RESEARCH ARTICLE

Three-dimensional nanoscale reduced-angle ptycho-tomographic imaging with deep learning (RAPID)

Ziling Wu[1,5*†] , Iksung Kang[2,6†], Yudong Yao[3], Yi Jiang[3], Junjing Deng[3], Jeffrey Klug[3], Stefan Vogt[3] and George Barbastathis[1,4*]

Abstract

X-ray ptychographic tomography is a nondestructive method for three dimensional (3D) imaging with nanometer-sized resolvable features. The size of the volume that can be imaged is almost arbitrary, limited only by the penetration depth and the available scanning time. Here we present a method that rapidly accelerates the imaging operation over a given volume through acquiring a limited set of data via large angular reduction and compensating for the resulting ill-posedness through deeply learned priors. The proposed 3D reconstruction method "RAPID" relies initially on a subset of the object measured with the nominal number of required illumination angles and treats the reconstructions from the conventional two-step approach as ground truth. It is then trained to reproduce equal fidelity from much fewer angles. After training, it performs with similar fidelity on the hitherto unexamined portions of the object, previously not shown during training, with a limited set of acquisitions. In our experimental demonstration, the nominal number of angles was 349 and the reduced number of angles was 21, resulting in a $\times 140$ aggregate speedup over a volume of $4.48 \times 93.18 \times 3.92\,\mu m^3$ and with $(14\,nm)^3$ feature size, i.e. $\sim 10^8$ voxels. RAPID's key distinguishing feature over earlier attempts is the incorporation of atrous spatial pyramid pooling modules into the deep neural network framework in an anisotropic way. We found that adjusting the atrous rate improves reconstruction fidelity because it expands the convolutional kernels' range to match the physics of multi-slice ptychography without significantly increasing the number of parameters.

Keywords X-ray ptychographic tomography, Deep learning, Reduced-angle, Rapid imaging, Atrous spatial pyramid pooling, Anisotropic

[†]Ziling Wu and Iksung Kang have equal contribution

*Correspondence:
Ziling Wu
zilingwu@mit.edu
George Barbastathis
gbarb@mit.edu
[1] Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2] Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[3] Argonne National Laboratory, Lemont, IL 60439, USA
[4] Singapore-MIT Alliance for Research and Technology (SMART) Centre, Singapore 138602, Singapore
[5] Present Address: Singapore-MIT Alliance for Research and Technology (SMART) Centre, Singapore 138602, Singapore
[6] Present Address: University of California, Berkeley, CA 94720, USA

1 Introduction

Three-dimensional (3D) imaging at the nanometer scale enables important insights in biology and material behaviors, including virus function [1], structural damage [2], nanoelectronics [3], etc. One way is to do this destructively, i.e. immobilize the specimen, etch the top layer finely with a particle beam, image the revealed features with a scanning electron microscope or similar high-resolution methods, and repeat this process until the entire specimen volume has been consumed [4, 5]. However, in many instances, it is preferable to operate non-destructively, and then a form of tomography is necessary. This is more challenging than, say, the medical case, for two main reasons: (i) as feature sizes approach the radiation wavelength, e.g. X-rays, diffraction and scattering effects influence image fidelity more strongly; and (ii) the number of voxels (resolvable 3D elements) within a macroscopic volume can become very large. For example, at $(10\,nm)^3$ 3D sampling rate, the number of voxels in a $1\,cm \times 1\,cm \times 100\,\mu m$ volume is $\sim 10^{16}$. In this paper, we use integrated circuits (IC) as an exemplar, because they present some practical conveniences—ICs are rigid and, thus, require no fixing—and it is also very useful, for example in manufacturing process verification, failure analysis and counterfeits detection [3, 6]. On the other hand, the challenge of 3D IC imaging grows with time due to Moore's law [7].

For nondestructive 3D IC imaging at the nanoscale, hard X-rays are ideal probes because of their long penetration depth and short wavelength. Unlike medical X-ray tomography, however, which operates almost always on the intensity of the projections, in the nanoscale case it is common to seek the complex field via ptychography [8] first, and then do tomography. This combined scheme is also known as X-ray ptychographic tomography (ptycho-tomography) [9]. There are several reasons to do this: for example, if the projection approximation is still applicable, then we can perform two tomographic reconstructions in parallel, one on the field amplitude yielding the imaginary part of the refractive index (attenuation) at each voxel and one on the field phase yielding the real part; most materials exhibit phase variations by 10 times larger than their respective absorption changes [10].

X-ray ptycho-tomography reconstructions are performed in the same sequence as experimental acquisition, i.e. in a two-step approach [9, 11]. First, 2D projections are retrieved from far-field diffraction patterns using phase-retrieval algorithms [12–14], and then, tomographic reconstructions are implemented to recover the real and/or imaginary parts of a 3D object from 2D projections [15–18]. Many applications have been successfully demonstrated with this two-step

approach: IC imaging [3, 19], microscopic organism imaging [9, 20] and studies of material properties such as fracture [21], percolation [22] and hydration [23]. However, both ptychography and tomography demand large redundancy in the data [24, 25], leading to long acquisition and processing times generally.

One way to reduce the acquisition time is through high-precision scanners that can reliably work with efficient scanning schemes [26–28] and at high scanning velocities [29, 30]. Reducing the data redundancy requirements in ptycho-tomography is an alternate way to speed up data acquisition but introduces ill-posedness. However, with reduced data, the conventional reconstruction algorithms are likely to produce artifacts and a general loss of fidelity.

Studies have coupled computationally the ptychography and tomography reconstruction processes to improve reconstruction qualities under limited data acquisition. One way is to split the whole problem into two sub-problems, as conventional two-step approaches, and perform them iteratively, to mildly relax the data redundancy requirements without sacrificing fidelity. For example, tomography naturally provides angular intersections of beams as they pass through the object, which is employed to coarsen the ptychographic sampling in each projection with iterative two-step algorithms [31–34]. The angular requirements in tomography could be eased as well [35–39] through physically modeling the interactions between X-ray and object with a multi-slice propagation model instead of projection approximation [40]. Depth information is resolved partially in individual projection planes to help relax the usual Crowther criterion for tomography. On the other hand, the coupling between ptychography and tomography could be expressed as a single optimization problem to reconstruct 3D objects from diffraction patterns directly instead of two separate cost functions [41–43] to further reduce the required number of measurements. Still, however, severe image artifacts are to be expected if data reduction is aggressive beyond a certain limit. Moreover, all the above mentioned variants of X-ray ptycho-tomographic reconstruction are computationally intensive and, hence, scale prohibitively with sample volume.

In general, regularization resolves ill-posedness by rejecting invalid objects and, hence, eliminating reconstruction artifacts that would be incompatible with our prior knowledge about the object. For example, handcrafted priors, such as sparsity, piecewise constancy, etc. are routinely incorporated in X-ray ptycho-tomography problems to improve reconstruction qualities to some extent [31–33, 43, 44]. Recently, deep neural networks (DNNs) have yielded even better

regularization performance under severe ill-posedness, e.g. 2D phase retrieval through scattering media [45, 46] and under extremely low light conditions [47, 48], digital staining [49, 50], limited-angle 3D volumetric reconstruction [51, 52], etc. These works are based on supervised learning, where the regularizing priors are learned from large datasets of available typical objects. Non-supervised approaches are also possible [53, 54] but not of interest for our present work. The purpose of this paper is to use DNN-based regularizers to radically increase the allowable angular reduction and associated gains in both acquisition and computation time in X-ray ptycho-tomography, which hasn't been explored in experimental X-ray ptycho-tomography yet [55].

Poor image fidelities resulting from severe ill-posedness in X-ray ptycho-tomography is difficult to improve if we naïvely apply a vanilla 3D DNN structure with the kernels size of $3 \times 3 \times 3$ [56], i.e., each layer's receptive field extends to the next layer by a single pixel away only. To perform image correction in hard cases (i.e., large angular reduction) one needs DNNs with many layers or large size kernels, and that is disadvantageous both because it introduces too many parameters, especially in 3D, and because training may saturate early, even with residuals [57]. On the other hand, the physics of tomographic image reconstruction suggests that larger receptive fields in the convolutional layers should be effective in a shallower network while still requiring a large number of parameters. The atrous convolution methods [58, 59] combat this problem by forcing all connections within the receptive field to be zero, except the ones at the outermost corners. Moreover, the implementation of atrous convolution in a Spatial Pyramid Pooling (ASPP) module is known to perform well in extracting long-range and multi-scale information [60, 61].

