Supplemental document accompanying submission to *Optica*

Title: Dose-efficient Automatic Differentiation for Ptychographic Reconstruction

Authors: Longlong Wu, Shinjae Yoo, Yong Chu, Xiaojing Huang, Ian Robinson

Submitted: 2/24/2024 5:47:56 PM

Dose-efficient Automatic Differentiation for Ptychographic Reconstruction: Supplemental Document

APPENDIX A: METHODS

1. Simulation of Photon Statistics

The simulation of photon statistics is performed based on the assumption that the ratio between the incoming amount of photons N_{in} and the scattered one N_{sc} equals the ratio between the incoming power W_{in} and the scattered power W_{sc}, given as:

$$\frac{N_{sc}}{N_{in}} = \frac{W_{sc}}{W_{in}}, \tag{S1}$$

The incoming power of an X-ray probe can be calculated as the integral on the incoming plane (*i.e.*, the XOY plane) of its intensity, given as:

$$W_{in} = \iint |P(\mathbf{r})|^2 dxdy\,, \tag{S2}$$

The scattered power at the scanning position $\mathbf{r}_i$ can be obtained through the integral over the scattered intensities $D_i(\mathbf{q})$, which is given as:

$$W_{sc,i} = \iint D_i(\mathbf{q}) dq_x dq_y = \iint |FT[P(\mathbf{r}-\mathbf{r}_i)\cdot O(\mathbf{r})]|^2 dq_x dq_y\,, \tag{S3}$$

Since it can be easily assumed that the incident X-ray probe is the same over all the scanning positions in a ptychographic measurement, the total amount of scattered photons at a scanning position $\mathbf{r}_i$ is given as:

$$N_{sc,\,i} = N_{in}\frac{\iint |FT[P(\mathbf{r}-\mathbf{r}_i)\cdot O(\mathbf{r})]|^2 dq_x dq_y}{\iint |P(\mathbf{r})|^2 dxdy}, \tag{S4}$$

Finally, by assuming that $\frac{1}{A^2}\iint |P(\mathbf{r})|^2 dxdy = \iint |P_N(\mathbf{r})|^2 dxdy = 1$, where A is a normalization coefficient and $P_N(\mathbf{r})$ is the corresponding normalized x-ray probe, the Eq. S4 can be further simplified as:

$$\begin{aligned} N_{sc,\,i} &= N_{in}\iint |FT[P_N(\mathbf{r}-\mathbf{r}_i)\cdot O(\mathbf{r})]|^2 dq_x dq_y \\ &= N_{in}\iint D_{N,i}(\mathbf{q}) dq_x dq_y\,. \end{aligned} \tag{S5}$$

Thus, based on Eq. S5, the amount of scattered photons $N_{sc,\,i}$ at the scanning position $\mathbf{r}_i$ is proportional to the incoming amount of photons N_{in} (*i.e.*, X-ray dose). When an X-ray probe is scanned on a sample, the scattered amount of X-ray photons $N_{sc,\,i}$ at the scanning position $\mathbf{r}_i$ is dependent on the normalized X-ray probe $P_N(\mathbf{r})$ and the scattering property of the sample. Here, it should be mentioned that the scattering efficiency of the sample, defined as the ratio between the scattering cross section and the geometric one, is included in the object $O(\mathbf{r})$ to simplify the derivation.

When simulating the coherent diffraction patterns with photon statistics, the number of the scattered photons $N_{sc,\,i}$ is determined firstly, based on Eq. S5. The normalized $D_{N,i}(\mathbf{q}) = |FT[P_N(\mathbf{r}-\mathbf{r}_i)\cdot O(\mathbf{r})]|^2$ will be treated as the corresponding probability distribution. Then

utilizing the acceptance-rejection method, which is a basic technique used to generate observations from a distribution in numerical analysis and computational statistics, the method works as follows:

1. Randomly sample a reciprocal space point $\mathbf{q} = [q_x, q_y]$ and calculate the corresponding scattering probability $D_{N,i}(\mathbf{q})$=$|FT[P_N(\mathbf{r} - \mathbf{r}_i) \cdot O(\mathbf{r})]|^2$.

2. Simulate a random number X with a uniform distribution on a scale of [0, max($D_{N,i}(\mathbf{q})$)].

3. If X is less than $D_{N,i}(\mathbf{q})$, then "accept" the candidate $\mathbf{q}$ and count one photon. Otherwise, "reject" X and go back to step 1.

4. Repeat the steps from 1 to 3 for a fixed number of times equal to the set amount of the scattered photons $N_{sc,\,i}$.

5. Construct a histogram in the reciprocal space for all the recorded positions $\{\mathbf{q}_1,...,\mathbf{q}_{N_{sc,\,i}}\}$ and the coherent diffraction pattern at the scanning position $\mathbf{r}_i$ will be obtained.

6. Repeat the steps from 1 to 5 for all different scanning positions $\{\mathbf{r}_1, ...,\mathbf{r}_i,..., \mathbf{r}_N\}$ and the corresponding ptychographic dataset with the set photon statistics will be obtained.

2. Maximum Likelihood Estimation (MLE)

The standard picture of photon counting statistics shows that measured pixel counter recordings (or intensity), $I_i(\mathbf{q})$, follow the Poisson probability distribution function (PDF). It leads to the following PDF:

$$f(I|\rho) = \prod_J e^{-D_i(\mathbf{q})} \times \frac{D_i(\mathbf{q})^{I_i(\mathbf{q})}}{I_i(\mathbf{q})!}, \tag{S6}$$

The corresponding negative log-likelihood estimation is given as,

$$\ell_{\mathcal{P},i}(\mathbf{q}) = \frac{1}{J}\sum_J D_i(\mathbf{q}) - I_i(\mathbf{q})\log[D_i(\mathbf{q})] + \log[I_i(\mathbf{q})!]\,, \tag{S7}$$

As the last part on the right of the above equation is a constant, to make the negative log-likelihood estimation nonnegative for the Poisson distribution and fit the variable-size mini-batch for our DAP algorithm, Eq. S7 is further modified as:

$$\begin{aligned}\ell_{\mathcal{P},i}(\mathbf{q}) &= \frac{1}{J}\sum_J D_i(\mathbf{q}) - I_i(\mathbf{q})\log[D_i(\mathbf{q})] - I_i(\mathbf{q}) + I_i\log[I_i(\mathbf{q})] \\ &= \frac{1}{J}\sum_J D_i(\mathbf{q}) - I_i(\mathbf{q}) + I_i(\mathbf{q})\log\left[\frac{I_i(\mathbf{q})}{D_i(\mathbf{q})}\right],\end{aligned} \tag{S8}$$

where the last two parts are additive constants to the estimation nonnegative.

