

NeuPh: scalable and generalizable neural phase retrieval with local conditional neural fields

Hao Wang,^a Jiabei Zhu,^a Yunzhe Li,^{a,†} Qianwan Yang^{①,a} and Lei Tian^{①,a,b,*}

^aBoston University, Department of Electrical and Computer Engineering, Boston, Massachusetts, United States

^bBoston University, Department of Biomedical Engineering, Boston, Massachusetts, United States

Abstract. Deep learning has transformed computational imaging, but traditional pixel-based representations limit their ability to capture continuous multiscale object features. Addressing this gap, we introduce a local conditional neural field (LCNF) framework, which leverages a continuous neural representation to provide flexible object representations. LCNF's unique capabilities are demonstrated in solving the highly ill-posed phase retrieval problem of multiplexed Fourier ptychographic microscopy. Our network, termed neural phase retrieval (NeuPh), enables continuous-domain resolution-enhanced phase reconstruction, offering scalability, robustness, accuracy, and generalizability that outperform existing methods. NeuPh integrates a local conditional neural representation and a coordinate-based training strategy. We show that NeuPh can accurately reconstruct high-resolution phase images from low-resolution intensity measurements. Furthermore, NeuPh consistently applies continuous object priors and effectively eliminates various phase artifacts, demonstrating robustness even when trained on imperfect datasets. Moreover, NeuPh improves accuracy and generalization compared with existing deep learning models. We further investigate a hybrid training strategy combining both experimental and simulated datasets, elucidating the impact of domain shift between experiment and simulation. Our work underscores the potential of the LCNF framework in solving complex large-scale inverse problems, opening up new possibilities for deep-learning-based imaging techniques.

Keywords: neural representation; phase retrieval; computational imaging; deep learning; computational microscopy.

Received Mar. 28, 2024; revised manuscript received Jun. 24, 2024; accepted for publication Jul. 30, 2024; published online Aug. 28, 2024.

© The Authors. Published by SPIE and CLP under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.

[DOI: [10.1117/1.APN.3.5.056005](https://doi.org/10.1117/1.APN.3.5.056005)]

1 Introduction

Deep learning (DL) has revolutionized the field of computational imaging,^{1,2} providing powerful solutions to enhance performance and address various challenges.^{3–11} The effectiveness of DL in computational imaging lies in its ability to capture the underlying imaging model and exploit object priors, enabling robust solutions to ill-posed inverse problems.¹ However, the most widely used reconstruction methods in computational imaging rely on discrete pixels to represent the objects. For instance, a convolutional neural network (CNN) for computational imaging is typically trained on a fixed pixel or voxel

grid.¹ This representation is inherently limited by the resolution and size of the grid and does not capture the continuous nature and multiscale details of the physical objects. Furthermore, the pixel grid representation poses challenges in scaling to process and store large-scale multi-dimensional computational imaging data.⁵

The neural field (NF) framework¹² has recently gained significant interest for its ability to represent continuous objects. Unlike CNN structures, NF uses a coordinate-based neural representation, where spatial coordinates are mapped to physical values using a multi-layer perceptron (MLP). This allows the NF to encode objects in a continuous representation, decoupled from a discrete grid. It enables on-demand synthesis of any part of the object by simply querying relevant coordinates across arbitrary dimensions and resolutions. Several NF-based DL techniques have been recently introduced in computational imaging

*Address all correspondence to Lei Tian, leitian@bu.edu

[†]Current address: Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, California, United States

for solving inverse problems.^{5,13–19} However, these methods are limited by the high computational cost and limited generalization ability. They either require retraining a new NF network for each object^{5,13–15,17–19} or suffer from the limited representation power of the latent space learned only on the global scale,¹⁶ restricting their ability to generalize to diverse objects.

To overcome these limitations, we propose a novel local conditional neural field (LCNF) framework for solving imaging inverse problems. We demonstrate LCNF's unique capabilities for solving the highly ill-posed phase retrieval problem in multiplexed Fourier ptychographic microscopy (FPM).^{20,21} We term this LCNF-based phase retrieval method as neural phase retrieval, or NeuPh [Fig. 1(a)]. NeuPh utilizes a CNN-based encoder to learn measurement-specific information from a set of low-resolution measurements and encode them into a compact latent-space representation. This CNN-based encoder effectively extracts the diffraction information that spreads over an area in the measurement by utilizing its extended receptive field, condensing them into latent vectors. Subsequently, an MLP-based decoder is employed to reconstruct the phase values of the object at specific locations based on the corresponding latent information. This decoder is conditioned on a learned latent vector that incorporates information across a region of the input images. This conditioning enables adaptation to different objects since each set of measurements is projected onto

a distinct latent space representation by the encoder. In addition, a crucial aspect of FPM phase retrieval is achieving “super-resolution” reconstruction, surpassing the diffraction limit of the input low-resolution measurements.²² To reach this goal, NeuPh's decoder further utilizes “super-resolved” latent information during training. By combining these components, NeuPh achieves scalable and generalizable DL-based continuous-domain high-resolution image reconstruction based on low-resolution measurements that is applicable to arbitrary objects with varying spatial scales and resolutions.

Our results highlight NeuPh's ability to apply continuous and smooth priors to the reconstruction and showcase more accurate reconstruction results compared to existing models. Using experimental datasets captured on Hela cells, we show that NeuPh can accurately reconstruct complex intricate subcellular structures. In addition, we highlight NeuPh's robustness when subjected to imperfect training datasets. Notably, NeuPh effectively eliminates common artifacts encountered in traditional model-based algorithms, such as residual phase unwrapping errors, noise, and background artifacts. Through comparison with a fully CNN-based model, we highlight the benefits of employing an MLP-based decoder and pixel-based training strategy in NeuPh over convolutional layer-based decoding and patch-wise training strategy in CNNs. In addition, we demonstrate the superior performance of NeuPh in multiplexed FPM