2 Results

In this study, we propose the novel deep learning-based pipeline for reduced-angle ptycho-tomography, RAPID. Our method works as follows: first, the far-field diffraction patterns obtained from reduced 21-angle acquisitions are pre-processed together to produce an Approximant [51]. This is a preliminary 3D reconstruction of the object's interior and generally exhibits low quality. The Approximant is obtained by gradient descent inversion on a multi-slice propagation model [40, 62]. Subsequently, the Approximant is fed into the RAPID network. During the training phase, matching the network's output to the corresponding golden standard is used to adjust the network weights in a standard stochastic gradient descent fashion. During testing, the network's output is the final reconstruction of the given volume. The procedure is schematically depicted in Fig. 1 and described in detail in the Methods section.
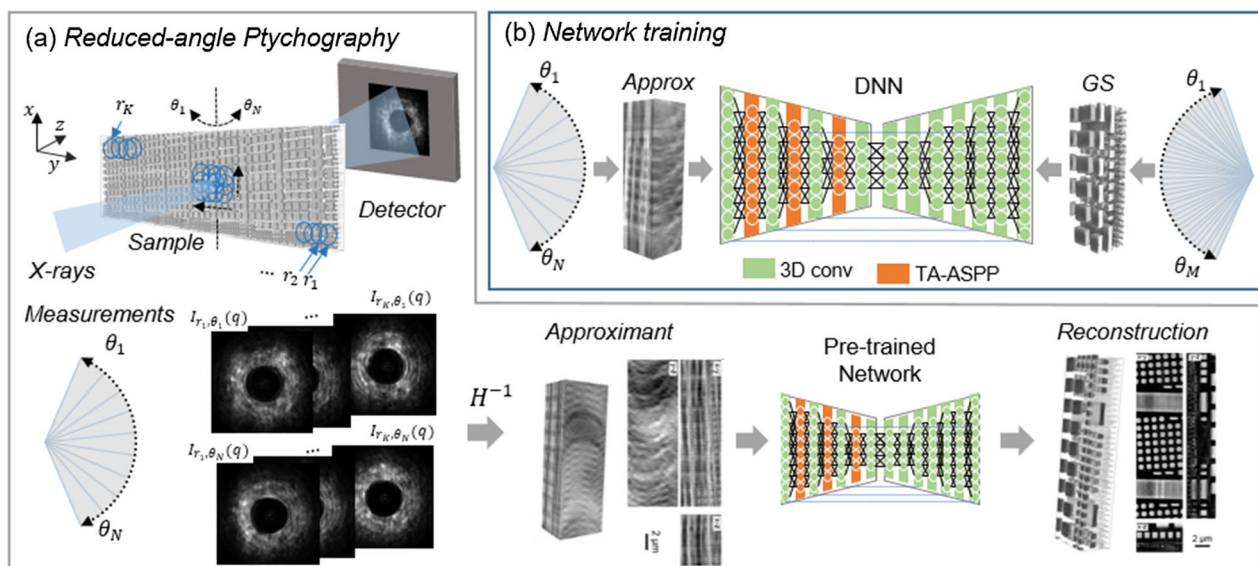

Fig. 1 Schematic of the proposed RAPID framework. **a** Reduced-angle ptycho-tomography experiment to collect diffraction pattern measurements via translational and rotational scanning. Raw diffraction patterns are pre-processed to generate the approximant as the input to the pre-trained network, and volumetric distribution are obtained as the final output. **b** Network training process. Diffraction patterns acquired from reduced-angle ptycho-tomography are pre-processed to get the approximant as the network input, and a two-step conventional approach is employed to generate the high-resolution golden standard (GS) as the ground truth to train the DNN

A new DNN structure is proposed by incorporating the atrous module in the 3D U-net structure [63, 64] to improve the image qualities. Here we modify the atrous module anisotropically to account for the 3D point spread function (PSF). We use the term "anisotropy" here in the sense that the atrous convolutional kernels are different along the x, y, and z axes.

In the experimental demonstration, the IC sample consists of 13 circuit layers. The layers have a different thickness each. The total thickness is unknown, but we estimated it based on the golden standard to be 3.92 μm. The area of each circuit layer is 25.10×93.18 μm^2. The upper part of the IC, relative to the optical axis, is used to pre-train the network. This training segment has a total volume of $20.60 \times 93.18 \times 3.92$ μm^3. The remaining part, with volume $4.48 \times 93.18 \times 3.92$ μm^3, serves for testing.

Better reconstruction performance can be expected as the number of rotation angles increases, at the expense of longer experimental and computational time. We explore the best scanning condition that results in the minimum feasible acquisition time. Starting from the extreme condition of a single angle, we gradually increase N to 349 by adding angular measurements uniformly within maximum angular range $\theta_{max} = 140.8°$. The improvement is noticeable quantitatively and visually when the total number of rotation angles is small, e.g. from 1 to 5, as shown in Fig. 2a, b; Additional file 1: Fig. S2. However, above 21 angles the improvement is marginal. Therefore,

this represents a good compromise between accuracy and acquisition cost.

We further explore the influence of maximum angular range θ_{max} with fixed number of rotation angles as $N = 21$. Figure 2c, d and Additional file 1: Fig. S3 show quantitative and qualitative performance when increasing θ_{max} from 8° to 140.8°. Small angles such as 8° and 16° perform badly. Increasing θ_{max} improves up to 32° and beyond the returns become diminishing again but without any added cost in computational time. Therefore, we can afford to use the full range of 140.8°.

Figure 3 describes typical testing results the $N = 349$ angles and $\theta_{max} = 140.8°$ for the golden standard compared with our optimal compromise, *i.e.* $N = 21$ angles and $\theta_{max} = 140.8°$. Parts (a) and (b) show quantitative metrics of image quality. We have chosen four: Pearson Correlation Coefficient (PCC), Multi-scale Structural Similarity Metric (MS-SSIM) [65], Bit Error Rate (BER) [66], and the Dice coefficient [67] (Detailed in Methods section). The first two are used often in statistics and image processing, while the third and fourth are information- and set-theoretic, respectively. The results indicate that the RAPID method indeed can learn to regularize better than the conventional filtered backprojection (FBP) and simultaneous algebraic reconstruction technique (SART) methods.

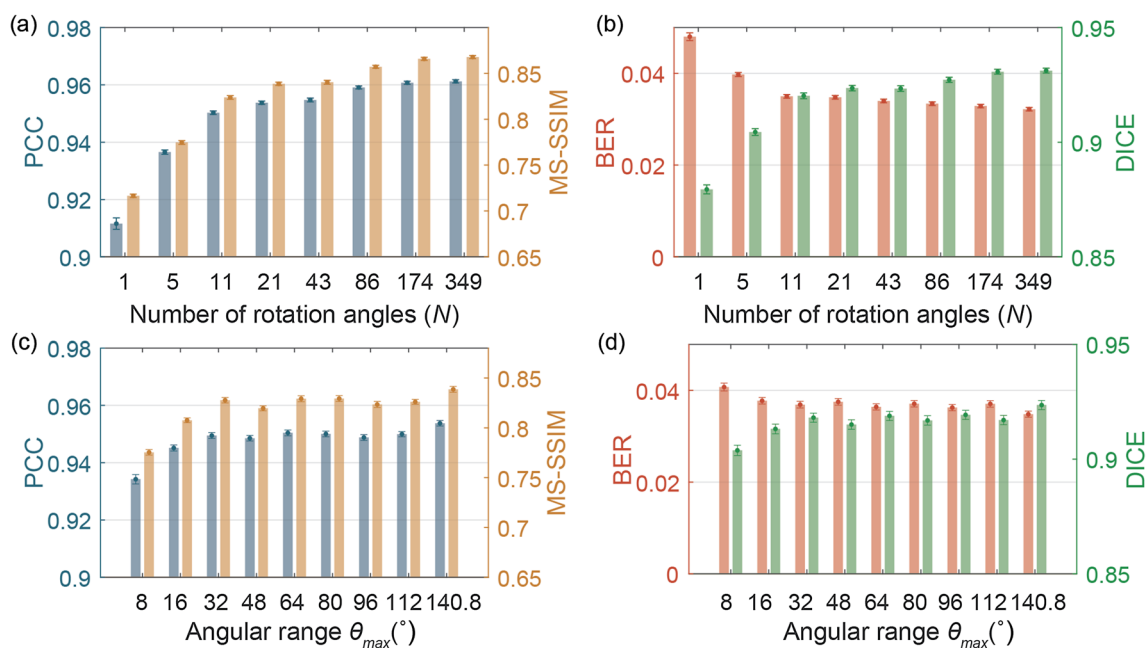Figure 3c–g show the golden standard and how well various reconstruction approaches come to approximate

