



## OPEN Physics constrained unsupervised deep learning for rapid, high resolution scanning coherent diffraction reconstruction

Oliver Hoidn , Aashwin Ananda Mishra & Apurva Mehta

By circumventing the resolution limitations of optics, coherent diffractive imaging (CDI) and ptychography are making their way into scientific fields ranging from X-ray imaging to astronomy. Yet, the need for time consuming iterative phase recovery hampers real-time imaging. While supervised deep learning strategies have increased reconstruction speed, they sacrifice image quality. Furthermore, these methods' demand for extensive labeled training data is experimentally burdensome. Here, we propose an unsupervised physics-informed neural network reconstruction method, PtychoPINN, that retains the factor of 100-to-1000 speedup of deep learning-based reconstruction while improving reconstruction quality by combining the diffraction forward map with real-space constraints from overlapping measurements. In particular, PtychoPINN gains a factor of 4 in linear resolution and an 8 dB improvement in PSNR while also accruing improvements in generalizability and robustness. This blend of performance and computational efficiency offers exciting prospects for high-resolution real-time imaging in high-throughput environments such as X-ray free electron lasers (XFELs) and diffraction-limited light sources.

Coherent diffractive imaging (CDI) is a state-of-the-art imaging technique that uses diffraction from a coherent beam of light or electrons to reconstruct an image of a specimen without the need for optics. This microscopy technique is useful because it overcomes the lens aberrations of traditional lens-based imaging, thus allowing imaging at finer scales than previously possible. As a general approach CDI has found application in a broad range of settings, including nanoscale imaging through Bragg coherent diffractive imaging reconstruction (BCDI), X-ray ptychography, optical super-resolution, and astronomical wavefront sensing<sup>1-3</sup>.

The main challenge of CDI, known as the phase retrieval problem, originates from the fact that detectors only record the intensity (i.e., squared amplitude) of the diffracted wave, not its phase. The phase carries essential information about the illuminated real-space object, and its loss stops the direct calculation of an image from the recorded diffraction. A breakthrough development over two decades ago outlined an iterative approach to solve the inverse problem of phase retrieval and made CDI image reconstruction possible<sup>4</sup>. Unfortunately, these iterative algorithms are slow and typically computationally expensive<sup>5</sup>, which consequently precludes CDI in high-throughput or in situ settings, such as at x-ray free electron lasers (XFELs). In addition, such methods often suffer from noise sensitivity and limitations in robustness.

A considerable body of literature focuses on applying deep learning (DL) to inverse problems, with significant success in employing neural networks (NNs) to solve the CDI phase problem much more rapidly than conventional iterative methods<sup>6-8</sup>. Early efforts used supervised learning to train these deep learning models and achieved several-orders of magnitude speed improvements, accompanied by two major drawbacks: degradation in reconstruction quality and the need for large volumes of high-quality labeled training data. More recently, strategies such as incorporating the diffraction forward map into deep learning models have been introduced to eliminate the requirement for labels.

Despite the promise of these developments, existing physics-informed neural networks (PINNs) for CDI have not explored physical priors or constraints beyond the diffraction forward map and discrete Fourier transform (DFT) sampling requirements, nor have they modeled any stochasticity of the relevant physics. Both these approaches introduce Inductive Biases in the learning algorithm<sup>9,10</sup>, enabling better generalization to test samples outside the explicit training set while also reducing the need for large corpora of training data. As an illustration, incorporation of physics based knowledge via hard constraints reduces the region in solution space to be explored

SLAC National Accelerator Laboratory, Menlo Park, CA, USA. ✉email: ohoidn@slac.stanford.edu

and thus reduces the need for large training datasets<sup>11</sup>. Similarly, incorporating uncertain or incomplete domain-specific knowledge via the use of probabilistic loss functions leads to improved convergence and generalizability<sup>11</sup>. In this light, it is conceivable that incorporating additional physics information in the model architecture and developing principled probabilistic loss functions may enhance both accuracy as well as generalizability, while also maintaining the speed of prior supervised learning approaches.

In this work, we develop the PINN computational approach to CDI, specifically for ptychographic reconstruction. Ptychography is defined as a conventional CDI technique with overlap, originally introduced to calculate the phase of Bragg reflections from crystals. It represents an amalgam between Scanning Transmission X-ray Microscopy (STXM) and Coherent Diffraction Imaging (CDI). Ptychography utilizes convolutions in Fourier space between an object's diffraction pattern and the transform of the illumination function. This involves the measurement of manifold diffraction patterns via the scanning of an X-ray probe over the specimen, while using the overlap between adjacent illuminations for over-determination. In this light, it is often also referred to as scanning CDI. When coupled with reconstruction approaches, this leads to robust computational imaging.

In this study, we integrate three novel elements for Machine Learning based approaches to ptychography: unsupervised training using the diffraction forward map, additional physics-based constraints informed by the ptychography setup, and an explicitly probabilistic treatment of photon counting (Poisson) statistics. We find that this unique combination of model features merges the advantages of a standard PINN approach, namely speed and unsupervised training, with significantly better generalizability and reconstruction accuracy than other NN-based solvers, including physics-informed approaches.

## Methods, models and tests

### Approach

Physics-based CDI reconstruction methods are accurate because they invert the physically correct forward map of far field diffraction, making them capable of finding the optimal solution, in principle, for any input. However, due to the difficult nonconvex optimization problem, these methods require computationally expensive iterative solution schemes. Additionally, iterative reconstruction is oblivious to regularities in the input data, so each new diffraction signal requires computation commencing from scratch. In contrast, NN-based reconstruction methods take a different approach: they may not incorporate domain knowledge of the diffraction physics but instead rely on a large amount of training data to fit a flexible black box model from scratch. The inductive biases introduced in the learning algorithm are chiefly from the choice of the model architecture. For instance, the use of convolutional and pooling layers renders the final model approximately translation invariant. The lack of domain knowledge-based inductive biases and physical consistency-enforcing constraints in conventional NN-based methods cause them to have reduced accuracy and generalization with respect to the underlying physics, although they may capture particular data regularities well. Additionally, conventional neural networks are trained via supervised learning approaches. Thus, they require large corpora of labeled data for training, and the limited diversity of the training data brings in bias. Nonetheless, the single forward pass nature of NN-based methods makes them intrinsically rapid at inference time.

