right|}\sqrt{I_O}\right\}, \\ S^{k_2} = C^{k_2} + \dfrac{M^*\odot\left[T^{k_2} - M\odot C^{k_2}\right]}{\left|M\right|_{\max}^2}, \\ \tilde{C}^{k_2+1} = D_\sigma\left[\mathrm{Re}\left(S^{k_2}\right)\right] + \mathrm{j}\cdot D_\sigma\left[\mathrm{Im}\left(S^{k_2}\right)\right], \\ C^{k_2+1} = \tilde{C}^{k_2} + \dfrac{2\left(Q_{k_2}-1\right)\left(\tilde{C}^{k_2+1} - \tilde{C}^{k_2}\right)}{1+\sqrt{1+4Q_{k_2}^2}}. \end{cases} \tag{S29}$$

Different from DrPR algorithm, TNRD filter is not imposed in Eq. (S29). Because the learning-based denoiser introduces a rich of convolution kernels to process the image, which requires more running time for calculation. In addition, the output of Eq. (S29) (C^{k_2}) includes high-frequency diffraction fringes, thus a strong denoiser easily wipes out detailed information. To keep a good balance, we utilize a guided filter with self-reference inputs for denoising. In the full version of guided filter, an input image and a reference image are required. If the input and

reference are used with one image, guided filter can serve as an edge-enhancing operator. With this self-reference setting, the high-angle diffraction fringe cannot be removed. Furthermore, the enhanced edge sparsity could increase the difference of Z-stacked refocused images in Eq. (S28).

In our method, we choose seven sharpness metrics [5-9], including GRA, LAP, SEPC, Tenengrad, SG, ToG, and NoG, to find the optimal metric for Z_1 estimation. A resolution target (Thorlabs) is selected as the sample, and the auto-focusing curves with the mask pixel from 2.5μm to 40μm are plotted in Figs. S1(a-e), where the peak values of the curves point to the sample-to-mask distance. It is noted that GRA, SPEC, and LAP metrics suffer from slope artifacts near 2mm for all mask pixels, and NoG, SG, and Tenengrad metrics are also affected by the slope artifact in Fig. S1(e). In addition, GRA, LAP, SEPC, Tenengrad, SG, and NoG metrics are undermined by multi-peak artifacts in Fig. S1(a). Only ToG metric keeps a good balance of robustness and unimodality. Therefore, we equip the mask-guided auto-focusing method with the ToG metric for Z_1 estimation.

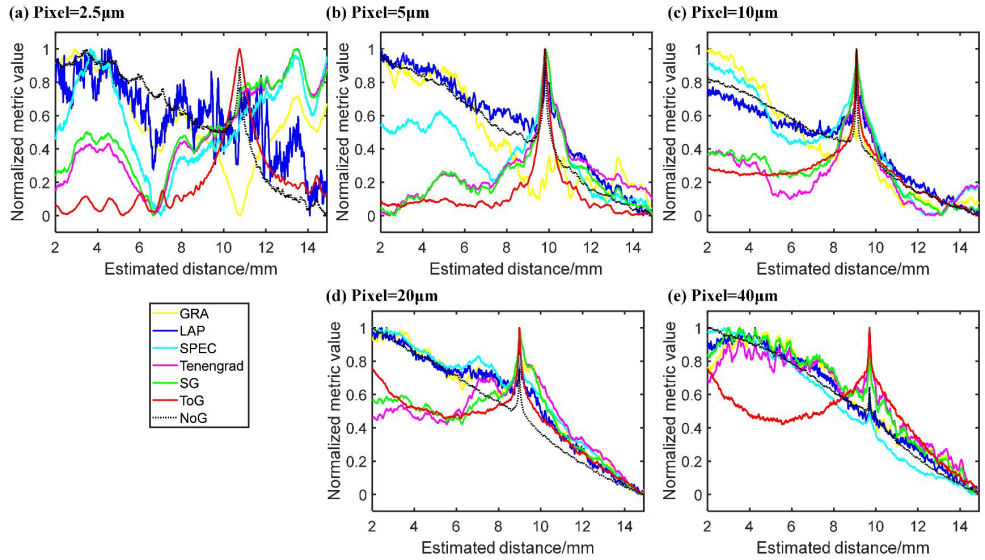

Fig. S1. Auto-focusing curves of different sharpness metrics for Z_1 estimation under the mask pixel of 2.5μm, 5μm, 10μm, 20μm, and 40μm.

4. Ablation study and parameter tuning of SrPR

SrPR algorithm consists of four modules, i.e., alternative projection, TV regularization, L1 regularization, and wavefield denoising. Here we provide an ablation study in Fig. S2 to show the effectiveness of each module, in which two masks (pixel: 5μm and 10μm) are chosen. As shown in Fig. S2(a), only performing alternative projection is incapable of reconstructing the mask function and its retrieved images are still covered with remarkable twin-image artifact. After TV regularization is executed, the edges of the retrieved images are enhanced in Fig. S2(b), but the twin-image artifact still exists. With the addition of L1 regularization, the artifact is mitigated and the target-to-background contrast is enhanced in Fig. S2(c) for the retrieved images. As shown in Fig. S2(c1) and Fig. S2(c2), the regions of the '1' pixels are unwrapped with the residue artifact. Accordingly, TNRD filter is loaded and the residue artifact is eliminated in Fig. S2(d). In summary, the TV term introduces an edge-preserving effect and the L1 term realizes a location of '0' pixels. The denoising term smooths the entire image and eliminates the residue background. The combination of these three penalty terms contributes to contrast-enhanced mask recovery.