Fig. 2 Quantitative comparison among different scanning strategies for testing volumes. **a**, **b** Show the performance change with the increase of the number of rotation angles. **c**, **d** Show the performance change with the increase of angular range
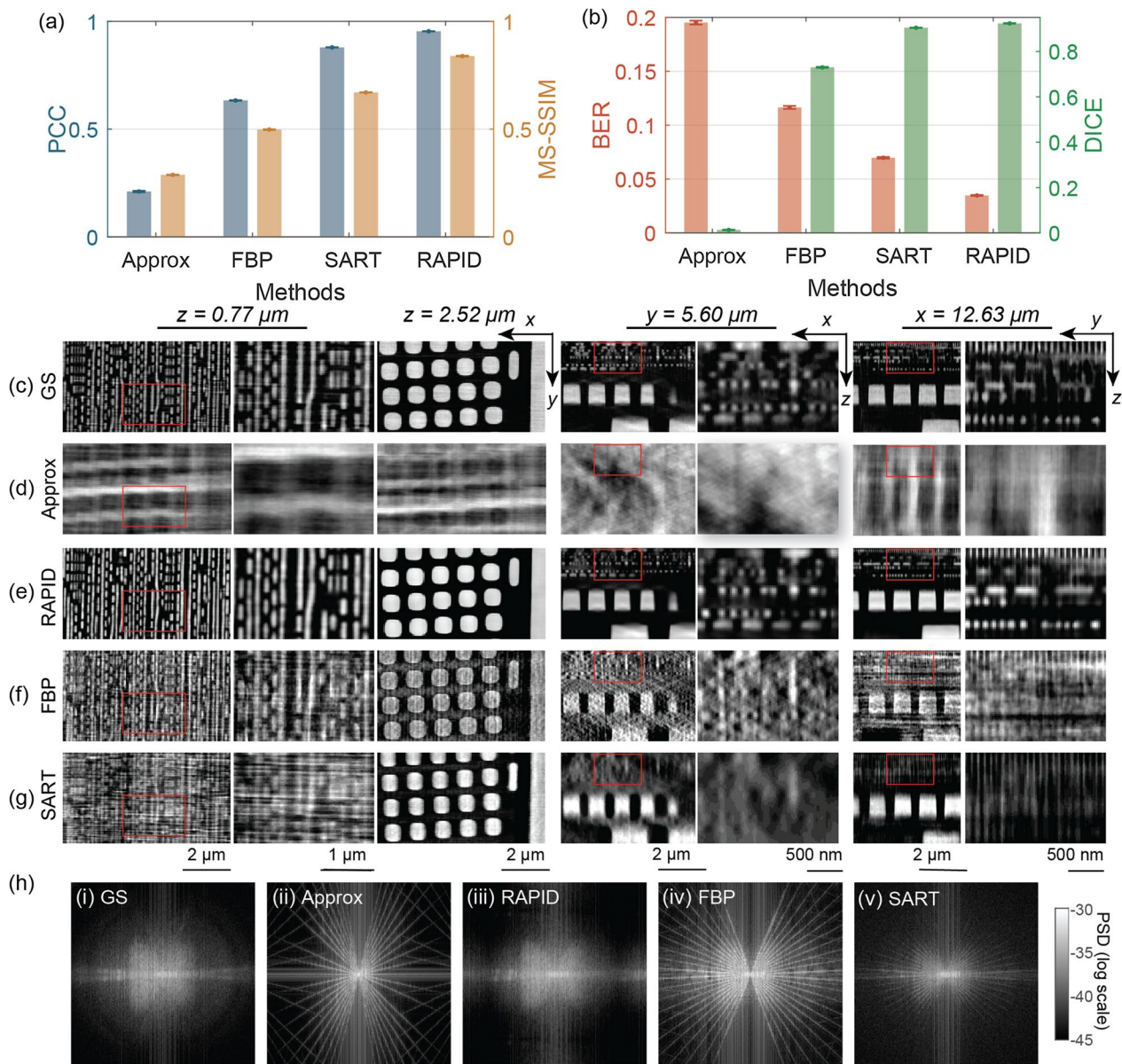
Fig. 3 Performance of RAPID under $N_\theta = 21$ acquisition within the range $\theta = 140.8°$. **a, b** Quantitative (PCC, MS-SSIM, BER, and DICE) comparison among approximant, RAPID, FBP, and SART. **c–g** Layer-wise visualization of the reconstruction results from different methods, including the golden standard reconstructed from $N_\theta = 349$ angles, Approximant, RAPID, FBP, and SART recovered from $N_\theta = 21$ angles. **h** PSD distribution in k_z-k_x plane of different methods. For k_z-k_y and k_x-k_y planes, refer to the Additional file 1: Fig. S4

it, for several circuit layers and orientations. As expected, the Approximant (Fig. 3d) is of rather poor quality because of the severe missing wedge problem in our reduced-angle configuration. The RAPID method does not fully eliminate the axial artifacts, but significantly reduces them—almost to the same extent as the golden standard (Fig. 3c). Part (h) shows the power spectral densities (PSD) of the whole testing volume in $k_x - k_z$ plane (the performances of $k_x - k_y$ and $k_y - k_z$ planes are shown in Additional file 1: Fig. S4), corresponding to

methods of (c–g). Notable are the differences in coverage of the space between the measured slices (emerging as radial spokes in the PSD) and of the missing wedges.