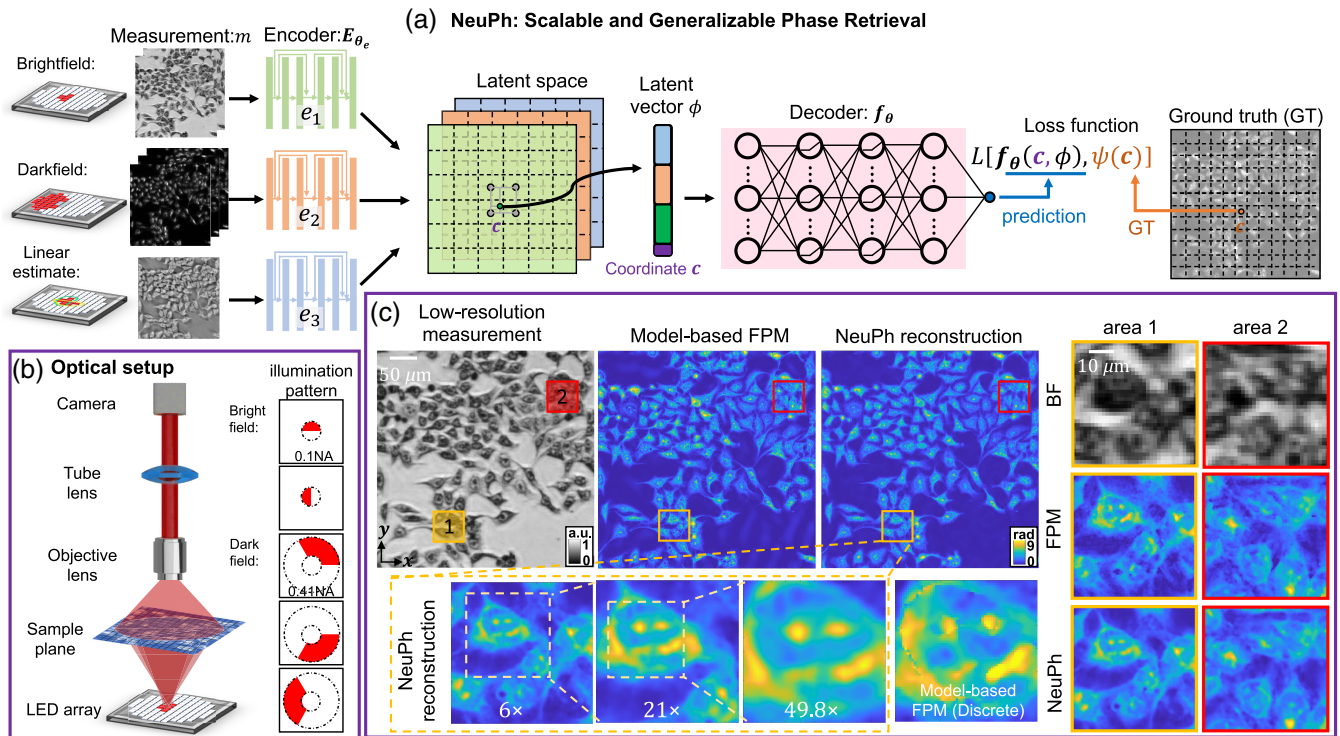


Fig. 1 Conceptual illustration of the NeuPh framework. (a) NeuPh employs a CNN-based encoder to learn measurement-specific information and encode them into a latent-space representation. The MLP decoder reconstructs the phase values at specific locations with an increased spatial resolution by synthesizing local conditional information from the corresponding latent vectors. (b) FPM experimental setup and illumination patterns for acquiring multiplexed BF and DF measurements. (c) Example low-resolution BF measurement and high-resolution phase reconstruction from the model-based FPM algorithm and NeuPh. NeuPh learns a continuous-domain representation and can infer phase maps on an arbitrary pixel grid (illustration with 6 $\times$, 21 $\times$, 49.8 $\times$ pixel density compared with the raw measurement).

reconstruction compared to state-of-the-art neural networks used in FPM phase retrieval problems.

Furthermore, we showcase NeuPh's strong generalization capabilities, which cannot be achieved by existing NF networks. First, we demonstrate that NeuPh can consistently perform high-resolution reconstruction even when trained with very limited data or under different experimental conditions. Moreover, we demonstrate that NeuPh can be trained using exclusively physics-model-simulated datasets composed of natural images. We show that this physics simulator-trained NeuPh generalizes well to experimental biological samples. We further investigate a hybrid training strategy combining both experimental and simulated datasets. By collectively analyzing the results from the pure-experimental-based, hybrid, and pure-simulation-based training methods, we elucidate the effects of domain shift and underscore the importance of aligning the data distribution between the simulation and experiment to ensure effective network training with simulation. Finally, we show that all NeuPh networks, regardless of the training dataset, can reliably reconstruct high-resolution images across a wide field of view (FOV). The results are robust to unknown spatially varying aberrations, benefiting from the smooth object priors provided by NeuPh.

In summary, we introduce a novel LCNF framework as a scalable, robust, accurate, and generalizable approach for solving highly ill-posed large-scale imaging inverse problems. By leveraging a continuous neural representation, NeuPh effectively captures multiscale object information from low-resolution measurements and provides robust resolution enhancement in the reconstruction. Furthermore, our LCNF framework exhibits superior reconstruction performance compared to existing models, highlighting the advantages of the MLP-based decoder and coordinate-based training strategy. The framework's ability to generalize with very limited training data and its capacity to leverage physics-based simulation further enhance its potential for advancing DL-based computational imaging techniques.

2 Methods

2.1 Experimental Setup and Data Preparation

We investigate different datasets for training NeuPh using both experimental and simulated datasets. The experimental data are taken on HeLa cells fixed in ethanol [HeLa(E)] or formalin [HeLa(F)] from Ref. 23. Due to different fixation procedures, the HeLa cells show different morphological features.

To briefly describe the experiment, our multiplexed measurements use five LED patterns (central wavelength of 630 nm), including two brightfield (BF) semi-circle and three darkfield (DF) 120 deg arc patterns [Fig. 1(b)], to efficiently encode high-resolution phase information across a wide FOV. Matching sequential FPM datasets are also captured by single-LED illumination with 185 LEDs. In both schemes, the maximum illumination numerical aperture (NA) is 0.41. Images are collected using a 4×, 0.1 NA objective (Nikon CFI Plan Achromat) and an sCMOS camera (PCO.edge 5.5) with 2560 pixel × 2160 pixel and a 6.5 μm pixel size. We collect 22 groups of HeLa(E) measurements and 20 groups of HeLa(F) measurements. The ground-truth images are computed using the model-based algorithm²¹ on the sequential FPM measurements to process the central 250 pixel × 250 pixel regions and reconstruct the high-resolution phase images yielding 1500 pixel × 1500 pixel. Accordingly, the full-pitch resolution²⁴ is theoretically improved from 6.3

(corresponding to 0.1 NA) to 1.235 μm (corresponding to 0.51 NA).

For the simulated dataset, we collect images from DIV2K dataset.²⁵ The whole simulated dataset consists of 900 cropped high-resolution natural images (600 pixel × 600 pixel) and corresponding simulated low-resolution multiplexed illuminated intensity images (100 pixel × 100 pixel). The forward model for the simulation, model-based reconstruction algorithms, data acquisition, and preparation are detailed in the [Supplementary Material](#).

2.2 NeuPh Framework

NeuPh's network structure is shown in Fig. 1(a). The encoder E_{θ_e} takes six low-resolution images as input and projects the learned information into a latent space. The input consists of two BF and three DF images, along with a low-resolution phase estimate using the two BF images and the differential phase contrast (DPC) method²⁰ (see details in the [Supplementary Material](#)). To handle the distinct features of BF, DF, and DPC images, three separate encoders $\{e_1, e_2, e_3\}$ are employed to extract the latent information. Each encoder utilizes convolution layers and residual blocks to extract spatial features. The lateral dimensions of the feature maps match those of the input image, allowing for direct coordinate-based latent information retrieval during decoding. The features learned from the three encoders are concatenated to form the final latent space representation $\Phi \in \mathbb{R}^{H \times W \times D}$, where H and W are the numbers of pixels in the lateral dimensions and D is the total number of concatenated feature maps in the latent space.