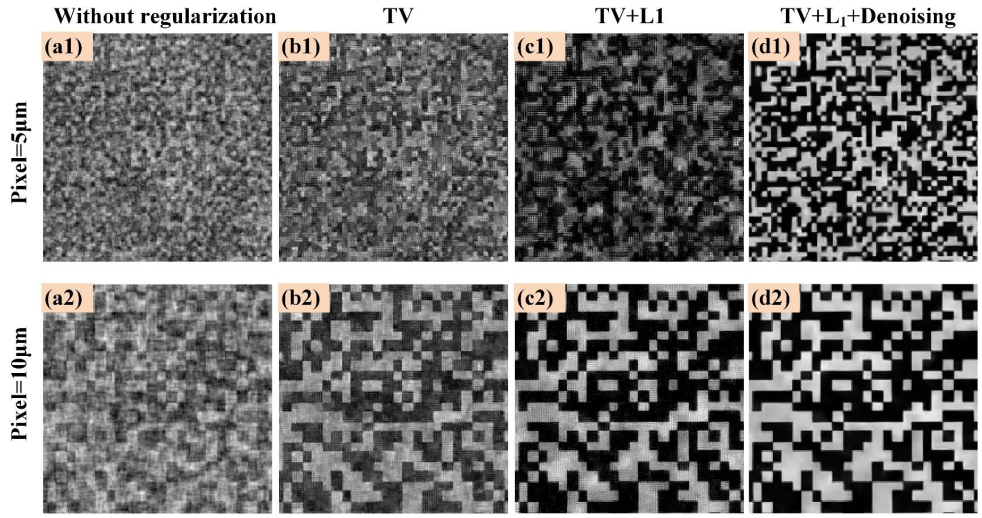

Fig. S2. Single-shot recovery performance of SrPR algorithm by using different penalty terms.

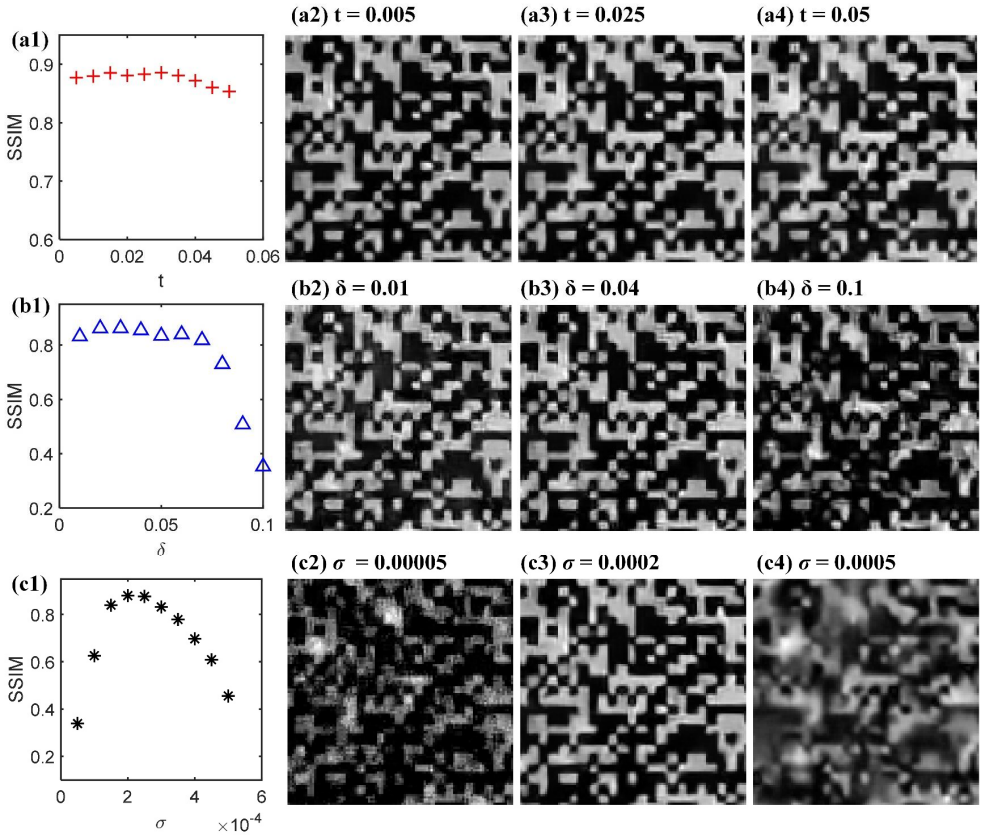

Fig. S3. Parameter tuning of SrPR algorithm with the variation of t, δ and σ.

In SrPR, two penalty parameters (t and δ) and noise variance (σ) are necessary to be tuned for optimal mask recovery. Fig. S3 is given to show the parameter tuning process of SrPR algorithm. The mask with a pixel of 5μm is used for test. 11 multi-height intensity images of

the mask are recorded and then a multi-distance phase retrieval is run to reconstruct the ground truth (GT) image. With the GT image, the SSIM values between GT and single-shot retrieved image are performed to quantitatively show the accuracy of mask recovery. As the parameters variate, the SSIM curves related to t, δ, and σ are plotted in Fig. S3(a1), Fig. S3(b1), and Fig. S3(c1). As an example, the retrieved images with different parameters are shown in Figs. S3(a2-a4), Figs. S3(b2-b4), and Figs. S3(c2-c4). The results of Fig. S3 show that the optimal values of t, δ, and σ are 0.03, 0.03, and 0.0002 for SrPR algorithm.