For the same configuration $N = 21$ angles and $\theta_{\max} = 140.8°$, Fig. 4 studies the influence of atrous anisotropy in our method and compares with different combinations of isotropic or partially anisotropic scheme. To make the comparison fair, all methods are designed with a similar number of total parameters and trained with the same strategy. Extending the anisotropic kernel range
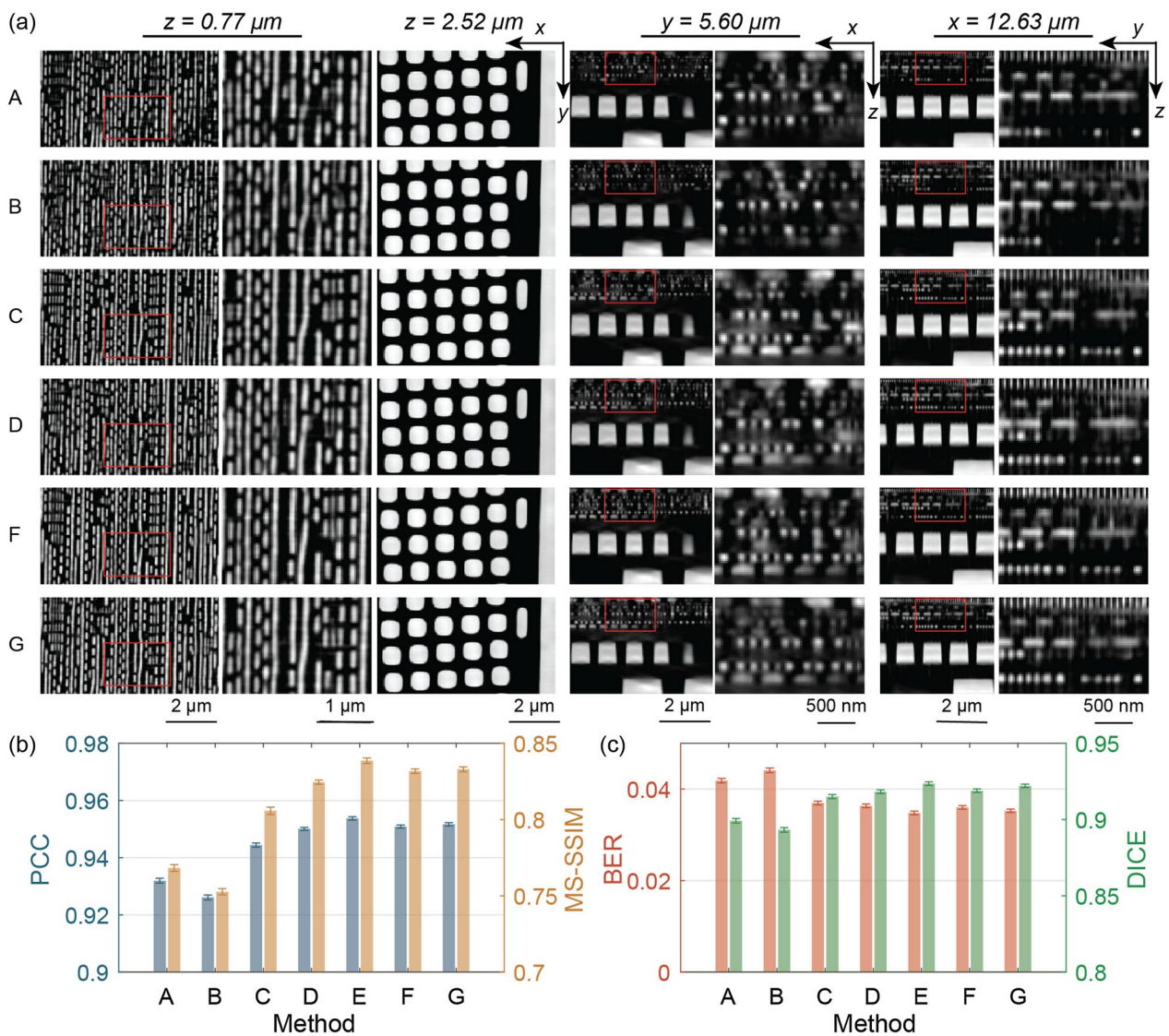
Fig. 4 Quantitative and qualitative comparison among the reconstruction results from different network architectures. **a** Layer-wise visualization of the reconstructions from different network architectures, A: 3D U-net structure as the baseline method; and modified 3D U-net by replacing the first convolutional kernels at each hierarchical level in the encoder as B: the combination of $x - y$, $y - z$, and $x - z$ convolution kernels without atrous; C: 3D isotropic atrous module, D: 3D anisotropic atrous module with the same max atrous rate $a_1 = a_2 = 18$, E–G: 3D anisotropic atrous module with different max atrous rates ($(a_1 = 24$ and $a_2 = 30)$, $(a_1 = 30$ and $a_2 = 36)$, and $(a_1 = 36$ and $a_2 = 42)$, respectively). The results of method E are shown in Fig. 3e. **b**, **c** Quantitative comparison of the testing volumes

Table 1 Experimental and computational time (hours) and reduction ratio of the whole testing volume $4.50 \times 93.18 \times 3.92\,\mu m^3$ from conventional two-step approaches with N=349 angles as the golden standard, RAPID, FBP, and SART methods with $N = 21$ angles

	Golden standard		RAPID		FBP		SART	
	Time	Ratio	Time	Ratio	Time	Ratio	Time	Ratio
Data acquisition	1.90	×1	0.110	×16	0.110	×16	0.110	×16
Computation	66.5	×1	0.380	×175	4.86	×14	9.53	×7
Total	68.4	×1	0.490	×140	14.0	×1	9.63	×7

along the axis of the missing wedge tends to effectively compensate for the axial artifacts. Results E, F, and G in the figure indicate that the choice of atrous parameters does not impact performance significantly.

Table 1 shows the data acquisition time, computational reconstruction time, and total pipeline duration for the techniques under comparison. RAPID is $\times 16$ faster in terms of data acquisition and $\times 175$ faster for image reconstruction compared to the golden standard. The aggregate acceleration for the entire pipeline is $\times 140$. The absolute durations for the golden standard and RAPID were $\sim 66\,\mathrm{h}\,30'$ and $\sim 30'$, respectively.

3 Discussion

To address severely ill-posed problems in X-ray imaging, we introduced anisotropic atrous spatial pyramid pooling modules which increase the size of the receptive field to enable long-range and multi-scale extraction of underlying features. This augmentation largely improves performance compared to non-atrous implementations. The max atrous rates in this novel module can be more rigorously determined by feature size, scattering potential, dataset sampling size, etc. For example, it would be worthwhile to investigate the relationship between max atrous rate and the anisotropy in the PSF of the imaging system. Alternatively, by means of global-range self-attention, transformer architectures [68, 69] have also been demonstrated for reduced-angle psycho-tomography [70]. Detailed comparison between these two methods is beyond the scope of the present paper.

Different from cylinder-shaped samples [9], the penetration path length of a plate-shaped sample increases significantly with the rotation angle. When the penetration path length is larger than the depth of field, multi-slice techniques are necessary to account for propagation effects within the sample when generating the approximant. In our implementation, we run a five-slice ptychographic algorithm under a reduced-angle framework for two iterations to speed up the computation, resulting in vague layer separation from each angle. The improvement in the reconstruction quality flattens after 21 projections as shown in Fig. 2a, b is related to this approximant generation algorithm. As reconstructions from adjacent angles are similar, adding more angles will not improve the approximant quality significantly and thus the final reconstruction. Besides, a multi-slice ptychographic algorithm can relax the Crowther criterion due to more frequency coverage in the Fourier domain for each projection angle [39], which also indicates the information from neighboring angles are similar. On the other hand, the plateau after 32° in Fig. 2c, d shows that measurements sampled from the angular range over 32° contribute similarly to the

approximant compared to the 32° case when fixing the total number of rotation angles as 21. More slices may be required to count for the diffraction effects at larger angles, but in turn, increase the computation burden. The theoretical proof for the turning point in terms of the number of projections and maximum angular range is out of the scope of this manuscript but is interesting for future study. On the other hand, the laminography technique [71, 72] compensates for uneven propagation lengths by scanning the illumination wavevectors along a conical surface. In either case, it may be possible to modify RAPID to further reduce the total scanning time by skipping steps in the ptychographic scan as well [73].

Supervised learning approaches often is a cause for concern regarding the generalization ability to new and unseen data. We propose a strategy to train on a subset of the sample, where a trustable but otherwise very slow alternative method can be used to obtain ground truths; and then use the train network on the rest of the sample, significantly speeding up the entire operation. This approach is appealing for integrated circuits or other large 3D specimens [74]. Besides, it is possible that transfer learning [75] might alleviate the efforts for training RAPID anew for new experiments. For even more general specimens, like viruses, nanoparticles, etc. comparable performance may be expected, but most likely at the cost of some redesign in the learning architecture.

4 Method

4.1 X-ray ptychographic tomography experiment of integrated circuits

X-ray ptychographic tomography experiment was carried out using the Velociprobe with a Dectris Eiger 500K detector (75 μm pixel size) positioned at a distance of 1.92 m from the sample at the Advanced Photon Source of the Argonne National Laboratory, USA. A schematic of the Velociprobe was shown in the previous paper [30]. The photon energy of 8.8 keV with a spectral bandwidth of 10^{-4} was selected using a double-crystal silicon monochromator. A Fresnel zone plate with 50 nm outmost zone width and 180 μm diameter was installed on the zone plate scanner. The first order diffracted beam from the zone plate was selected by the combined use of a 60 μm diameter tungsten central stop and a 30 μm diameter order-sorting aperture placed ~ 62 mm downstream of the zone plate. The illumination spot size on the sample is about 1.4 μm. The sample was fly-scanned in a snake pattern [30] with a 100-nm and 500-nm step size in the horizontal and vertical directions, respectively.

A total number of 349 rotation angles with the angle spacing 0.4° within the angular range of

$\theta_{\max} = 140.8°$ from the reference axis was acquired for an IC produced with 16 nm technology with the size of $25.09 \times 93.18 \times 3.92\,\mu m^3$. $\sim 60k$ diffraction patterns were captured at each angle. The field of view of the projection at each angle was $30 \times 100\,\mu m^2$ with the detector frame rate of 500 Hz, giving 2 ms exposure time per scan. It took about 129 s for each rotation angle and the total data acquisition time for 349 angles was ~ 13 h. For reduced-angle ptycho-tomography, we increased the angular spacing proportionally. The experiment time for the whole testing volume was estimated linearly according to the ratio of testing volume to the whole volume and the number of reduced angles to the whole angle, which is reasonable as the translational and angular scanning scheme of ptycho-tomography.