From this perspective, physics-informed neural networks (PINNs) attempt to unite the best of both worlds (Archetypal PINNs utilize a soft-constraint on the solution space by using the residual of the governing Partial Differential Equation as a regularization term. Our PINN model incorporates domain physics information in the architecture of the model). By strongly constraining a neural network model's hypothesis space to exclude parameter combinations that generate unphysical solutions, we can predispose a model towards physically correct solutions while also considerably reducing the need for training data. As a concrete starting point, defining the model's loss function over the forward-mapped (i.e., far field-diffracted) NN output – instead of the immediate NN output – forces the NN to learn diffraction physics rather than merely fit *a priori* arbitrary input/output pairs. This is the foundation for prior PINN approaches for unsupervised CDI reconstruction<sup>6,7</sup>.

### Map formulation

With this background in place, we develop a new approach for ptychographic reconstruction based on a combination of NN layers, explicit constraints, and the forward map of far-field diffraction.

To begin, the reconstruction problem requires approximating a mapping  $G : X \rightarrow Y$  from the diffraction/reciprocal-space domain  $X$  to the real-space domain  $Y$ . Because we wish to avoid supervised training, we rely on an autoencoder formalism that composes  $G$  with a second mapping  $F : Y \rightarrow X$ . The output of the autoencoder is then  $\hat{x} = F(G(x))$ , where  $x$  is the (complex) diffracted wave field (The measured diffraction intensity is therefore  $I \propto |x|^2$ ).

In prior PINN approaches to CDI reconstruction,  $F$  is typically the forward map of far-field diffraction. Here, we instead start by subdividing  $F$  into two parts: a *constraint* map  $F_c : Y \rightarrow Y$  and a diffraction map  $F_d : Y \rightarrow X$ , such that  $F(Y) = F_d(F_c(Y))$ . In most practical settings, diffraction amplitudes are centrosymmetric even for non-centrosymmetric objects. To break this centrosymmetry and make the inversion well-posed,  $F_c$  imposes real-space constraints derived from overlapping diffraction measurements.  $F$  depends on parameters  $\theta$  of the experimental geometry (including the probe position), so we relate it to a functional  $\mathbf{F}$ , such that  $F = \mathbf{F}(\theta)$ . Notably, this formulation omits explicit dependence on the probe illumination function  $P(r)$  of the diffraction map  $F_d$  because we assume  $P(r)$  to be known. In the analysis of simulated data we must therefore provide the ground truth  $P(r)$  to the model, while for experimental data it is necessary to estimate  $P(r)$  using an iterative solver such as ePIE.

In our concrete implementation, an entire training set is composed of a collection of diffraction images grouped into samples of four  $64 \times 64$  diffraction images each. We use  $i$  to index individual samples in a dataset and  $k$  ( $k \in \mathcal{K} = \{0, 1, 2, 3\}$ ) to index individual images in a sample; samples and images are noted  $x_i$  and  $x_i^k$ ,

respectively. In general the samples may be overlapping (i.e., an individual image appears in more than one sample), but the only requirement for training is that images within a single  $64 \times 64 \times 4$  sample,  $\{x_i\}$ , overlap in real space. (For the particular datasets presented in this paper the measurement pattern of probe positions over the object is a grid, and the assignment of images to  $64 \times 64 \times 4$  samples corresponds to sliding a  $2 \times 2$  window over the grid of scan point positions).

Each sample  $x_i$  is paired with its matching 2D Euclidean probe coordinates,  $r_i$  ( $r_i \in \theta$ ). Correspondingly, the reconstruction  $y_i = G(x_i)$  is a 3D tensor of shape  $32 \times 32 \times 4$ . (The factor-of-two difference in size between  $x_i$  and  $y_i$  satisfies the oversampling requirement for invertibility of the discrete Fourier transform<sup>12</sup>). Finally, to distinguish between the two real-space representations we let  $\hat{y}_i = F_c(y_i)$ . We denote the amplitude and phase of  $\hat{y}$  as  $A$  and  $\phi$ , respectively, throughout this paper.

The inverse map  $G$  consists of an encoder-decoder architecture with a similar structure to PtychoNN and  $F_d$  is mostly defined by diffraction physics. The element of most interest is  $F_c$ , which we detail below.

#### Real-space constraints

In CDI reconstruction the phase problem manifests as invariances of the diffraction amplitude to coordinate inversion and translation of the real-space object. In the case of scanning CDI this must be solved using real-space constraints based on overlapping measurements.

In order to detail how the real-space constraint map,  $F_c(r_i)$ , is applied to  $y_i$ , we can break it down into two steps. Initially, we sum the individual  $y_i^k$  to create a unified reconstruction  $\hat{y}_i$  with a shape of  $64 \times 64$ :

$$\hat{y}_i = \sum_{k \in \mathcal{K}} T(r_i^k - \mu_i, \text{Pad2d}(y_i^k)) \oslash \sum_{k \in \mathcal{K}} T(r_i^k - \mu_i, \text{Pad2d}(\mathbf{1})), \quad (1)$$

where  $\text{Pad2d}$  denotes zero-padding a  $32 \times 32$  tensor to  $64 \times 64$  and  $T(\delta r, y)$  is the translation of a 2D tensor  $y$  by the vector  $\delta r$ .  $\mu_i$  is the origin of a local coordinate system for the group of scan points and is normally set to their centroid position, for convenience.  $\oslash$  represents elementwise division, while  $\mathbf{1}$  stands for a  $32 \times 32$  tensor composed entirely of ones. Therefore, the divisor  $\sum_{k \in \mathcal{K}} T(r_i^k - \mu_i, \text{Pad2d}(\mathbf{1}))$  has the role of a normalizing tensor.

The purpose of this first transformation is to construct a larger real-space image  $\hat{y}_i$  that encompasses the entire solution region. The second, and arguably more critical, step enforces translational symmetry of the reconstruction via shifts in the relative probe position. This step also computes the exit wave, illuminating the shifted 2D object  $\hat{y}_i$  with  $P(r)$  and transforming it into a 3D object that matches the required input format of  $F_d$ :

$$\hat{y}_i^k = T(\mu_i - r_i^k, \hat{y}_i) P(r) \approx O(r) P(r - r_i^k). \quad (2)$$

Here,  $O(r)$  represents the ground-truth object, with the domain of its discrete approximation (left hand side) confined to a  $32 \times 32$  patch centered at  $r_i^k$ .

