

An improved physics-constrained network for fast, single-shot coherent diffractive imaging

August 28, 2025

Abstract

Ptychographic coherent diffractive imaging (CDI) underpins nanoscale imaging at synchrotron and X-ray free-electron laser facilities, but computational bottlenecks prevent real-time reconstruction. We present a physics-informed neural network that reconstructs experimental diffraction data from the Linac Coherent Light Source (LCLS) and Advanced Photon Source (APS) 100-1000 \times faster than conventional iterative algorithms. The method handles challenging experimental realities—stage jitter, low photon flux (10^4 photons/frame), and sparse scanning (sub-10% overlap)—where conventional algorithms fail to converge or require careful position correction. Our self-supervised framework trains directly on experimental diffraction patterns without requiring ground-truth images and can generalize from as few as 250 training samples compared to several thousand for comparable quality with supervised approaches. We also find that the framework produces capable single-shot, overlap-free reconstructions in Fresnel CDI configurations—a surprising capability that is absent in conventional ptychography solvers. These advances will make real-time coherent imaging feedback feasible at high-repetition-rate light sources such as LCLS 2-HE.

1 Introduction

Ptychographic coherent diffractive imaging has emerged as a cornerstone technique for nanoscale imaging at synchrotron and XFEL facilities, achieving sub-10 nm resolution without the fabrication limits of X-ray optics [?]. However, classical PIE-style iterative phase retrieval typically targets \sim 60–70% scan overlap to ensure robust convergence [?, ?], yet still processes only \sim 0.1–1 diffraction patterns per second on standard hardware; at high-repetition-rate sources (e.g., LCLS-II, MHz-class), acquisition outpaces reconstruction by orders of magnitude [?]. As a result, even optimized GPU/HPC solvers (e.g., SHARP) require substantial infrastructure and careful engineering, limiting low-latency experimental feedback [?, ?].

Recent machine learning approaches have attempted to accelerate ptychographic reconstruction, achieving large speedups [?]. However, many supervised methods struggle to transfer to new experimental settings due to reliance on curated ground truth and sensitivity to experimental artifacts (probe variation, position errors, noise statistics) [?, ?, ?]. *TODO: cite ptychonn nat com ‘machine learning at the edge’ natcom paper as example of the impracticality of these approaches* Our previous PtychoPINN framework [?] demonstrated that physics-informed neural networks can achieve rapid and generalizable self-supervised reconstruction, but was limited to synthetic data with idealized probe functions and regular scan patterns.

In this work, we present practical extensions to PtychoPINN that enable robust reconstruction of experimental data from LCLS and APS. The key advances include: (i) accommodation of realistic probe functions with extended tails, (ii) support for arbitrary scan patterns through

nearest-neighbor clustering, accommodating irregular sampling, (iii) characterization of a Poisson photon-counting likelihood that correctly models dose-dependent statistics [?, ?], and (iv) cross-facility generalization, allowing models trained on one beamline to reconstruct data from another. These modifications transform PtychoPINN from a proof-of-concept to a *practical* tool – perhaps the first in that category – for real-time reconstruction at high-repetition rate light sources.

Beyond acceleration, the enhanced framework enables new imaging modalities previously inaccessible to iterative solvers. In particular, we demonstrate single-shot, overlap-free reconstruction in Fresnel CDI configuration [?, ?]—a capability that could dramatically increase throughput by eliminating the demanding overlap requirements of traditional scanning ptychography [?]. This opens opportunities for high-speed imaging of dynamic processes and dose-sensitive samples at modern light sources.

2 Methods and Architecture

2.1 Formulation

Ptychographic reconstruction inverts the mapping from diffraction intensities to real-space objects. We formulate this as learning an inverse map $G : X \rightarrow Y$ composed with a forward model $F : Y \rightarrow X$, enabling self-supervised training on measured diffraction patterns.

Following the approach developed previously, we decompose $F = F_d \circ F_c$ where $F_c : Y \rightarrow Y$ enforces overlap constraints between adjacent scans and $F_d : Y \rightarrow X$ implements diffraction physics. Each training sample x_i contains C_g diffraction patterns x_i^k at probe positions r_i^k , $k \in \{1, \dots, C_g\}$.

The map $G(x_i, r_i)$ outputs C_g object patches which are then merged into a consistent reconstruction \bar{y}_i . This real-space reduction $N \times N \times C_g \rightarrow N \times N$ is done by F_c . The diffraction operator F_d then computes expected intensities given \bar{y}_i , the probe function $P(r)$, and the scan positions r_i .

The main contributions of the original framework [?] were a neural network implementation and training procedure for G , along with proof of concept results using synthetic datasets.

TODO: right now the transition from mathematical formulation of the problem to our approach is abrupt

2.2 Key Improvements

This original framework was limited to regular scan grids and compact probes. To solve these limitations we now handle the scan positions r_i as explicit input and group diffraction images into multi-image samples based on proximity, relaxing the rigid geometric assumptions embedded in the original forward model (section 2.3). This allows reconstruction of datasets collected in arbitrary (instead of merely rectangular) scan patterns.