4.2 Multi-slice forward and inverse models for reduced-angle acquisition

We applied the multi-slice propagation method to model the measurements exiting the object. In the multi-slice propagation model, the object f is divided into L slices along the beam propagation direction, as $[f_1, f_2, ..., f_L]$. Each slice is with the thickness of Δz. The wave field $u_{l,j}(x, y, z_l)$ from probe position j entering lth slice is modulated by the slice f_l to yield a wave field $u'_{l+1,j}(x, y, z_{l+1})$ as $u'_{l,j}(x, y, z_l) = u_{l,j}(x, y, z_l)f_l(x, y, z_l)$. The wavefront is then propagated to the next slice according to the Fresnel diffraction integral given by

$$u_{l+1,j}(x, y, z_{l+1}) = \mathcal{P}_{\Delta z} u'_{l+1,j}(x, y, z_{l+1})$$
$$= \mathcal{F}^{-1}\{\mathcal{F}\{u'_{l+1,j}(x, y, z_{l+1})\}h_{\Delta z}\}. \tag{1}$$

Here $h_{\Delta z} = \exp(-i(q - \sqrt{q^2 - q_x^2 - q_y^2}))$ and q is the reciprocal domain coordinate. This process is repeated for all L layers until one obtains the exit wave leaving the object $\psi_{L,j}$, represented as $\psi_{L,j} = f_L \mathcal{P}_{\Delta z} f_{L-1} ... \mathcal{P}_{\Delta z} f_2 \mathcal{P}_{\Delta z} f_1 u_{1,j}$. Here $u_{1,j}$ is the incident probe at scan position j. Then we apply the far-field propagation operator $\mathcal{P}_d$ to take the exit wave $\psi_{L,j}$ from the object to the plane of the detector, which is performed with a simple Fourier transform as $u_j(q) = \mathcal{P}_d \psi_{L,j} = \mathcal{F}\{\psi_{L,j}\}$.

In this experiment, the quasi-coherent X-ray illumination was modeled as the combination of multiple coherent modes with the index of $m = [1, 2, ..., M]$ to improve the accuracy. Thus, the far-field diffraction measurements were represented as the sum of each coherent mode. In addition, reduced-angle ptycho-tomography requires the illumination of the object from several rotation angles θ. Here we rotated the object according to a constant wave propagation direction, which was performed with

a rotation operation $f_\theta = \mathcal{R}_\theta f$, and the rotated object f_θ is further sliced into L different layers $[f_{\theta,1}, f_{\theta,2}, ..., f_{\theta,L}]$. This leads to a combined forward operation of

$$H_{\theta,j} = \sum_m |u_{\theta,j}^{(m)}(q)|^2$$
$$= \sum_m |\mathcal{P}_d f_{\theta,L} \mathcal{P}_{\Delta z} f_{\theta,L-1} ... \mathcal{P}_{\Delta z} f_{\theta,1} \mathcal{P}_{\Delta z} f_{\theta,1} u_{1,j}^{(m)}|^2. \tag{2}$$

In order to apply the gradient descent updates to find optimal f, we start with the data fidelity term of the loss function

$$\mathcal{L} = \frac{1}{2}\frac{1}{N_\theta N_j}\sum_{\theta,j}\|\sum_m |u_{\theta,j}^{(m)}(q)|^2 - g_{\theta,j}\|_2^2 = \frac{1}{N_\theta}\sum_\theta \mathcal{L}_\theta. \tag{3}$$

The gradient of $\mathcal{L}$ with respect to f is derived as

$$\nabla_f \mathcal{L} = \frac{1}{N_\theta}\sum_\theta \frac{\partial \mathcal{L}_\theta}{\partial f}$$
$$= \frac{1}{N_\theta}\sum_\theta \left[\frac{\partial \mathcal{L}_\theta}{\partial f_{\theta,1}}, \frac{\partial \mathcal{L}_\theta}{\partial f_{\theta,2}}, ... \frac{\partial \mathcal{L}_\theta}{\partial f_{\theta,L}}\right]\frac{\partial f_\theta}{\partial f}. \tag{4}$$

The term $\frac{\partial f_\theta}{\partial f}$ could be obtained with the rotation matrix. We derive the formula of $\frac{\partial \mathcal{L}_\theta}{\partial f_\theta}$ to get the $\nabla_f \mathcal{L}$ with the chain rule as

$$\frac{\partial \mathcal{L}_\theta}{\partial f_\theta} = \sum_{m,j}\frac{\partial \mathcal{L}}{\partial |u_{\theta,j}^{(m)}|}\frac{\partial |u_{\theta,j}^{(m)}|}{\partial f_\theta}$$
$$= \frac{1}{N_j}\sum_{m,j}\left(\sum_m |u_{\theta,j}^{(m)}(q)|^2 - g_{\theta,j}\right)2|u_{\theta,j}^{(m)}|\frac{\partial |u_{\theta,j}^{(m)}|}{\partial f_\theta} \tag{5}$$

Following the similar notation as ref. [76], we employ the auxiliary variable $\chi_{\theta,j}^{(m)}$ as

$$\chi_{\theta,j}^{(m)} = \mathcal{F}^{-1}\left\{2\left(\sum_m |u_{\theta,j}^{(m)}(q)|^2 - g_{\theta,j}\right)u_{\theta,j}^{(m)}\right\}. \tag{6}$$

In this way, the gradient of the loss function $\mathcal{L}_\theta$ with respect to the object f_θ is defined as

$$\frac{\partial \mathcal{L}_\theta}{\partial f_{\theta,l}} = \begin{cases} \frac{1}{N_j}\sum_{m,j}\chi_{\theta,j}^{(m)*}u_{\theta,j,L}^{(m)}; & l = L \\ u_{\theta,j,l}^{(m)}\mathcal{P}_{-\Delta z}\left\{\frac{\partial \mathcal{L}_\theta}{\partial u_{\theta,j,l+1}^{(m)}}\right\}^*; & 1 \le l < L \end{cases}. \tag{7}$$

where the asterisk represents the complex conjugate. Here $\frac{\partial \mathcal{L}_\theta}{\partial u_{\theta,j,l}^{(m)}}$ was derived as follows

$$\frac{\partial \mathcal{L}_\theta}{\partial u_{\theta,j,l}^{(m)}} = \begin{cases} \frac{1}{N_j}\chi_{\theta,j}^{(m)*}f_{\theta,L}; & l = L \\ f_{\theta,l}\mathcal{P}_{-\Delta z}\{\frac{\partial \mathcal{L}_\theta}{\partial u_{\theta,j,l+1}^{(m)}}\}^*; & 1 \le l < L \end{cases} \quad (8)$$

4.3 Computation of the approximant

The performance of the DNN is significantly improved if the raw measurements are preprocessed by considering the imaging formation as an approximation of the solution, which is also known as Approximant [47]. Here we treated the reconstruction of 3D refractive index of an object $f(r) = \exp[\alpha(r) + i\phi(r)]$ from multi-angle ptychographic diffraction measurements g as a nonlinear optimization problem by minimizing the loss function

$$\hat{f} = \underset{f}{\mathrm{argmin}}\{\|H(f) - g\|^2 + \gamma\Phi(f)\}, \quad (9)$$

where the first component is known as the data fidelity term, which models the physical relationship between the object f and the measurements g in a reduced-angle setting; Φ is the regularizer expressing the prior knowledge of the object, which is learned from the volumetric pairs of golden standard IC patterns reconstructed from 349 rotation angles with a two-step reconstruction method, and the approximant retrieved from the one-step multi-slice preprocessor with 21 rotation angles; and γ is the regularization parameter controlling the competition between the data fidelity term and regularization term. We assume that the sample is a pure phase object, i.e., $\alpha(r) = 0$, which is reasonable as the phase contrast is about 10 times larger than the absorption contrast in X-ray experiments for IC samples.