To enable high-resolution reconstruction independent of a fixed grid, a coordinate-based decoder is employed using an MLP f_{θ} . For local conditioning, a latent vector $\phi \in \mathbb{R}^{1 \times 1 \times D}$ is concatenated with the corresponding coordinate $\mathbf{c}$ before being input to f_{θ} . This ensures that the learned mapping by the MLP is informed by the input, allowing for model generalization. The decoder's output is a predicted phase value at location $\mathbf{c}$. NeuPh is trained end-to-end by minimizing a loss function $L[f_{\theta}(\mathbf{c}, \phi), \psi(\mathbf{c})]$ that measures the difference between the predicted $[f_{\theta}(\mathbf{c}, \phi)]$ and the ground-truth phase $\psi(\mathbf{c})$ at $\mathbf{c}$. To accelerate the training, we randomly select N coordinates from small training batches at each step. Correspondingly, we perform the minimization:

$$\min_{\theta_e, \theta} \frac{1}{N} \sum_{n=1}^N \|f_{\theta}(\mathbf{c}_n, \phi_n) - \psi(\mathbf{c}_n)\|_1, \quad (1)$$

where θ_e and θ are the weights of the encoder and decoder, respectively, $\mathbf{c}_n$ is the n 'th selected coordinate, $\phi_n = E_{\theta_e}(m, \mathbf{c}_n)$ is the latent vector encoded from input m for the queried coordinate $\mathbf{c}_n$, $\psi(\mathbf{c}_n)$ is the ground-truth phase value at position $\mathbf{c}_n$, and $\|\cdot\|_1$ denotes the L_1 norm. The ground-truth images are obtained from the paired sequential FPM measurements and reconstructed by a model-based algorithm (see [Supplementary Material](#)).

A key aspect of FPM reconstruction is "super resolution." To facilitate learning high-resolution information beyond the low-resolution $H \times W$ grid, NeuPh is also trained on "off-the-grid" high-resolution data queried from a denser grid $H' \times W'$. However, the corresponding off-the-grid latent vector is not readily available from the encoded latent space. To solve this issue, the nearest latent vector is used for the decoder.

In addition, to inform the decoder about the relative position of the queried off-the-grid location with respect to the nearest latent vector location, the implementation of Eq. (1) utilizes their relative coordinate $\Delta \mathbf{c}$ instead of the absolute coordinate $\mathbf{c}$, following the LIIF method.²⁶ Furthermore, to utilize the information provided by the neighboring latent vectors and improve the continuity of the reconstruction, enhancement techniques including feature unfolding, local ensemble, and cell decoding²⁶ are applied.

After training, NeuPh allows for querying arbitrary points on the object by inputting the corresponding measurements and the queried coordinates. This eliminates the need for a fixed input grid in traditional model-based and CNN methods. The high-resolution phase reconstruction can be achieved on any desired grid. This feature is demonstrated in Fig. 1(c), where reconstructions are queried at three distinct pixel densities, showcasing smooth transitions across these diverse spatial scales without any artifacts. More details about NeuPh's structure, reconstruction enhancement techniques, and ablation studies of its architecture can be found in Secs. 2, 3, 6, Fig. S6, and Table S2 in the [Supplementary Material](#).

2.3 Training Strategies

To comprehensively evaluate NeuPh's generalization capability and investigate the impact of domain shift between the simulated and experimental datasets on reconstruction result, in total, we explored four different training strategies, including training with the full experimental dataset (**NeuPh_{E(18)}**, **NeuPh_{F(16)}**), training with a single pair of experimental datasets (**NeuPh_{E(1)}**, **NeuPh_{F(1)}**), training with a purely simulated natural image dataset (**NeuPh_{Sim(18)}**, **NeuPh_{Sim(16)}**, **NeuPh_{Sim}**), and training with mixed experimental and simulated images (**NeuPh_{E:Sim(9:9)}**, **NeuPh_{E:Sim(1:17)}**, **NeuPh_{F:Sim(8:8)}**, and **NeuPh_{F:Sim(1:15)}**).

For all four scenarios, the training typically converges at 500 epochs. Training the network with a single pair of experimental datasets takes ~ 3 h, while training with the full experimental dataset, a blend of an experimental and simulated datasets, or the simulated dataset takes ~ 24 to 48 h to converge on an NVIDIA Tesla P100 GPU. More details about the implementation of network training and datasets used for training different networks can be found in the Sec. 5 in the [Supplementary Material](#).

2.4 Network Inference

To perform inference, we input the preprocessed measurements of the desired FOV and configure the pixel resolution for the reconstruction. During the inference, NeuPh predicts the phase value for each queried position. In our main results, we perform inference with a $6\times$ denser grid compared to the raw measurement. We first divide the measurement into small patches with $250 \text{ pixel} \times 250 \text{ pixel}$. Next, we reconstruct each patch individually, producing high-resolution images with $1500 \text{ pixel} \times 1500 \text{ pixel}$. This process takes ~ 25 s, resulting in an average rate of $\sim 1 \times 10^{-5}$ s/pixel on an NVIDIA Quadro RTX8000 GPU. In comparison, the model-based FPM reconstruction with sequential illumination takes around 8 min. In addition, the model-based multiplexed FPM reconstruction algorithm fails to generate high-resolution phase images from multiplexed illumination with only five intensity images.²³ To create the final wide-FOV reconstruction, we employed alpha blending²³ to stitch together the individual reconstructions,

forming a high-resolution image with a diameter of 12,960 pixel. It is worth noting that NeuPh is inherently capable of directly inferring the entire FOV image without any stitching. However, due to the GPU memory (48 GB) limit on our GPU, we apply this patch-wise inference method.

3 Results

3.1 Continuous-Domain Phase Reconstruction by NeuPh

We first demonstrate NeuPh's capability on experimentally imaged Hela cells fixed with ethanol or formalin. Here, we show results using the networks that are trained separately for these two data types (**NeuPh_{E(18)}** and **NeuPh_{F(16)}**) to showcase NeuPh's best performance since different cell fixations induce distinctive morphological features within single cells and thus a domain shift in the data (which we explore in more detail in Sec. 3.5). In Fig. 1(c), we present the raw low-resolution BF intensity image, the model-based FPM reconstruction, and the NeuPh reconstruction of ethanol-fixed Hela cells. In Figs. 2(a) and 2(b), we show additional reconstruction results for Hela cells fixed with ethanol or formalin, respectively, with a few zoomed-in regions. It is evident that NeuPh successfully reconstructs high-resolution phase images from low-resolution measurements, accurately recovering intricate subcellular structures and achieving much better results compared with the linear DPC estimate.