We can compare this approach to iterative scanning CDI reconstruction schemes, which enforce real-space constraints in a different way. In methods such as ePIE, the optimization loop is organized around alternating error corrections in real and reciprocal space. In the 'backpropagation' step of ePIE, the error in the reconstructed diffraction amplitude for one measurement – equivalent to our formulation's  $|x_i^k| - |\hat{x}_i^k|$  – induces an update of the  $O(r)$  guess via inverse DFT. In the training of PtychoPINN, errors from the  $|\mathcal{K}|$  individual patterns are converted into model parameter updates instead of reconstruction updates, and this is done in parallel instead of sequentially.

#### Probabilistic output and loss function

We choose the model output to be a collection of independent Poisson distributions parameterized by the forward-mapping of the final-layer output of the NN. This reproduces the correct photon-counting statistics of the diffraction measurement and allows us to calculate a likelihood, with respect to the Poisson parameters, for the distribution of per-pixel detected photon counts. The negative log over these Poisson likelihoods is *a priori* a more principled loss function than the typical choice of mean absolute average (MAE) deviation between target and predicted pixel values.

Explicitly, the loss function is

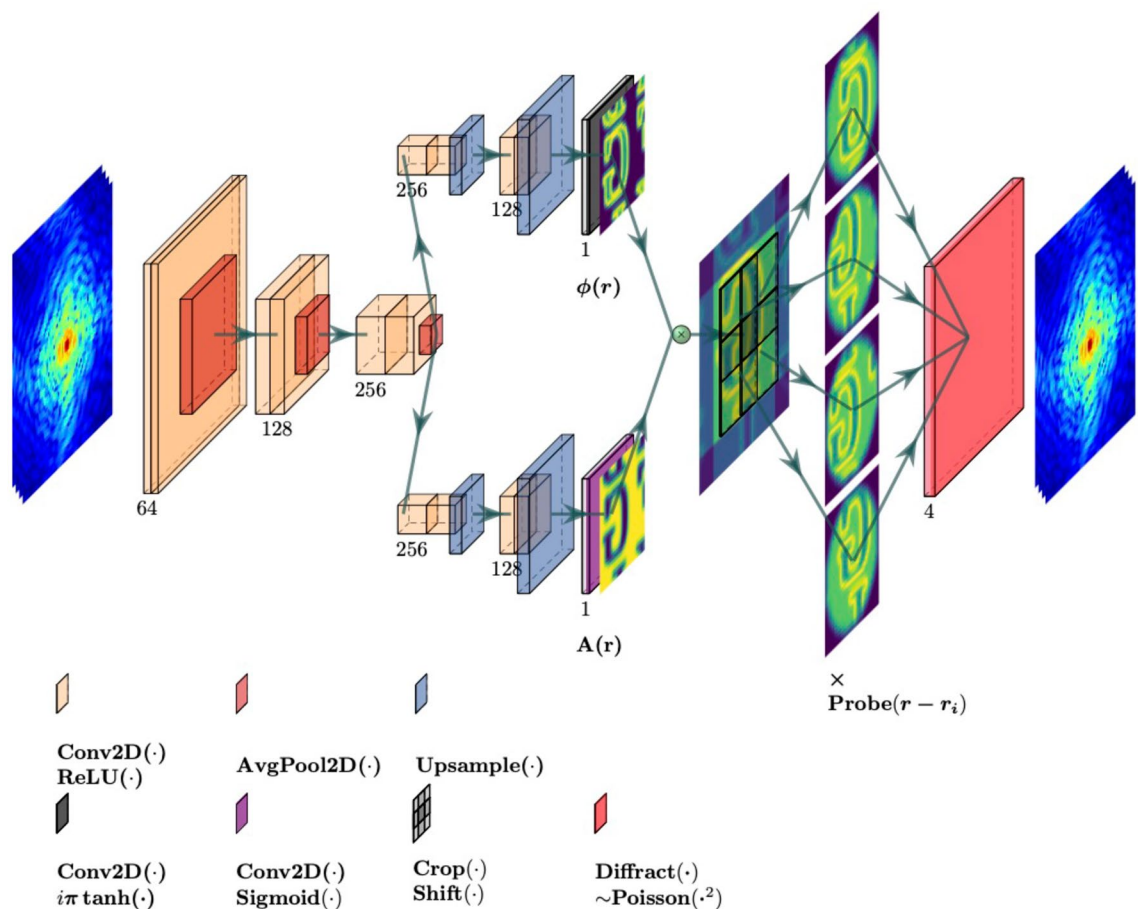
$$\text{Loss}(x, \lambda(\hat{x})) = \sum_{i,j,k} \log f_{\text{Pois}}(x_{ijk}^2; \lambda_{ijk}),$$

where  $\lambda_{ijk}(\hat{x}) = \hat{x}_{ijk}^2$ ,  $x_{ijk}^2$  is the number of photons detected in a single pixel (its square root  $x_{ijk}$  is the associated target amplitude),  $\lambda_{ijk}$  is the matching final-layer output of the CNN,  $i$  and  $j$  index the detector pixel coordinates, and  $k$  indexes separate images within a diffraction set.

For the above probabilistic formulation of the data, model output, and loss function to be self-consistent it is necessary for the units of the diffraction pixel values to be (unscaled) photon counts. A typical diffraction intensity is  $10^9$  photons per exposure whereas the magnitude of activations within the NN – including, in particular, the reconstructed real-space amplitude – should be of order unity. To invertibly scale the input (and output) we define a global normalization parameter that we initialize using a simple heuristic based on the mean photon count of images in the training dataset and unitarity property of the Fourier transform. This normalization parameter can optionally be either fixed or optimized during training.

#### PtychoPINN architecture

With the approach now laid out, we introduce its implementation in the form of the deep learning framework PtychoPINN. As illustrated in Fig. 1, PtychoPINN uses an autoencoder architecture that incorporates 2D



**Figure 1.** Neural network architecture and training configuration of the PtychoPINN model.

convolutional, average pooling, and upsampling layers in addition to custom layers that scale the network input (and output) and apply the constraint transformations (1) and (2). All convolutional layers use rectified linear unit (ReLU) activations, and as in the BCDI model AutophaseNN we incorporate sigmoid and tanh activations to limit the domain of the phase and amplitude of  $\hat{y}_i$  to  $[0, 2\pi]$  and  $[0, 1]$ , respectively<sup>7</sup>.

