Supplementary material for

Complex-domain enhancing neural network for large-scale coherent imaging

Xuyang Chang,[1,2] Rifa Zhao,[1,2] Shaowei Jiang,[4] Cheng Shen,[6] Guoan Zheng,[5] Changhuei Yang[6] and Liheng Bian[1,2,3,*]

[1] *MIIT Key Laboratory of Complex-field Intelligent Sensing, Beijing Institute of Technology, Beijing 100081, China*

[2] *School of Information and Electronics & Advanced Research Institute of Multidisciplinary Sciences, Beijing Institute of Technology, Beijing 100081, China*

[3] *Yangtze Delta Region Academy of Beijing Institute of Technology (Jiaxing), Jiaxing 314019, China*

[4] *School of Communication Engineering, Hangzhou Dianzi University, Hangzhou, 310018, China*

[5] *Department of Biomedical Engineering, University of Connecticut, Storrs, Connecticut 06269, USA*

[6] *Department of Electrical Engineering, California Institute of Technology, Pasadena, California 91125, USA*

[*] *bian@bit.edu.cn*

Contents

1 Supplementary Note 1: Training details.

Training dataset. We used 20000 images from ImageNet and Pascal VOC datasets, then each pair of them was combined randomly to make up 10000 complex-domain images. Specifically, two images in each pair were first resized to 256×256 pixels as the amplitude and phase, respectively. The synthesized complex-domain image can be indicated as

$$s(x', y') = A(x', y') \cdot e^{j\varphi(x', y')}, \tag{S1}$$

where A is the amplitude image, φ presents the phase image. The coordinates in the sample plane, pupil plane and camera plane are indicated as (x', y'), (u, v) and (x, y), respectively.

Then, we added multi-source noise to the real and imaginary parts of the clear complex-domain images with a random shuffle strategy. The noisy image is modeled as

$$s_{noisy}(x', y') = s_R + j * s_I + \omega_R + j * \omega_I, \tag{S2}$$

where s_R, s_I are the real and imaginary parts of the clear image, ω_R and ω_I are the real and imaginary parts of the multi-source noise. Finally, the noisy image $s_{noisy}(x', y')$ was resized to 512×512 pixels to simulate the super-resolution reconstruction noise.

Noise map. The noise map was a quantified index controlling the denoising degree and details maintaining. A larger noise map corresponds to more smooth results. Because the Gaussian noise can be quantified and was certainly added to the training data, thus we employed Gaussian noise variance as the noise map value. In our training, the noise map was padded to the same size as the input complex-domain images and put into the network as another channel.

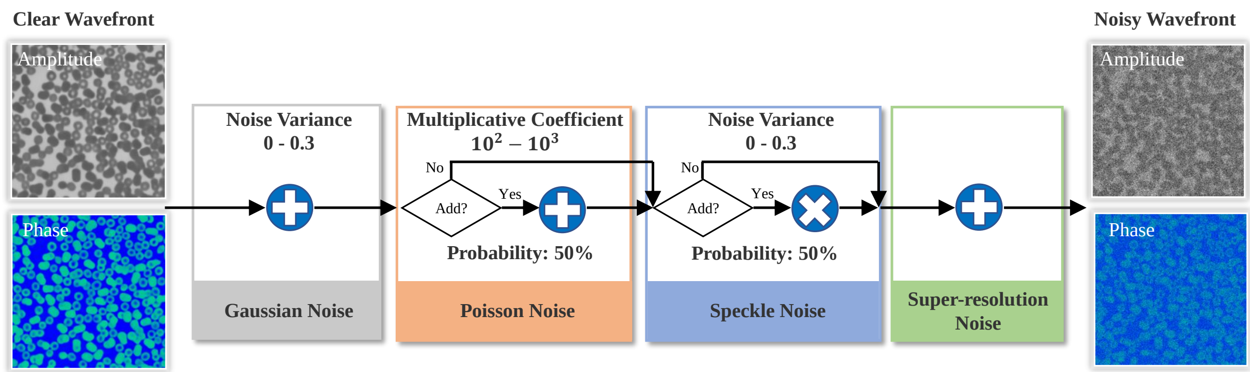

Figure S1: Noise model and parameters for constructing training data.

The generalized complex chain rule for a real-domain loss function. Assuming that L is a real-domain loss function, and $Z = Z_R + j * Z_I$ is a complex value. The generalized complex chain rule is

$$\nabla_L(z) = \frac{\partial L}{\partial z} = \frac{\partial L}{\partial Z_R} + j\frac{\partial L}{\partial Z_I} = \Re\left(\nabla_L(z)\right) + i\Im\left(\nabla_L(z)\right). \tag{S3}$$

2 Supplementary Note 2: Regularization optimization framework for ptychographic reconstruction.

Theory derivation. Recently developed regularization optimization frameworks in computer vision show great advantages for image reconstruction [32,33]. We introduced the reported CI-CDNet technique and other comparison algorithms as the prior regularization terms during the iterations.

Considering the computational complexity, reconstruction quality and generalization, we employed the efficient generalized-alternating-projection (GAP) framework which has been applied for large-scale phase retrieval tasks in our previous work [33]. Specifically, the reconstruction of ptychography (Fourier ptychographic microscopy and lensless coded ptychography) can be modeled as a generalized optimization function

$$\hat{s} = \arg\min_{s} f(s) + g(s), \tag{S4}$$

where s is the reconstruction objective, $f(s)$ is the data fidelity term and $g(s)$ is the prior regularization term. Following the GAP framework, with a introduced auxiliary variable κ, Eq. (S4) can be derived as

$$\begin{gathered}(s, \kappa) = \operatorname{argmin} 1/2\|s - \kappa\|_2^2 + \eta g(\kappa) \\ \text{s.t. } I = |As|^2,\end{gathered} \tag{S5}$$

where κ is an auxiliary variable, η is a weight coefficient to balance the data fidelity term and prior regularization, A denotes the forward model, and I represents intensity-only measurements. Then, Eq. (S5) can be solved by the following two subproblems.

- Solving s: given $\kappa^{(k)}$, $s^{(k+1)}$ is updated via a Euclidean projection of $\kappa^{(k)}$ on the manifold

$I = |As|^2$ as

$$s^{k+1} = \kappa^{(k)} + \eta \cdot PR\left(I - |A\kappa|^2\right), \tag{S6}$$

where PR is a phase retrieval solver. We employ the alternating projection (AP) framework as this solver due to its great generalization and low computational complexity. It alternates between the object and imaging planes and imposes constraints.

- Updating κ: given $s^{(k+1)}$, $\kappa^{(k+1)}$ is updated by different denoising solvers as

$$\kappa^{k+1} = DE\left(s^{k+1}\right). \tag{S7}$$

After initialization, the variables are updated alternatively following Eq. (S6) and Eq. (S7). Since both the two solvers PR and EN are highly efficient and flexible, the entire reconstruction maintains low computational complexity and strong generalization. We summarized the reconstruction algorithms of Kramers-Kronig-relations holography (KKR), Fourier ptychographic microscopy (FPM) and lensless coded ptychography (LCP) in Algorithm 1, Algorithm 2 and Algorithm 3, respectively.

Algorithm 1: KKR reconstruction with CI-CDNet.

Input: Intensity-only measurements I_i, Scanning aperture $D(u-u_i, v-v_i)$.

Output: Recovered wavefront s.

$r_i = e^{-j(u_i \cdot x + v_i \cdot y)}$ ▷ Hypothetical reference waves

$H_i = -j\,\mathrm{sgn}\,(v_i) \cdot \mathrm{sgn}(v)$ ▷ Defined Hilbert Kernels

$X = \ln\left[\mathcal{F}^{-1}\left\{S_i\right\}/r_i\right]$ ▷ Create a analytic function

$\mathrm{Re}\{X\} = \frac{1}{2}\ln\left[I_i/\left|r_i\right|^2\right]$ ▷ Recover the real part of X

$\mathrm{Im}\{X\} = \mathcal{F}^{-1}\left\{\mathcal{F}\{\mathrm{Re}\{X\}\} \cdot H_i\right\}$ ▷ Recover the imaginary part of X

$S_i = \mathcal{F}\left\{e^{\mathrm{Re}\{X\}+j\,\mathrm{Im}\{X\}} \cdot r_i\right\}$ ▷ Recover the subregions of Fourier spectrum

$S = \sum_{i=1}^{4} S_i/\left[\sum_{i=1}^{4} D\left(u-u_i, v-v_i\right) + \varepsilon\right]$

▷ Recover the available Fourier spectrum by aperture synthesis

$s = \mathcal{F}^{-1}\left\{S\right\}$

$s \leftarrow$ CI-CDNet$[s]$

Algorithm 2: FPM reconstruction with CI-CDNet regularizer.

Input: Intensity-only measurements I, Initialization s_0 ($\kappa_0 = s_0$),

Coherent transfer function C, Forward model A.

Output: High-resolution wavefront s.