To showcase NeuPh's continuous object representation capability, we conduct inference on arbitrary coordinates. As shown in Fig. 1(c) and Fig. S5 in the [Supplementary Material](#), we perform queries for subarea 1 at pixel densities of $6\times$, $21\times$, $49.8\times$, $73.5\times$, and $105.9\times$ compared to the input low-resolution intensity image. NeuPh successfully reconstructs the phase at these density grids. For comparison, we also include the model-based reconstruction of the same area. Due to predefined grids, the reconstruction exhibits discrete grid artifacts in the enlarged image. Furthermore, it may suffer from phase unwrapping artifacts (see Sec. 3.2 for more details). In contrast, NeuPh provides continuous object reconstruction without any discrete or other phase artifacts.

3.2 Robustness to Phase Artifacts

Next, we highlight NeuPh's robustness to various phase artifacts that arise from practical FPM experiments, including measurement noise, phase unwrapping errors, and artifacts resulting from an imperfect imaging model. As shown in Figs. 1(c) and 3(a), model-based FPM reconstruction may exhibit discontinuous artifacts due to imperfect phase unwrapping when dealing with samples with a phase range exceeding 2π . Moreover, Fig. 3(a) illustrates rippling artifacts in the background region of the model-based reconstruction, possibly resulting from the model-mismatch in the reconstruction.²⁷

In contrast, NeuPh (**NeuPh_{E(18)}**) effectively eliminates these artifacts and achieves accurate, smooth, and continuous reconstructions without any explicit regularization term. We quantitatively evaluate the artifact-suppression capability of NeuPh by measuring the residual random fluctuations in the background regions.²⁴ Our analysis shows that NeuPh can reduce the background artifacts by several folds compared with the model-based FPM reconstruction, as shown in Fig. 3(a). Additional

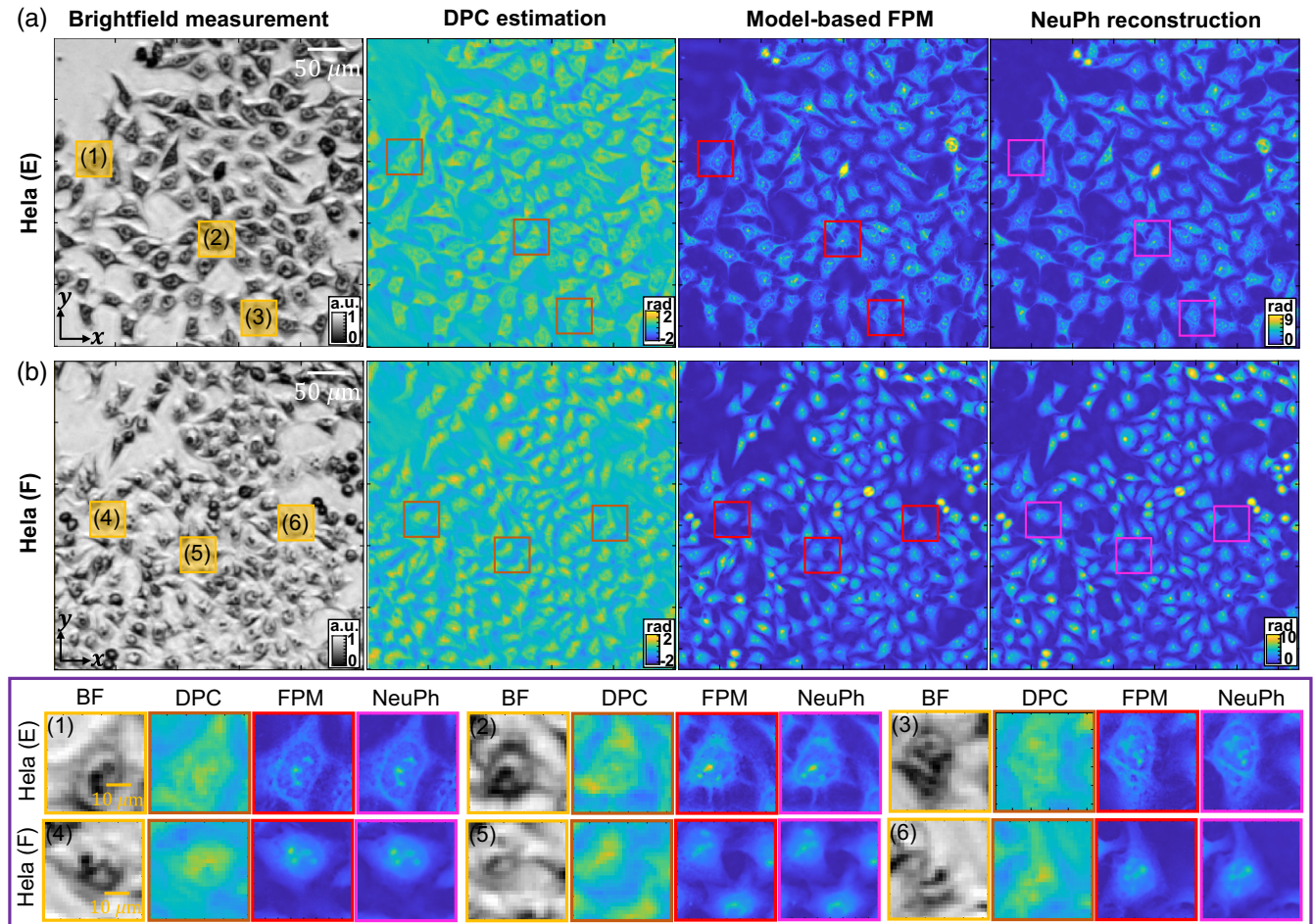


Fig. 2 Reconstruction results using NeuPh trained with the experimental dataset. An example of the BF low-resolution intensity image, DPC estimation, model-based FPM reconstruction, and NeuPh reconstruction for (a) HeLa(E) and (b) HeLa(F). Subareas (1)–(6) highlight specific regions of interest, demonstrating NeuPh’s capacity to accurately reconstruct subcellular high-resolution features without any artifacts.

results on phase artifact removal using NeuPh are shown in Fig. S7(a) in the [Supplementary Material](#).

This result is particularly intriguing because the NeuPh network was trained using this type of ‘noisy’ images from real experiments, and the network shows remarkable robustness in both the training and testing stages. We attribute this robustness to the continuity of the learned latent space and the continuous representation imposed by the coordinate-based decoder, both of which implicitly regularize the ill-posed inverse problem. First, NeuPh encodes the input images into a continuous latent space representation, effectively filtering out noisy information. Second, NeuPh decodes the phase value by a coordinate-based network, which implicitly learns a continuous neural representation of the object.