The approximant was generated by iteratively updating the data fidelity term (Eq. 3) via gradient descent $f^{(k+1)} = f^{(k)} - s(\nabla_f\mathcal{L})_{f^{(k)}}$. Here k denotes the iteration step, s is the step size, and $(\nabla_f\mathcal{L})_{f^{(k)}}$ is the gradient of $\mathcal{L}$ with respect to f evaluated at $f^{(k)}$, as shown in Eq. 4. Raw diffraction patterns g of $256 \times 256\,\mathrm{px}^2$ were downsampled by $\times 2$ to accelerate the computation, which results in the Approximant $\times 2$ smaller in x and y directions compared to the golden standard. In this work, we chose $k = 2$ to further speed up the computation, $L = 5$ by considering the depth of focus of our system, and $M = 12$ coherent modes of the synchrotron X-ray for the reconstruction. The number of desired reconstruction slices is much larger, i.e. 280, so we simply dilated the generated slices to match it. As a result, the quality of network input is poor.

4.4 Network architecture and implementation details

RAPID is an encoder-decoder network architecture based on 3D U-net via including a special convolution module,

the anisotropic atrous module, at each hierarchical level in the encoding branch, as shown in Additional file 1: Fig. S1a. The original 2D ASPP module contains one 1×1 convolution and three 3×3 convolutions with isotropic atrous rates = (6, 12, 18). Here we extend it to a 3D version by incorporating the 3D atrous convolution, which is defined as

$$g(i,j,k) = \sum_{l=1}^{L}\sum_{m=1}^{M}\sum_{n=1}^{N} f(i - r_1 l, j - r_2 m, k - r_3 n)h(l,m,n). \quad (10)$$

Here, (i, j, k) is the voxel in the original volume f and filtered volume g, (l, m, n) is the voxel in the convolutional kernel h, r_1, r_2, and r_3 are atrous rates in x, y and z axes. r_1, r_2 and r_3 are generally the same, which is also known as isotropic atrous convolution.

Additional file 1: Figure S1b shows the design of the anisotropic atrous module including three anisotropic ASPP modules in $x - y$, $y - z$, and $x - z$ planes and an additional 3D convolution module to capture features in 3D. The features extracted from these four branches are fused via concatenation and passed through another standard $3 \times 3 \times 3$ convolution kernels. Inter-slice cross-talking in the Approximant, originating from the nature of multi-slice reconstruction model, makes feature separation in the z direction difficult. To emphasize this artifact and achieve isotropic volumetric resolution, we include two more anisotropic atrous convolutions $a = a_1, a_2$ to address the severe feature residuals along the z direction for atrous convolution operated in $y - z$ and $x - z$ planes.

The whole volumes were split into 2781 examples with the size of $128 \times 128 \times 280$ for the golden standard, and $64 \times 64 \times 280$ for the Approximant. The split volumes overlapped 50% between each other. As mentioned before, the upper part with respect to the beam propagation direction was used for training, which contained 2060 volumes; and the lower part was used for testing, which included 618 volumes. There was no overlap between the training and testing volumes. We employed a negative Pearson correlation coefficient (NPCC = −PCC) as the loss function and the training runs for 200 epochs with a batch size of 2. The PCC is defined as

$$PCC(A, B) = \frac{\sum_i (A_i - \bar{A})(B_i - \bar{B})}{\sqrt{\sum_i (A_i - \bar{A})^2 \sum_i (B_i - \bar{B})^2}}, \quad (11)$$

for two volumes A and B. *Adam* optimizer for stochastic optimization [77] with a polynomial learning rate schedule was used to update a learning rate as

$$lr(\text{step}) = (lr(0) - lr(\text{end})) \times \left(1 - \frac{\text{step}}{T}\right)^p, \quad (12)$$

where the initial learning rate $lr(0) = 2e - 4$, the end learning rate $lr(\text{end}) = 5e - 5$, the total decay steps $T = 3e4$, and $p = 0.5$. The rest parameters of *Adam* optimizer were set as default values.

For training processes, we used the MIT Supercloud with an Intel Xeon Gold 6248 CPU with 384 GB RAM and dual NVIDIA Volta V100 GPUs with 32 GB VRAM. Once the network was trained, it took less than one minute for giving predictions over each test volume with a single NVIDIA Volta V100 GPU. And the whole testing area took around 125 seconds to get the final result. Our scripts for training and testing are publicly available in *https://github.com/Ziling-Wu*.

4.5 Two-step reconstruction as the ground truth and comparison algorithms

The ptychography reconstruction was conducted with the multi-slice least square maximum likelihood algorithm [78] for 600 iterations to generate the phase projections at each tomographic angle in PtychoShelves [79]. In total, it took $\sim$ 360 h for ptychographic reconstruction for all 349 angles with 8 Tesla V100 GPUs in parallel. We further aligned all 349 projections in the form of a phase ramp removal process and post-process with a pre-trained super-resolution network to refine the projections, which took about $\sim$ 5 h with a single Tesla V100 GPU. The final tomographic reconstruction was performed with 10 iterations of SART [17] to generate a 3D reconstruction of the IC sample with the isotropic 14 nm voxel size, which took about 1 h using 8 Tesla V100 GPUs.

We compared our algorithm with two conventional approaches for 21-angle X-ray ptycho-tomography . The experiment time was reduced by $\times 16$ as a result of the angular reduction. In terms of computation, ptychography reconstruction and projection refinement were performed as the golden standard for diffraction measurements from $N = 21$ angles. The final tomographic reconstruction was conducted with FBP and SART (10 iterations) algorithms implemented in TomoPy with the ASTRA toolbox [80, 81]. The computation time for ptychographic reconstruction and refinement was also reduced linearly according to angular reduction compared to the golden standard. The tomographic reconstructions were performed with the same configuration as our method, which took 116 s and 1678 s for the whole volume with FBP and SART (10 iterations), respectively. The computation time for the whole testing volume was estimated proportionally.

4.6 Quantitative comparison metrics

We used PCC, MS-SSIM, BER, and DICE metrics to quantify our proposed method. PCC and MS-SSIM are used to quantify the correlation between two volumes. PCC is defined in the Eq. (12), and MS-SSIM [65] is a weighted similarity metric with fixed weights on SSIM values from different scales.

The remaining two metrics DICE and BER are used to quantify the segmentation volumes. DICE, also known as F1 score, is broadly used to compare the similarity among two segmented volumes via calculating overlapping size over their total size. Similarly, BER quantifies the ratio of erroneously classified voxels. Both of them involve the derivation of probability distribution functions, thus probabilistic, and are based on binary classification as follows

$$\text{DICE} = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FN} + \text{FP}}, \quad (13)$$

and

$$\text{BER} = \frac{\text{FP} + \text{FN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad (14)$$

where TP, TN, FP, and FN indicate the number of true positives, true negatives, false positives, and false negatives, respectively. For the gold standard, the binary thresholds and prior probabilities $p(0)$, $p(1)$ required for these quantities were estimated by an Expectation Maximization (EM) algorithm. For testing, we used Bayes' rule $p(x|0)p(0) = p(x|1)p(1)$ with $p(0)$, $p(1)$ same as for the gold standard.

Supplementary Information

Additional file 1: Fig S1. Network architecture of the proposed RAPID method. **Fig. S2.** Layer-wise visualization in terms of x−y plane at two different depths, y−z, and x−z planes for the reconstructions acquired from different number of angular scans. **Fig. S3.** Layer-wise visualization in terms of x−y plane at two different depths, y−z, and x−z planes for the reconstructions acquired from different angular ranges. **Fig. S4.** Power spectral density distributions in terms of k_x−k_y, k_y−k_z, and k_x−k_z planes for different approaches.