For $k = 1, 2, \ldots,$ **do**

  $I^{(k)} = I - |A\kappa^{(k)}|^2$

  $S^{(k)} = \mathcal{F}\{s^{(k)}\}$

  For $i = 1, 2, \ldots$, **do**

    $L_i \leftarrow S^{(k)}_{subregion}(u_l : u_h, v_l : v_h)$ ▷ Sub-region in Fourier plane

    $l_i = \mathcal{F}^{-1}\{C \odot L_i\}$

    $l'_i = \sqrt{I^{(k)}_i} \odot \frac{l_i}{|l_i|}$ ▷ Update wavefront with intensity constraint

    $L'_i = \mathcal{F}\{l'_i\}$

    $S^{(k)}(u_l : u_h, v_l : v_h) \leftarrow L'_i$ ▷ Fourier plane synthesis

  End

  $s^{(k)}_{update} = \mathcal{F}^{-1}\{S^{(k)}\}$

  $s^{(k+1)} = \kappa^{(k)} + \eta \cdot s^{(k)}_{update}$

  $\kappa^{(k+1)} \leftarrow$ CI-CDNet$[s^{(k+1)}]$

End

Algorithm 3: LCP reconstruction with CI-CDNet regularizer.

Input: Intensity measurements I, Initialization sample s_0 and diffuser D_0,

Propagation function $PSF(d)$, Propagation distance d_1 and d_2,

Forward model A.

Output: High-resolution wavefront s and diffuser's profile D.

For $k = 1, 2, \ldots,$ **do**

  $I^{(k)} = I - |A\kappa^{(k)}|^2$

  $W = s^{(k)} * PSF_{free}(d_1)$ ▷ Propagation d_1 to diffuser plane

  For $i = 1, 2, \ldots,$ **do**

    $W_i \leftarrow W\left(u + u_i, v - v_i\right)$ ▷ Wavefront shift

    $\phi_i = W_i \odot D$ ▷ Diffuser modulation

    $\psi_i = \phi_i * PSF_{free}\left(d_2\right)$ ▷ Propagation d_2 to imaging plane

    $\psi_i' = \sqrt{I_i^{(k)}} \odot \frac{\psi_i}{|\psi_i|}$ ▷ Update wavefront with intensity constraint

    $\phi_i' = \psi_i' * PSF_{free}\left(-d_2\right)$ ▷ Propagation $-d_2$ to diffuser plane

    $W = W + \frac{\mathrm{conj}(D) \odot \left[\phi_i' - \phi_i\right]}{(1-\alpha_1)|D|^2 + \alpha_1 |D|^2_{\max}}$ ▷ Updata wavefront

    $D = D + \frac{\mathrm{conj}(W) \odot \left[\phi_i' - \phi_i\right]}{(1-\alpha_2)|W|^2 + \alpha_2 |W|^2_{\max}}$ ▷ Updata diffuser

    $W \leftarrow W_i\left(u - u_i, v + v_i\right)$ ▷ Wavefront shift back

  End

  $s^{(k)}_{update} = W * PSF_{free}(-d_1)$ ▷ Propagation $-d_1$ to sample plane

  $s^{(k+1)} = \kappa^{(k)} + \eta \cdot s^{(k)}_{update}$

  $\kappa^{(k+1)} \leftarrow$ CI-CDNet$\left[s^{(k+1)}\right]$

End

3 Supplementary Note 3: Principle and setups of different coherent imaging modalities.

Kramers-Kronig-relations holography. The principle of aperture modulation Kramers-Kronig-relations (KKR) holography is presented in Fig. S2. In order to satisfy the analyticity [36], we used a circular binary aperture $D(u,v)$ to scan in the Fourier plane, as shown in the first line of Fig. S2 (b). A spatial light modulator (SLM) was employed to generate the modulation aperture, and the aperture's edge strictly crosses the objective pupil center. We implemented modulation four times to ensure the whole objective numerical aperture (NA) is covered. The available spectrum in the Fourier plane is

$$S(u,v) = \mathcal{F}\left\{s\left(x',y'\right)\right\} \cdot C(u,v), \tag{S8}$$

where S is the available Fourier spectrum, s is the sample, C is the coherent transfer function and $\mathcal{F}$ presents the 2D Fourier transform. Then, the intensity-only measurements I in the imaging plane can be modeled as

$$I_i(x,y) = \left|\mathcal{F}^{-1}\left\{S(u,v) \cdot D(u-u_i, v-v_i)\right\}\right|^2, \tag{S9}$$

where $D(u-u_i, v-v_i)$ is the scanning aperture ($i = 1,2,3,4$) and $\mathcal{F}^{-1}$ presents the 2D inverse Fourier transform.

The complex wavefront $S(u,v)$ can be recovered using the four measurements. Specifically, the hypothetical reference waves were first generated by

$$r_i(x,y) = \mathcal{F}^{-1}\left\{\delta\left(u+u_i, v+v_i\right)\right\} = e^{-j(u_i \cdot x + v_i \cdot y)}, \tag{S10}$$

where $\delta(u+u_i, v+v_i)$ is the Dirac delta function. We defined Hilbert kernels which depend on the

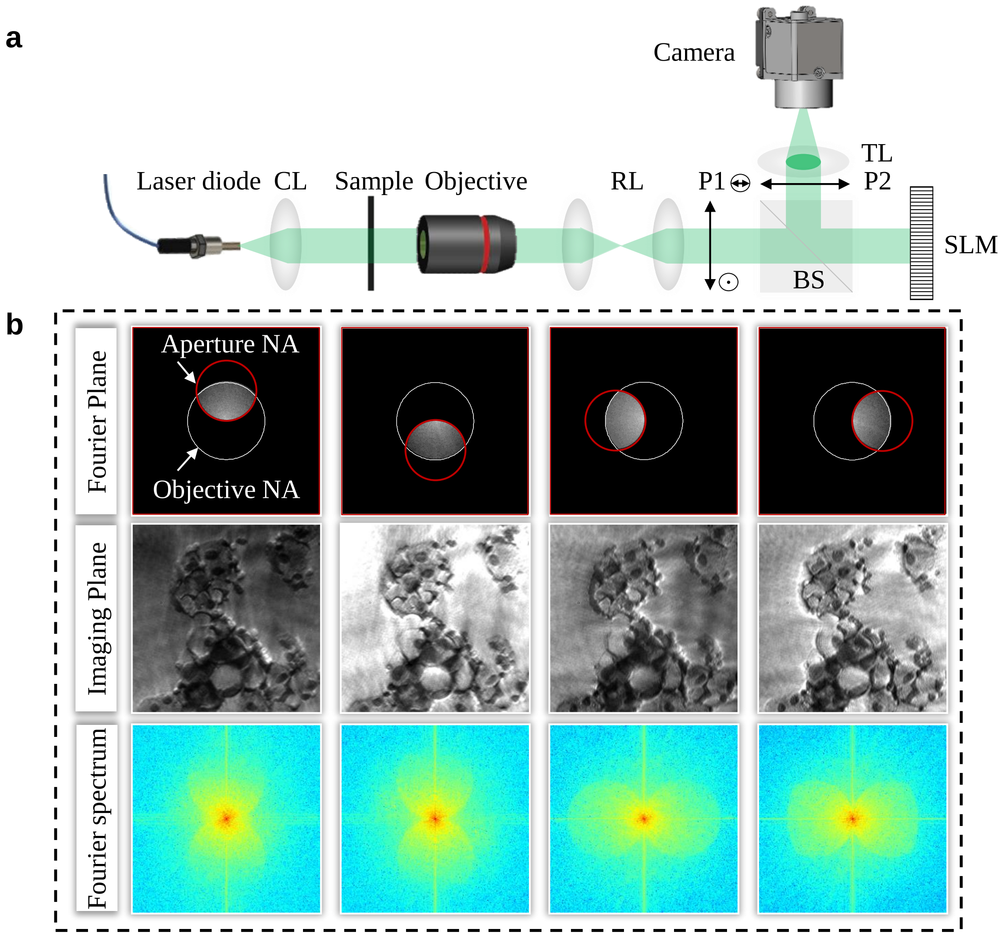

Figure S2: Principle of aperture modulation Kramers-Kronig-relations holography. (a) System diagram and setups. (b) Aperture modulation strategies (first line), intensity-only measurements corresponding to different subregions (second line) and Fourier spectrums of different measurements (third line).

positions of the scanning aperture

$$H_i(u,v) = -j\,\mathrm{sgn}\,(v_i) \cdot \mathrm{sgn}(v), \tag{S11}$$

where $\mathrm{sgn}(v)$ is the sign function. Then, an intermediate variable X is defined as

$$X = \ln\left[\mathcal{F}^{-1}\left\{S_i\left(u+u_i, v+v_i\right)\right\}/r_i(x,y)\right]. \tag{S12}$$

We can recover the real part of X using the measurements I and hypothetical reference waves r. Its imaginary part can be obtained using Kramers-Kronig relations and the defined Hilbert kernels.

$$\begin{aligned} \mathrm{Re}\{X\} &= \frac{1}{2}\ln\left[I_i(x,y)/\left|r_i(x,y)\right|^2\right] \\ \mathrm{Im}\{X\} &= \mathcal{F}^{-1}\left\{\mathcal{F}\{\mathrm{Re}\{X\}\}\cdot H_i(u,v)\right\}. \end{aligned} \tag{S13}$$

Then, the shifted spectrum subregions $S_i\left(u+u_i, v+v_i\right)$ can be recovered using the intermediate variable X and hypothetical reference waves r

$$S_i\left(u+u_i, v+v_i\right) = \mathcal{F}\left\{e^{\mathrm{Re}\{X\}+j\,\mathrm{Im}\{X\}}\cdot r_i(x,y)\right\}. \tag{S14}$$

Finally, the sample's Fourier spectrum can be recovered through the synthetic aperture technique

$$S(u,v) = \sum_{i=1}^{4} S_i(u,v)/\left[\sum_{i=1}^{4} D\left(u-u_i, v-v_i\right)+\varepsilon\right], \tag{S15}$$

where $\varepsilon = 10^{-5}$ is a small constant for numerical stability. The recovered sample $s(x', y')$ is the inverse Fourier transform of $S(u,v)$.