Further, when dealing with counting statistics, if the data are corrupted by an additive Gaussian (thermal) noise to the square root (or amplitude) of an expected intensity[1, 2], by considering an Gaussian asymptotic form of the noise model, the PDF, in this case, can be given as:

$$f(I|\rho) = \prod_J (2\pi\sigma^2)^{-1/2}\exp\left[-\frac{1}{2}\left(\frac{I_i^{1/2}(\boldsymbol{q}) - D_i^{1/2}(\boldsymbol{q})}{\sigma}\right)^2\right], \tag{S9}$$

where σ^2 is the constant variance for the Gaussian noise. Thus, through negative log-likelihood estimation, its averaged estimation is given as:

$$\ell_{\mathcal{G},i}(\boldsymbol{q}) = \frac{1}{J}\sum_{J}\left[I_i^{1/2}(\boldsymbol{q}) - D_i^{1/2}(\boldsymbol{q})\right]^2, \tag{S10}$$

Additionally, if the Gaussian noise is additive to the expected intensity directly with its corresponding variance approximated by the measured intensity, the PDF in this situation can be expressed as[2, 3]:

$$f(I|\rho) = \prod_{J}\prod [2\pi I_i(\mathbf{q})]^{-1/2}\exp\left[-\frac{1}{2}\left(\frac{I_i(\mathbf{q}) - D_i(\mathbf{q})}{I_i^{1/2}(\mathbf{q})}\right)^2\right], \tag{S11}$$

Applying the negative log-likelihood estimation to Eq. S11 and averaging over all the pixels in a diffraction pattern leads to:

$$\ell_{\mathcal{R},i}(\boldsymbol{q}) = \frac{1}{2J}\sum\nolimits_{I_i(\mathbf{q})\neq 0}\left[\frac{I_i(\boldsymbol{q}) - D_i(\boldsymbol{q})}{I_i^{1/2}(\boldsymbol{q})}\right]^2 + \frac{1}{J}\sum\nolimits_{I_i(\mathbf{q})=0} D_i(\boldsymbol{q}), \tag{S12}$$

where the last part in the above equation is added to constrain the zero intensity for reducing noisy solutions.

3. Maximum Likelihood Estimation for Mixed State Ptychographic Reconstruction

Assuming that the physical object and X-ray probe can be effectively modeled by M independent object states with $O^{(m)} = \{O^{(1)}, O^{(2)}, ..., O^{(m)}\}$ and N independent probe states with $P^{(n)} = \{P^{(1)}, P^{(2)}, ..., P^{(n)}\}$, the existing wave for the *m*-th object and *n*-th probe at the scanning position $\mathbf{r}_i$ can be expressed as:

$$\psi_i^{(m,n)}(\mathbf{r}) = P^{(n)}(\mathbf{r} - \mathbf{r}_i)\cdot O^{(m)}(\mathbf{r}). \tag{S13}$$

Therefore, the expected coherent X-ray diffraction intensity on the detector can be estimated:

$$D_i^{(m,n)}(\mathbf{q}) = \left|FT\left[\psi_i^{(m,n)}(\mathbf{r})\right]\right|^2. \tag{S14}$$

Thus, the corresponding negative log-likelihood estimation for this state is given as $\ell_{\alpha,i}\left[D_i^{(m,n)}, I_i^{(m,n)}\right]$, where $I_i^{(m,n)}$ is the corresponding measured coherent X-ray diffraction intensity. Here, the subscript α stands for $\mathcal{P}$, $\mathcal{G}$, and $\mathcal{R}$ (*i.e.*, different statistical models under consideration). Since all the different states are independent from each other [4, 5], we can obtain that $I_i^{(m,n)} \propto D_i^{(m,n)}$ and $I_i \propto D_i = \sum_{m,n} D_i^{(m,n)}$. Thus, the experimental coherent X-ray intensity $I_i^{(m,n)}$ for this state can be estimated as $I_i^{(m,n)} = I_i\frac{D_i^{(m,n)}}{\sum_{m,n} D_i^{(m,n)}}$. Consequently, for the mixed states ptychographic reconstruction at the scanning position $\mathbf{r}_i$, the negative log-likelihood estimation is a sum over all different states, given as:

$$\sum_{m=1}^{\mathrm{M}}\sum_{n=1}^{\mathrm{N}} \ell_{\alpha,i}\left[D_i^{(m,n)}, I_i^{(m,n)}\right]. \tag{S15}$$

Finally, by further considering the ATV constraint in a mini-batch, it leads to:

$$\mathcal{L}_{\alpha,\,l} = \frac{1}{LMN}\sum_{i\in\Omega}\sum_{m=1}^{M}\sum_{n=1}^{N}\ell_{\alpha,\,i}\left[D_i^{(m,n)},I_i^{(m,n)}\right] + \frac{\gamma}{M}\sum_{m=1}^{M}ATV(O_m), \tag{S16}$$

where Ω contains the indices of coherent diffraction patterns in a mini-batch with the size of L.

4. Details of DAP Algorithm

Algorithm 1: DAP algorithm

Input: N measurements $\{I_i\}_{i=1}^{N}$, number of epochs *K,* number of mix states M and N for the probe and object

Initialize: O^0, P^0

for $k = 1,\ldots,K$ **do**

if update *P* **is** True **then**

P.require_grad = True

else

P.require_grad = False

adjust the size *b and* random permutate the index in $\{l\}$.

for $l_b = l_1,\ldots,l_b$ **do**

for i **in** l_b **do**

calculate $D_i^{(m,n)} \leftarrow \left|FT\left[\psi_i^{(m,n)}(\mathbf{r})\right]\right|^2$

end for

if switch_noise_model **then**

$\ell_{\alpha,\,i} = \ell_{\mathcal{G},i}, \ell_{\mathcal{P},i}$ *or* $\ell_{\mathcal{R},i}(\boldsymbol{q})$

$\mathcal{L}_{\alpha,\,l_b} = \frac{1}{LMN}\sum_{i\in\Omega}\sum_{m=1}^{\mathrm{M}}\sum_{n=1}^{\mathrm{N}}\ell_{\alpha,\,i}\left(D_i^{(m,n)},I_i^{(m,n)}\right) + \frac{\gamma}{M}\sum_{m=1}^{\mathrm{M}}ATV(O_m)$

Calculate gradients with respect to all optimizable parameters.