### Data generation and training

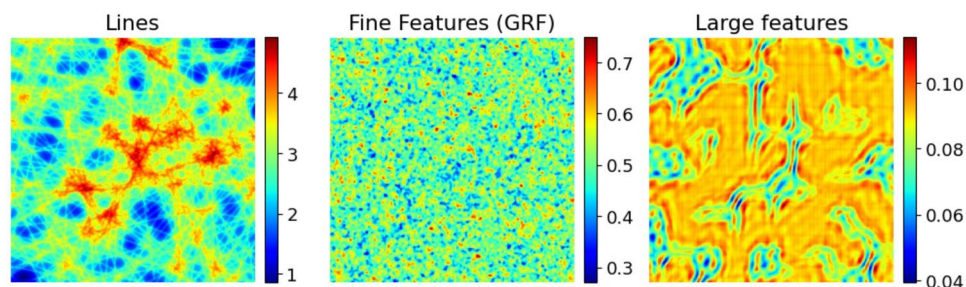
To prepare the model's training and evaluation data we begin with a collection of complex-valued images. We consider three types of images: simulated compositions of randomly-oriented lines (high aspect ratio features); samples with fine, isotropic features from a simulated Gaussian random field; and large features sampled from experimentally-derived phase and amplitude from x-ray ptychographic measurements of an etched tungsten test sample, which we retrieved from a publicly available dataset<sup>13</sup>. Representative samples of each dataset type are shown in Fig. 2, labeled respectively as Lines for randomly-oriented lines, GRF for Gaussian Random Field and 'Large features' for the experimentally-derived data of Ref.<sup>13</sup>.

Each of these datasets is characterized based on isotropy, sharpness, and the characteristic lengths in their real-space structure. Isotropy refers to uniformity of the statistical structure across all directions, i.e., invariance to rotation, reflection, and translation. The characteristic length signifies the spatial scale at which the correlation between two points diminishes. Sharpness refers to the granularity of the distinguishing features.

As summarized in Table 1, the GRF dataset is both isotropic and sharp, and has an extremely small characteristic length. In comparison, the 'Lines' objects, which consist of overlays of oriented edge features, are sharp but lack isotropy. As is the case in anisotropic media, their characteristic lengths have a direction-dependence. Finally, the 'Large features' dataset presents a distinct combination of anisotropy and coarseness. Like 'Lines', it lacks isotropy, and the characteristic length of its features is large in two dimensions—whereas that of 'Lines' is large in only one dimension.

For each real-space object dataset we simulate a collection of diffraction patterns corresponding to a rectangular grid of scan points on the sample and a known (complex-valued) probe function. Given the real-space object  $O(r)$  and far-field diffraction forward map  $F_d$ , the simulated diffraction pixel values are random samples from  $f_{\text{Pois}}(F_d(O(r))^2)$ , where  $f_{\text{Pois}}$  is the Poisson distribution.

For all object types except 'Large features', each training dataset contains 49,284 diffraction patterns densely sampled from 9 simulated objects (5476 patterns per object), and the size of each simulated object is  $392 \times 392$  pixels. (The 'Large features' dataset is limited to a single object with 16,100 diffraction patterns, corresponding



**Figure 2.** *Datasets.* Examples of amplitude images from three distinct dataset types used in this study. The ‘Lines’ and ‘Fines Features’ (Gaussian Random Field - GRF) datasets provide contrasting conditions of local symmetry (i.e., isotropy). The Gaussian random field process produces characteristic speckle, while ‘Lines’ consists of sharp, oriented edges. ‘Large Features’ is anisotropic but coarser than both ‘Lines’ and ‘Fine Features’.

	Isotropy	Sharpness	Characteristic length
GRF	Yes	Yes	Small
Lines	No	Yes	Mixed
Large Features	No	No	Large

**Table 1.** Comparing the characteristics of the three dataset types used in this study.

to its smaller underlying experimental dataset.) The probe positions were sampled on a nested rectangular grid with a spacing of 8 pixels between solution regions and a spacing of 4 pixels probe positions within a solution region. (As defined in 2.1.2, a solution region comprises a  $2 \times 2$  local grid of overlapping diffraction measurements.) To generate diffraction from an object patch we assume the ground truth probe illumination function  $P(r)$  and take as a simulation parameter the expected number of photons incident on-detector, which we set to  $10^9$  photons for all simulated data presented in this paper.

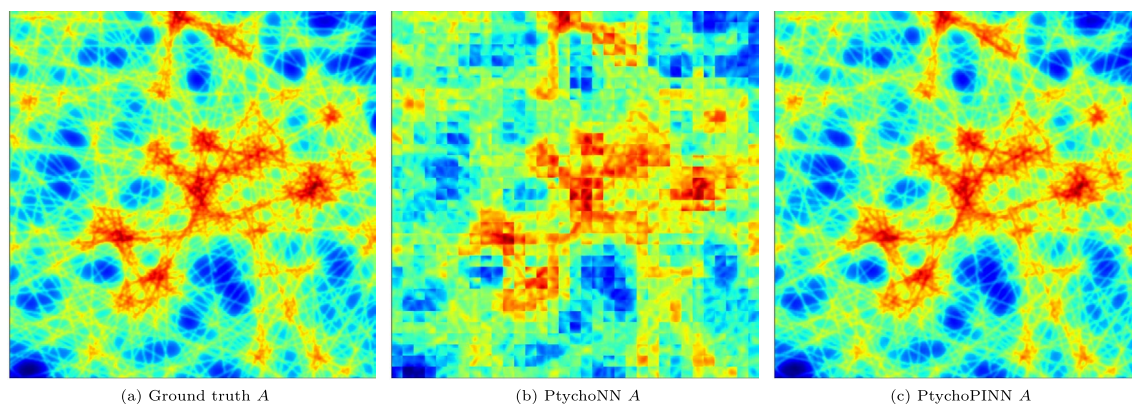
To train the model parameters we use the Adaptive Moment Estimation (ADAM) optimizer with an initial learning rate of 0.001<sup>14</sup>. We train for 50 epochs and with a batch size of 16, which takes approximately 10 minutes on an Nvidia RTX 3090 GPU.