Fourier ptychographic microscopy. Figure S3 (a) represents the schematic diagram of Fourier ptychographic microscopy. Assuming that the incident light through the sample is a plane wave, thus the light field transmitted from the sample can be described as $s(x', y')e^{\left(jx'\frac{2\pi}{\lambda}\sin\theta_{x'}, jy'\frac{2\pi}{\lambda}\sin\theta_{y'}\right)}$, where $s(x', y')$ is the sample, λ is the wavelength, and $\theta_{x'}$ and $\theta_{y'}$ are the illumination angles. Then the light wave interacts with the pupil in the Fourier plane, which can be expressed as

$$C\left(u,v\right)\cdot\mathcal{F}\left(s(x', y')e^{\left(jx'\frac{2\pi}{\lambda}\sin\theta_{x'}, jy'\frac{2\pi}{\lambda}\sin\theta_{y'}\right)}\right). \tag{S16}$$

Then, the light wave passed through a tube lens to the imaging plane, and a detector captured

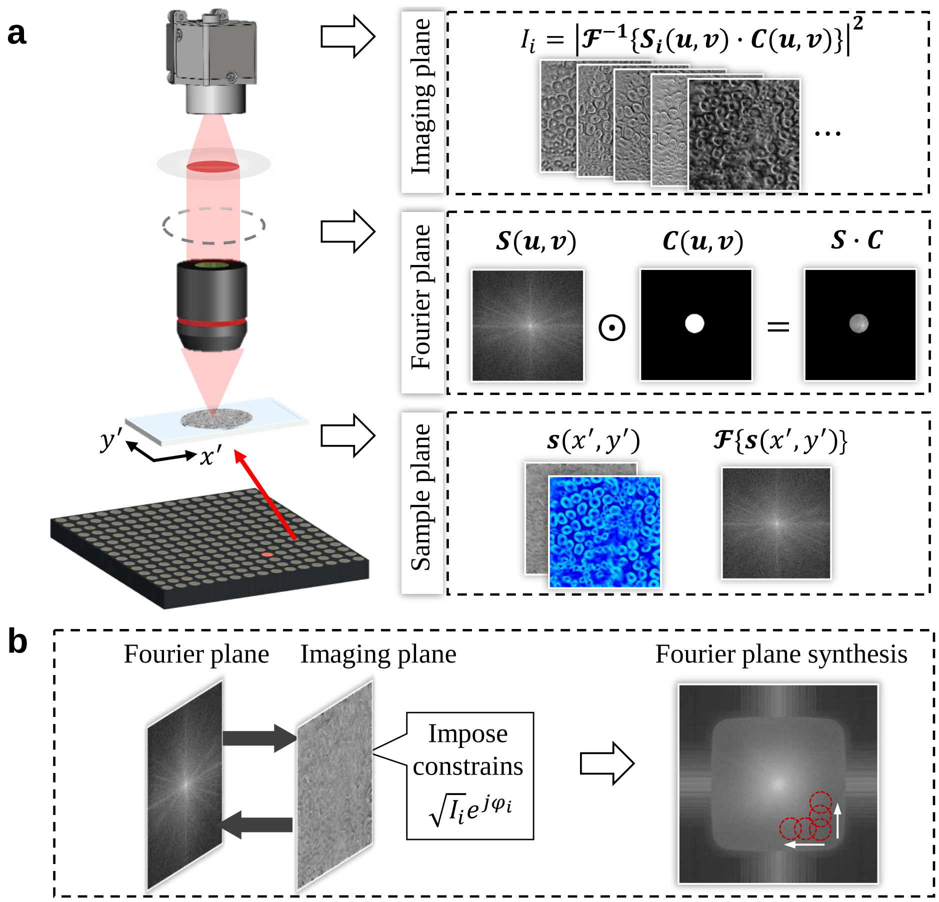

Figure S3: Principle of Fourier ptychographic microscopy (FPM). (a) Schematic diagram and setups. (b) FPM reconstruction principle.

the light's intensity. The final formation of FPM can be indicated as

$$\begin{aligned} I(x,y) &= \left|\mathcal{F}^{-1}\left[C\left(u,v\right)\cdot\mathcal{F}\left\{s(x',y')e^{\left(jx'\frac{2\pi}{\lambda}\sin\theta_{x'},jy'\frac{2\pi}{\lambda}\sin\theta_{y'}\right)}\right\}\right]\right|^2 \\ &= \left|\mathcal{F}^{-1}\left[C\left(u,v\right)\cdot S\left(u-\frac{2\pi}{\lambda}\sin\theta_{x'},v-\frac{2\pi}{\lambda}\sin\theta_{y'}\right)\right]\right|^2. \end{aligned} \tag{S17}$$

The FPM reconstruction is a phase retrieval task. We employed the alternating projection (AP) algorithm as the baseline, as shown in Fig. S3 (b). AP started with an initial guess, then alternates between the Fourier plane and detector plane to impose constraints. The final available Fourier spectrum is the synthesis of multiple lowpass spectrums.

Lensless coded ptychography. Lensless coded ptychography (LCP) combines blind ptychography and scattering multiplexing techniques. LCP has two significant advantages. First, the diffuser encodes the high-frequency information to the measurements through scattering multiplexing. The final achievable resolution is not limited by the optical elements (e.g. pixel size of the detector), but by the feature size of the diffuser. Second, LCP is a low-cost technique to realize super-resolution coherent imaging, without any expensive modulators (spatial light modulators or digital mirror devices). The forward model of the LCP platform can be expressed as

$$\begin{cases} W = s\left(x', y'\right) * PSF_{free}\left(d_1\right) \\ I_i = \left|\left[W_i\left(u + u_i, v - v_i\right) \odot D\right] * PSF_{free}\left(d_2\right)\right|^2 \end{cases} \tag{S18}$$

where $s(x', y')$ is the exit wavefront of the sample plane, $PSF_{free}(d)$ denotes the point spread function for free space propagation over distance d, W is the wavefront in the diffuser plane and $W(u+u_i, v-v_i)$ presents the wavefront shift, D is the diffuser's profile, I_i is the ith (i = 1,2,. . . ,I) intensity measurements, $*$ represents the convolution operation and $\odot$ represents Hadamard product.

The conventional LCP reconstruction method is based on the ePIE technique which is following the AP framework, as shown in the inner loop of Algorithm 3. It started with a high-resolution initial wavefront, then propagates to the diffuser plane (line 3) and shifts the wavefront (line 5). After the diffuser's modulation (line 6), the wavefront propagates to the imaging plane (line 7) and imposes intensity constraints (line 8). Then, the wavefront propagates to the diffuser plane again and updates the wavefront and diffuser using line 10 and line 11. Finally, the wavefront is shifted back (line 12) and propagates back to the sample plane (Line 14).