Update all the optimizable parameters by using Adam optimizer.

end for

with no_grad

Apply Constraints:

Recenter O^k, P^k

Rescale O^k, P^k

Amplitude constraint O^k, P^k

Phase constraint O^k, P^k

end for

Output: O^k, P^k

APPENDIX B: SUPPLEMENTAL FIGURES

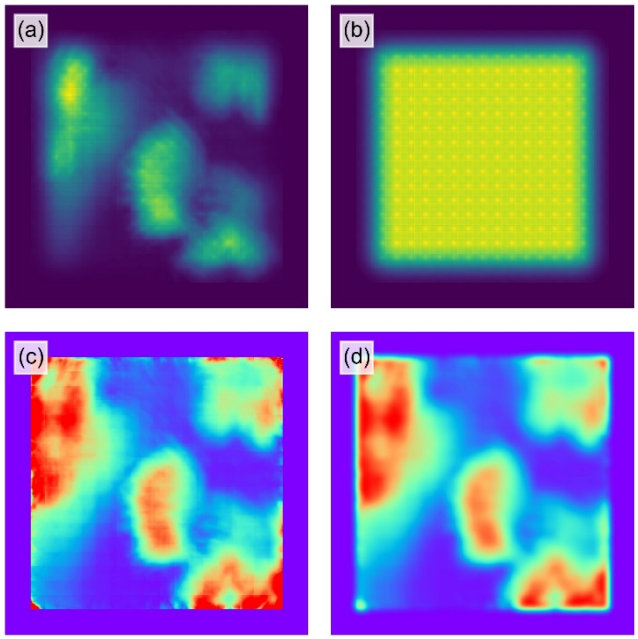

Fig. S1 Example of Autocorrelation initialization to obtain the initial object. (a) Calculated $\sum_{i=1}^{N}\varphi_i(\mathbf{r}+\mathbf{r}_i)$, where $\varphi_i(\mathbf{r}) = |FT^{-1}(I_i)|$. (b) Calculated $\sum_{i=1}^{N}\Phi(\mathbf{r}+\mathbf{r}_i)$, where $\Phi(\mathbf{r}) = \left|FT^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} I_i\right)\right|$. (c) Calculated $\frac{\sum_{i=1}^{N}\varphi_i(\mathbf{r}+\mathbf{r}_i)}{\sum_{i=1}^{N}\Phi(\mathbf{r}+\mathbf{r}_i)}$. (d) obtained $g(\mathbf{r})\otimes\frac{\sum_{i=1}^{N}\varphi_i(\mathbf{r}+\mathbf{r}_i)}{\sum_{i=1}^{N}\Phi(\mathbf{r}+\mathbf{r}_i)}$.

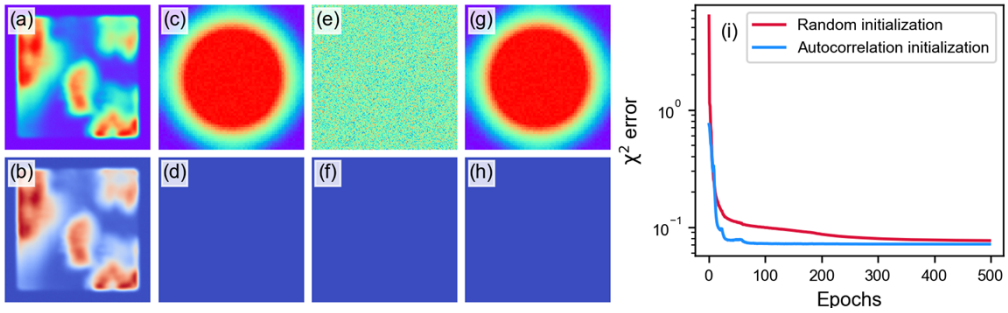

Fig. S2 Performance of the DAP method when using different initialization. (a) Initialized amplitude of the object using the autocorrelation initialization. (b) Corresponding estimated phase. (c). Initialized amplitude of the x-ray probe using the autocorrelation method. (e-f) Random initialized amplitude and phase of the object. (g-h) Initialized amplitude of the x-ray probe with a flat phase using the autocorrelation method. (i) Performance of the DAP when using different initialization. For the blue line, the (a-d) are used. For the red line, (e-h) are used.

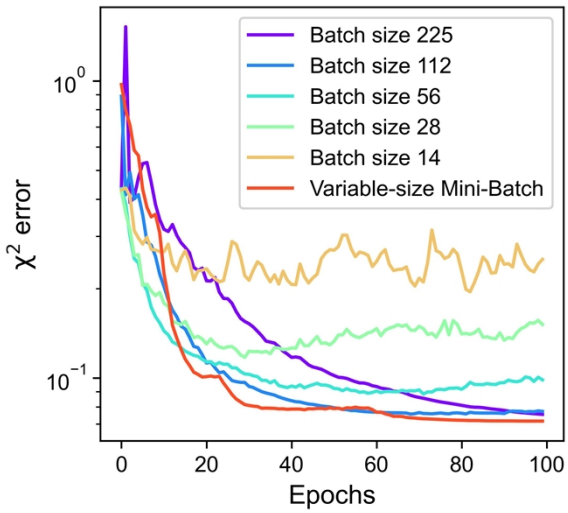

Fig. S3 Effect of batch size on the performance of the AD method. Here, the corresponding ptychographic data comes from Fig. 3b of the main paper.