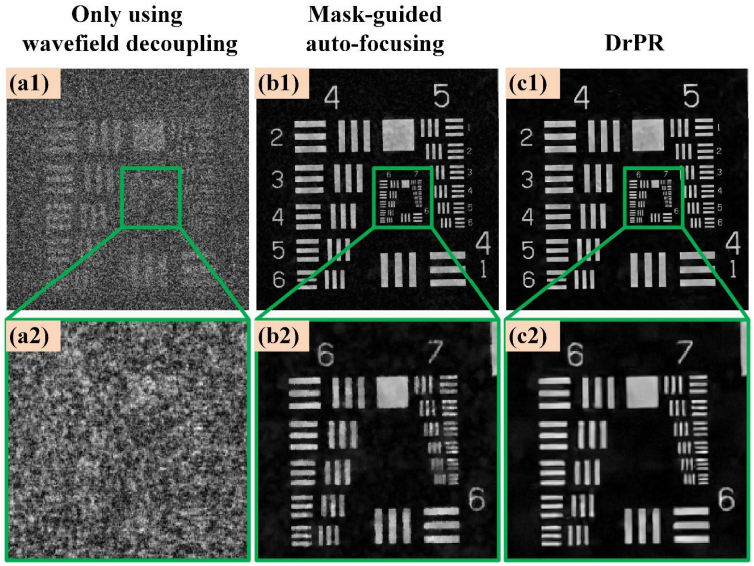

Fig. S4. Single-shot recovery performance of DrPR algorithm by using different settings.

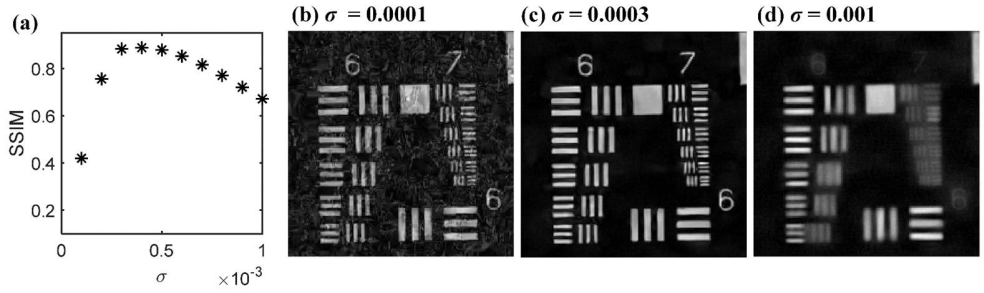

Fig. S5. Parameter tuning of DrPR algorithm with the variation of σ.

5. Ablation study and parameter tuning of DrPR

The key steps of DrPR algorithm are wavefield decoupling and denoising, thus the recovery performance with/without wavefield denoising needs to be discussed. In addition, the refocused result from mask-guided auto-focusing method can also output a clear sample image. Thus, the above three conditions are performed in Fig. S4 to find which one is most useful for sample recovery. Fig. S4(a) is the retrieved image by DrPR without the use of TNRD filter, in which only alternative projection and wavefield decoupling are used. It is noted that DrPR without denoising cannot reconstruct the sample and its retrieved image is undermined by noisy backgrounds. As the refocusing operation is used as $A_{Z_1}^{-1}C^{K_2}$, the output of mask-guided auto-focusing method is shown in Fig. S4(b). With the addition of guided filtering, the quality of the retrieved image is improved. As a comparison, the retrieved image of the full-version DrPR is

given in Fig. S4(c), which removes the residue artifact and achieves contrast-enhanced sample recovery. The bad performance of Fig. S4(b) is attributed to the use of guided filter. The used guided filter adopts a self-reference input and it serves as an edge-enhancing operator, thus the background noise could be accordingly enhanced. Obviously, guided filter can be directly replaced with TNRD filter for auto-focusing, but the corresponding running time will be tremendously increased. Also, TNRD filter possesses a strong denoising capability, which easily wipes out the detailed information for auto-focusing. Based on the above analysis, we select a self-reference guided filter for Z_1 estimation and TNRD filter for final sample recovery. As the noise variance of TNRD filter ranges from 0.0001 to 0.01, the retrieved results of DrPR are shown in Fig. S5. Here the retrieved image from multi-distance phase retrieval is also used as a GT image, and the SSIM curve with different noise variances is plotted in Fig. S5(a). The retrieved image with the variance of 0.0001, 0.0003, and 0.001 are displayed in Figs. S5(b-d). It is noted that Fig. S5(b) is covered with residue noise and Fig. S5(d) outputs an oversmoothed result. Therefore, the optimal variance σ is 0.0003 for DrPR algorithm.