## Results and discussion

### Numerical experiments

Figure 3 compares the reconstruction of artificially generated objects—specifically, those from the ‘lines’ data type in Fig. 2—using PtychoPINN and the supervised learning baseline, PtychoNN. PtychoPINN exhibits minimal degradation in the real-space amplitude and phase. In contrast, the supervised learning baseline experiences significant blurring. The amplitude Fourier ring correlation at the 50 percent threshold (FRC50)—a useful measure of linear resolution—is  $165.4 \text{ pixel}^{-1}$  for PtychoPINN, a notable improvement to the baseline’s  $22.0 \text{ pixel}^{-1}$ . Additionally, PtychoPINN’s peak signal-to-noise ratio (PSNR), a standard super-resolution metric, is 13 dB higher.

We further assessed the model’s performance across diverse data by including Gaussian random field (GRF) generated images and the ‘Large features’ object derived from experimental measurements (Table 2). For each image type, PtychoPINN significantly outperforms the baseline PtychoNN in both phase and amplitude, as per



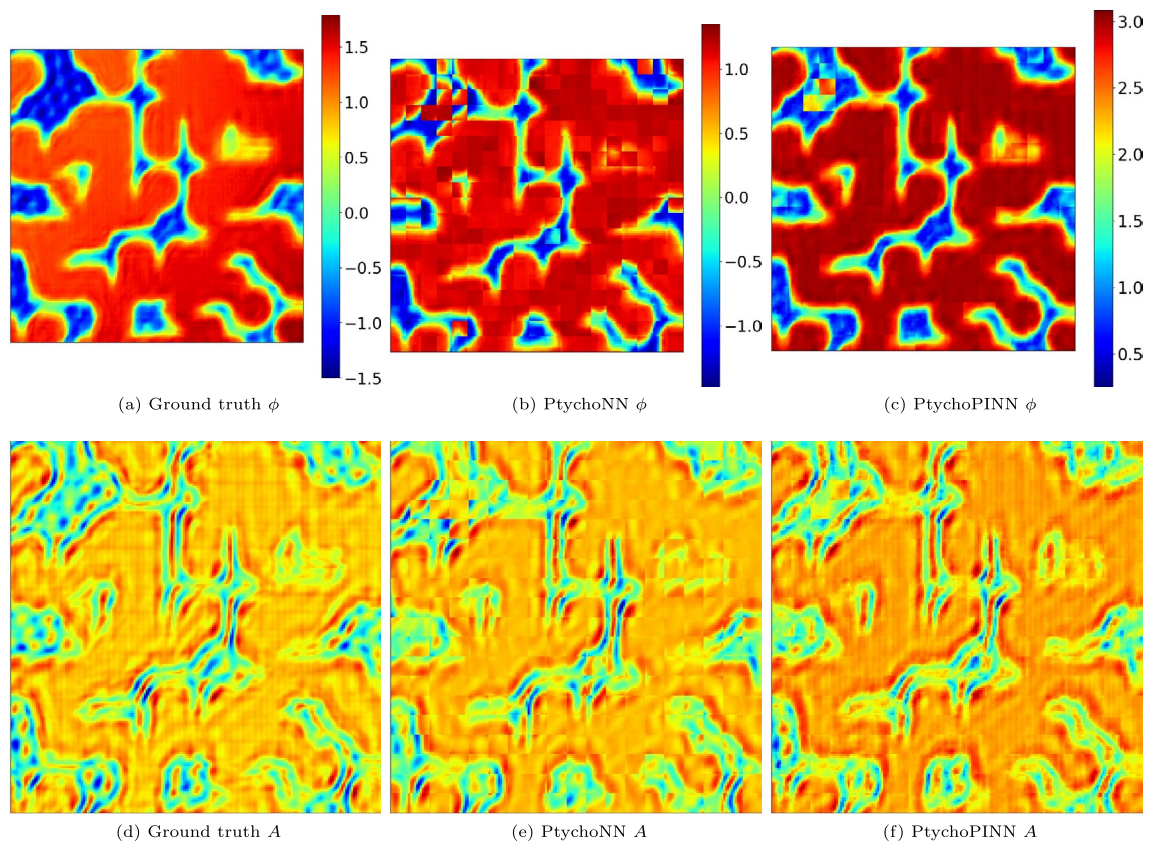
**Figure 3.** Representative Amplitude reconstruction comparison for the Lines dataset.

Model	Metric	Lines*		GRF		Large features	
		$\phi$	A	$\phi$	A	$\phi$	A
PtychoNN	MAE	–	0.201	0.0335	0.0153	0.219	0.0038
	PSNR (dB)	–	59.6	75.6	82.4	56.7	92.9
	FRC50 (pixel <sup>-1</sup> )	–	22.0	64.0	65.2	23.4	34.0
PtychoPINN	MAE	–	<b>0.0473</b>	<b>0.0109</b>	<b>0.00507</b>	<b>0.149</b>	<b>0.00303</b>
	PSNR (dB)	–	<b>72.6</b>	<b>85.2</b>	<b>91.9</b>	<b>60.6</b>	<b>95.0</b>
	FRC50 (pixel <sup>-1</sup> )	–	<b>165.4</b>	<b>171.5</b>	<b>171.3</b>	<b>93.7</b>	<b>38.7</b>