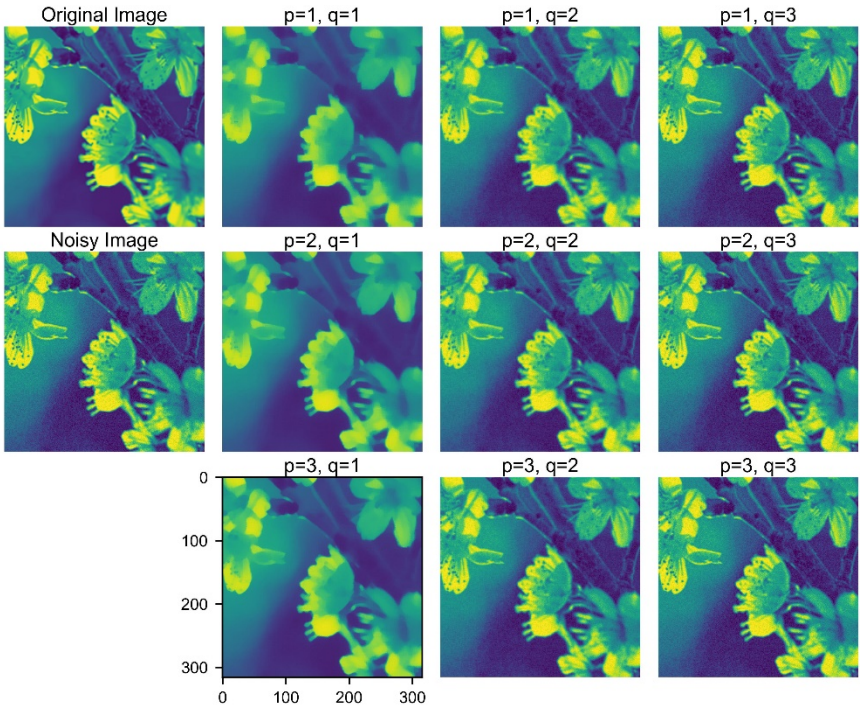

Fig. S4 Effect of p and q on the reconstructed image. To demonstrate the effect of p and q on the reconstructed image, we have performed a minimization of the objective function, defined as $\min_X \|X - Y\|^2 + \gamma ATV(X)$. Here, X is the image needed to be optimized, and Y is the observed noisy image. γ is a weight parameter and the ATV represents the adaptive L_p-norm based TV, defined by Eq. 6 in the main paper. The γ is set to 0.6 for all minimizations.

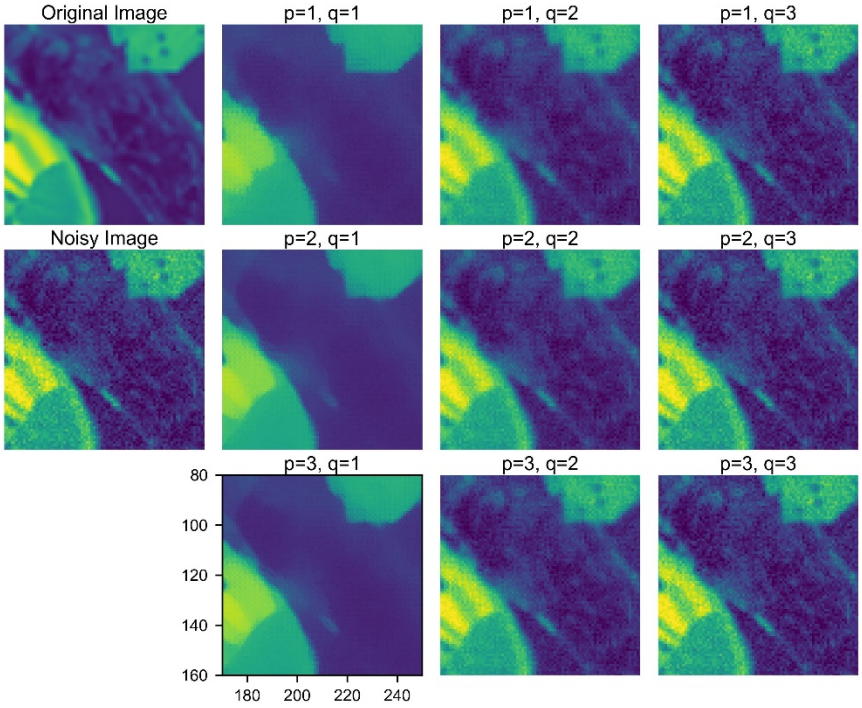

Fig. S5 Enlarged views of Fig. S4

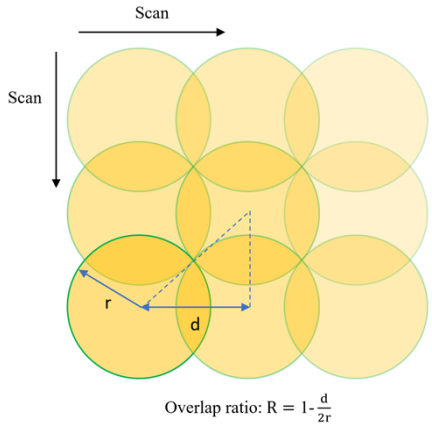

Fig. S6. Definition of the overlap ratio for a grid mesh scan. Here, d is the scan step size and r is the radius of the X-ray probe. As shown, the minimum overlap ratio for a circular probe with a grid mesh scan is $R = 1 - \frac{2r\frac{1}{\sqrt{2}}}{2r} \approx 29.3\%$.

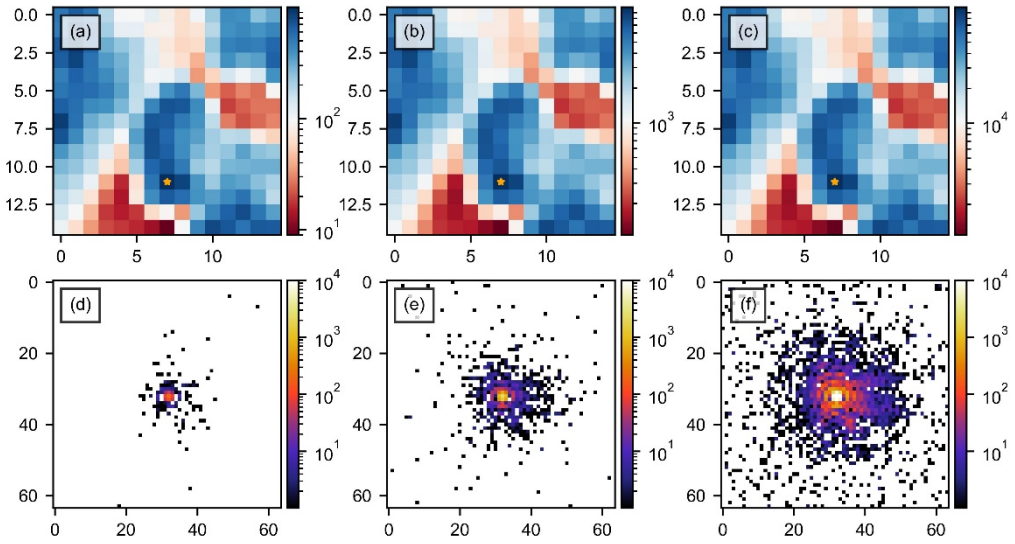

Fig. S7. Simulated ptychographic datasets with different photon statistics. (a) With the max number of photons equal to 10^3. (b) Corresponding diffraction pattern with its position marked by the orange star in (a). (b) With the max number of photons equal to 10^4. (c) Corresponding diffraction pattern with its position marked by the orange star in (b). (c) With the max number of photons equal to 10^5. (d) Corresponding diffraction pattern with its position marked by the orange star in (c).