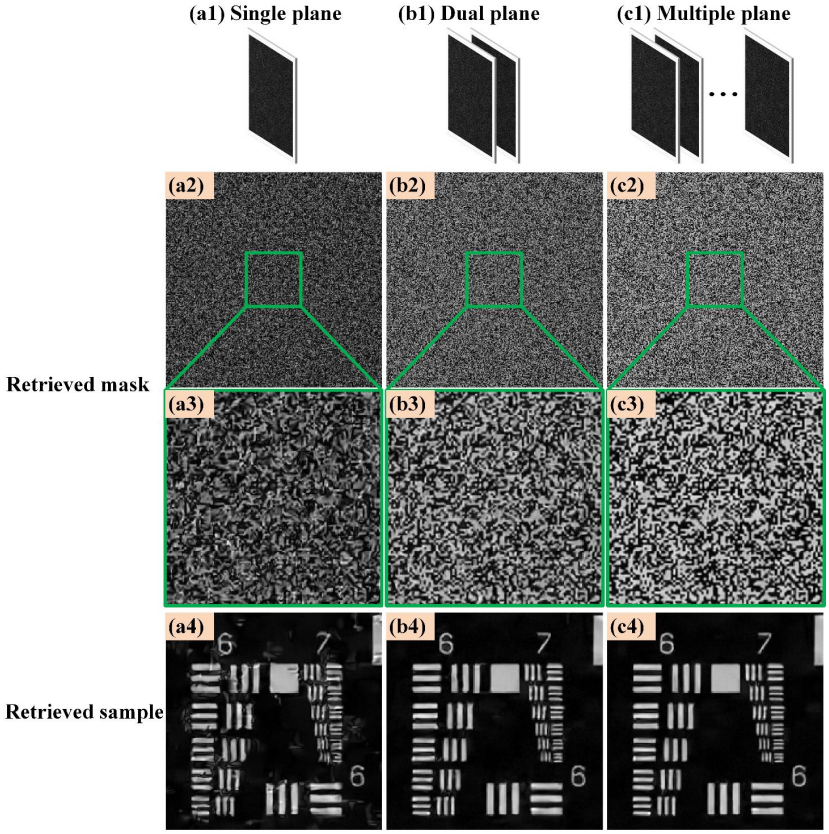

Fig. S6. The retrieved results of our method under the mask pixel of 2.5μm. (a1-a2), (b1-b2), and (c1-c2) are reconstructed by SrPR algorithm with single-frame, dual-plane, and multi-plane intensity patterns. (a3), (b3), and (c3) are reconstructed by DrPR algorithm using single-frame masked intensity image.

6. Discussion of joint recovery using the mask pixel of 2.5μm

In this section, we discuss the reconstructed results of our method under the mask pixel of 2.5μm. Fig. S6(a1) is the captured intensity image of the mask for single-shot measurement. With Fig. S6(a1), the retrieved image of SrPR algorithm is given in Fig. S6(a2), and its central green region is expanded in Fig. S6(a3). It is noted from Fig. S6(a3) that the retrieved mask is

slightly blurred and the detailed distribution of '0' and '1' cannot be completely reconstructed. As this blurred mask is used, the sample image is reconstructed by DrPR in Fig. S6(a4). It is noted that the retrieved sample image is impaired by the residue artifact and parts of line pairs in Group 7 are distorted. This bad performance originates from the mismatch of the mask distribution. The results of Figs. S6(a1-a4) indicate that our method cannot accomplish single-shot joint recovery for the mask with a pixel of 2.5μm.

By analyzing Fig. S6(a), we can infer that the unsuccessful recovery of our method is attributed to that SrPR algorithm cannot realize single-frame recovery for the mask of 2.5μm pixel. Considering that the pixel of the sensor chip is 1.34μm, the pixel-level location of '0' pixels provides a highly ill-posed inverse problem for mask recovery. To further exploit how SrPR can function well, we impose multi-height defocused measurement, and offer the retrieved results of dual-plane and multi-plane measurement in Fig. S6(b) and Fig. S6(c), respectively. Fig. S6(b1) corresponds to dual-plane measurement, in which two intensity images of the mask with different diffractive distances are captured. As dual-plane datasets are input into SrPR algorithm, the amplitude image of the mask is retrieved in Figs. S6(b2-b3). In Fig. S6(c1), 11 intensity images of the mask are recorded under different distances, and the corresponding retrieved image is shown in Figs. S6(c2-c3). As the obtained masks of Fig. S6(b2) and Fig. S6(c2) are used as the physical constraint for DrPR, the reconstructed results of dual-plane and multi-plane are shown in Fig. S6(b4) and Fig. S6(c4) by using single-frame masked intensity image. The comparison of Figs. S6(a4-c4) indicates that the dual-plane measurement of the mask's intensity patterns could result in an efficient recovery, which proves that the minimal measurement number of SrPR is 2 for the recovery of the mask (pixel: 2.5μm).