References
1. K. Grunewald, P. Desai, D.C. Winkler, J.B. Heymann, D.M. Belnap, W. Baumeister, A.C. Steven, Three-dimensional structure of herpes simplex virus from cryo-electron tomography. Science **302**(5649), 1396–1398 (2003)
2. P. Ercius, O. Alaidi, M.J. Rames, G. Ren, Electron tomography: a three-dimensional analytic tool for hard and soft materials research. Adv. Mater. **27**(38), 5638–5663 (2015)
3. M. Holler, M. Guizar-Sicairos, E.H. Tsai, R. Dinapoli, E. Müller, O. Bunk, J. Raabe, G. Aeppli, High-resolution non-destructive three-dimensional imaging of integrated circuits. Nature **543**(7645), 402–406 (2017)
4. B. Parkinson, Layer-by-layer nanometer scale etching of two-dimensional substrates using the scanning tunneling microscope. J. Am. Chem. Soc. **112**(21), 7498–7502 (1990)
5. C. Harrison, M. Park, P. Chaikin, R.A. Register, D.H. Adamson, N. Yao, Layer by layer imaging of diblock copolymer films with a scanning electron microscope. Polymer **39**(13), 2733–2744 (1998)
6. S.H. Lee, K.-N. Chen, J.J.-Q. Lu, Wafer-to-wafer alignment for three-dimensional integration: a review. J. Microelectromech. Syst. **20**(4), 885–898 (2011)
7. R.R. Schaller, Moore's law: past, present and future. IEEE Spectr. **34**(6), 52–59 (1997)
8. R. Hegerl, W. Hoppe, Dynamische theorie der kristallstrukturanalyse durch elektronenbeugung im inhomogenen primärstrahlwellenfeld. Ber. Bunsenges. Phys. Chem. **74**(11), 1148–1154 (1970)
9. M. Dierolf, A. Menzel, P. Thibault, P. Schneider, C.M. Kewish, R. Wepf, O. Bunk, F. Pfeiffer, Ptychographic x-ray computed tomography at the nanoscale. Nature **467**(7314), 436–439 (2010)
10. H. Takano, Y. Wu, J. Irwin, S. Maderych, M. Leibowitz, A. Tkachuk, A. Kumar, B. Hornberger, A. Momose, Comparison of image properties in full-field phase x-ray microscopes based on grating interferometry and Zernike's phase contrast optics. Appl. Phys. Lett. **113**(6), 063105 (2018)
11. M. Guizar-Sicairos, A. Diaz, M. Holler, M.S. Lucas, A. Menzel, R.A. Wepf, O. Bunk, Phase tomography from x-ray coherent diffractive imaging projections. Opt. Express **19**(22), 21345–21357 (2011)
12. H.M.L. Faulkner, J. Rodenburg, Movable aperture lensless transmission microscopy: a novel phase retrieval algorithm. Phys. Rev. Lett. **93**(2), 023903 (2004)
13. A.M. Maiden, J.M. Rodenburg, An improved ptychographical phase retrieval algorithm for diffractive imaging. Ultramicroscopy **109**(10), 1256–1262 (2009)
14. F. Pfeiffer, X-ray ptychography. Nat. Photonics **12**(1), 9–17 (2018)
15. M.J. Willemink, P.B. Noël, The evolution of image reconstruction for ct-from filtered back projection to artificial intelligence. Eur. Radiol. **29**(5), 2185–2195 (2019)
16. P. Gilbert, Iterative methods for the three-dimensional reconstruction of an object from projections. J. Theo. Biol. **36**(1), 105–117 (1972)
17. A.H. Andersen, A.C. Kak, Simultaneous algebraic reconstruction technique (sart): a superior implementation of the art algorithm. Ultrason. Imaging **6**(1), 81–94 (1984)
18. C. Bouman, K. Sauer, A generalized Gaussian image model for edge-preserving map estimation. IEEE Trans. Image Process. **2**(3), 296–310 (1993)
19. A. Schropp, P. Boye, A. Goldschmidt, S. Hönig, R. Hoppe, J. Patommel, C. Rakete, D. Samberg, S. Stephan, S. Schöder, Non-destructive and quantitative imaging of a nano-structured microchip by ptychographic hard x-ray scanning microscopy. J. Microsc. **241**(1), 9–12 (2011)
20. A. Diaz, B. Malkova, M. Holler, M. Guizar-Sicairos, E. Lima, V. Panneels, G. Pigino, A.G. Bittermann, L. Wettstein, T. Tomizaki, Three-dimensional mass density mapping of cellular ultrastructure by ptychographic x-ray nanotomography. J. Struct. Biol. **192**(3), 461–469 (2015)
21. P. Trtik, A. Diaz, M. Guizar-Sicairos, A. Menzel, O. Bunk, Density mapping of hardened cement paste using ptychographic x-ray computed tomography. Cement Concr. Compos. **36**, 71–77 (2013)
22. B. Chen, M. Guizar-Sicairos, G. Xiong, L. Shemilt, A. Diaz, J. Nutter, N. Burdet, S. Huo, J. Mancuso, A. Monteith, Three-dimensional structure analysis and percolation properties of a barrier marine coating. Sci. Rep. **3**(1), 1–5 (2013)
23. M. Esmaeili, J.B. Fløystad, A. Diaz, K. Høydalsvik, M. Guizar-Sicairos, J.W. Andreasen, D.W. Breiby, Ptychographic x-ray tomography of silk fiber hydration. Macromolecules **46**(2), 434–439 (2013)
24. O. Bunk, M. Dierolf, S. Kynde, I. Johnson, O. Marti, F. Pfeiffer, Influence of the overlap parameter on the convergence of the ptychographical iterative engine. Ultramicroscopy **108**(5), 481–487 (2008)
25. R.A. Crowther, D. DeRosier, A. Klug, The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. Proc. R. Soc. London A Math. Phys. Sci. **317**(1530), 319–340 (1970)
26. P.M. Pelz, M. Guizar-Sicairos, P. Thibault, I. Johnson, M. Holler, A. Menzel, On-the-fly scans for x-ray ptychography. Appl. Phys. Lett. **105**(25), 251101 (2014)
27. J.N. Clark, X. Huang, R.J. Harder, I.K. Robinson, Continuous scanning mode for ptychography. Opt. Lett. **39**(20), 6066–6069 (2014)
28. Y. Yao, Y. Jiang, J.A. Klug, M. Wojcik, E.R. Maxey, N.S. Sirica, C. Roehrig, Z. Cai, S. Vogt, B. Lai, Multi-beam x-ray ptychography for high-throughput coherent diffraction imaging. Sci. Rep. **10**(1), 1–8 (2020)
29. J. Deng, Y.S. Nashed, S. Chen, N.W. Phillips, T. Peterka, R. Ross, S. Vogt, C. Jacobsen, D.J. Vine, Continuous motion scan ptychography: characterization for increased speed in coherent x-ray imaging. Opt. Express **23**(5), 5438–5451 (2015)
30. J. Deng, C. Preissner, J.A. Klug, S. Mashrafi, C. Roehrig, Y. Jiang, Y. Yao, M. Wojcik, M.D. Wyman, D. Vine, The velociprobe: an ultrafast hard x-ray nanoprobe for high-resolution ptychographic imaging. Rev. Sci. Instrum. **90**(8), 083701 (2019)
31. D. Gürsoy, Direct coupling of tomography and ptychography. Opt. Lett. **42**(16), 3169–3172 (2017)
32. M. Kahnt, J. Becher, D. Brückner, Y. Fam, T. Sheppard, T. Weissenberger, F. Wittwer, J.-D. Grunwaldt, W. Schwieger, C.G. Schroer, Coupled ptychography and tomography algorithm improves reconstruction of experimental data. Optica **6**(10), 1282–1289 (2019)

33. H. Chang, P. Enfedaque, S. Marchesini, Iterative joint ptychography-tomography with total variation regularization. in 2019 IEEE International Conference on Image Processing (ICIP), pp. 2931–2935 (2019). IEEE

34. S. Aslan, V. Nikitin, D.J. Ching, T. Bicer, S. Leyffer, D. Gürsoy, Joint pytycho-tomography reconstruction through alternating direction method of multipliers. Opt. Express **27**(6), 9128–9143 (2019)

35. A. Suzuki, S. Furutaku, K. Shimomura, K. Yamauchi, Y. Kohmura, T. Ishikawa, Y. Takahashi, High-resolution multislice x-ray ptychography of extended thick objects. Phys. Rev. Lett. **112**(5), 053903 (2014)

36. K. Shimomura, A. Suzuki, M. Hirose, Y. Takahashi, Precession x-ray ptychography with multislice approach. Phys. Rev. B. **91**(21), 214114 (2015)

37. K. Shimomura, M. Hirose, T. Higashino, Y. Takahashi, Three-dimensional iterative multislice reconstruction for ptychographic x-ray computed tomography. Opt. Express **26**(24), 31199–31208 (2018)

38. P. Li, A. Maiden, Multi-slice ptychographic tomography. Sci. Rep. **8**(1), 1–10 (2018)

39. C. Jacobsen, Relaxation of the crowther criterion in multislice tomography. Opt. Lett. **43**(19), 4811–4814 (2018)

40. A.M. Maiden, M.J. Humphry, J. Rodenburg, Ptychographic transmission microscopy in three dimensions using a multi-slice approach. JOSA A **29**(8), 1606–1614 (2012)