Second, we increase the output size of G from $N/2 \times N/2$ to $N \times N$ to avoid truncation artifacts from extended probe tails, which otherwise foils the reconstruction of realistic datasets. We do this in such a way that the well-posedness of the inverse problem – which typically requires zero / constant-padding in real space – is preserved. (Section 2.5.1)

2.3 Solution Region Sampling via Nearest-Neighbor Assignment

The framework handles arbitrary scan patterns by grouping neighboring scan points into overlapping sets. For each scan position in $\{\vec{r}_i\}$, we identify its K nearest neighbors (where $K \geq C_g - 1$) and randomly sample $C_g - 1$ of them to form a training group with the original position. Repeating

this sampling n_{samples} times per position generates multiple training examples, transforming N_{scan} scan points into $N_{\text{scan}} \times n_{\text{samples}}$ samples.

Each sample's coordinates are stored relative to their centroid $\vec{r}_{\text{global}}^i = \frac{1}{C_g} \sum_{k=1}^{C_g} \vec{r}_k$, with local offsets $\vec{r}_{\text{relative},k} = \vec{r}_k - \vec{r}_{\text{global}}^i$ for each patch within the sample.

2.4 Data Preprocessing

A dataset consists of diffraction images from one or more objects measured with a fixed probe illumination P . After grouping images into overlapping samples (Section 2.3), we normalize the raw diffraction amplitudes to ensure numerical stability during training.

The normalization rescales each diffraction pattern's amplitude:

$$x = x' \cdot \sqrt{\frac{(N/2)^2}{\langle \sum_{i,j} |x'_{ij}|^2 \rangle}} \quad (1)$$

where x' denotes raw measurements and the expectation is taken over all images in the dataset. This choice ensures order-unity activations in the neural network: by Parseval's theorem, unit-amplitude real-space objects produce diffraction power of approximately $N^2/4$, so this normalization maps experimental amplitude images to the optimal range of internal activations for gradient-based optimization.¹

Additionally, we introduce a trainable scalar α_{\log} that converts between the dimensionless internal model activations and per-pixel integrated amplitudes. As discussed in Section 2.8, the role of α_{\log} is to convert the output *intensity* into physical units of photons per pixel. The final, scaled, network input is $x_{\text{in}} = x \cdot e^{-\alpha_{\log}}$.

2.5 Neural Network Architecture

The reconstruction network follows the encoder-decoder architecture established in our previous work [?]. It learns a mapping $G : X \rightarrow Y$ from reciprocal- to real-space, inverting a stack of diffraction amplitude images into a stack of complex real-space object patches. The NN structure is unaltered from the original framework, with the exception of the following adaptation:

2.5.1 Handling Extended Probe Illumination

A challenge in CNN-based ptychographic reconstruction arises from the mismatch between probe extent and reconstruction grid size. CNN architectures are limited to modest dimensions ($N \leq 128$) and we must furthermore restrict high-resolution reconstruction to the central $N/2 \times N/2$ region per Nyquist requirements [?]. However, experimental probes commonly have long intensity tails and, given the modest magnitude of N , it is not possible to fully inscribe the probe within the central $N/2 \times N/2$ pixels without excessive binning. This causes a dilemma, as enforcing zero amplitude at the $N/2 \times N/2$ support boundary introduces edge artifacts.

We resolve this by approximating the object across the full $N \times N$ grid using a dual-resolution reconstruction: high-resolution for the central $N/2 \times N/2$ region and low-resolution for the periphery illuminated by probe tails. Specifically, we split the decoder's penultimate layer, using most channels ($C-4$) of the final latent space representation for the central region and the remaining 4 channels to reconstruct the periphery.

¹The factor of 4 arises from the oversampling requirement—the reconstructed object occupies the central $N/2 \times N/2$ region of an $N \times N$ grid.

Explicitly:

$$O_{\text{amplitude}} = \text{Pad}_{N/4}(\sigma_A(\text{Conv}(H_A^{\text{central}}))) + \sigma_A(\text{ConvUpBlock}(H_A^{\text{boundary}})) \odot M_{\text{border}} \quad (2)$$

$$O_{\text{phase}} = \text{Pad}_{N/4}(\pi \cdot \tanh(\text{Conv}(H_\phi^{\text{central}}))) + \pi \cdot \tanh(\text{ConvUpBlock}(H_\phi^{\text{boundary}})) \odot M_{\text{border}} \quad (3)$$

$$O_{\text{patch}} = O_{\text{amplitude}} \cdot \exp(i \cdot O_{\text{phase}}) \quad (4)$$

where H_A^{central} , H_A^{boundary} are the first $C - 4$ and last 4 channels of amplitude decoder features, H_ϕ^{central} , H_ϕ^{boundary} are the corresponding phase decoder feature splits, and M_{border} is a binary mask that isolates the boundary contributions to the outer region. This modification avoids artifacts from artificially truncating the exit wave, leading to superior reconstruction quality with experimentally realistic probes. *should we have a results figure that demonstrates this?*

2.6 Physics-Informed Constraints

2.6.1 Real-Space Overlap Constraints

The constraint map F_c enforces consistency between overlapping measurements through translational average pooling. For a collection of patches $\{O_k\}_{k=1}^{C_g}$ with positions $\{r_k\}_{k=1}^{C_g}$, the pooled reconstruction is:

$$O_{\text{region}}(r) = \frac{\sum_{k=1}^{C_g} \mathcal{T}_{-r_k}[\text{Pad}(O_k)]}{\sum_{k=1}^{C_g} \mathcal{T}_{-r_k}[\text{Pad}(\mathbf{1}_k)] + \epsilon} \quad (5)$$

where Pad denotes zero-padding, \mathcal{T}_r denotes translation by vector r , and $\epsilon = 10^{-3}$ prevents division by zero.