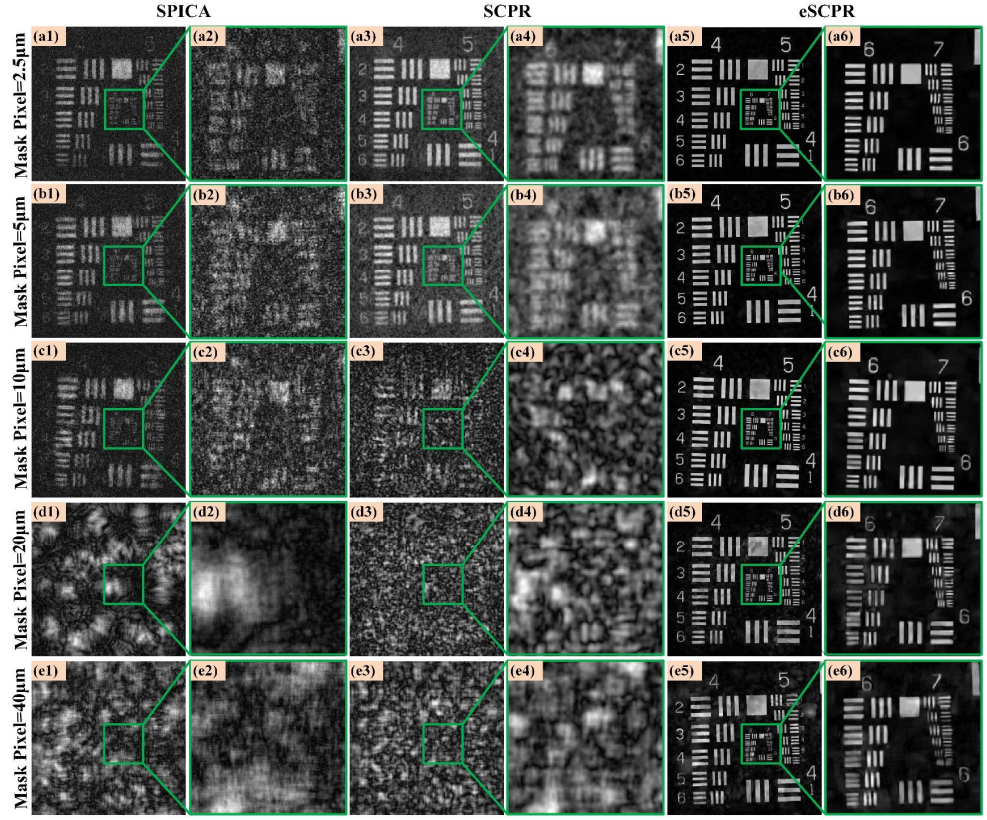

Fig. S7. The reconstructed resolution target of SPICA, SCPR, and our method with the mask pixels of 2.5μm, 5μm, 10μm, 20μm, 40μm.

7. Performance comparison of SPICA, SCPR, and our method on different mask's pixel sizes

In the manuscript (Fig. 3), we only provide the results of our method for different mask pixels. As a supplement, the full results retrieved by SPICA, SCPR, and our method are presented and compared in Fig. S7. Here the mask function of 2.5μm pixel is obtained by dual-plane measurement, and other mask functions are acquired by single-frame measurement. SPICA and SCPR cannot realize single-frame mask recovery, thus the masks retrieved by our SrPR algorithm are provided to SPICA and SCPR for sample recovery. As shown in Figs. S7(a1-a4) and Figs. S7(b1-b4), SPICA, and SCPR only reconstruct the amplitude images of resolution target for the mask pixels of 2.5μm and 5μm. Also, the retrieved images of SPICA and SCPR are undermined by messy backgrounds. In contrast, our method can reconstruct the target for all mask sizes, which proves that our method is insensitive to mask pixels. The robustness of our method provides powerful and practical technical support for the design of LMI systems.

8. Extra experimental results of Fig. 4

In the manuscript (Fig. 4), we only offer the results that cropped from a small region. To show a full-field recovery capability, we supplement the results from other regions in this section. Also, the auto-focusing results of different samples are supplemented.

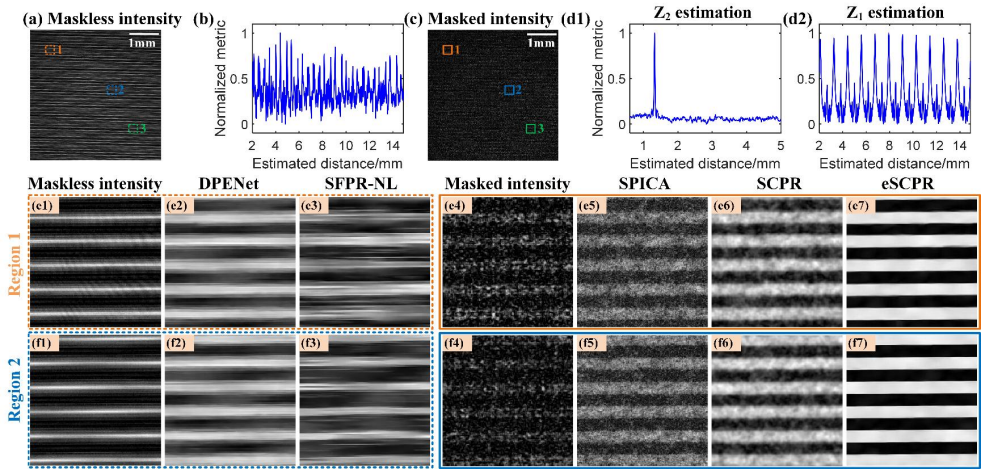

Fig. S8. Single-shot reconstructed results of Rongchi grating with different methods for maskless and masked imaging systems. (a) Full-field maskless intensity image. (b) Auto-focusing curve obtained using (a). (c) Full-field masked intensity image. (d1) and (d2) are auto-focusing curves of Z_2 and Z_1 estimation for our method. (e1-e7) and (f1-f7) are the results cropped from Region 1 and Region 2. The results of Region 3 have been given in the manuscript.

8.1 Rongchi grating

The supplemented results of Rongchi grating are given in Fig. S8. Fig. S8(a) is the captured full-field maskless intensity image (4.6mm×4.6mm). As the Z-stacked images refocused from Fig. S8(a) are assessed by the ToG metric, the auto-focusing curve of sample-to-sensor distance is plotted in Fig. S8(b) for maskless imaging conditions. With the addition of the mask (pixel: 5μm), the captured masked intensity image is displayed in Fig. S8(c). In our method, the mask is inserted between sample and sensor, thus the distance is divided into two parts: sample-to-mask distance (Z_1) and mask-to-sensor distance (Z_2). With the use of the ToG metric and mask-guided auto-focusing method, the auto-focusing curves of Z_2 and Z_1 are plotted in Fig. 8S(d1) and Fig. 8S(d2). Interestingly, the curve of Fig. 8S(d2) exhibits periodic multi-peak distribution, which accords with Talbot effect of a period grating, i.e., the diffraction pattern of a period

grating will be self-imaged under Talbot distances. However, this phenomenon cannot be observed in Fig. S8(b). Furthermore, the main peaks of Fig. 8S(d1) and Fig. 8S(d2) point at Z_2=1.32 mm and Z_1= 7.93mm. This estimated sample-to-sensor distance (Z_1+Z_2) accords with the real distance of our system. Unfortunately, the accurate sample-to-sensor distance cannot be acquired from the main peak of Fig. S8(b). As cropped from Region1 and Region2, the retrieved amplitude images of the grating are shown in Fig. 8S(e) and Fig. 8S(f) for different methods. The results of Region3 have been given in the manuscript. It is noted that our method enables contrast-enhanced images for all regions.