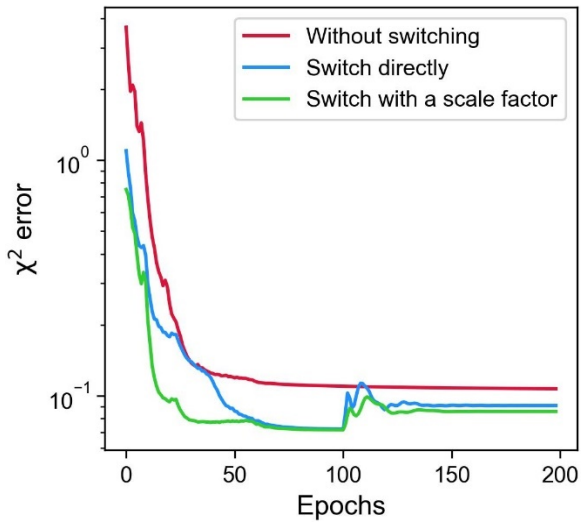

Fig. S8 The effect of the optimization strategy on the convergence of the algorithm. The red line was obtained by using the Poisson MLE. The blue line was obtained by switching the loss function from Gaussian amplitude MLE to Poisson MLE without using the scale factor. The green line was obtained by using the proposed approach, *i.e.*, when switching, a scale factor of 2 was applied to the loss function. Here, the corresponding ptychographic data comes from Fig. 3b.

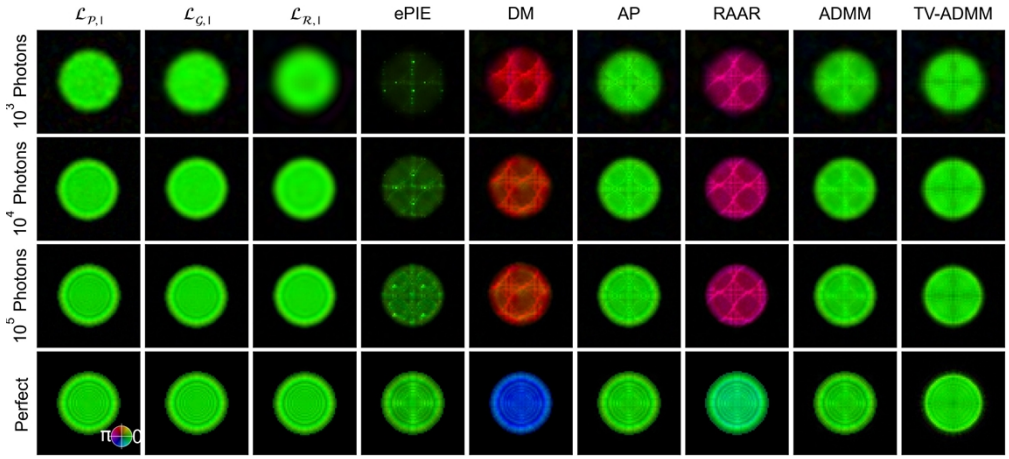

Fig. S9 Reconstructed X-ray probe with different photons statistics using simulated datasets. The corresponding overlap ratio is 50%.

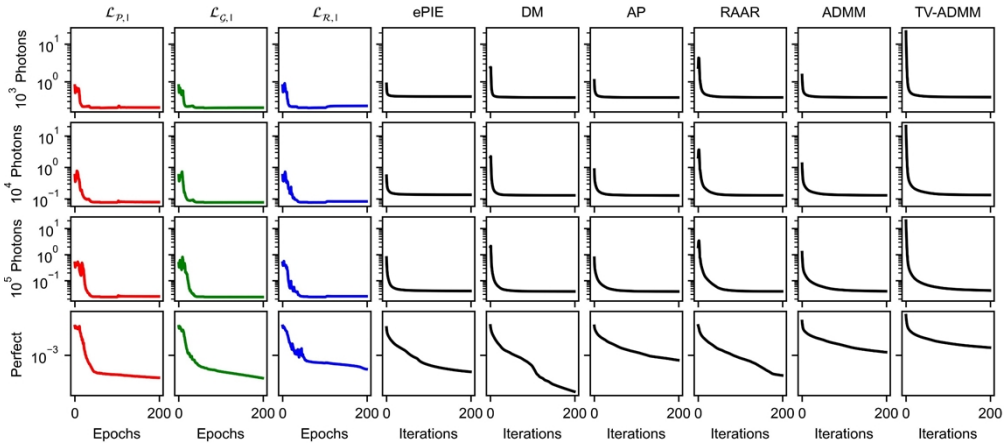

Fig. S10 Calculated χ^2 error for different approaches on simulated ptychographic dataset with different photons statistics. The corresponding overlap ratio is 50%.

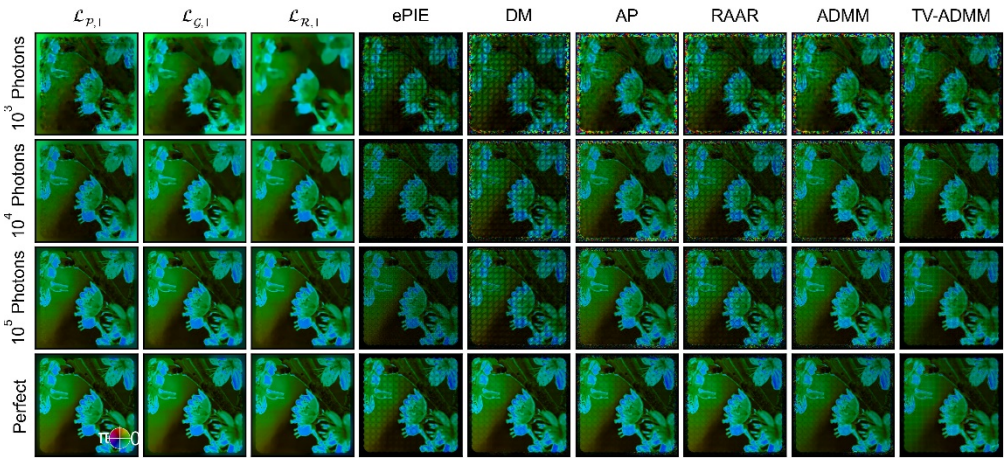

Fig. S11 Results of recovered images by DAP with the three different noise models, extended Ptychographic Iterative Engine (ePIE), Difference Map (DM), Alternating Projection (AP), Relaxed Averaged Alternating Reflections (RAAR), Alternating Direction Method of Multiplier (ADMM)[6], and ADMM with total variation constraint (TV-ADMM)[7] from left to right, where different photon statistics are also applied. The scanning overlap ratio is 50% for all images. Here, for each ptychography dataset, there is a total of 225 diffraction patterns in a 15×15 array.