3.3 Superior Phase Retrieval Performance Compared to the CNN-Based Model

We conduct a thorough ablation study to illustrate NeuPh’s superior performance compared to CNN-based models. The detailed procedure of the ablation study is provided in the

[Supplementary Material](#), and the reconstruction results of ethanol-fixed HeLa cells are shown in Fig. 3(b). Specific subareas of interest are zoomed in, with subcellular details highlighted by red circles. In addition, spatial spectra are included at the bottom left of each figure, along with quantitative metrics, including the mean square error (MSE) and frequency measurement (FM) metric noted at the bottom. The FM quantifies the recovery of frequency components^{28,29}—a metric for measuring the resolving power of the reconstruction algorithm—where higher FM values represent the recovery of more frequency components. We also compute the structural similarity index measure (SSIM) and peak signal-to-noise ratio (PSNR), as detailed in Table S3 in the [Supplementary Material](#). Additional comparisons are shown in Figs. S7(b) and S10 in the [Supplementary Material](#). The metrics for an additional 100 testing image patches outside the training FOV region are provided in Table S4 in the [Supplementary Material](#). Our results show that the CNN-based network produces more reconstruction artifacts and fails to recover as many high-frequency components as NeuPh, as evident in the red circled regions and the reconstruction spectra in Fig. 3(b), as well as the FM scores.

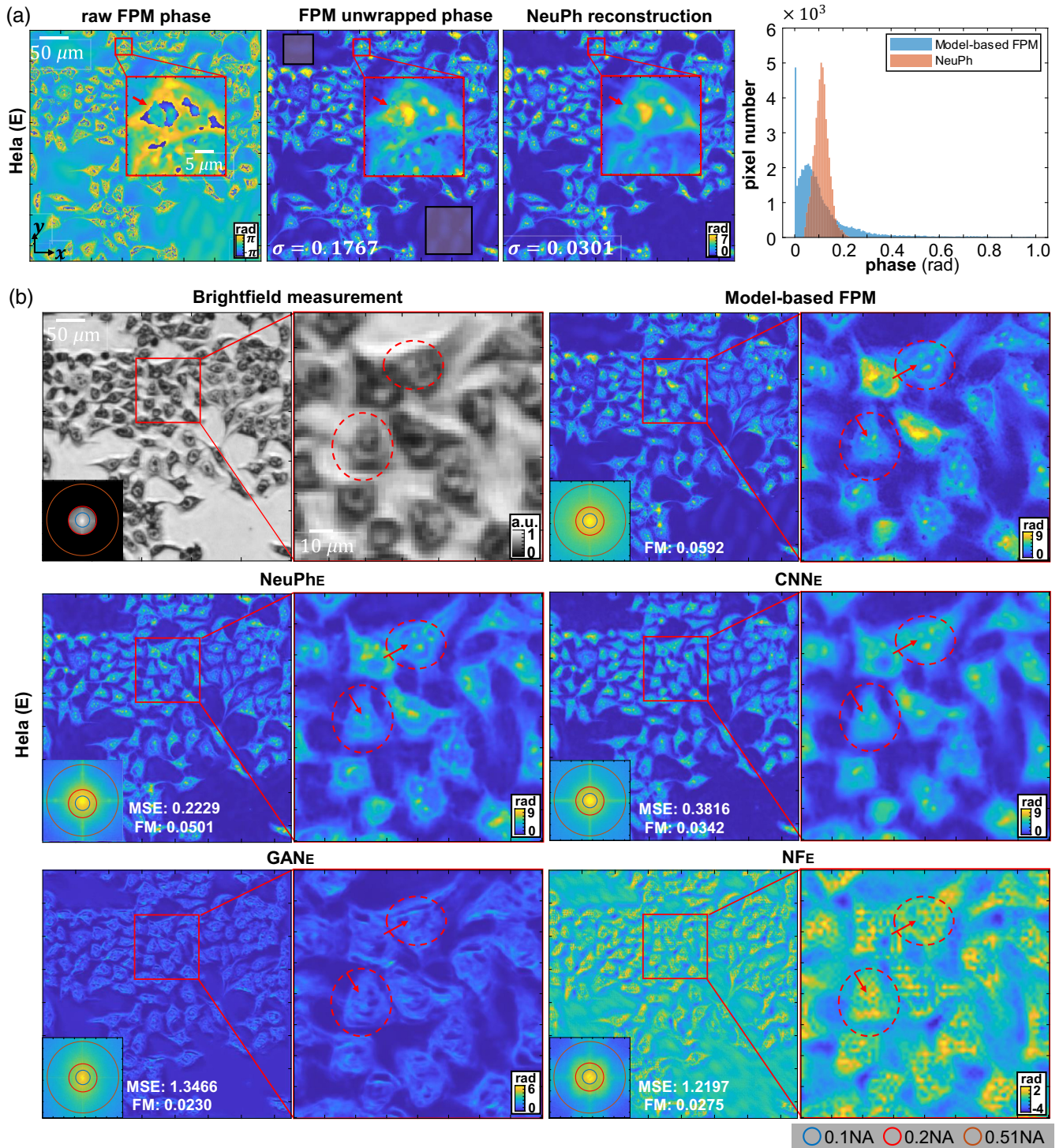


Fig. 3 (a) NeuPh's robustness to phase artifacts. NeuPh eliminates discontinuous phase-unwrapping errors (marked by red arrows) and background rippling artifacts (noted by the block box). The phase histogram of the background areas, measuring the residual background fluctuations, is shown in the rightmost column. The standard deviations (σ) are shown at the bottom of the reconstructions. (b) NeuPh outperforms the CNN-based reconstruction method and existing neural networks. Comparison between the reconstructions by NeuPh (**NeuPh_E**), CNN-based (**CNN_E**) networks, GAN (**GAN_E**), and traditional NF networks (**NF_E**), benchmarked by the ground-truth model-based reconstruction. Zoomed-in regions showcase intricate subcellular features that can be reconstructed with better resolution by NeuPh than other networks, as highlighted by the red circles and arrows. The reconstructed spectra are shown at the bottom left of each image, with blue, red, and brown circles indicating the bandwidth of the objective (0.1 NA), BF measurements (0.2 NA), and theoretically achievable reconstruction bandwidth (0.51 NA), respectively.

3.4 Superior Phase Retrieval Performance Compared to Existing Neural Networks

To further evaluate our network's phase retrieval performance, we compared NeuPh with existing state-of-the-art networks used in FPM reconstructions. Specifically, we compared our network's performance with phase retrieval results provided by a GAN-based network²³ and a traditional NF network.¹⁹ We trained the GAN-based network using the same single-paired experimental dataset as the CNN-based network and NeuPh in Sec. 3.3. The training took around 5 h with 500 epochs using an NVIDIA Quadro RTX 8000 GPU, and the inference for a single image took around 1 min. For the traditional NF method, we directly used the five captured raw intensity images since it is a self-supervised learning method and does not need paired images for training. The training (and reconstruction) for a single image took around 30 min using an NVIDIA Quadro RTX 8000 GPU. The reconstruction results for the GAN network and traditional NF network on ethanol-fixed Hela cells are shown in Fig. 3(b) GAN_E and NF_E , respectively. As shown in the figure, both the GAN-based network and traditional NF methods failed to recover high-resolution phase images, whereas our NeuPh successfully recovered the high-resolution phase image. Additional reconstructions for formalin-fixed Hela cells are shown in Fig. S7(b) in the [Supplementary Material](#). Since both the GAN-based network and the traditional NF network failed to reconstruct high-resolution images, and the traditional NF network lacks generalization ability and must be retrained for different objects, we did not test them on the additional 100 test samples.