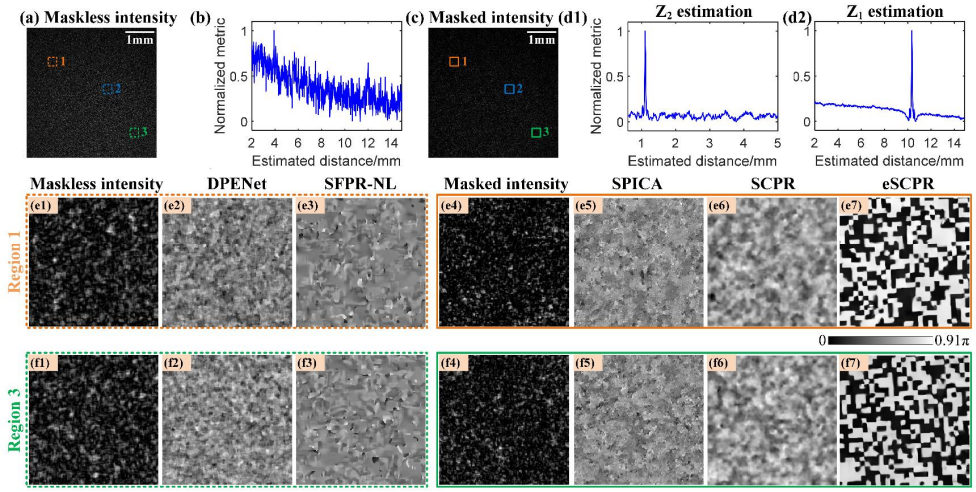

Fig. S9. Single-shot reconstructed results of random phase mask with different methods for maskless and masked imaging systems. (a) Full-field maskless intensity image. (b) Auto-focusing curve obtained using (a). (c) Full-field masked intensity image. (d1) and (d2) are auto-focusing curves of Z_2 and Z_1 estimation for our method. (e1-e7) and (f1-f7) are the results cropped from Region1 and Region3. The results of Region2 have been given in the manuscript.

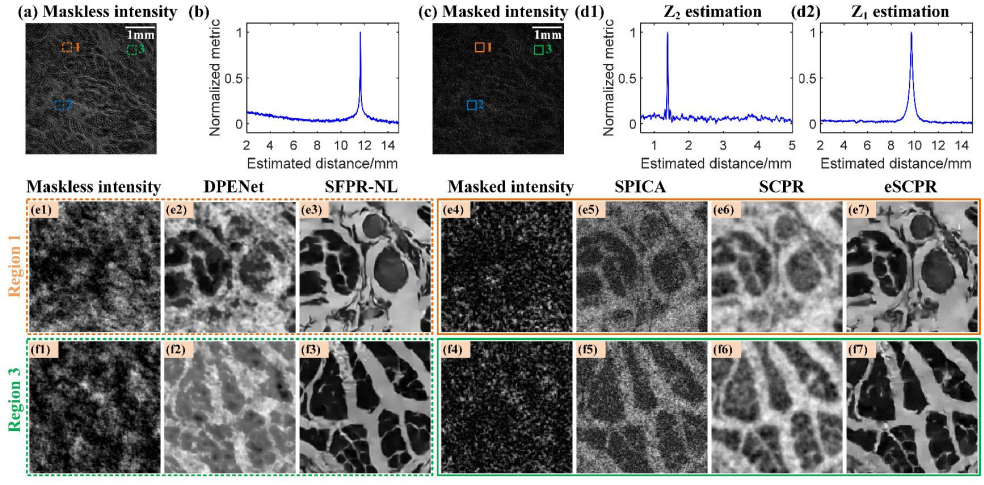

Fig. S10. Single-shot reconstructed results of a stained pathological slide with different methods for maskless and masked imaging systems. (a) Full-field maskless intensity image. (b) Auto-focusing curve obtained using (a). (c) Full-field masked intensity image. (d1) and (d2) are auto-focusing curves of Z_2 and Z_1 estimation for our method. (e1-e7) and (f1-f7) are the results cropped from Region1 and Region3. The results of Region2 have been given in the manuscript.

8.2 Random phase mask

The supplemented results of the random phase mask are given in Fig. S9. In Fig. S9(b), the auto-focusing curve of the maskless intensity is heavily fluctuated and a unimodal curve is not obtained. Also, the estimated distance from the peak of Fig. S9(b) is ~4mm, but this distance is not accurate. With the mask modulation, the auto-focusing curves of Z_2 and Z_1 are plotted in Fig. S9(d1) and Fig. S9(d2), in which the unimodality of the curves are remarkably enhanced and the distances can be easily found. The retrieved phase images of Region1 and Region3 are given in Fig. S9(e) and Fig. S9(f) for different methods. The comparison of Fig. S9(e) and Fig. S9(f) indicates that our method can realize high-quality phase recovery but other methods are still nonconvergent.