41. T. Ramos, B.E. Grønager, M.S. Andersen, J.W. Andreasen, Direct three-dimensional tomographic reconstruction and phase retrieval of far-field coherent diffraction patterns. Phys. Rev. A **99**(2), 023801 (2019)

42. S. Barutcu, P. Ruiz, F. Schiffers, S. Aslan, D. Gursoy, O. Cossairt, A.K. Katsaggelos, Simultaneous 3d x-ray ptycho-tomography with gradient descent. in 2020 IEEE International Conference on Image Processing (ICIP), pp. 96–100 (2020). IEEE

43. Z. Fabian, J. Haldar, R. Leahy, M. Soltanolkotabi, 3d phase retrieval at nano-scale via accelerated wirtinger flow. in 2020 28th European Signal Processing Conference (EUSIPCO), pp. 2080–2084 (2021). IEEE

44. V. Nikitin, S. Aslan, Y. Yao, T. Biçer, S. Leyffer, R. Mokso, D. Gürsoy, Photon-limited ptychography of 3d objects via Bayesian reconstruction. OSA Continuum **2**(10), 2948–2968 (2019)

45. S. Li, M. Deng, J. Lee, A. Sinha, G. Barbastathis, Imaging through glass diffusers using densely connected convolutional networks. Optica **5**(7), 803–813 (2018)

46. Y. Li, Y. Xue, L. Tian, Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media. Optica **5**(10), 1181–1190 (2018)

47. A. Goy, K. Arthur, S. Li, G. Barbastathis, Low photon count phase retrieval using deep learning. Phys. Rev. Lett. **121**(24), 243902 (2018)

48. M. Deng, S. Li, A. Goy, I. Kang, G. Barbastathis, Learning to synthesize: robust phase retrieval at low photon counts. Light Sci. Appl. **9**(1), 1–16 (2020)

49. Y. Rivenson, T. Liu, Z. Wei, Y. Zhang, K. de Haan, A. Ozcan, Phasestain: the digital staining of label-free quantitative phase microscopy images using deep learning. Light Sci. Appl. **8**(1), 1–11 (2019)

50. H. Majeed, A. Keikhosravi, M.E. Kandel, T.H. Nguyen, Y. Liu, A. Kajdacsy-Balla, K. Tangella, K.W. Eliceiri, G. Popescu, Quantitative histopathology of stained tissues using color spatial light interference microscopy (cslim). Sci. Rep. **9**(1), 1–14 (2019)

51. A. Goy, G. Rughoobur, S. Li, K. Arthur, A.I. Akinwande, G. Barbastathis, High-resolution limited-angle phase tomography of dense layered objects using deep neural networks. Proc. Natl. Acad. Sci. **116**(40), 19848–19856 (2019)

52. I. Kang, A. Goy, G. Barbastathis, Dynamical machine learning volumetric reconstruction of objects' interiors from limited angular views. Light Sci. Appl. **10**(1), 1–21 (2021)

53. F. Wang, Y. Bian, H. Wang, M. Lyu, G. Pedrini, W. Osten, G. Barbastathis, G. Situ, Phase imaging with an untrained neural network. Light Sci. Appl. **9**(1), 1–7 (2020)

54. H. Chung, J. Huh, G. Kim, Y.K. Park, J.C. Ye, Missing cone artifact removal in odt using unsupervised deep learning in the projection domain. IEEE Trans. Comput. Imag. **7**, 747–758 (2021)

55. S. Aslan, Z. Liu, V. Nikitin, T. Bicer, S. Leyffer, D. Gürsoy, Joint ptycho-tomography with deep generative priors. Mach. Learn. Sci. Technol. **2**(4), 045017 (2021)

56. L. Alzubaidi, J. Zhang, A.J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M.A. Fadhel, M. Al-Amidie, L. Farhan, Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. J. Big Data **8**(1), 1–74 (2021)

57. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

58. S. Mallat, *A Wavelet Tour of Signal Processing* (Elsevier, Amsterdam, 1999)

59. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Semantic image segmentation with deep convolutional nets and fully connected crfs. arXiv preprint arXiv:1412.7062 (2014)

60. K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. **37**(9), 1904–1916 (2015)

61. L.-C. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587 (2017)

62. U.S. Kamilov, I.N. Papadopoulos, M.H. Shoreh, A. Goy, C. Vonesch, M. Unser, D. Psaltis, Learning approach to optical tomography. Optica **2**(6), 517–522 (2015)

63. O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation. in International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234–241 (2015). Springer

64. Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-Net: learning dense volumetric segmentation from sparse annotation. in International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 424–432 (2016). Springer

65. Z. Wang, E.P. Simoncelli, A.C. Bovik, Multiscale structural similarity for image quality assessment. In: The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, vol. 2, pp. 1398–1402 (2003). IEEE

66. J. Lim, Is ber the bit error ratio or the bit error rate? EDN. Retrieved, 02–16 (2015)

67. L.R. Dice, Measures of the amount of ecologic association between species. Ecology **26**(3), 297–302 (1945)

68. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł Kaiser, I. Polosukhin, Attention is all you need. Adv. Neural Informat. Process. Syst. **30**, (2017)

69. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)

70. I. Kang, Z. Wu, Y. Jiang, Y. Yao, J. Deng, J. Klug, N. Weisse-Bernstein, S. Vogt, G. Barbastathis, Attentional Ptycho-Tomography (APT) for three-dimensional nanoscale X-ray imaging with minimal data acquisition and computation time (Submitted)

71. L. Helfen, T. Baumbach, P. Mikulik, D. Kiel, P. Pernot, P. Cloetens, J. Baruchel, High-resolution three-dimensional imaging of flat objects by synchrotron-radiation computed laminography. Appl. Phys. Lett. **86**(7), 071915 (2005)

72. M. Holler, M. Odstrcil, M. Guizar-Sicairos, M. Lebugle, E. Müller, S. Finizio, G. Tinti, C. David, J. Zusman, W. Unglaub, Three-dimensional imaging of integrated circuits with macro-to nanoscale zoom. Nat. Electron. **2**(10), 464–470 (2019)

73. Z. Wu, I. Kang, T. Zhou, V. Coykendall, B. Ge, M.J. Cherukara, G. Barbastathis, Photon-starved x-ray ptychographic imaging using spatial pyramid atrous convolution end-to-end reconstruction (ptychospacer). in Computational Optical Sensing and Imaging, pp. 1–6 (2022). Optica Publishing Group

74. M.J. Cherukara, T. Zhou, Y. Nashed, P. Enfedaque, A. Hexemer, R.J. Harder, M.V. Holt, Ai-enabled high-resolution scanning coherent diffraction imaging. Appl. Phys. Lett. **117**(4), 044103 (2020)

75. K. Weiss, T.M. Khoshgoftaar, D. Wang, A survey of transfer learning. J. Big data **3**(1), 1–40 (2016)

76. E.H. Tsai, I. Usov, A. Diaz, A. Menzel, M. Guizar-Sicairos, X-ray ptychography with extended depth of field. Opt. Express **24**(25), 29089–29108 (2016)

77. D.P. Kingma, J. Ba, Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

78. M. Odstrčil, A. Menzel, M. Guizar-Sicairos, Iterative least-squares solver for generalized maximum-likelihood ptychography. Opt. Express **26**(3), 3108–3123 (2018)

79. K. Wakonig, H.-C. Stadler, M. Odstrčil, E.H. Tsai, A. Diaz, M. Holler, I. Usov, J. Raabe, A. Menzel, M. Guizar-Sicairos, Ptychoshelves, a versatile high-level

framework for high-performance analysis of ptychographic data. J. Appl. Crystallogr. **53**(2), (2020)

80. D. Gürsoy, F. De Carlo, X. Xiao, C. Jacobsen, Tomopy: a framework for the analysis of synchrotron tomographic data. J. Synchrotron Radiat. **21**(5), 1188–1193 (2014)

81. D.M. Pelt, D. Gürsoy, W.J. Palenstijn, J. Sijbers, F. De Carlo, K.J. Batenburg, Integration of tomopy and the astra toolbox for advanced processing and reconstruction of tomographic synchrotron data. J. Synchrotron Radiat. **23**(3), 842–849 (2016)