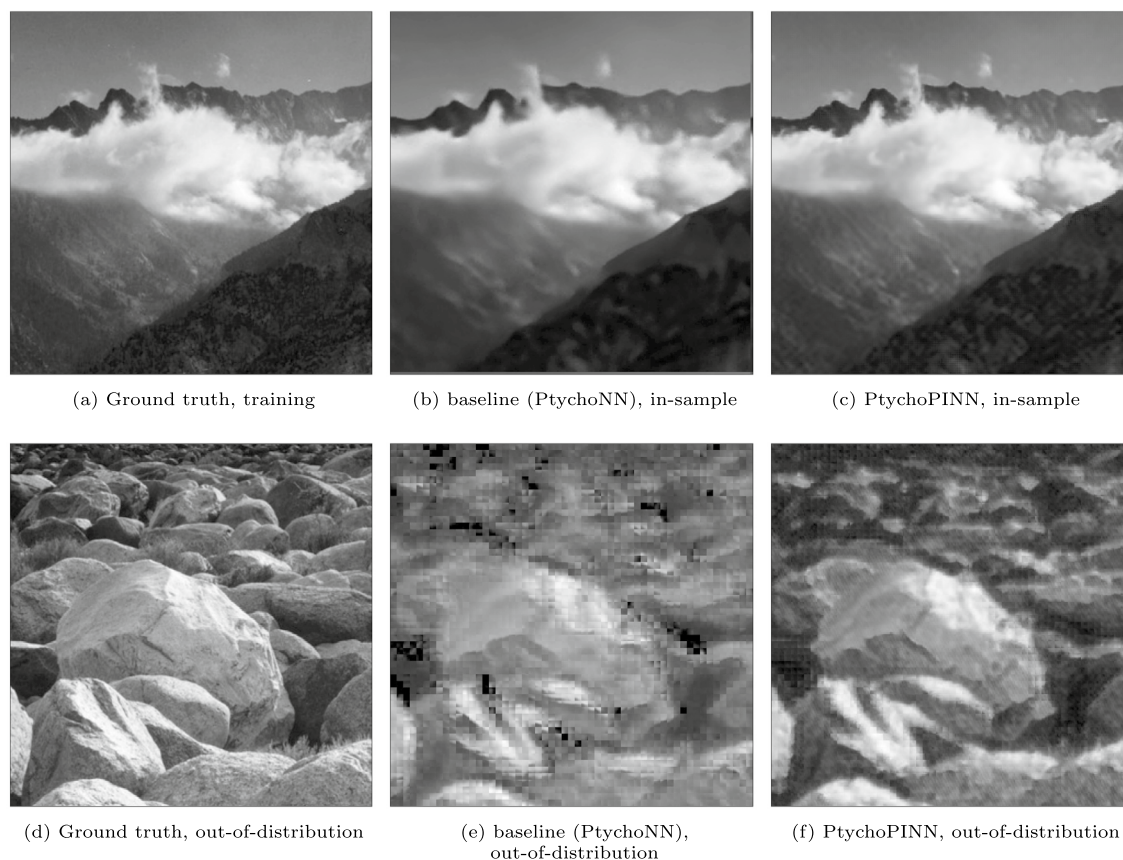
**Table 2.** Three reconstruction metrics for the baseline supervised-learning model (PtychoNN) and PtychoPINN, repeated for datasets with contrasting asymmetry and feature sharpness: Lines, GRF and ‘Large features’. For each combination of dataset (top column header) and reconstruction metric, the best amplitude and phase reconstructions are bolded. Lines, with its sharp, asymmetric features, yields the largest improvements in reconstruction of the amplitude image (bottom left of table). The baseline model recovers large amplitude features quite well (rightmost column), but PtychoPINN is more robust in its phase reconstruction across all datasets. \*We omit metrics for phase images of the ‘Lines’ datasets, as these are all zero-valued.

the MAE metric. However, enhancements in PSNR and FRC50 were comparatively modest for images featuring larger (‘Large Features’) or more equiaxed (‘GRF’) features, when juxtaposed with the improvements observed for the lines dataset.

As shown in Fig. 4, both PtychoPINN and PtychoNN provide good reconstructions of the ‘Large features’ amplitude. However, the baseline model’s phase reconstruction is less favorable, with an FRC50 of 23.4 pixel<sup>-1</sup> compared to PtychoPINN’s 93.7 pixel<sup>-1</sup>. In regions of the object with low scattering amplitude, phase reconstruction is less reliable than that of the amplitude under both models. However, PtychoPINN is more robust, producing significantly fewer artifacts in those regions compared to PtychoNN (Fig. 4b and c). Crucially, improving



**Figure 4.** Amplitude and phase reconstruction. Comparison of PtychoPINN to a supervised learning method, PtychoNN, on the ‘Large features’ object. As the amplitude object lacks sharp features, PtychoNN reconstructs it nearly as well as PtychoPINN. However, it tends to invert the phase structure in low-amplitude regions.



**Figure 5.** Model Generalization. This presents a comparison of the reconstructions produced by PtychoPINN and the baseline model, PtychoNN. Both models are trained using data derived from the amplitude object depicted in panel (a). Panels (b) and (c) show the training image as reconstructed by PtychoNN and PtychoPINN, respectively. The trained models then reconstruct diffraction from an out-of-distribution object (d), resulting in images (e) and (f) for PtychoNN and PtychoPINN, respectively. Both models degrade out-of-distribution, but PtychoPINN proves more robust.

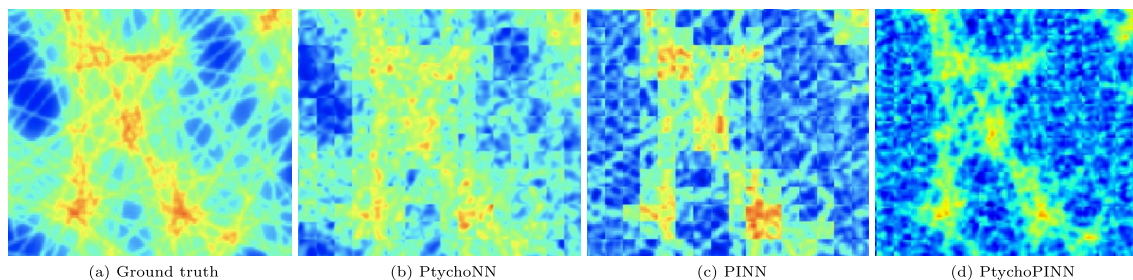
	Dataset		PSNR	FRC50
PtychoPINN	Train	A	78.41	160
		$\phi$	69.11	158
	Test	A	61.11	42
		$\phi$	59.87	41
Baseline	Train	A	73.72	76
		$\phi$	72.11	74
	Test	A	57.91	22
		$\phi$	55.77	18

**Table 3.** Numerical study of out-of-distribution robustness corresponding to the data of Fig. 5.

the estimation of the object in phase space not only results in a higher-quality final image, but also ensures the problem is well-posed<sup>15</sup> and leads to a unique solution.

The real-space image is the full, complex-valued  $O(r)$  – not only its phase or amplitude, as portrayed for convenience in Fig. 3 and elsewhere. A competent image reconstruction requires both the amplitude and phase of  $O(r)$  and, moreover, the phase image sometimes proves more informative than the amplitude image, such as in the case of biological materials that have small variations in the differential scattering cross-section but good contrast in  $\phi(r)$ .

At a high level, we summarize the resolution enhancement as a factor of 4, based on the FRC50 metric from the ‘Lines’ and ‘GRF’ datasets in Table 2, both chosen for their high spatial-frequency features. For PSNR the average improvement is approximately 8 dB, derived from the aggregate of all datasets in Table 3.