The failure of previous state-of-the-art networks may arise from several factors. For the GAN-based network, changing the training dataset from hundreds of paired images to a single paired image, similar to our NeuPh training setup, made training the GAN-based network more challenging. As for the traditional NF network, we address the multiplex-FPM reconstruction problem, reducing the number of captured images from hundreds to only five, which significantly increases the ill-posedness of the problem compared to the one solved in Ref. 19.

3.5 Strong Generalization Capability of NeuPh

A notable advantage of our LCNF framework is its superior generalization capability, overcoming the limitations of existing NF frameworks⁵ that require retraining for different objects. To thoroughly evaluate NeuPh's generalizability, we perform phase reconstruction for ethanol-fixed Hela cells with different dataset-trained networks. The reconstruction results are presented in Fig. 4 with the corresponding MSE and FM scores provided in the figures. We also compute the PSNR and SSIM scores in Table S5 in the [Supplementary Material](#). In addition, to gain insights into NeuPh's robustness against realistic spatially varying aberrations in our experiment, we evaluate the reconstructions on an additional 100 image patches outside the training region, whose performance metrics are provided in Table S6 in the [Supplementary Material](#). In Fig. S8 and Tables S5 and S6 in the [Supplementary Material](#), we repeat the same evaluation on formalin-fixed Hela cells.

Our NeuPh successfully reconstructs high-resolution phase images regardless of the training dataset, showcasing its strong generalization. The MSE generally increases, while the PSNR, SSIM, and FM decrease when training NeuPh with a very limited dataset or a different type of data compared to the network trained with the same cell type and the full experimental dataset. This indicates that the network's generalization performance generally degrades when it is trained on a smaller training dataset or the distribution of the testing data is shifted from that of the training data, which is expected. However, the changes in the metric scores are small and hardly noticeable in the visualizations in Fig. 4, even for the network trained with a single paired training dataset. This highlights NeuPh's robust generalization. When NeuPh is trained with ethanol-fixed Hela cells ($\text{NeuPh}_{E(18)}$) and applied to formalin-fixed Hela cells, the FM is slightly higher than those of the network trained with formalin-fixed Hela cells ($\text{NeuPh}_{F(16)}$), as indicated in Fig. S8 and Tables S5 and S6 in the [Supplementary Material](#). We attribute this unusual result to the fact that ethanol-fixed Hela cells contain more structural details and provide a broader spectrum compared to formalin-fixed Hela cells [see the spectra shown in

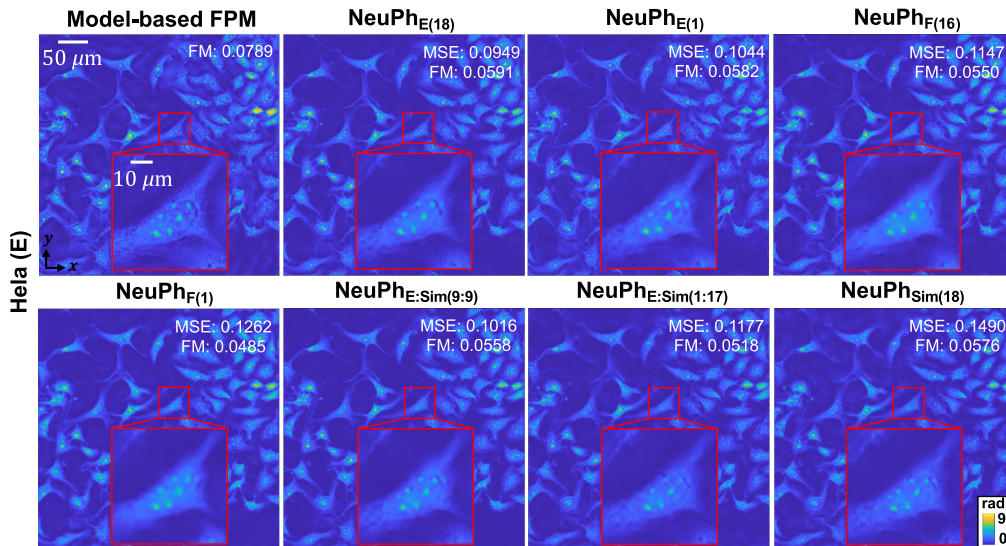


Fig. 4 Strong generalization capability of NeuPh. Reconstructions of ethanol-fixed Hela cells with different dataset-trained networks.

Fig. 3(b) and Fig. S7(b) in the [Supplementary Material](#)]. Therefore, **NeuPh_{E(18)}** may reconstruct more frequency components.

Furthermore, we demonstrate NeuPh's robust generalization by employing pure simulator-based training (**NeuPh_{Sim(18)}**, **NeuPh_{Sim(16)}**) to reconstruct the experimental Hela cell dataset, showcasing the straightforward application of simulator-based training for generating subcellular phase images from low-resolution intensity images. Our approach overcomes the limitations of existing networks when obtaining paired experimental training data is challenging, as they either require retraining^{5,13–15,17} or are hindered by the limitations of generative adversarial networks (GANs).²³ Although the quantitative results indicate that the simulator-trained NeuPh performs slightly worse than the experimental-data-trained networks, this difference is expected due to the significantly different image features between the natural images used in training and the cells in the experiment.

To better understand the influence of domain shift between simulation and experiment, Fig. 4 compares inference results in **NeuPh_{E(18)}**, **NeuPh_{E:Sim(9:9)}**, **NeuPh_{E:Sim(1:17)}**, and **NeuPh_{Sim(18)}** which are trained by different blends of experimental and simulated datasets. From the reconstructed images and the metrics in Table S5 in the [Supplementary Material](#), it is evident that incorporating experimental datasets consistently leads to better performance compared to networks trained solely on simulated datasets. Generally speaking, an increase in the proportion of the simulated data in the training set results in slightly degraded reconstruction performance on experimental measurements, indicating the adverse impact of the domain shifts between the simulation and experimental datasets. However, this performance degradation is minimal by implementing our distribution matching procedure, as detailed in the [Supplementary Material](#) and Fig. S3 in the [Supplementary Material](#). In addition, we show that the NeuPh trained solely on the experimentally captured biological images can successfully reconstruct simulated measurements from natural images in Fig. S9 in the [Supplementary Material](#).