This formulation naturally handles arbitrary overlap patterns.

2.6.2 Diffraction Forward Model

The diffraction map F_d captures the physics of coherent scattering:

For each scan position k , the predicted diffraction amplitude is:

$$O'_k(r) = \text{Crop}_N[\mathcal{T}_{r_k}[O_{\text{region}}]] \quad (6)$$

$$\Psi_k = \mathcal{F}\{O'_k(r) \cdot P(r)\} \quad (7)$$

$$\hat{A}_k = |\Psi_k| \cdot e^{\alpha_{\log}} \quad (8)$$

where $P(r)$ is the probe function, \mathcal{F} is the 2D Fourier transform, and $e^{\alpha_{\log}}$ is an optionally trainable scale parameter that links the model's normalized output to the physical photon counts required by the Poisson loss function. It is derived from the target average total photon count per diffraction image, N_{photons} , and the normalization scheme (Eq. 1) which fixes the average squared L2 norm of the normalized diffraction amplitude, $\langle \|\Psi_k\|/N \rangle_2^2 \approx (N/2)^2$. Setting the total expected intensity to this target, $N_{\text{photons}} = \langle \|\hat{A}_{ik}\|_2^2 \rangle_{i,k} \approx (e^{\alpha_{\log}})^2 (N/2)^2$, yields the following initialization for the scale parameter:

$$e^{\alpha_{\log}} \leftarrow 2\sqrt{N_{\text{photons}}}/N. \quad (9)$$

TODO if we dont know nphotons this is meaningless and we fall back to MAE etc.

2.7 Boundary Constraints and Physical Priors

2.7.1 Object Boundary Handling

The framework provides two approaches for handling object boundaries:

- **Adaptive boundary learning** (`pad_object=True`): Neural network predicts the amplitude and phase of O_{patch} across the entire $N \times N$ real-space grid, using low-resolution estimates for pixels outside the well-sampled $N/2 \times N/2$ central region.
- **Zero padding** (`pad_object=False`): Assumes zero amplitude outside the central patches (typically not robust on real data).

Adaptive boundary learning accommodates experimental probes with intensity profiles extending beyond the central $N/2 \times N/2$ region, a common occurrence in practical experimental settings.

2.8 Loss Function Formulation

The training procedure optimizes the inverse map G using a negative log-likelihood loss under Poisson statistics:

$$\mathcal{L}_{\text{NLL}} = - \sum_{i,j,k} \log f_{\text{Poiss}}(N_{ijk}; \lambda_{ijk}) \quad (10)$$

where $N_{ijk} = |x'_{ijk}|^2$ is the measured photon count and $\lambda_{ijk} = |\hat{A}_{ijk}|^2$ is the predicted count.

Since the network operates on normalized inputs (Eq. 1) for numerical stability, a scale parameter $e^{\alpha_{\log}}$ bridges normalized and physical units. When the mean photon flux N_{photons} is known, we initialize:

$$e^{\alpha_{\log}} = 2\sqrt{N_{\text{photons}}}/N \quad (11)$$

ensuring predicted intensities match measurement statistics. For datasets lacking absolute photon counts, we employ mean absolute error on normalized amplitudes: $\mathcal{L}_{\text{MAE}} = \sum_{i,j,k} |x_{ijk} - \hat{x}_{ijk}|$.

Appendix: Key Configuration Parameters

These parameters control critical aspects of the reconstruction process and should be tuned based on experimental conditions and computational constraints.

3 Results

We evaluate PtychoPINN on experimental data from the Advanced Photon Source (APS) and Linac Coherent Light Source (LCLS), comparing against iterative solvers and supervised neural network baselines. We also provide quantitative comparisons based on reconstructions of simulated datasets, which allows a genuine comparison to ground truth for both classes of reconstruction methods (neural network and conventional / iterative).

3.1 Reconstruction Quality

Figure ?? compares PtychoPINN and PtychoNN on APS Siemens star data. For this evaluation, we used the top half of the Siemens star pattern for training and the bottom half for testing, demonstrating in-distribution generalization. PtychoPINN achieves SSIM = for amplitude and for phase, compared to the supervised baseline's and respectively. The improvement is most pronounced in phase reconstruction, where PtychoNN produces artifacts in low-scattering regions.

Table 1: Critical implementation parameters and their default values

Parameter	Default	Description
N	64	Patch dimension (pixels)
C_g	4	Number of patterns per group
K	7	Nearest neighbors for grouping
nsamples	1	Random samplings per scan point
pad_object	True	Enable adaptive boundary learning
probe.mask	True	Apply circular probe support
gaussian_smoothing_sigma	0.0	Probe boundary smoothing
intensity_scale.trainable	False	Learnable intensity scaling
n_filters_scale	1	Network width multiplier
amp_activation	sigmoid	Amplitude decoder activation
offset	4	Scan step size (pixels)
d	3-5	Encoder depth (resolution-dependent)
C	132	Total decoder channels (before split)
C _{latent}	128	Latent channels at bottleneck

3.2 Overlap-Free Reconstruction

Focused probe configurations wherein the [probe waveform is spherically curved] make single-shot reconstruction possible, in principle. Conventional solvers don't support this (COMMENT does the comparison to conventional solvers belong in discussion only, not here?) but one can make use of this in PtychoPINN by simply configuring $C_g = 0$. Figure ?? shows this [hybrid vs idealized with and wo overlap, wo overlap is only a bit worse].