8.3 Stained pathological slide

The supplemented results of the stained pathological slide are given in Fig. S10. Different from random phase mask or Rongchi grating, the commercial pathological slide (tissue thickness is ~7μm) can be viewed as a weakly-scattering object, which results in isotropic diffraction for intensity patterns. In this case, the back-propagated result is only covered with the twin-image artifact. Thus, the auto-focusing curve of Fig. S10(b) can output an accurate distance. However, for random phase or grating, the shape and energy of the diffraction field is tremendously changed and thus a simple backpropagation cannot be used for distance estimation. As shown in Fig. S10(e) and Fig. S10(f), the retrieved amplitude images of DPENet, SPICA, and SCPR are degraded by messy backgrounds. SFPR-NL and our method eliminate the artifact and output clear tissue morphology.

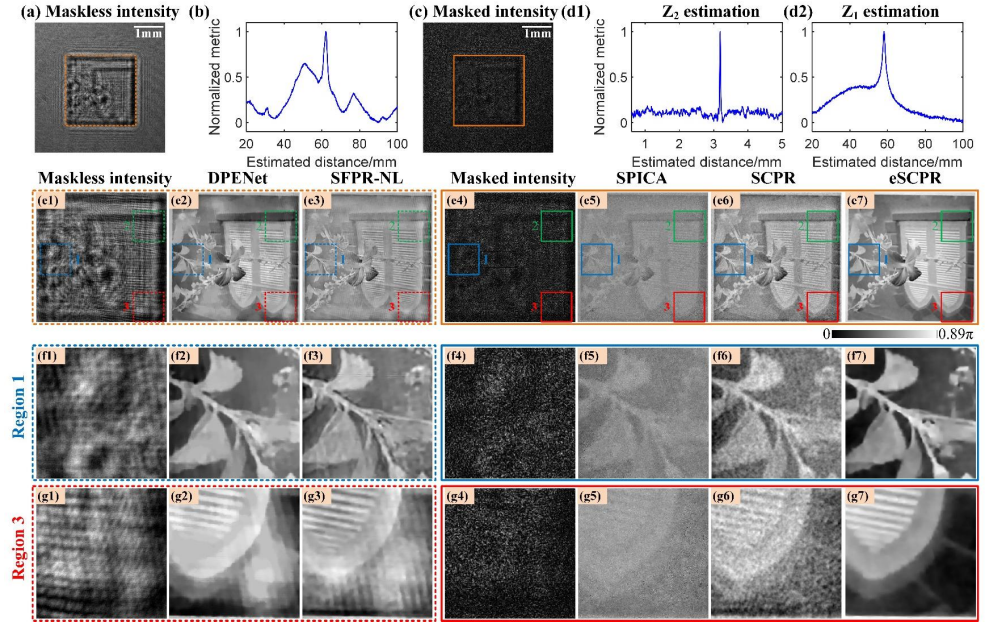

Fig. S11. Single-shot reconstructed results of SLM target with different methods for maskless and masked imaging systems. (a) Full-field maskless intensity image. (b) Auto-focusing curve obtained using (a). (c) Full-field masked intensity image. (d1) and (d2) are auto-focusing curves of Z_2 and Z_1 estimation for our method. (e1) and (e4) are maskless and masked intensity images cropped from the orange boxes of (a) and (c). (e2), (e3) (e5-e7) are retrieved phase images by DPENet, SFPR-NL, SPICA, SCPR, and our method. (f1-f7) and (g1-g7) are the results cropped from Region1 and Region3. The results of Region 2 have been given in the manuscript.

8.4 SLM target

Resolution target, Rongchi grating, phase mask, and stained slide are tested in a transmitted lensless system. The phase-only spatial light modulator (SLM) is built in a reflective lensless system, where a natural image is loaded in the SLM to generate a phase-only target. The captured maskless and masked intensity images are shown in Fig. S11(a) and Fig. S11(c). Considering that a beam splitter mirror is required for the reflective system, the sample-to-sensor distance is increased to dozens of millimeters. Thus, we select the distance range from 20mm to 100mm for distance estimation. The auto-focusing curve of Fig. S11(a) is plotted by the ToG metric in Fig. S11(b). The auto-focusing curve of Fig. S11(c) is plotted in Fig. S11(d1) and Fig. S11(d2) for the estimation of Z_2 and Z_1. In comparison with Fig. S11(b), the mask modulation eliminates the multi-peak artifact for Fig. S11(d). As the orange region is cropped for display, the retrieved phase images of DPENet, SFPR-NL, SPICA, SCPR, and our method are presented in Fig. S11(e2), Fig. S11(e3), Fig. S11(e5), Fig. S11(e6), and Fig. S11(e7). The local results cropped from Region1 and Region3 are displayed in Fig. S11(f) and Fig. S11(g). The results of Region 2 have been given in the manuscript. It is noted that the retrieved phase images of DPENet and SFPR-NL are affected by fringe-like backgrounds. The retrieved phase images of DPENet and SFPR-NL are contaminated by noisy backgrounds. Our method is free of this artifact and outputs a clear phase image.