We attribute NeuPh's accurate reconstruction and strong generalization capability to our local conditional neural representation and the coordinate-based training strategy. Unlike existing NF frameworks, NeuPh employs a CNN-based encoder to encode measurements and then provide object information to the decoder, enabling adaptability to different objects' reconstruction. In addition, the coordinate-based training strategy treats pixels as training pairs [see Eq. (1)], offering two potential advantages. First, by considering coordinates, the training data effectively expand from a single paired image to a diverse set of pixels. This allows NeuPh to learn from a vast and varied dataset, contributing to its superior generalization capabilities. Second, the coordinate-based training strategy helps NeuPh mitigate overfitting to specific image features in the training data. As a result, NeuPh can achieve high-quality reconstructions even when trained on a different cell type or completely different objects. These attributes not only reduce the necessity for a large number of training samples but also expedite the entire experimental process. This is particularly beneficial in challenging scenarios where collecting experimental training data is labor-intensive and expensive.

3.6 Robust Wide-FOV High-Resolution Phase Reconstruction

Finally, we employ NeuPh for wide-FOV high-resolution reconstructions, as shown in Fig. 5. The network **NeuPh_{E(18)}** trained

on experimental data utilizes measurements solely from the central 250 pixel $\times$ 250 pixel region, as indicated by the dashed black square in Fig. 5(a). The simulator-trained network **NeuPh_{Sim}** assumes an ideal imaging system without any aberrations. Subsequently, we perform phase reconstruction across a substantially larger FOV, covering a circular region with a 2160-pixel diameter in the raw measurements (3.51 mm FOV). The resulting wide-FOV reconstructions (12,960 pixel in diameter) are shown in Figs. 5(b) and 5(c) for **NeuPh_{E(18)}** and **NeuPh_{Sim}**, respectively. Furthermore, Figs. 5(d)–5(g) highlight specific subareas ranging from the central to the periphery of the FOV.

Overall, the NeuPh networks achieve high-quality reconstructions, with intricate subcellular details distinctly visible and minimal artifacts. However, some distortions are noticeable at the FOV's extreme periphery, as seen in the subarea (iv). These distortions can be attributed to the unaddressed spatially varying aberrations in our setup, which become more pronounced at the periphery of the FOV.³⁰ This model mismatch leads to a distributional shift between the training and testing datasets, worsening as the aberrations grow more severe. Addressing this limitation will require novel training strategies that integrate spatially variant imaging models.

Furthermore, we present additional wide-FOV reconstructions in Figs. S11–S13 in the [Supplementary Material](#) for experimental Hela cells using NeuPh networks trained with different datasets based on experimental or simulated datasets. The results further underscore NeuPh's reliability in achieving wide-FOV high-resolution phase reconstructions, regardless of the training dataset employed. Notably, our framework demonstrates excellent performance even when trained with very limited data, including a single paired image at the extreme case or when utilizing simulated training data.

4 Discussion and Conclusion

We have introduced LCNF, a scalable and generalizable DL framework for solving imaging inverse problems. Unlike traditional model-based or CNN frameworks, LCNF employs a continuous neural representation, facilitating flexible reconstruction of multiscale information. A novel local conditioning technique combined with a coordinate-based training strategy is incorporated, improving its accuracy compared with existing state-of-the-art deep learning models and significantly enhancing its generalization capability compared to existing NF frameworks.

By applying LCNF to solve the multiplexed FPM phase retrieval problem, we demonstrate a novel deep neural network, NeuPh, for continuous-domain large-scale super-resolution phase reconstruction from a sparse set of low-resolution intensity measurements. Notably, NeuPh exhibits robustness against noisy training data. Reconstructions by NeuPh are free from typical artifacts, such as residual phase unwrapping errors, noise, and background ripples, which contaminate the training data obtained by traditional FPM reconstructions. In addition, we observe the superior reconstruction performance of NeuPh, which recovers more frequency components with fewer artifacts compared to the CNN-based model and shows better reconstruction performance on multiplex FPM than existing state-of-the-art networks. This observation suggests that NeuPh has the potential to mitigate the inherent spectral bias associated with CNN-based models.

Furthermore, NeuPh demonstrates remarkable generalization across different object types and experimental conditions, surpassing existing NF frameworks. We demonstrate that NeuPh