3.3 Data Efficiency

Figure ?? plots reconstruction quality versus training set size. PtychoPINN exhibits superior reconstruction quality at approximately one-tenth the training set size compared to PtychoNN. [TODO it should be explained, maybe in a footnote, that this factor of 10 comes from averaging the horizontal distance between the two curves in fig:ssim_training_size. "The horizontal shift between the two curves in Figure ?? quantifies this advantage: at any given SSIM threshold, PtychoPINN requires an order of magnitude less training data than the supervised baseline."]. This demonstrates our self-supervised framework's significantly more favorable in-distribution generalization from limited data.

3.3.1 Sparse Sampling robustness

Figure ?? presents a three-way comparison of PtychoPINN, the supervised ML baseline (PtychoNN), and conventional rPIE on a very small dataset comprising a subsample of 512 diffraction images from the Siemens star test pattern [TODO again, this is a vertical train test split so the ml models were trained on just 256 patterns. this should be mentioned]. This qualitative test reveals the failure modes of different reconstruction methods when faced with extremely small, sparsely sampled datasets. At this sampling density, conventional rPIE is insufficiently constrained in real

Table 2: Mathematical symbols and conceptual descriptions

Symbol	Type / Structure	Description
x'	Set of C_g real images	Raw diffraction patterns for one sample
x	Set of C_g real images	Normalized diffraction patterns for one sample
\vec{r}_k	2D Position Vector	Absolute scan position for the k -th image within a sample
\vec{r}_{global}	2D Position Vector	Centroid of a solution region (group of scans)
$\vec{r}_{\text{relative},k}$	2D Offset Vector	Relative scan offset within a solution region
$e^{\alpha_{\log}}$	Scalar (trainable)	Internal log-intensity scale parameter
N_{photons}	Scalar	Target average total photons per diffraction pattern
$P(\vec{r})$	$N \times N$ complex array	Effective probe function
O_k	$N \times N$ complex array	k -th object patch decoded by the network G
O_{region}	$M \times M$ complex array	Merged object representation for a solution region
O'_k	$N \times N$ complex array	Object patch extracted from O_{region} for forward model
Ψ_k	$N \times N$ complex array	Predicted complex wavefield at the detector
\hat{A}_k	$N \times N$ real array	Predicted final diffraction amplitude for one patch
λ_{ijk}	Scalar	Poisson rate parameter for a single pixel

N: patch dimension, C_g : patches per group, M : merged region size

Table 3: Reconstruction quality metrics at maximum training set size (16,384 images). Values shown are mean \pm standard deviation across 5 trials. Best values per dataset are highlighted in green.

Dataset	Method	PSNR (dB)		MS-SSIM	
		Amplitude	Phase	Amplitude	Phase
fly001-synthetic	Baseline (supervised)	84.83 ± 0.23	68.62 ± 0.02	0.930 ± 0.002	0.912 ± 0.003
	PINN (self-supervised)	85.53 ± 0.02	70.54 ± 0.06	0.955 ± 0.001	0.962 ± 0.001

space and fails to converge entirely, producing only noise. The supervised PtychoNN baseline also fails, generating extremely blurred images. [TODO language, flow] In contrast, PtychoPINN successfully reconstructs both amplitude and phase with clearly resolved features, demonstrating robust generalization from extremely sparse data.

3.4 Out-of-distribution Generalization

[TODO: abrupt, flow] We trained models on APS data and tested on LCLS measurements without retraining. Notably, the underlying illumination profiles are completely different [TODO refer to insets in fig:cross_facility or a separate figure]. The resulting reconstructions provide a qualitative, but stringent, test of out of distribution generalization. [TODO I'll extend the fig by adding an in-distribution leg (train on LCLS, recon on LCLS. so it'll be top row in distribution, bottom row out of distribution, left column ptychopinn, center column ptychonn, right column ground truth]. The supervised PtychoNN baseline, trained on APS, fails completely on LCLS data (Figure 2), producing . This out-of-distribution robustness stems from physics-informed training.

Table 4: Reconstruction metrics at 16,384 training images (mean \pm std, n=5)

Dataset	Method	PSNR (dB)		MS-SSIM	
		Amp.	Phase	Amp.	Phase
fly001-syn.	Base. (sup.)	84.83(23)	68.62(2)	0.930(2)	0.912(3)
	PINN (self.)	85.53(2)	70.54(6)	0.955(1)	0.962(1)

3.5 Photon-Limited Performance

Table ?? compares Poisson NLL versus MAE loss across photon doses. At 10^3 photons/pattern, NLL maintains SSIM = while MAE drops to . Figure ?? shows the NLL-trained model preserves features that MAE loses to noise.

3.6 Computational Performance

PtychoPINN processes 2000 diffraction patterns in 1 second on an NVIDIA RTX 3090, versus [TODO: find typical times for ePIE / rPIE and similar algorithms on GPU and insert supporting cite(s)], a [TODO speedup] speedup.

rPIE fails to converge due to

To compare the failure modes

In this separate study, we wanted to see Looking at this other fig

4 Discussion

4.1 New Imaging Modalities

PtychoPINN’s ability to reconstruct from sparse, low-overlap data fundamentally changes the constraints on ptychographic experiments. Conventional iterative solvers require 60-70% overlap between adjacent scan positions to converge reliably [?, ?]. This requirement directly limits throughput—most of the beam time is spent re-measuring the same sample regions. Our results demonstrate successful reconstruction from 5% overlap and even overlap-free configurations, potentially increasing area coverage rates by an order of magnitude.