**Figure 6.** Cross-training comparison. Several models are first trained on locally-symmetric GRF objects, then reconstruct strongly-contrasting, asymmetric compositions of lines. The supervised model (b) struggles to reconstruct the oriented features in ground truth image (a) while the PINN – both without (c) and with (d) ptychographic overlap constraints – fares better.

#### Ablation study

To better understand the specific aspects of the PtychoPINN architecture that contribute to its improved performance, we conducted an ablation study evaluating the impact of PINN-based training and real-space constraints. The comparison of reconstruction accuracy (MAE), resolution (Fourier ring correlation, FRC), and peak signal-to-noise ratio (PSNR) in Supplementary Table S1 suggests that the combination of overlap constraints and PINN structure is the most significant factor in improving performance. *Combination* is an important qualifier, as neither the PINN architecture nor the overlap constraints yields a substantial improvement in reconstruction accuracy when used individually.

#### Out-of-distribution generalization

To complement these standard evaluation metrics we tested PtychoPINN's out-of-distribution generalization. For this purpose, we extracted two distinct  $392 \times 392$  patches from a larger image and used 26, 896 diffraction patterns, simulated from just one patch for training both the PtychoPINN model and the baseline model (Fig. 5).

In order to provide a rigorous test of the models' out-of-distribution performance, we intentionally selected two image patches with considerable divergence in their features. This choice of markedly different images creates a more stringent test of out-of-distribution behavior than the other datasets that we have presented.

Evaluating the models' ability to reconstruct both in-distribution and out-of-distribution patches provides insight into their generalization capacity. As shown in Fig. 5, a comparison between panels (c) and (f) (representing PtychoPINN) and panels (b) and (e) (representing the baseline - PtychoNN - model) indicates that PtychoPINN experiences a less severe drop in out-of-distribution reconstruction quality, implying better generalizability compared to the baseline. This observation is further backed by numerical comparison (Table 3). As expected, both models perform less well out-of-distribution, with a more pronounced decline in the FRC50 (resolution) metric than the PSNR metric. However, PtychoPINN outperforms the baseline model, with almost double the resolution.

#### Performance comparison to Iterative Solvers

The reconstruction of a dataset of 1024 diffraction images using the trained model takes 0.3 seconds. In comparison, a CPU implementation of ePIE (without position correction) completes in 165 seconds.

## Discussion

In summary, we present an unsupervised learning approach for scanning coherent diffraction imaging (CDI) that integrates real-space constraints with a physics-informed neural network (PINN) architecture (PtychoPINN) to yield substantial improvements in reconstruction accuracy and resolution while preserving the inherent speed of the previous NN-based approaches.

In the ensuing discussion, our primary objective is to highlight features of PtychoPINN that would be most important to a domain scientist intending to apply it. When possible, we will also take the perspective of a machine-learning researcher to understand *why* PtychoPINN exhibits these characteristics. We specifically concentrate on three essential attributes: unsupervised training, generalizability, and the interplay between resolution and accuracy.

#### Unsupervised training

The use of unsupervised training in our approach offers three important advantages over supervised learning methods in an experimental setting. Firstly, supervised methods require labeled data, placing the onus on the experimenter to first gather or simulate numerous ptychographic datasets, and then painstakingly reconstruct the real-space images using iterative methods. Only then can a faster neural network-based method be trained.

Secondly, the demand for labeled data in supervised learning restricts the ability to retrain or update the model 'on-the-fly'. Unsupervised training, on the other hand, allows the use of training samples drawn from the same distribution as the test-time samples. This reduces the risk of any issues related to out-of-distribution generalization that the model may otherwise encounter, which aids robustness – particularly in real-time settings.



Lastly, supervised learning methods can introduce biases due to the lack of diversity in the image types within the labeled training data, which in turn reduces the model's generalizability. We delve deeper into addressing this issue in the following sections.

#### Generalizability

The full PINN model has another inherent advantage: superior generalizability. Even the basic PINN (that is, PtychoPINN stripped of overlap constraints) exhibits better out-of-distribution generalization compared to the supervised learning baseline in a qualitative out-of-distribution benchmark, shown in Fig. 6. With full PtychoPINN, this advantage becomes more pronounced (Fig. 6d).

From the point of view of the experimenter, the benefit of generalizability complements that of unsupervised training. An unsupervised architecture facilitates each instance of training; conversely, a more *generalizable* model is more robust and does not need to be as frequently retrained. We attribute PtychoPINN's superior generalizability to the PINN structure, which necessitates the inverse map  $G(X)$  to learn diffraction physics through its connection to the far-field diffraction map  $F_d$  and diffraction-space loss function  $L$ . This connection enforces an approximation of physical consistency between the real-space reconstruction and the measured diffraction. Conversely, supervised training typically leads to violations of even basic conservation rules, such as the FT's unitarity (i.e.,  $\|x\| \approx \|F_c(G(x))\| = \|F_d(\bar{y})\| = \|\hat{x}\|$ , by Parseval's identity and  $\bar{y} \equiv F_c(G(x))$ ).

Furthermore, given that  $\hat{x} = F_d(\bar{y})$  is not bijective and the loss function does not directly rely on  $\bar{y}$ , the PINN structure inherently overlooks variations in real-space structure that the diffraction measurement can't discern. Conversely, supervised training is susceptible to wasting model capacity by memorizing the training data's real-space structure. All told, supervised learning approaches face challenges in generalizability due to the difficult task of reconstructing an *a priori* arbitrary map  $X \rightarrow Y$  with less guidance by useful inductive biases.

As an aside, generalizability is closely connected to the concept of model *robustness*. Robustness depends on not only the network architecture but also the choice of loss function and properties of the data. Experimental data such as CDI images are often plagued by noise and distortions that must be modeled or explicitly removed for good results<sup>16</sup>. While our framework takes a step in this direction by modeling experimental Poisson noise via its training objective, we leave a full treatment of the topic outside of the paper's scope.

#### Reconstruction quality

The advances in reconstruction quality presented by PtychoPINN have dual significance: first, in their practical implications for scientific applications and second, for potential insights into the source of improvement within the model structure. To explore the latter we find it useful to distinguish between real-space and reciprocal-space reconstruction accuracy. Both the full-featured version of PtychoPINN and the 'basic PINN' variation, which lacks overlap constraints, produce good reconstructions of the diffraction measurement. Only the full model, however, produces an accurate real-space reconstruction.