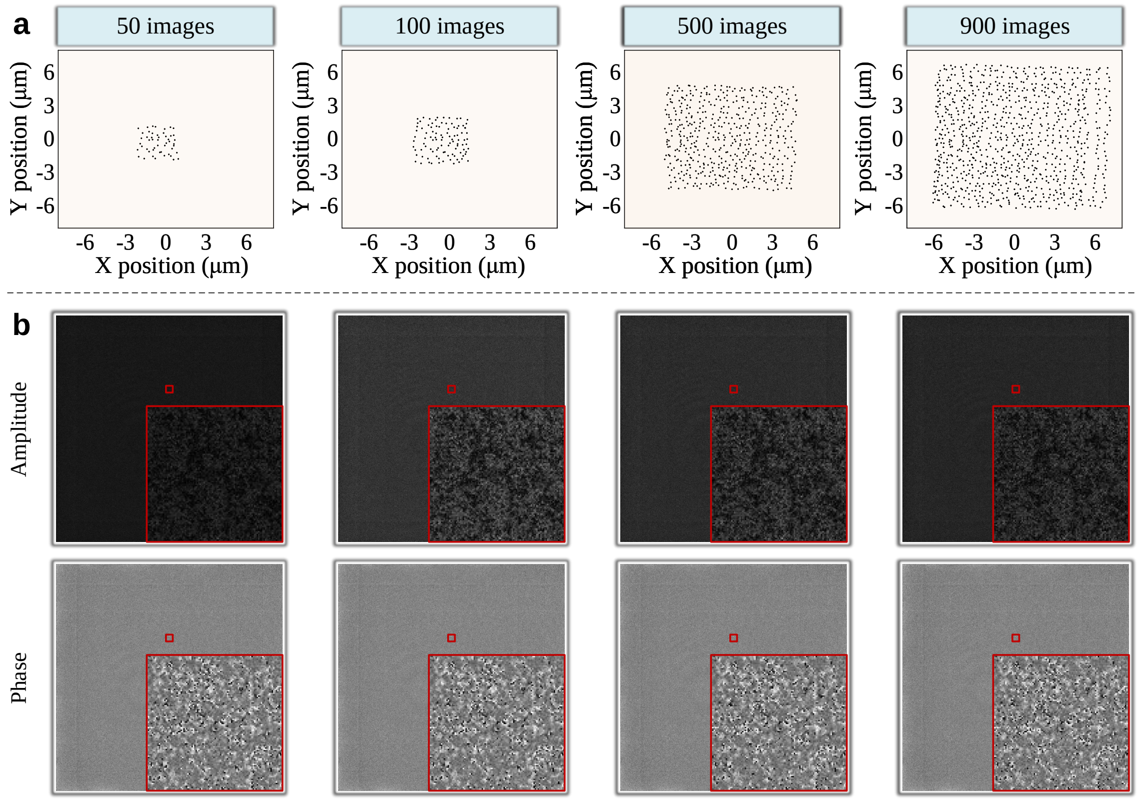

Figure S4: Principle of lensless coded ptychography (LCP). (a) Positions of the sample's shifts under different data volumes. (b) Reconstructed diffuser's profile under different data volumes.

Scattering layer preparation. We employed glass-etching chemicals to create a cover slip. This involved applying a solution comprising 17% barium sulfate, 11% sulfuric acid, 8% sodium bifluoride, 5% ammonium bifluoride to the microscope coverslip for a duration of 1-3 seconds, followed by thorough washing with water. This etching and cleaning process was repeated 5-10 times to generate dense phase scatters on the surface. Subsequently, the etched surface was gently rubbed with silk cloth to impart a positive charge. Carbon nanoparticles were then deposited onto the etched surface using a negatively charged printer roller.

4 Supplementary Note 4: Details of the high-level semantic analysis.

Cell segmentation. We employed U-net to implement white blood cell segmentation, as shown in Fig. S5. The training dataset was from the Jiangxi Tecom Science Corporation [46], and we extended the dataset to 2700 through translation and rotation. We used L1 and L2 mixed loss with equal weight and Adam optimizer to update parameters. The batch size was 128, and the epoch was 500 with a learning rate from 2×10^{-4} to 6×10^{-6}. We implemented the training in PyTorch 1.8.1 and NVIDIA 2080ti GPU for about one day.

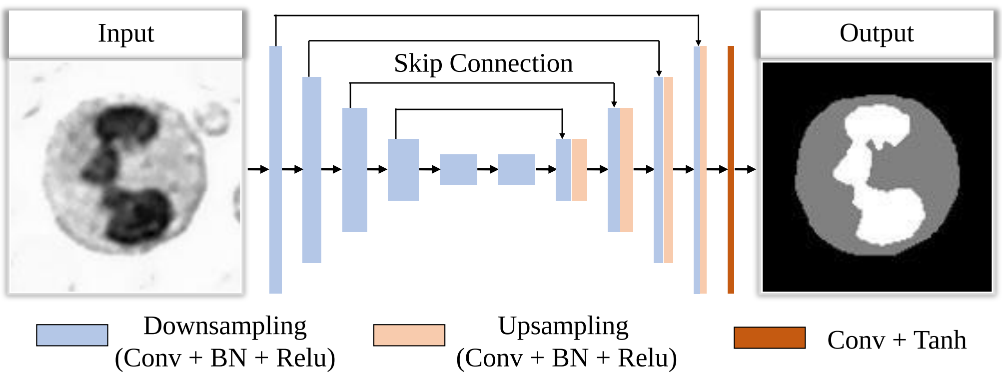

Figure S5: U-net architecture for white blood cell segmentation.

Cell virtual staining. We used the cycleGAN with phase attention guide to realize cell virtual staining [19], as indicated in Fig. S6. It contains two generators G_{AB} and G_{BA} to realize the prediction between the reconstructed wavefront A and the virtually stained image B. Each generator has a discriminator (D_A and D_B) to distinguish the real and fake images. We utilized dual-channel images comprising both amplitude and phase as the input for CycleGAN. Specifically, the generator G_{AB} in our network consists of nine pairs of downsampling blocks followed by nine up-sampling blocks. The down-sampling blocks in the phase path serve as multiscale attention guidance for the feature maps in the intensity path. This approach allows us to leverage the phase information to guide the virtual staining process effectively. The loss function is

$$
\begin{aligned}
\operatorname{loss}\left(G_{\mathrm{AB}}, G_{\mathrm{BA}}, D_{\mathrm{A}}, D_{\mathrm{B}}\right) = {} & L_{\mathrm{GAN-AB}}\left(G_{\mathrm{AB}}, D_{\mathrm{B}}, A, B\right) \\
& + L_{\mathrm{GAN-BA}}\left(G_{\mathrm{BA}}, D_{\mathrm{A}}, B, A\right) + \lambda_1 \times L_{\mathrm{cyc}}\left(G_{\mathrm{AB}}, G_{\mathrm{BA}}\right) \\
& + \lambda_2 \times \left(1 - \mathrm{msSSIM}_g\left(G_{\mathrm{AB}}(A), A\right)\right) \\
& + \lambda_2 \times \left(1 - \mathrm{msSSIM}_g\left(G_{\mathrm{BA}}(B), B\right)\right),
\end{aligned} \tag{S19}
$$

where L_{GAN-AB} and L_{GAN-BA} are the adversarial losses, L_{cyc} is the cycle consistency loss. $msSSIM_g$ presents the multiscale SSIM loss between the green channel of the stained images and the reconstructed amplitude. The generators are U-Net with nine upsampling and downsampling scales. The discriminators are PatchGAN classifiers.

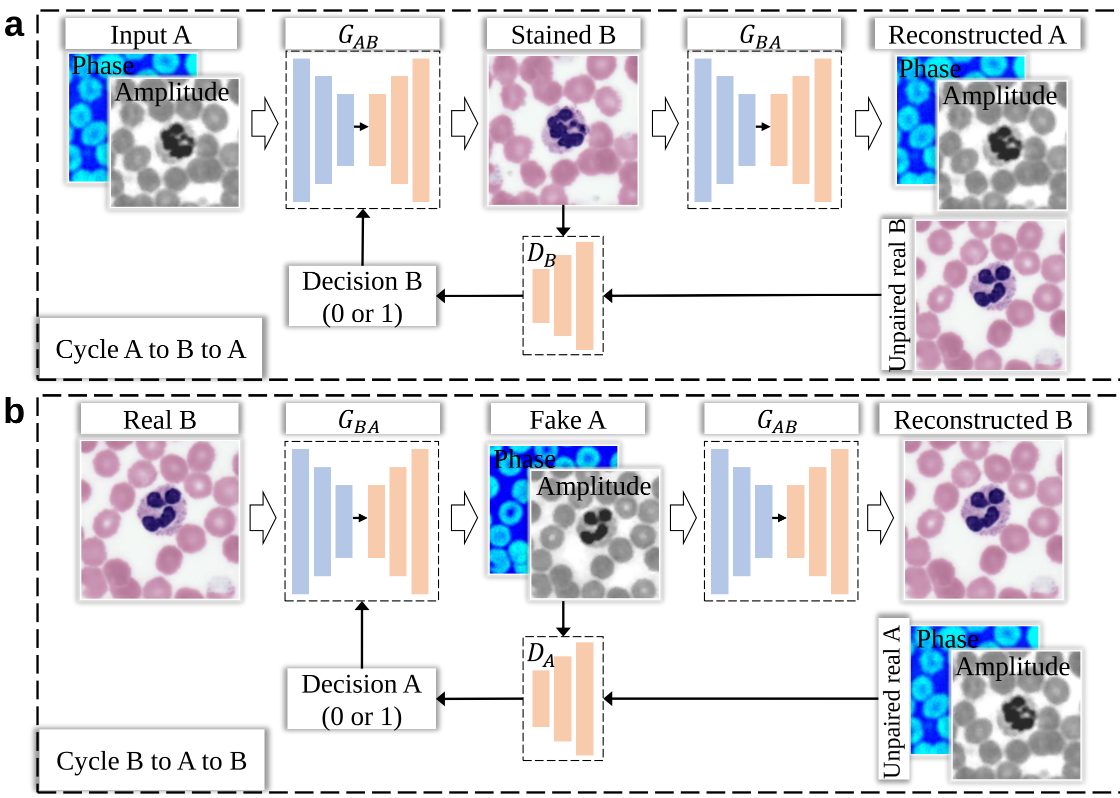

Figure S6: Architecture of cycleGAN for virtual staining.