The single-shot reconstruction capability extends PtychoPINN’s applicability to Fresnel CDI [?, ?], where each diffraction pattern corresponds to a unique sample position without overlap. While Fresnel CDI inherently operates without overlap constraints, it traditionally requires prior knowledge or constraints to resolve phase ambiguities. PtychoPINN leverages its learned understanding of diffraction physics to reconstruct from individual Fresnel diffraction patterns, effectively bridging ptychographic and single-shot CDI modalities. This unification opens new experimental designs that can switch between scanning and single-shot modes based on sample requirements.

4.2 Dose-Efficient Imaging

The Poisson loss formulation provides principled handling of photon statistics, with practical implications for dose-sensitive samples. At doses of 10^4 photons per pattern Poisson-loss training results in reconstructions of comparable quality to 10^5 photons per pattern with the more typical L1 reconstruction error objective (i.e., MAE). This improvement stems from correct accounting of photon-counting shot noise, without which pixelwise reconstruction errors will be incorrectly scaled.

The MAE loss, in contrast, implicitly assumes homoscedastic noise. This causes an insensitivity of the training loss to the weak signal in high- q regions of the diffraction image.

4.3 Computational Implications for Modern Light Sources

Recent advances in X-ray source technology have dramatically increased data generation rates. LCLS-II-HE, most notably, will deliver up to TODO repetition rates [?]. Parallel developments in experimental techniques—including multi-probe ptychography using diffraction gratings—further multiply data rates *TODO cites*. The throughput limitations of existing approaches force either reduced acquisition rates or batched offline processing that precludes experimental steering.

TODO transition and summary / take-home message

4.4 Limitations and Future Directions

TODO: stage jitter, lack of stochastic modeling, resolution limitations

The cross-facility generalization demonstrates that physics-informed training captures fundamental diffraction physics rather than dataset-specific features. This principle should extend to related inverse problems in computed imaging, such as Bragg CDI and reflection ptychography *TODO cites*. Future work could leverage this generalization by training on combined datasets from multiple facilities and modalities.

TODO: decide how much to emphasize comparison to ML baseline. This was already covered pretty well in the first paper. On the other hand it's good to emphasize the relative generalizability and self-supervised training, since this is what makes ptychopinn practical to use where previous ml-based solvers were not.

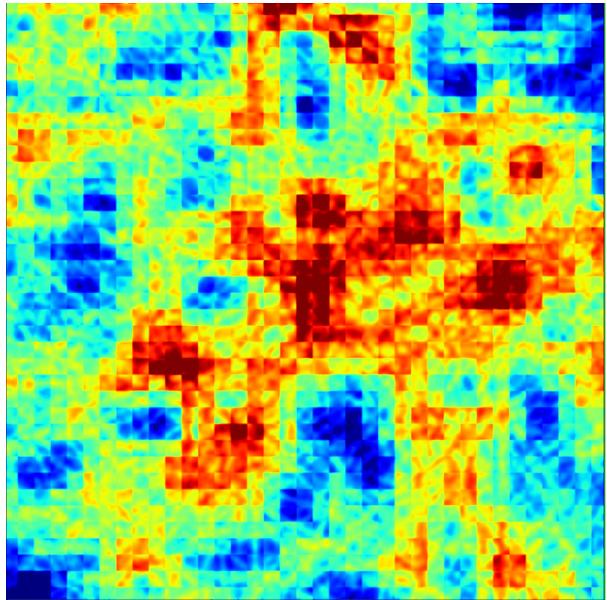
cites and discussion: papers recommended by Junjin Deng. We expect adoption at APS / velociprobe, which might influence the project's technical direction.

results:

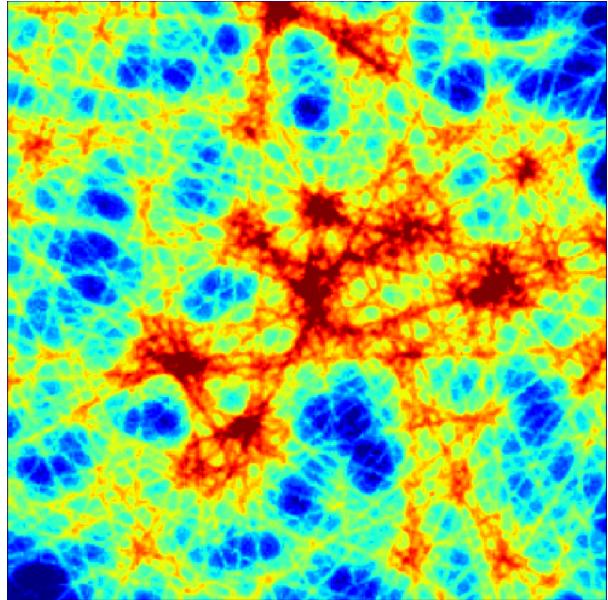
References

- [1] Aswin Varghese Babu, Tao Zhou, Swarnendu Kandel, Vinayak Banakha, Kang Li, Sha Li, Junjing Deng, Youssef S. G. Nashed, Ross Harder, Mathew J. Cherukara, and Charudatta Phatak. Deep learning at the edge enables real-time streaming ptychographic imaging. *Nature Communications*, 14:7059, 2023.
- [2] Oliver Bunk, Martin Dierolf, Søren Kynde, Ian Johnson, Olof Marti, and Franz Pfeiffer. Influence of the overlap parameter on the convergence of the ptychographical iterative engine. *Ultramicroscopy*, 108(5):481–487, 2008.
- [3] Mathew J. Cherukara, Tao Zhou, Youssef S. G. Nashed, Alexander Hexemer, Ross Harder, Oleg G. Shpyrko, Jun Wang, Sunil K. Sinha, and Subramanian K. R. S. Sankaranarayanan. Ai-enabled high-resolution scanning coherent diffraction imaging. *Applied Physics Letters*, 117(4):044103, 2020.
- [4] Manuel Guizar-Sicairos and Pierre Thibault. Ptychography: A solution to the phase problem. *Physics Today*, 74(9):42–48, 2021.
- [5] Oliver Hoidn, Aashwin Ananda Mishra, and Apurva Mehta. Physics constrained unsupervised deep learning for rapid, high resolution scanning coherent diffraction reconstruction. *Scientific Reports*, 13:22789, 2023.

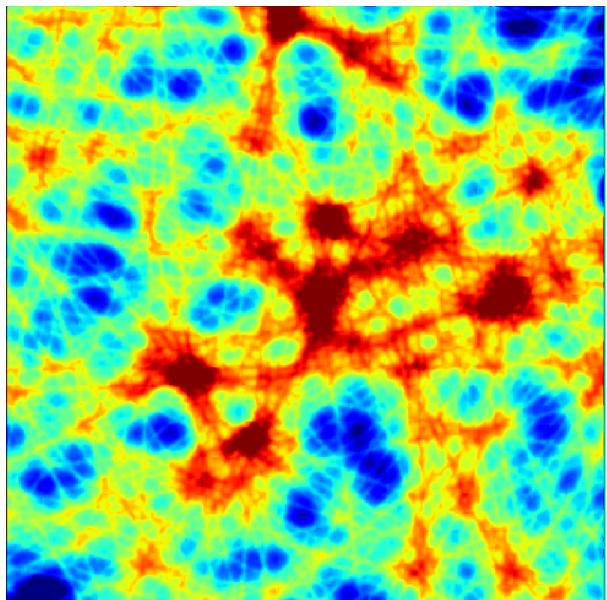
- [6] Andrew M. Maiden, Matthew J. Humphry, Matthew C. Sarahan, Bernd Kraus, and John M. Rodenburg. An annealing algorithm to correct positioning errors in ptychography. *Ultramicroscopy*, 120:64–72, 2012.
- [7] Andrew M. Maiden and John M. Rodenburg. An improved ptychographical phase retrieval algorithm for diffractive imaging. *Ultramicroscopy*, 109(10):1256–1262, 2009.
- [8] Stefano Marchesini, Hari Krishnan, Benedikt J. Dauner, David A. Shapiro, Talita Perciano, James A. Sethian, and Filipe R. N. C. Maia. Sharp: a distributed gpu-based ptychographic solver. *Journal of Applied Crystallography*, 49(4):1245–1252, 2016.
- [9] Jianwei Miao, Pambos Charalambous, Janos Kirz, and David Sayre. Extending the methodology of x-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens. *Nature*, 400(6742):342–344, 1999.
- [10] Joachim P. Seifert, Zhaoning Chen, Myung-Jae Yoon, Andrew Ulvestad, Mathew J. Cherukara, Youssef S. G. Nashed, Savannah M. Wild, and Ross Harder. Maximum-likelihood ptychography in the presence of poisson-gaussian noise. *Optics Letters*, 48(18):4897–4900, 2023.
- [11] SLAC National Accelerator Laboratory. LCLS-II-HE: Design and Performance. <https://lcls.slac.stanford.edu/lcls-ii-he/design-and-performance>, 2023. Accessed: 2025-08-14.
- [12] Pierre Thibault and Manuel Guizar-Sicairos. Maximum-likelihood refinement for coherent diffractive imaging. *New Journal of Physics*, 14:063004, 2012.
- [13] G. J. Williams, H. M. Quiney, B. B. Dhal, C. Q. Tran, K. A. Nugent, A. G. Peele, D. Patterson, and M. D. de Jonge. Fresnel coherent diffractive imaging. *Physical Review Letters*, 97(2):025506, 2006.
- [14] Fan Zhang, Irène Peterson, Joan Vila-Comamala, Ana Diaz, François Berenguer, Richard Bean, Bo Chen, Andreas Menzel, Ian K. Robinson, and John M. Rodenburg. Phase retrieval by coherent modulation imaging. *Nature Communications*, 7:13367, 2016.
- [15] Fei Zhang, Irène Peterson, Joan Vila-Comamala, Ana Diaz, François Berenguer, Richard Bean, Bo Chen, Andreas Menzel, Ian K. Robinson, and John M. Rodenburg. Translation position determination in ptychographic coherent diffraction imaging. *Optics Express*, 21(11):13592–13606, 2013.