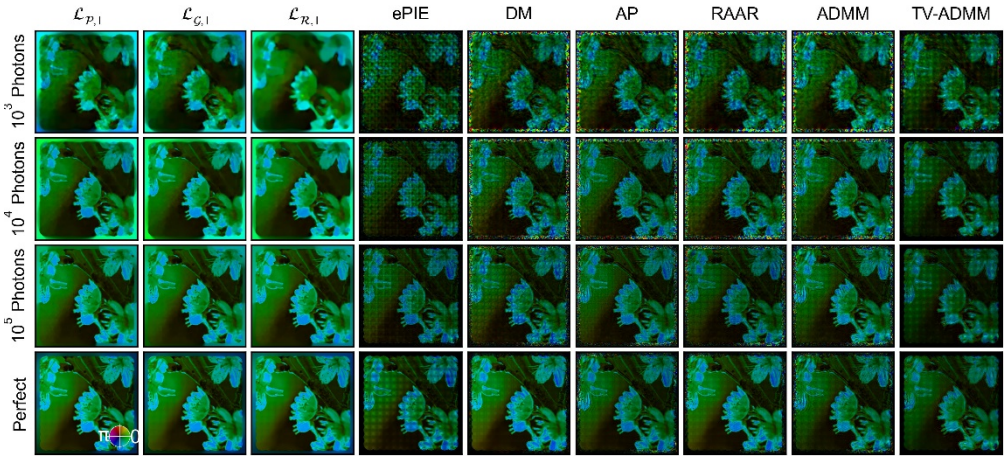

Fig. S12 Results of recovered images by DAP with the three different noise models, ePIE, DM, AP, RAAR, and ADMM, from left to right, where different photon statistics are also applied. The scanning overlap ratio is 40% for all images. Here, for each ptychography dataset, there is a total of 169 diffraction patterns in a 13×13 array.

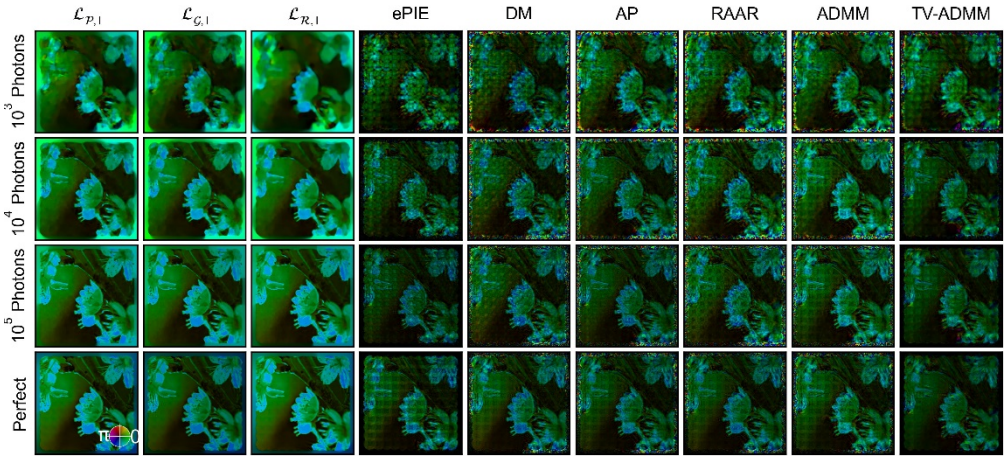

Fig. S13 Results of recovered images by DAP with the three different noise models, ePIE, DM, AP, RAAR, and ADMM, from left to right, where different photon statistics are also applied. The scanning overlap ratio is 30% for all images. Here, for each ptychography dataset, there is a total of 121 diffraction patterns in a 11×11 array.

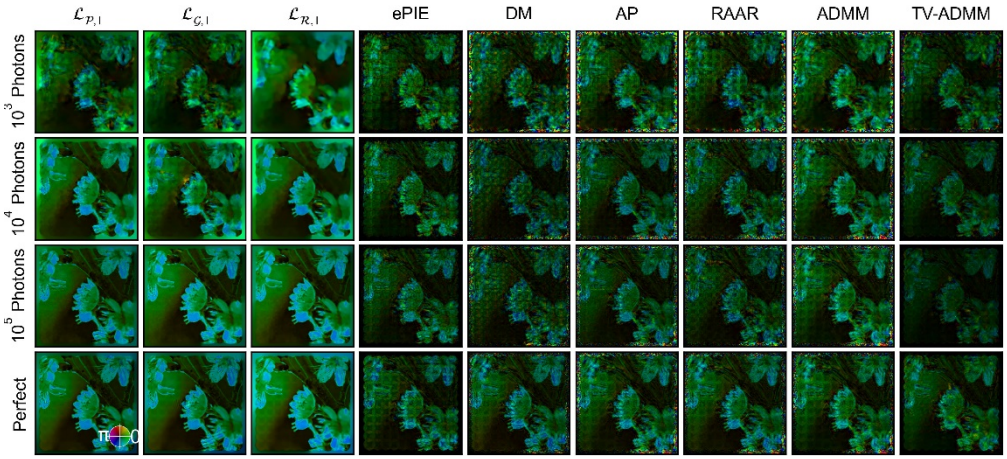

Fig. S14 Results of recovered images by DAP with the three different noise models, ePIE, DM, AP, RAAR, and ADMM, from left to right, where different photon statistics are also applied. The scanning overlap ratio is 25% for all images. Here, for each ptychography dataset, there is a total of 121 diffraction patterns in a 11×11 array , where the scanning region is slightly larger.