5 Supplementary Note 5: Quantitative comparison of data volume requirement and exposure time.

Exposure times of Kramers-Kronig-relations holography. Figure S7 shows the result of CI-CDNet under 1 ms exposure time and the results of KKR direct reconstruction under different exposure times. The PSNR and SSIM indexes validate that the reconstruction quality of CI-CDNet under 1 ms exposure time is better than the result of KKR under 10 ms exposure time, and is close to the result of KKR under 50 ms exposure time. Thus, CI-CDNet reduced more than one order of magnitude in exposure time, which is more practical in low-light imaging.

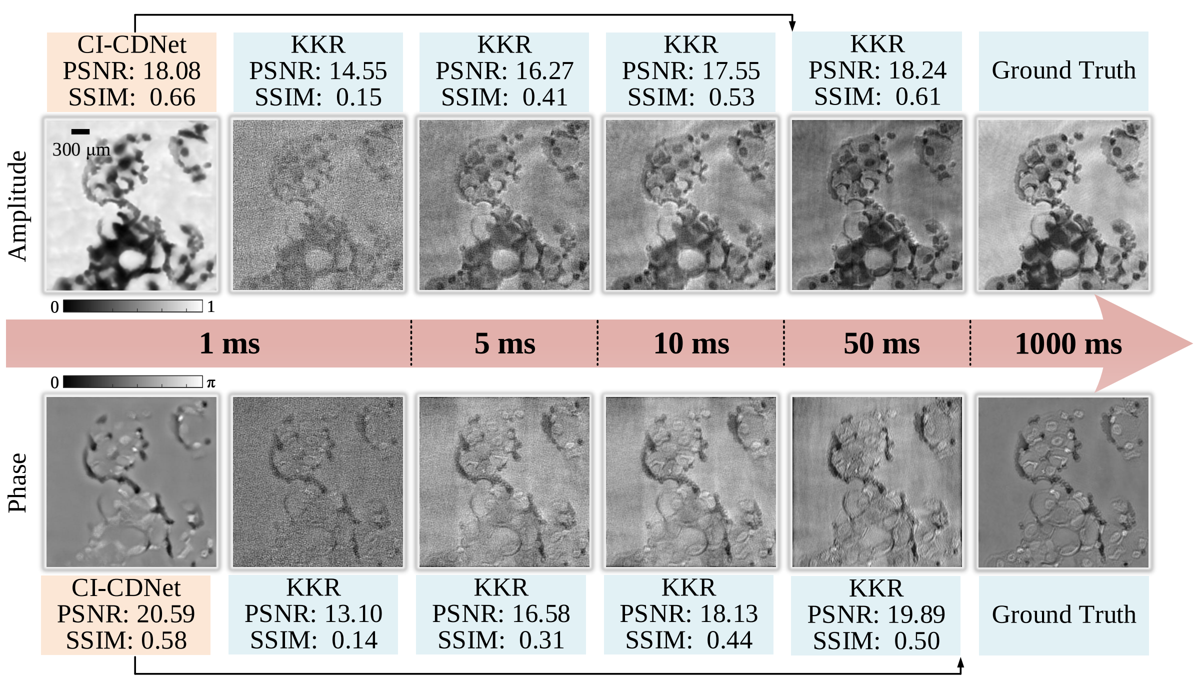

Figure S7: Quantitative results of Kramers-Kronig-relations holography under different exposure times. The reconstruction results of CI-CDNet under 1 ms exposure time are close to the results of KKR under 50 ms exposure time.

Data volume requirement of lensless coded ptychography. We also compared the data volume requirement in lensless coded ptychography. Figure S8 is the results of different data volumes (quantified by the number of captured images). The PSNR index indicates that the reconstruction results of CI-CDNet using 50 images are better than the results of ePIE using 500 images. Thus, CI-CDNet reduced about one order of magnitude in data volume requirement.

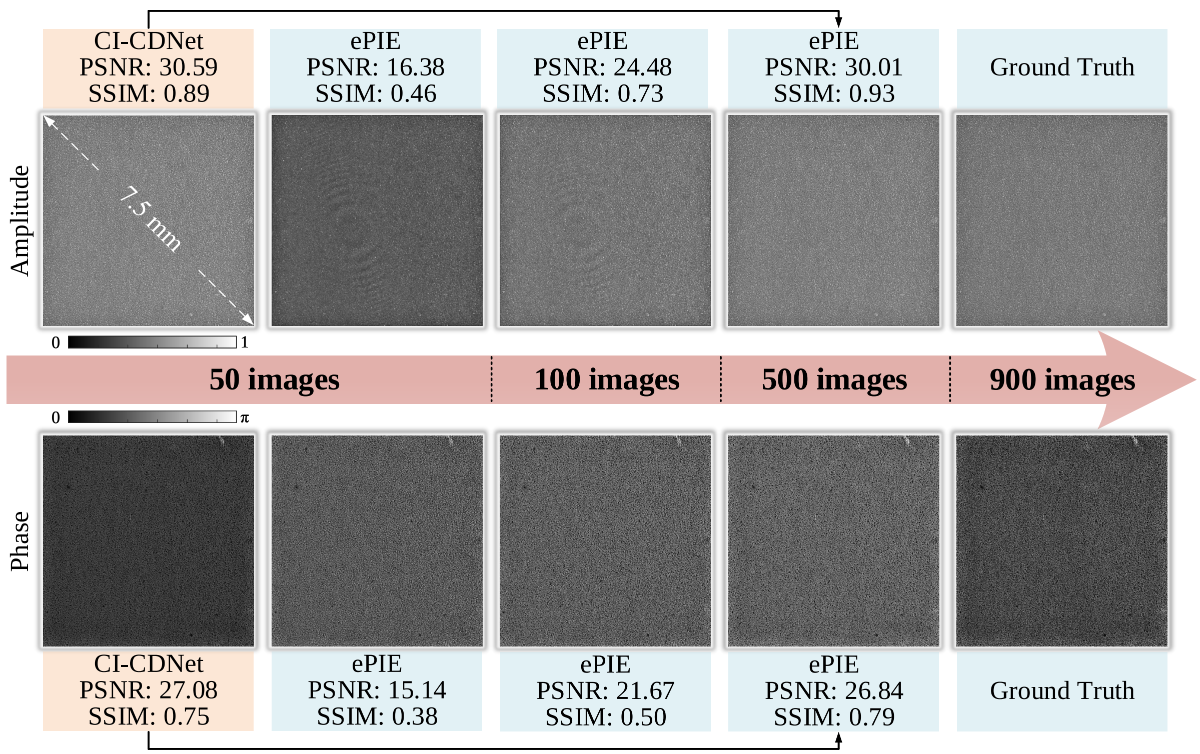

Figure S8: Quantitative results of lensless coded ptychography using different image numbers. The reconstruction results of CI-CDNet using 50 images are close to the results of ePIE using 500 images.

6 Supplementary Note 6: Additional simulation results.

We make a series of simulations to explore the latent coupling information between amplitude and phase and demonstrate the effectiveness of CI-CDNet. We used the DIV2K dataset to compose 450 complex-domain test datasets. These data were resized to 1024×1024 pixels, and added random noise to the real and imaginary parts, respectively. The general complex wavefront can be indicated in Eq. S1.

Simulation results with multi-source noise. In order to explore the latent couping information, we first select three paired data in the 450 test dataset and show their visual comparison in Fig. S10. We can see that there are obvious same features in the amplitude and phase, as shown in the red arrows of the first data, which means they are correlated. The conventional real-domain denoising algorithms (BM3D and Real-NN) can not remove the crosstalk, resulting in unsatisfied denoising performance. Although the phase image of CD-BM3D is better than real-domain techniques, the amplitude has no advantage. In contrast, the reported CI-CDNet is able to take full advantage of the amplitude-phase correlations, obtaining the best performance of both amplitude and phase.

Then, we quantitatively compared different enhancing methods in all 450 test datasets, as shown in Figure S9. Figure S9 (a) is the amplitude results of different methods. Figure S9 (b) is the phase results. Figure S9 (c) is the running times (s). The quantitative results of PSNR and SSIM further validate the satisfactory performance of CI-CDNet. The running efficiency of CI-CDNet also outperforms the competitive algorithms. It only requires less than 10 s to process the 450 test data, which is a thousandth of BM3D.