We can reconcile the seeming contradiction between the basic PINN's accurate diffraction reconstruction and inaccurate real-space images by examining the latter. When we compare  $32 \times 32$  real-space reconstruction patches from PtychoPINN, PtychoNN, and a basic PINN (Supplementary Fig. S1), we observe that the basic PINN generates superposed pairs of line features that are inversions of one another through the solution patch's origin. Despite the basic PINN's poor real-space accuracy, its reconstructed line features have similar sharpness to the full model equivalents. Furthermore, the real-space reconstructions of the basic PINN and the full model are highly similar up to the mapping  $O(r) \rightarrow (O(r) + O(-r))/2$ .

The above observations demonstrate that the basic PINN recovers most of the information in the ground truth image but has poor real-space accuracy because it cannot distinguish between pairs of features that are equivalent to one another under coordinate inversion. (In particular,  $O(r)$ ,  $O(-r)$ , and  $(O(r) + O(-r))/2$  have identical diffraction amplitudes for this zero-phase demonstration dataset.) The role of the real-space overlap constraints in the full model is to correct the alignment of real-space features by breaking the degeneracy associated with inversion symmetry, which leads to a large improvement in real-space accuracy.

## Conclusion and future directions

In conclusion, we present an autoencoder framework for scanning CDI into which we incorporate physical principles using PINN training and real-space constraints, thus gaining greatly improved accuracy, resolution, and generalizability compared to existing supervised deep learning-based lensless imaging methods. This new model, named PtychoPINN, inherits the intrinsic speed of NN approaches and is trained without labels. These attributes combine to make it a promising candidate for practical real-time, high-resolution imaging that transcends the resolution of lens-based systems without sacrificing imaging throughput.

To train PtychoPINN, we incorporated a probabilistic (Poisson) model output and corresponding negative log likelihood (NLL) objective, thus modeling the Poisson noise intrinsic in experimental data. Our preliminary studies show that the NLL objective significantly improves reconstruction quality in photon-limited scenarios. This suggests that selecting an appropriate loss function is important in deep learning models that incorporate physics. In a similar vein, sources of stochasticity such as probe jitter introduce uncertainties into the reconstruction that may degrade image quality. To mitigate this, we advocate for the exploration of probabilistic methods to improve performance and provide a principled estimation of model uncertainties. A more immediate omission in our framework is its lack of a mechanism for error inference in probe positioning – an important feature for robust experimental use. We are developing this capability in ongoing work.

## Data availability

All data and code used in this study are available on GitHub. The repository, which includes both raw datasets and scripts for reproducing the simulations and analyses, can be accessed at [github.com/hoidn/PtychoPINN](https://github.com/hoidn/PtychoPINN).

Received: 23 June 2023; Accepted: 25 November 2023

Published online: 21 December 2023

## References

- Dean, B.H., Aronstein, D.L., Smith, J.S., Shiri, R., & Acton, D.S.: Phase retrieval algorithm for jwst flight and testbed telescope. In: Space Telescopes and Instrumentation I: Optical, Infrared, and Millimeter, vol. 6265, pp. 314–330 (2006). SPIE
- Heintzmann, R. Answers to fundamental questions in superresolution microscopy. *Phil. Trans. R. Soc. A* **379**(2199), 20210105 (2021).
- Miao, J., Ishikawa, T., Robinson, I. K. & Murnane, M. M. Beyond crystallography: Diffractive imaging using coherent x-ray light sources. *Science* **348**(6234), 530–535 (2015).
- Miao, J., Charalambous, P., Kirz, J. & Sayre, D. Extending the methodology of x-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens. *Nature* **400**(6742), 342–344 (1999).
- Maiden, A. M. & Rodenburg, J. M. An improved Ptychographical phase retrieval algorithm for diffractive imaging. *Ultramicroscopy* **109**(10), 1256–1262 (2009).
- Ratner, D. *et al.* Recovering the phase and amplitude of x-ray fel pulses using neural networks and differentiable models. *Opt. Express* **29**(13), 20336–20352 (2021).
- Yao, Y. *et al.* Autophasenn: unsupervised physics-aware deep learning of 3d nanoscale bragg coherent diffraction imaging. *npj Computat. Mater.* **8**(1), 124 (2022).
- Chang, D. J. *et al.* Deep-learning electron diffractive imaging. *Phys. Rev. Lett.* **130**(1), 016101 (2023).
- Mitchell, T.M.: The need for biases in learning generalizations. Cbm-tr-117, Computer Science Department, Rutgers University (1980)
- Baxter, J. A model of inductive bias learning. *J. Artif. Intell. Res.* **12**, 149–198 (2000).
- Baker, N. *et al.* *Workshop Report on Basic Research Needs for Scientific Machine Learning: Core Technologies for Artificial Intelligence* (Technical report, USDOE Office of Science (SC), 2019).
- Miao, J., Kirz, J. & Sayre, D. The oversampling phasing method. *Acta Crystallogr. D Biol. Crystallogr.* **56**(10), 1312–1315 (2000).
- Cherukara, M. J. *et al.* Ai-enabled high-resolution scanning coherent diffraction imaging. *Appl. Phys. Lett.* **117**(4), 044103 (2020).
- Kingma, D.P., & Ba, J.: Adam: A method for stochastic optimization. Preprint at [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014).
- Millane, R. P. Phase retrieval in crystallography and optics. *JOSA A* **7**(3), 394–411 (1990).
- Lee, S. Y. *et al.* Denoising low-intensity diffraction signals using k-space deep learning: Applications to phase recovery. *Phys. Rev. Res.* **3**(4), 043066 (2021).

## Acknowledgements

The authors would like to acknowledge insightful discussions with Daniel Ratner and Yousef Nashed. Additionally, we would like to acknowledge invaluable assistance and guidance from Matt Seaberg.

## Author contributions

A.M. provided the project's initial suggestion and guided the overall scientific direction. O.H. proposed the approach, built the model, and performed the training and analyses. A.A.M. guided model development and elucidated connections to the scientific ML literature. O.H. and A.A.M. wrote the manuscript with input from A.M. All authors discussed results, edited the manuscript, and gave final approval for publication.

## Funding

This work was performed and partially supported by the US Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences Data, Artificial Intelligence and Machine Learning at the DOE Scientific User Facilities program under the MLEExchange Project (award No. 107514). Aashwin Mishra was partially supported by the SLAC ML Initiative.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-48351-7>.

**Correspondence** and requests for materials should be addressed to O.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023