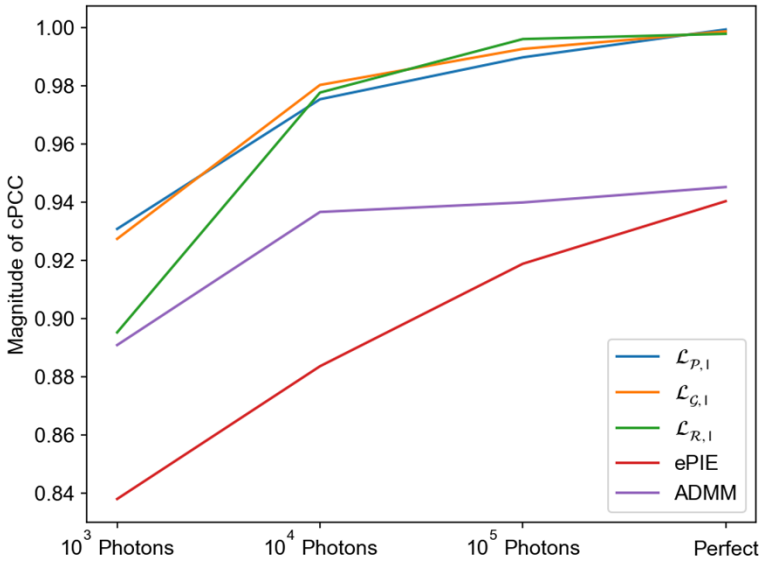

Fig. S15 Magnitude of cPCC of the reconstructed objects in Fig. 3 of the main paper by comparing with its original image.

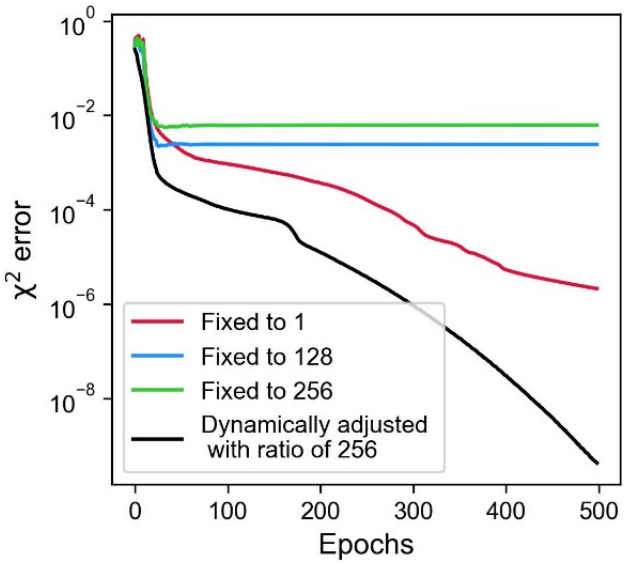

Fig. S16 Effect of the tuning parameter γ on the performance of DAP. When the γ is set to 1, thought the DAP can converge, it has negligible effect on the periodic artifact. When the γ is set to 256, it will prevent the convergence of the DAP algorithm, resulting in a low-quality image. Finally, when γ is set to keep the ratio between MLE and ATV to 256 during the reconstruction, the best error was obtained while the grid artifact is removed. Here, we show this effect by using the data from Fig. 3d in the main paper.

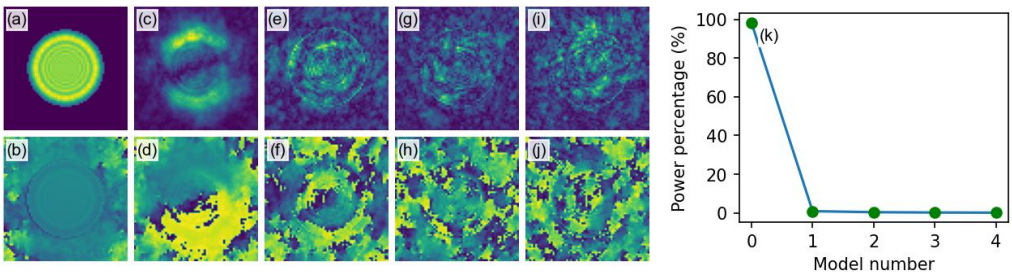

Fig. S17 Application of the mixed states ptychographic reconstruction (*i.e.*, using Eq. 6) for simulated data with five modes. Here, the ptychographic data comes from Fig. 3c. (a-b) Amplitude and phase of the first mode. (c-d) Amplitude and phase of the second mode. (e-f) Amplitude and phase of the third mode. (g-h) Amplitude and phase of the fourth mode. (i-j) Amplitude and phase of the fifth mode. (k) The power percentage of the recovered modes.

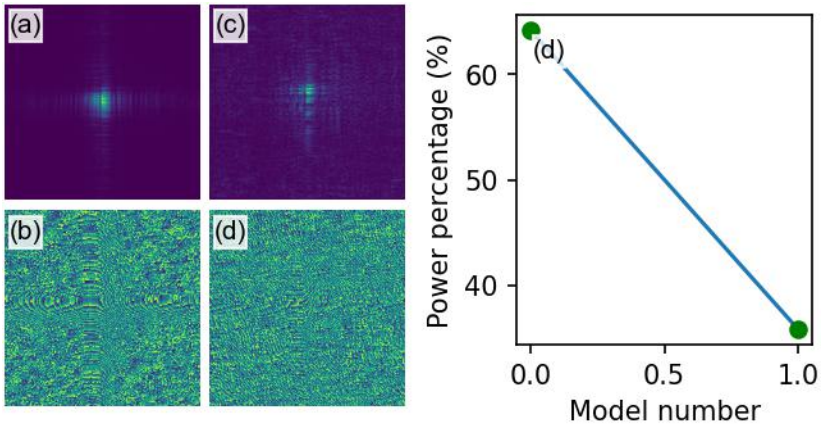

Fig. S18 Recovered modes from the experimental data in Fig. 4a. The data were obtained from an MLLs focusing system. (a-b) Amplitude and phase of the first mode. (c-d) Amplitude and phase of the second mode. (e) The power percentage of the recovered modes.