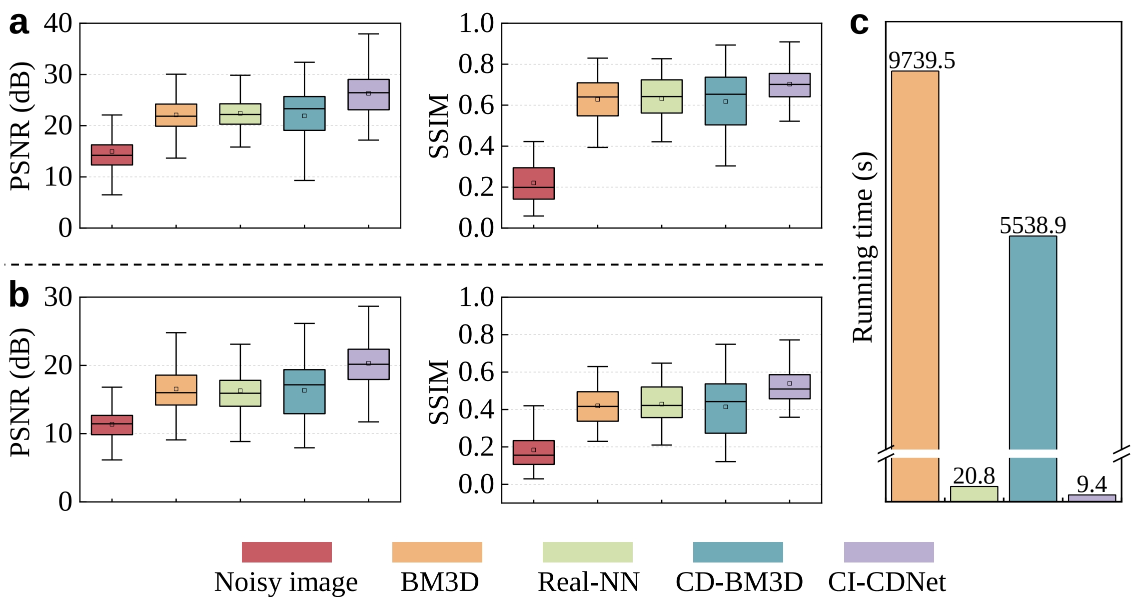

Figure S9: Quantitative results of 450 test dataset which added random multi-source noise. (a) - (b) PSNR and SSIM indexes of amplitude and phase. (c) Running time (s) of different methods.

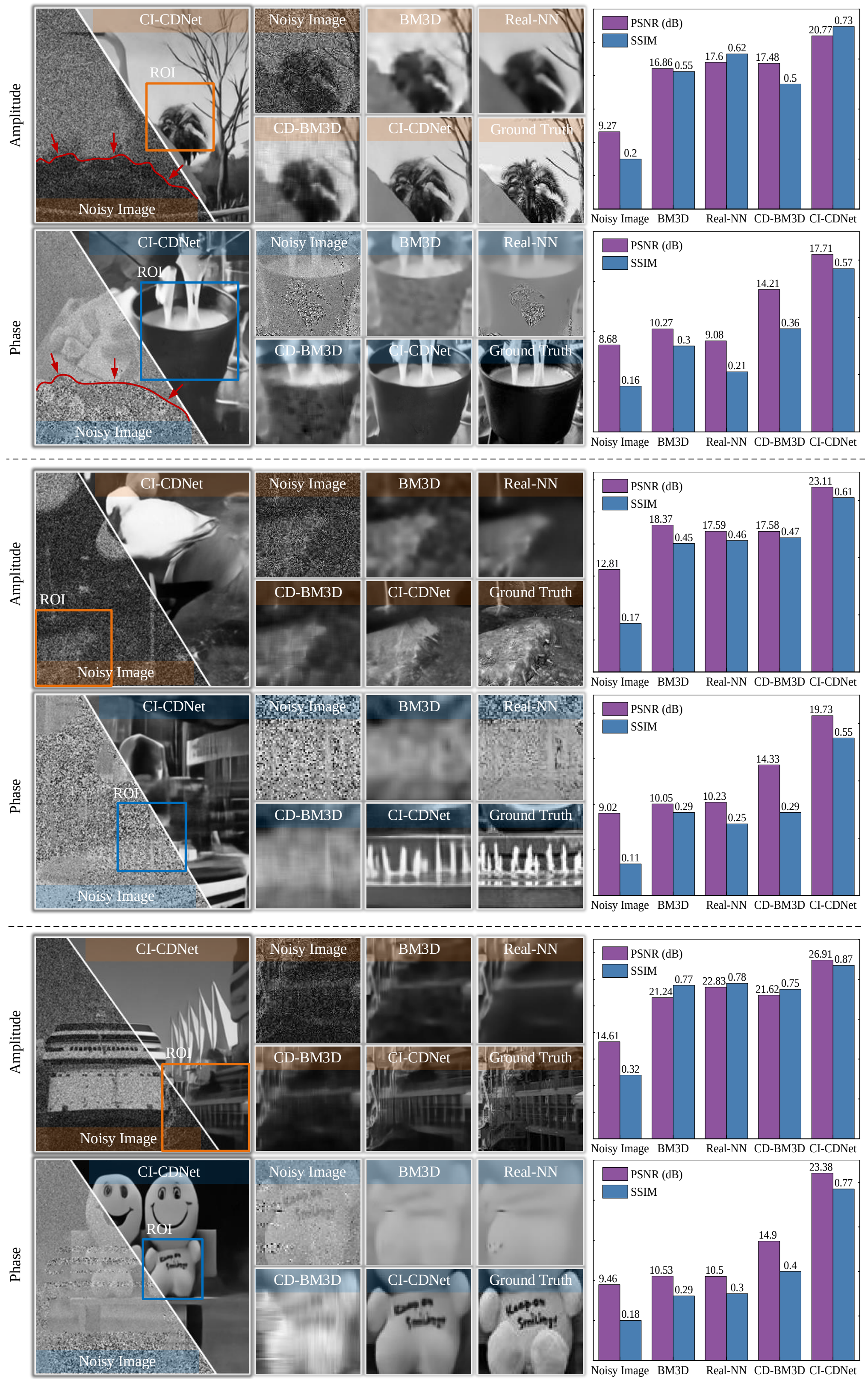

Figure S10: Visual comparison of different methods under random multi-source noise.

Simulation results with Gaussian noise. Then, we added three levels of Gaussian noise to the 450 test dataset which is quantitated by noise variance (30/255, 50/255 and 70/255). Figure S11 presents the average PSNR and SSIM results of different methods using 450 test datasets. Figure S12 and S13 are the visual comparison. We can see that the reported CI-CDNet still outperforms other methods in both noise suppression and detail maintenance.

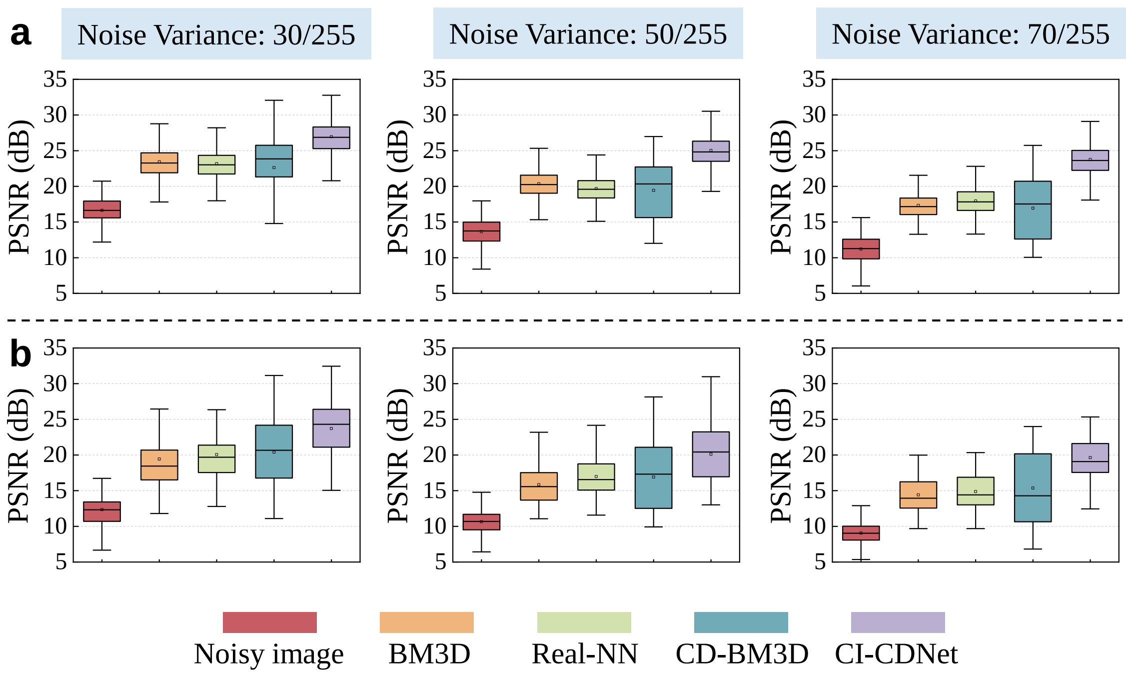

Figure S11: Quantitative results of 450 test dataset which added Gaussian noise (quantified by variance 30/255, 50/255 and 70/255). (a) and (b) are the results of amplitude and phase, respectively.