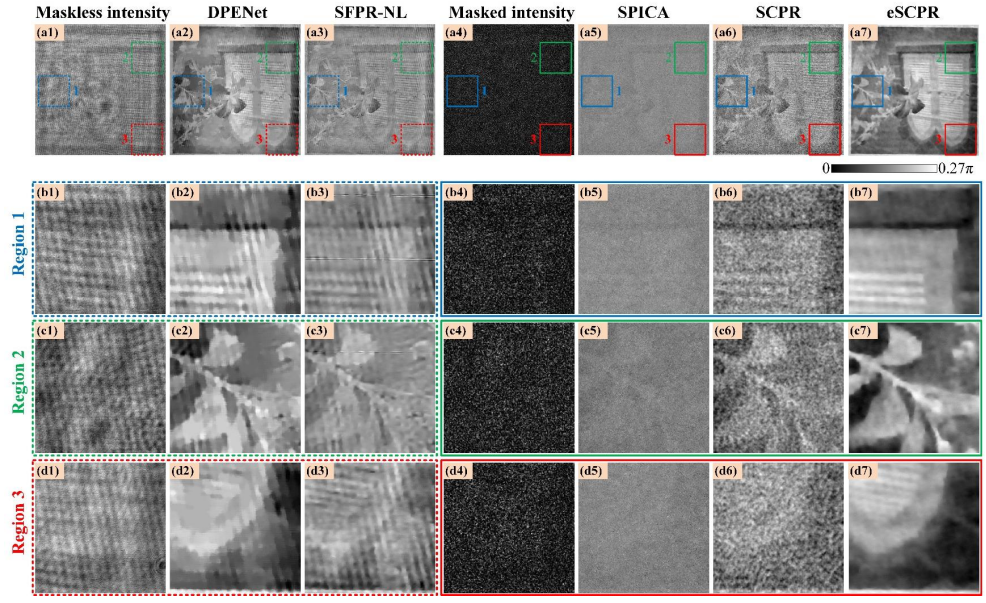

Fig. S12. Single-shot reconstructed results of SLM target when object signal is weak.

9. Performance comparison on noisy conditions

We provide two noisy conditions to further test the noise-robustness of our method. The first experiment is conducted for the reflective SLM target, in which the object signal is decreased so that the gray level of the captured intensity image keeps ~50. The captured maskless and masked intensity images are shown in Fig. S12(a1) and Fig. S12(a4), where the exposure time remains the same for both maskless and masked systems. The retrieved phase images of different methods are shown in Figs. S12(b-d). It is noted that the noise affection on imaging performance is obvious under the weak signal setting. The fringe backgrounds are strengthened for the retrieved images of DPENet and SFPR-NL. For SPICA and SCPR, the retrieved phase images are highly corrupted by noisy backgrounds and their corresponding targets cannot be

distinguished. Our method is robust to the background noise and its retrieved phase image is still clean.

The second experiment is conducted in the transmitted system, where the photon of the incident plane wave is decreased to generate a low-light illumination. In this condition, the exposure time is still the same with/without the loading of the mask. The maskless and masked intensity images are shown in Fig. S13(a1-c1) and Fig. S13(a4-c4). The retrieved amplitude images of breast tissue are shown in Fig. S13 for different methods. It is noted that our method outperforms other methods with high-quality performance.

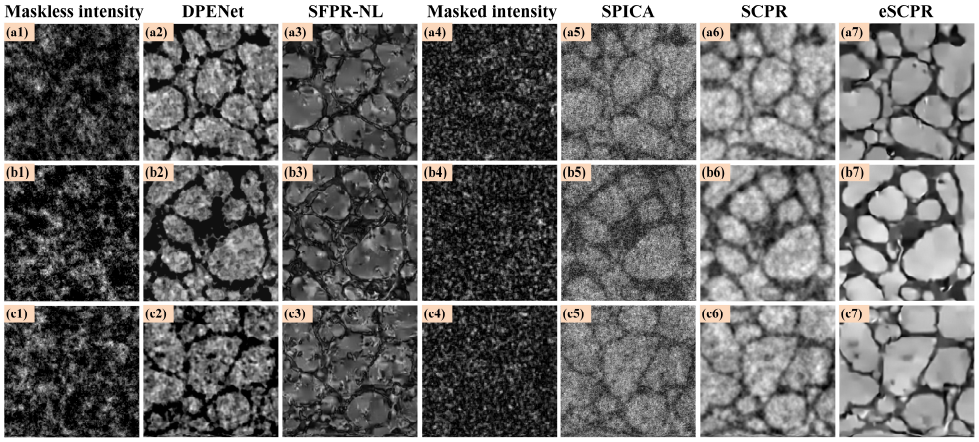

Fig. S13. Single-shot reconstructed results of stained breast tissue under low light conditions.

References

1. Z. Li, Q. Yan, Y. Qin, W. Kong, G. Li, M. Zou, D. Wang, Z. You, and X. Zhou, "Sparsity-based continuous wave terahertz lens-free on-chip holography with sub-wavelength resolution," Opt. Express 27, 702–713 (2019).
2. 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).
3. S. Zheng, Z. Ding, R. Jiang, and C. Guo, "Lensless masked imaging with self-calibrated phase retrieval," Opt. Lett. 48, 3279-3282 (2023).
4. J. M. Rodenburg and H. M. Faulkner, "A phase retrieval algorithm for shifting illumination," Appl. physics letters 85, 4795–4797 (2004).
5. P. Langehanenberg, B. Kemper, D. Dirksen, and G. von Bally, "Autofocusing in digital holographic phase contrast microscopy on pure phase objects for live cell imaging," Appl. Opt. 47(19), D176-182 (2008).
6. Y. S. Choi, and S. J. Lee, "Three-dimensional volumetric measurement of red blood cell motion using digital holographic microscopy," Appl. Opt. 48(16), 2983-2990 (2009).
7. E. Krotkov, "Focusing," Int. J. Comput. Vis. 1(3), 223-237 (1988).
8. 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).
9. 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).