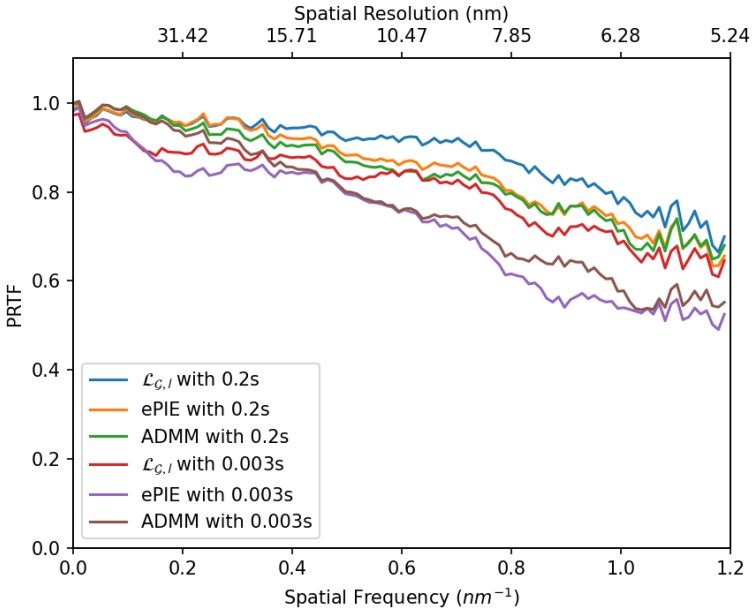

Fig. S19 Estimated phase retrieval transfer functions (PRTFs) after azimuthal averaging[8], which are obtained with the results from different algorithms shown in Fig. 4 using the two ptychographic datasets.

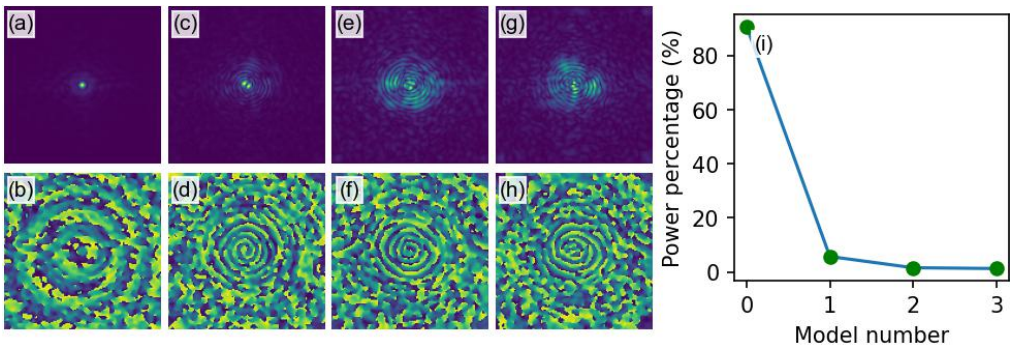

Fig. S20 Recovered modes from the experimental data in Fig. 5a. The data were obtained from an FZP focusing system. (a-b) Amplitude and phase of the first mode. (c-d) Amplitude and phase of the second mode. (e-f) Amplitude and phase of the third mode. (g-h) Amplitude and phase of the fourth mode. (i) The power percentage of the recovered modes.

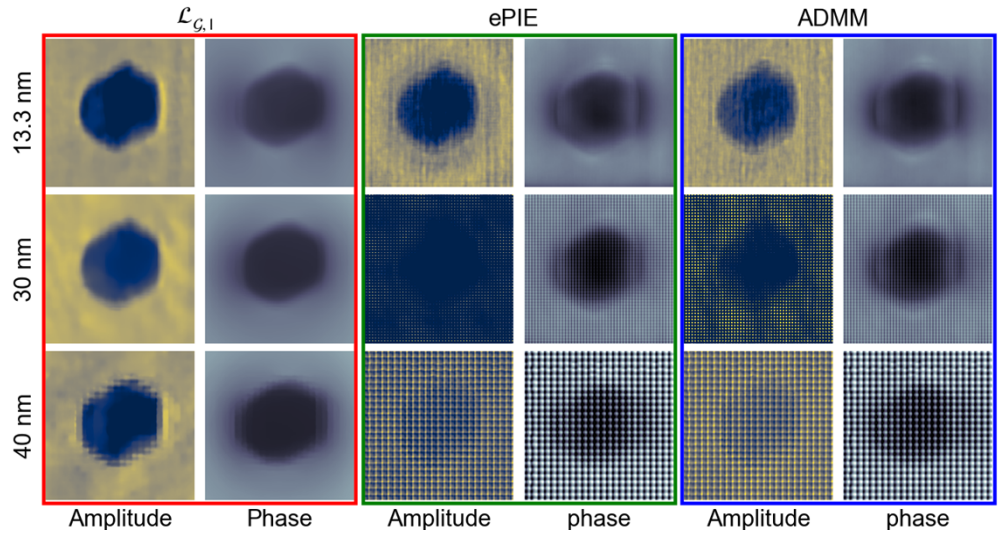

Fig. S21 Ptychography imaging of a gold nanoparticle measured in a raster ptychographic scan with different scanning step sizes. Here the scanning step sizes are 13.3 nm, 30 nm and 40 nm, respectively. As marked separately by the red, green, and blue box, the DAP with $\mathcal{L}_{\mathcal{G},l}$ loss function, ePIE and ADMM are used to compare the performance of the ptychographic reconstructions.

References

1. M. F. Freeman, and J. W. Tukey, "Transformations related to the angular and the square root," The Annals of Mathematical Statistics, 607-611 (1950).
2. P. Godard, M. Allain, V. Chamard, and J. Rodenburg, "Noise models for low counting rate coherent diffraction imaging," Opt. Express **20**, 25914-25934 (2012).
3. C. A. Bouman, and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," IEEE Trans Image Process **5**, 480-492 (1996).
4. P. Li, T. Edo, D. Batey, J. Rodenburg, and A. Maiden, "Breaking ambiguities in mixed state ptychography," Opt. Express **24**, 9038-9052 (2016).
5. P. Thibault, and A. Menzel, "Reconstructing state mixtures from diffraction measurements," Nature **494**, 68-71 (2013).
6. H. Yan, "Ptychographic phase retrieval by proximal algorithms," New Journal of Physics **22** (2020).
7. H. Chang, Y. Lou, Y. Duan, and S. Marchesini, "Total Variation--Based Phase Retrieval for Poisson Noise Removal," SIAM Journal on Imaging Sciences **11**, 24-55 (2018).
8. R. N. Wilke, M. Vassholz, and T. Salditt, "Semi-transparent central stop in high-resolution X-ray ptychography using Kirkpatrick–Baez focusing," Acta Crystallographica Section A Foundations of Crystallography **69**, 490-497 (2013).