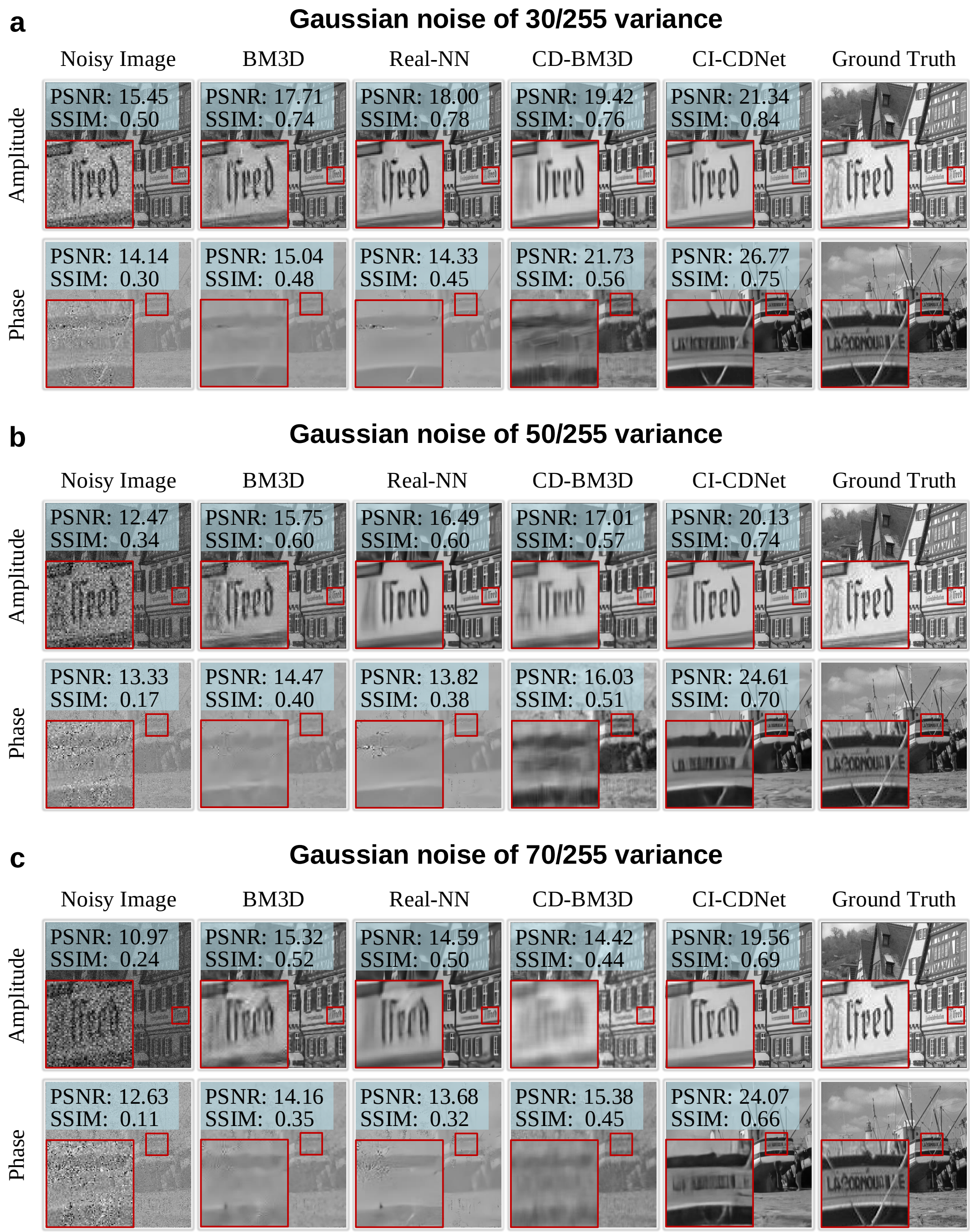

Figure S12: Visual comparison of different methods under different Gaussian noise levels (Group 1).

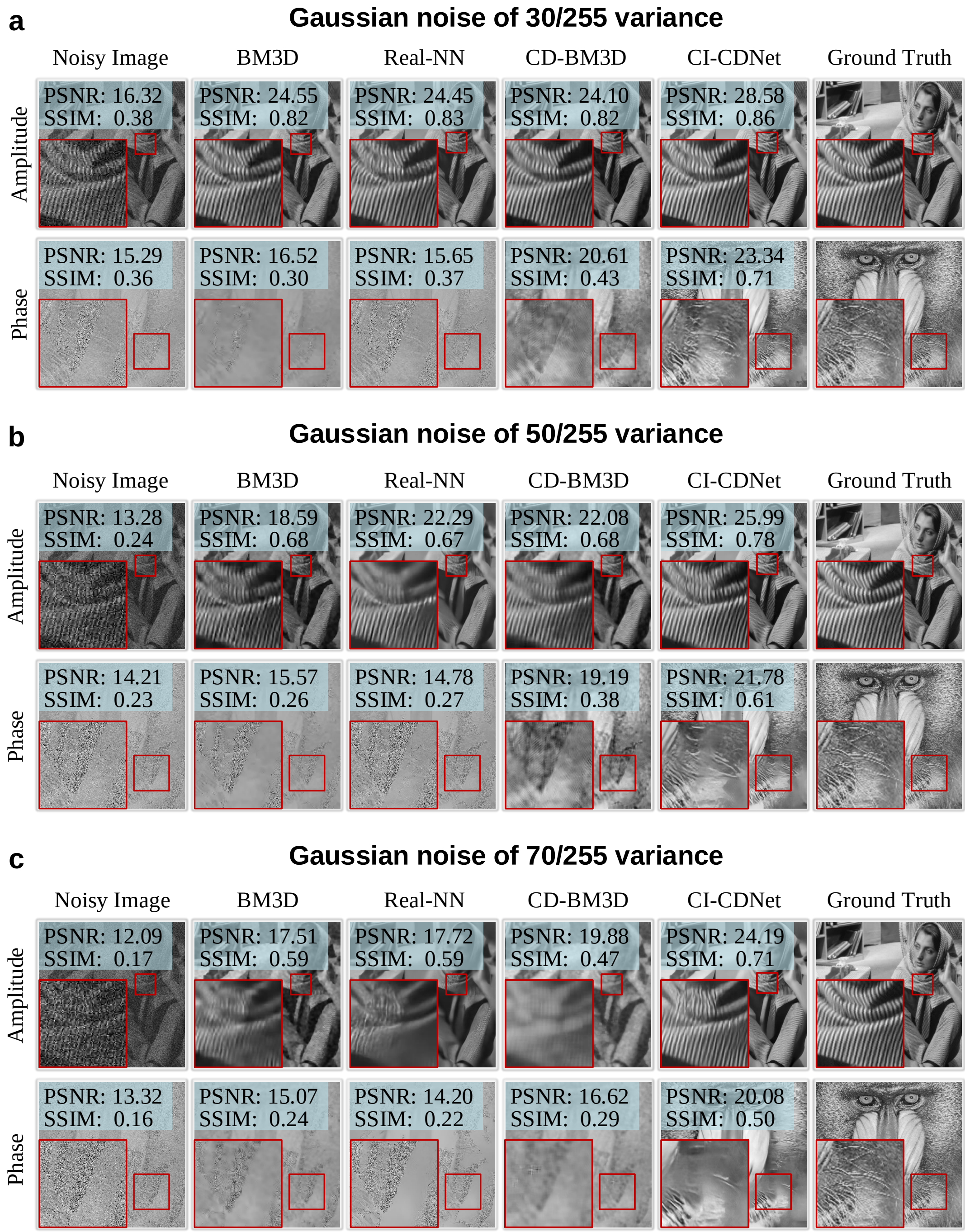

Figure S13: Visual comparison of different methods under different Gaussian noise levels (Group 2).

7 Supplementary Note 7: Additional experiment results.

Additional results of Kramers-Kronig-relations holography (5 ms exposure time).

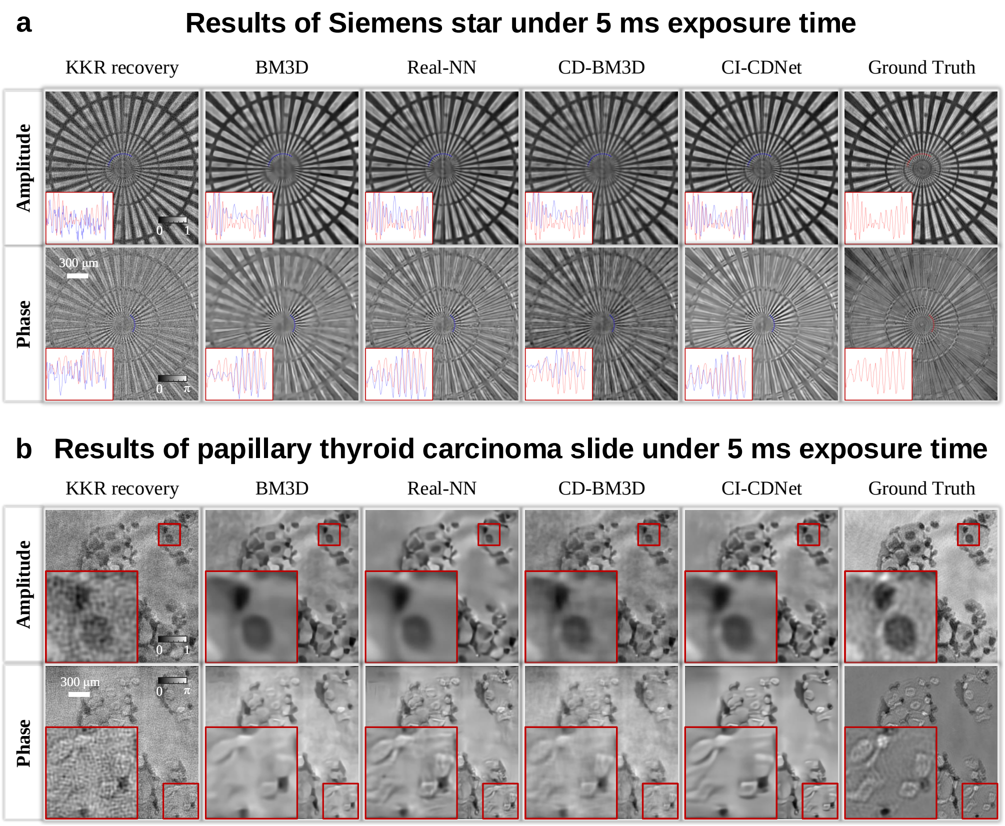

Figure S14: Additional results of Kramers-Kronig-relations holography under 5 ms exposure time. (a) Results of Siemens star. (b) Results of papillary thyroid carcinoma slide.