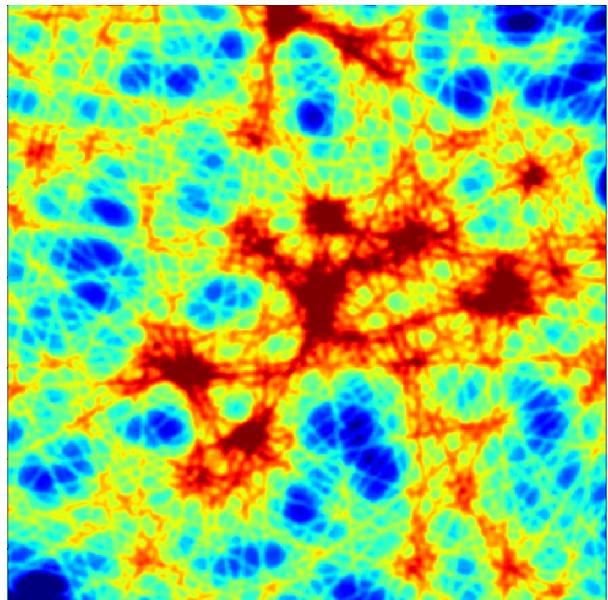
(a) Idealized — CDI



(b) Idealized — Ptycho



(c) Hybrid — CDI



(d) Hybrid — Ptycho

Figure 1: Reconstruction comparison. Rows: Idealized vs Hybrid probe. Columns: CDI vs Ptycho.

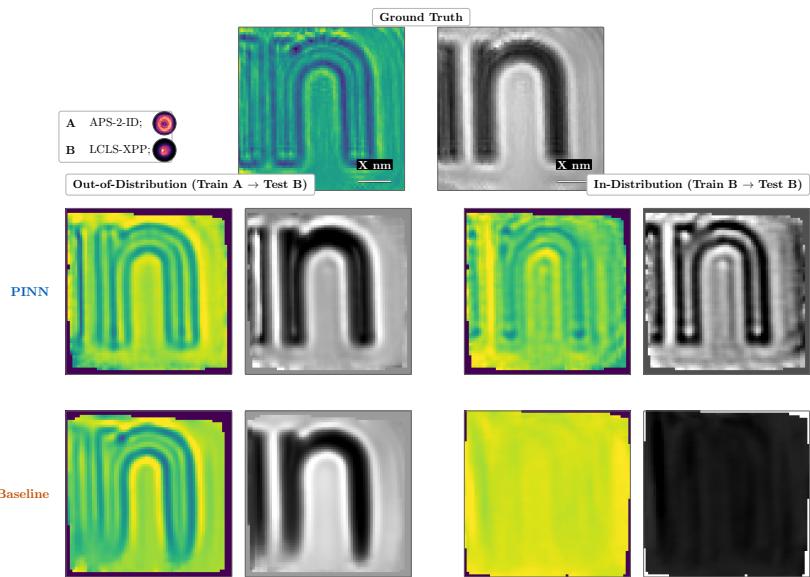
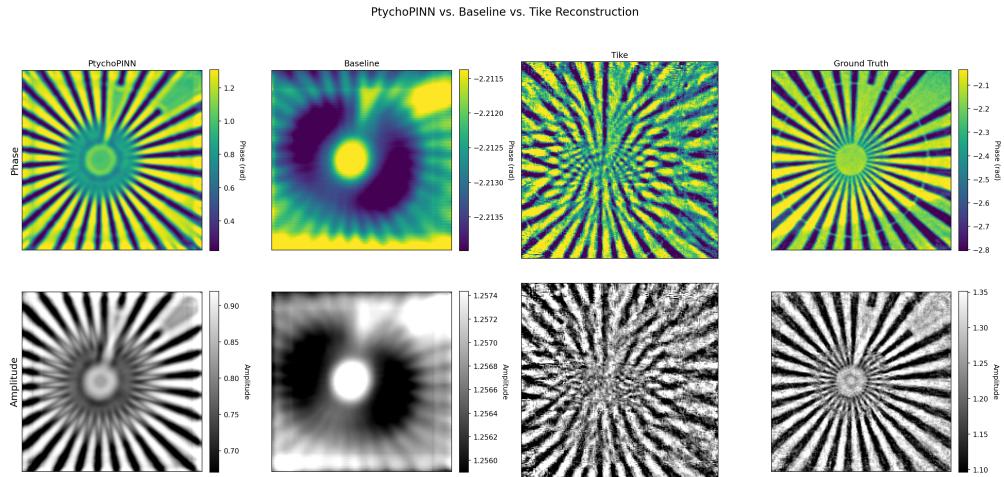
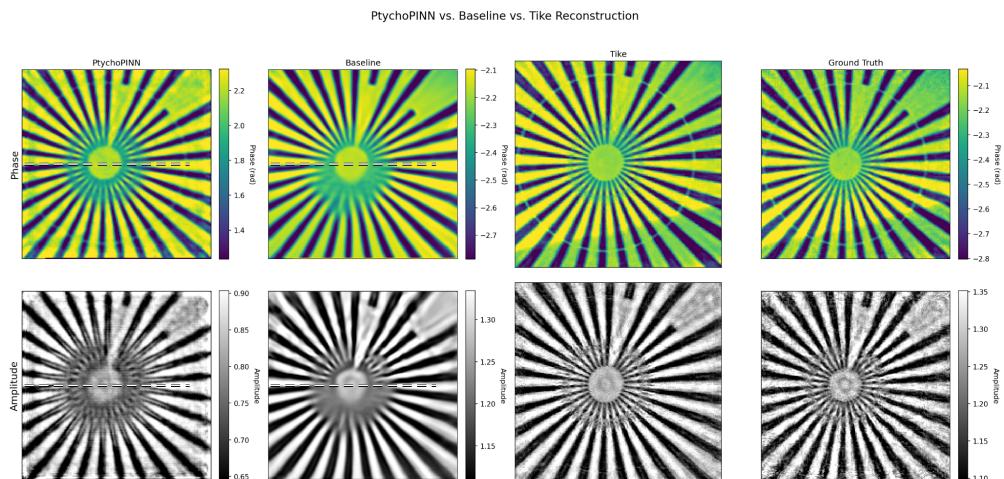