Additional results of Fourier ptychographic microscopy (0.25 ms exposure time). The FPM reconstruction utilized 225 low-resolution images, and the abundant images provided robustness to fight short exposure time. Thus, the exposure time of 0.25 ms undermined the advantages of CI-CDNet compared with the exposure time of 0.15 ms.

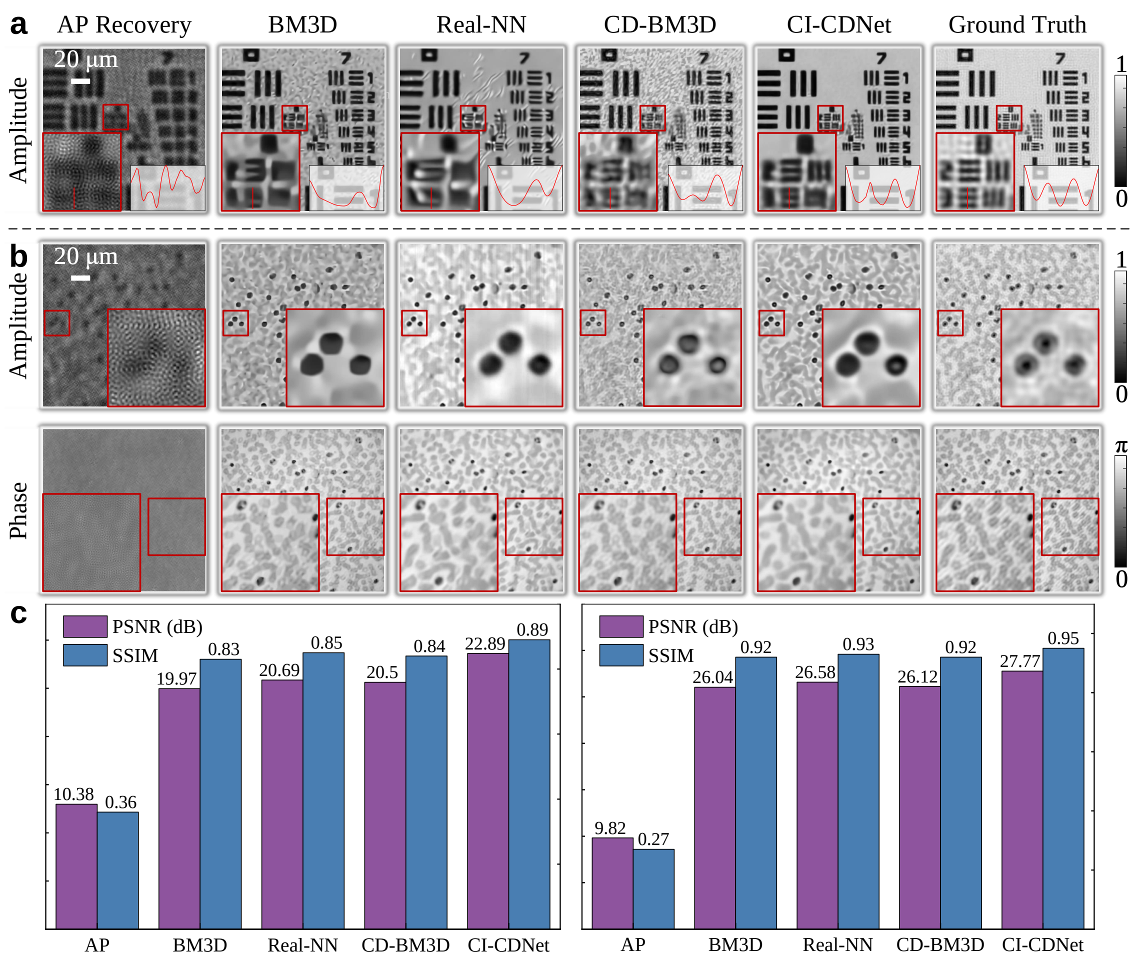

Figure S15: Additional results of Fourier ptychographic microscopy under 0.25 ms exposure time. (a) Amplitude results of the USAF resolution test chart. (b) Amplitude and phase results of blood smear. (c) Qualitative results of blood smear. The left histogram shows the PSNR and SSIM results of amplitude. The right histogram shows the PSNR and SSIM results of the phase.

8 Supplementary Note 8: Comparison between CI-CDNet and dual-channel real-domain neural networks.

We compared the reported CI-CDNet with the dual-channel real-domain neural network. Specifically, the dual-channel real-domain neural network had the same U-net architecture as CI-CDNet. The amplitude and phase images were input to a dual-channel network as two independent channels without any connection. Besides, we further compared CI-CDNet to the dual-channel network with double the parameter to eliminate the effect of parameter numbers.

Figure S16: Comparison to dual-channel neural networks on simulation data.

Figure S16 presents the simulation results using the data of Fig. S10. Figure S17 presents the experiment results using the data of KKR holography under 1 ms exposure time. We can see that the dual-channel network obtains better performance for phase images. However, it has little improvement for amplitude images. Besides, there are no obvious advantages to using double the parameter in the dual-channel network. In contrast, the reported CI-CDNet outperform dual-channel network methods with higher fidelity and resolution in both simulations and experiments. Quantitatively, the PSNR index shows that CI-CDNet has ∼6.2 dB and ∼3.2 dB improvement of amplitude and phase compared to dual-channel networks.

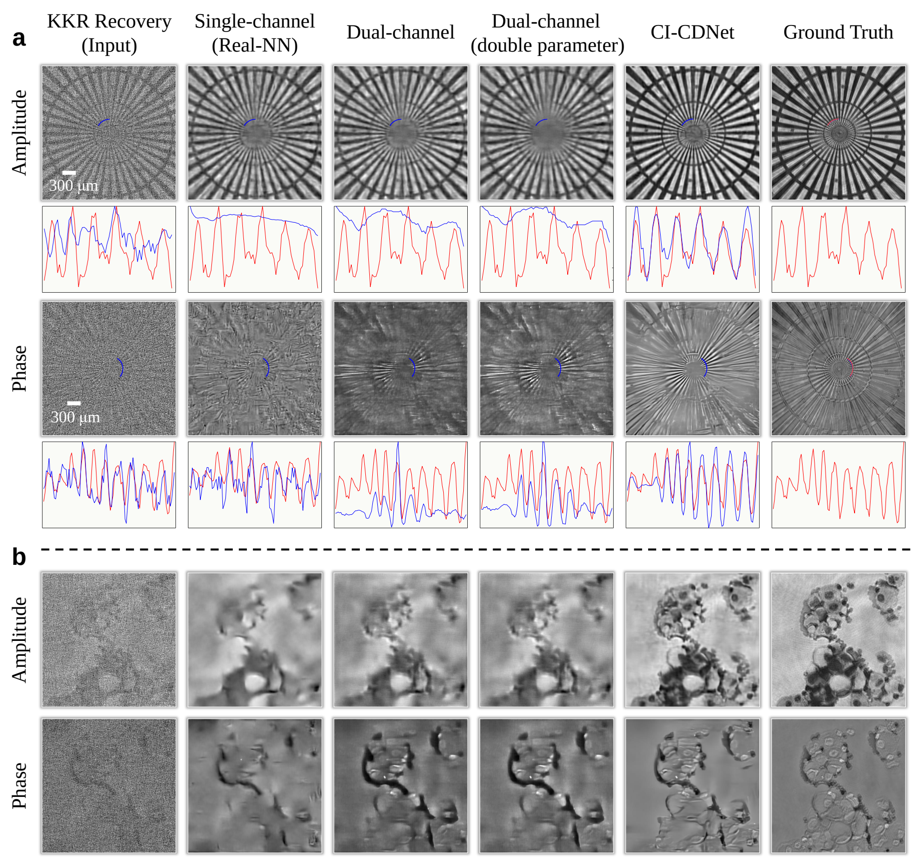

Figure S17: Comparison between CI-CDNet and dual-channel neural networks on experiment data (Kramers-Kronig-relations holography). (a) Results of Siemens star under 1 ms exposure time. (b) Results of papillary thyroid carcinoma slide under 1 ms exposure time.