Figure 2: Comparison of methods across in- and out-of-distribution cases.



(a) 512 diffraction patterns of the Siemens star test pattern.



(b) 8192 diffraction patterns of the Siemens star test pattern.

Figure 3: Comparison of reconstruction quality with different numbers of diffraction patterns.

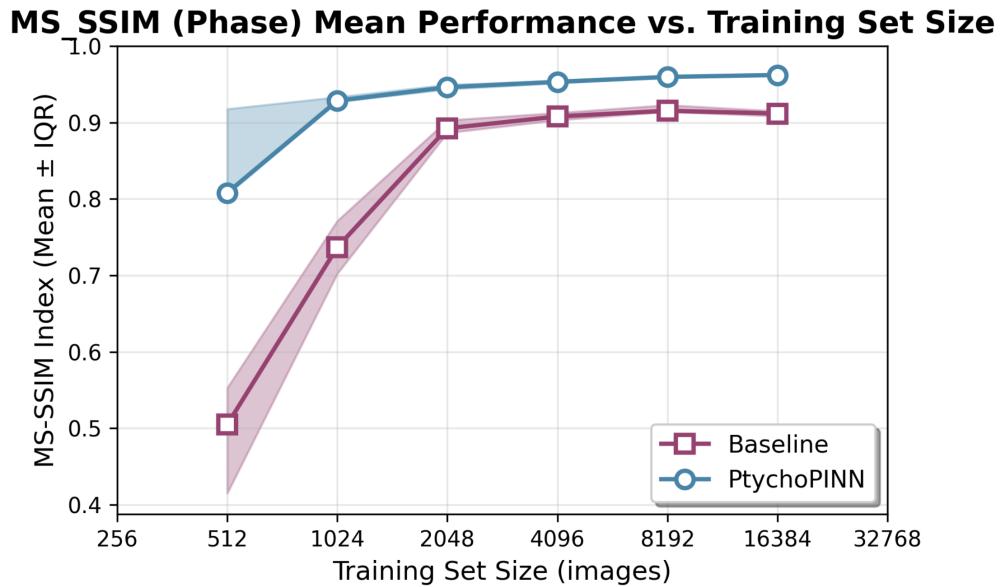


Figure 4: Structural similarity of PtychoPINN, conventional reconstruction (rPIE in Tike), and baseline as a function of training set size.

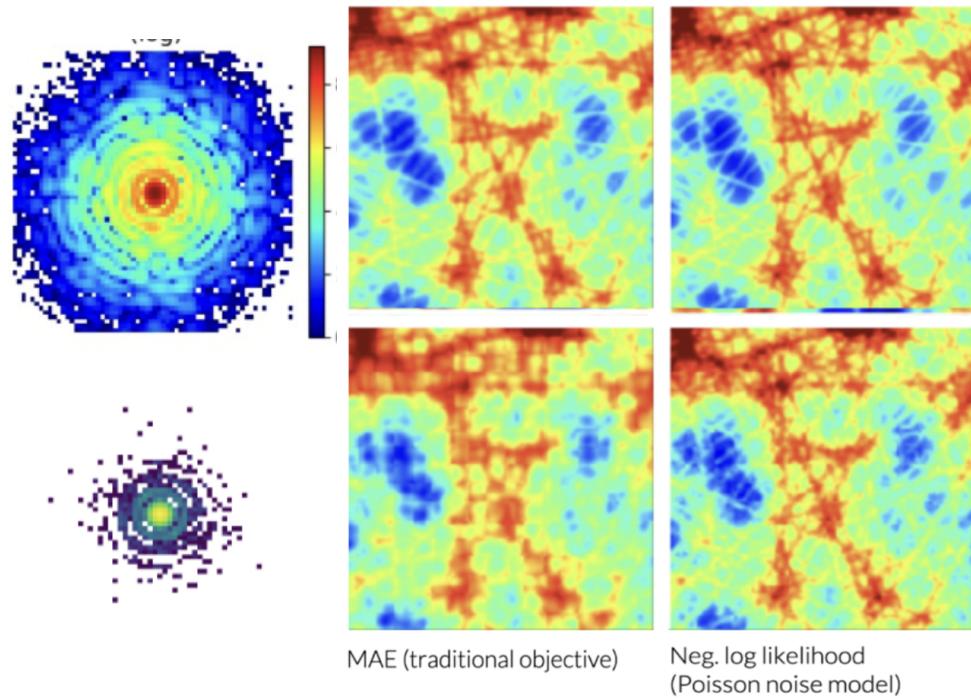


Figure 5: An example figure demonstrating how to insert images in LATEX. *TODO: remake this figure with better resolution and layout*