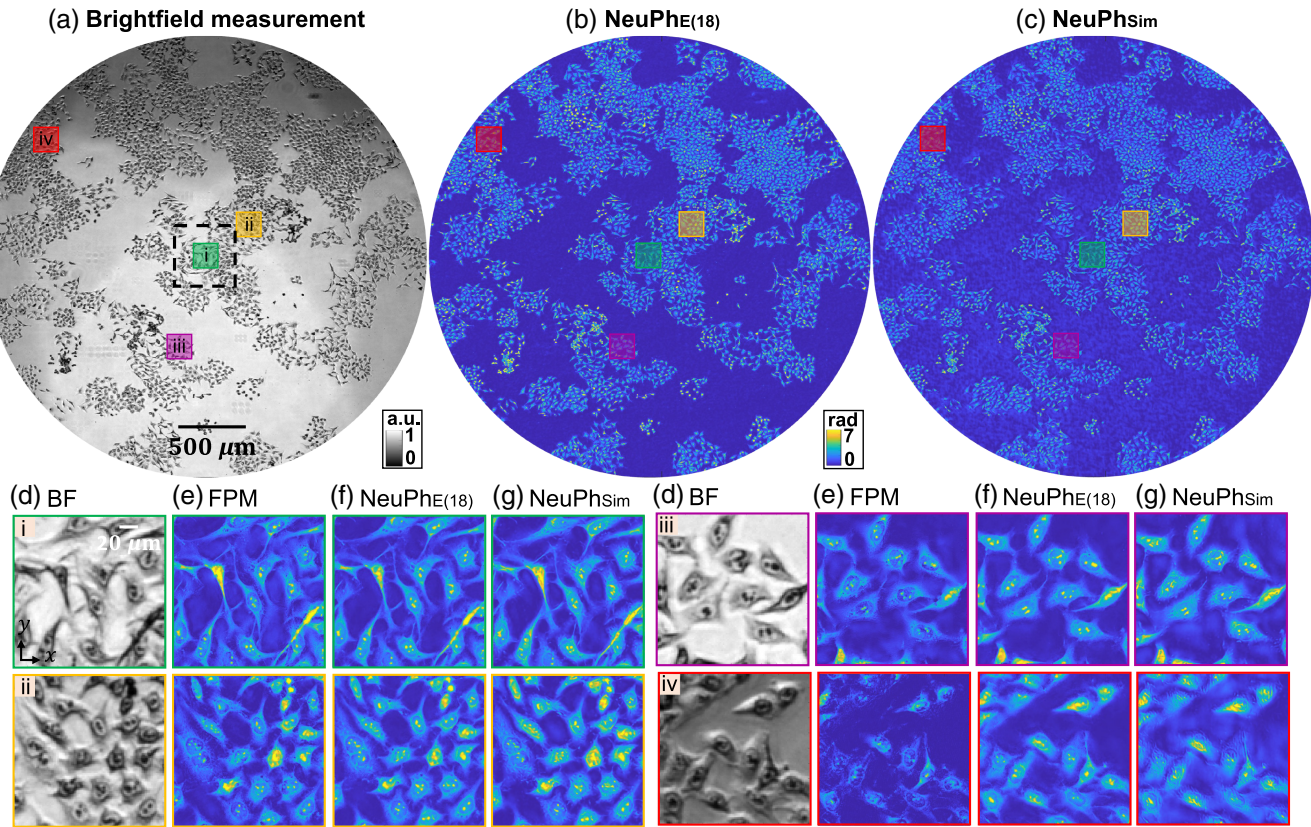


Fig. 5 Wide-FOV high-resolution phase reconstruction by NeuPh. (a) BF image captured over a 2160-pixel diameter (3.51 mm) FOV. Wide-FOV reconstruction by training NeuPh with the (b) experimental dataset (**NeuPh_{E(18)}**) and (c) simulated dataset (**NeuPh_{Sim}**). (d)–(g) Selected subareas extracted from the central to the edge of the FOV, identified as (i)–(iv), and enclosed within different colored boxes. (d) BF image. (e) Model-based reconstruction. (f) **NeuPh_{E(18)}** reconstruction. The experimental dataset used for training **NeuPh_{E(18)}** is obtained from the central region, indicated by the dashed black square. (g) **NeuPh_{Sim}** reconstruction.

trained on various datasets including purely simulated ones can reliably generalize to biological samples and reconstruct subcellular details with super-resolution without requiring network re-training or transfer learning. This highlights an effective strategy to circumvent the collection of experimental training data altogether, leveraging the knowledge of imaging physics to simulate training datasets. In addition, we emphasize the importance of aligning dataset distributions between the experimental and simulated datasets to optimize reconstruction results.

As indicated by our wide-FOV reconstruction results in Sec. 3.6, the unknown spatially varying aberrations in the imaging system present a main limitation that needs to be addressed to further improve NeuPh's performance. Both advanced physical modeling methods^{7,8} and novel network designs^{8,31} have been effective in addressing this issue in fluorescence microscopy. Adapting these ideas to phase microscopy offers a promising direction for future research. In addition, NF has demonstrated its unique capability in effectively modeling multi-dimensional spatiotemporal information in several dynamic imaging applications.^{14,17,18} By integrating these spatiotemporal NF frameworks with our local conditioning technique, one may achieve broad generalization and significantly reduce the computational cost in dynamic image reconstruction, making it another promising direction for future advancements.

In conclusion, we present LCNF as a scalable, robust, accurate, and generalizable DL-based continuous-domain image reconstruction framework. While our main focus in this work centers on addressing large-scale, super-resolution phase retrieval based on the multiplexed FPM technique, we envision that the LCNF framework holds potential for broader applications. It can be adapted to various computational imaging techniques for solving the underlying ill-posed inverse problems in areas, such as holographic imaging, imaging through complex media, and light-field microscopy.

Disclosures

The authors declare no competing interests.

Code and Data Availability

The neural network and example dataset used in this work are available at <https://github.com/bu-cisl/LCNF>.

Acknowledgment

The authors acknowledge the Boston University Shared Computing Cluster for providing the computational resources. This work was supported by the National Science Foundation (Grant No. 1846784).

References

1. G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," *Optica* **6**(8), 921–943 (2019).
2. J. Dong et al., "Phase retrieval: from computational imaging to machine learning: a tutorial," *IEEE Signal Process. Mag.* **40**(1), 45–57 (2023).
3. Y. Rivenson, Y. Wu, and A. Ozcan, "Deep learning in holography and coherent imaging," *Light: Sci. Appl.* **8**(1), 85 (2019).
4. A. Saba et al., "Physics-informed neural networks for diffraction tomography," *Adv. Photonics* **4**(6), 066001 (2022).
5. R. Liu et al., "Recovery of continuous 3D refractive index maps from discrete intensity-only measurements using neural fields," *Nat. Mach. Intell.* **4**(9), 781–791 (2022).
6. Y. Rivenson et al., "Deep learning microscopy," *Optica* **4**(11), 1437–1443 (2017).
7. Y. Xue et al., "Deep-learning-augmented computational miniature mesoscope," *Optica* **9**(9), 1009–1021 (2022).
8. K. Yanny et al., "Deep learning for fast spatially varying deconvolution," *Optica* **9**(1), 96–99 (2022).
9. Y. Li, Y. Xue, and L. Tian, "Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media," *Optica* **5**(10), 1181–1190 (2018).
10. K. Wang et al., "On the use of deep learning for phase recovery," *Light: Sci. Appl.* **13**(1), 4 (2024).
11. J. Park et al., "Artificial intelligence-enabled quantitative phase imaging methods for life sciences," *Nat. Methods* **20**(11), 1645–1660 (2023).
12. Y. Xie et al., "Neural fields in visual computing and beyond," *Comput. Graph. Forum* **41**(2), 641–676 (2022).
13. A. W. Reed et al., "Dynamic CT reconstruction from limited views with implicit neural representations and parametric motion fields," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, pp. 2258–2268 (2021).
14. L. Lozanski, M. A. Anastasio, and U. Villa, "A memory-efficient self-supervised dynamic image reconstruction method using neural fields," *IEEE Trans. Comput. Imaging* **8**, 879–892 (2022).
15. L. Shen, J. Pauly, and L. Xing, "NeRP: implicit neural representation learning with prior embedding for sparsely sampled image reconstruction," *IEEE Trans. Neural Netw. Learn. Syst.* **35**(1), 770–782 (2022).
16. E. D. Zhong et al., "CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks," *Nat. Methods* **18**(2), 176–185 (2021).
17. B. Y. Feng et al., "NeuWS: neural wavefront shaping for guide-star-free imaging through static and dynamic scattering media," *Sci. Adv.* **9**(26), eadg4671 (2023).
18. R. Cao et al., "Dynamic structured illumination microscopy with a neural space-time model," in *IEEE Int. Conf. Comput. Photogr. (ICCP)*, IEEE, pp. 1–12 (2022).
19. H. Zhou et al., "Fourier ptychographic microscopy image stack reconstruction using implicit neural representations," *Optica* **10**, 1679–1687 (2023).
20. L. Tian and L. Waller, "Quantitative differential phase contrast imaging in an led array microscope," *Opt. Express* **23**, 11394–11403 (2015).
21. L. Tian et al., "Multiplexed coded illumination for Fourier ptychography with an led array microscope," *Biomed. Opt. Express* **5**(7), 2376–2389 (2014).
22. G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution Fourier ptychographic microscopy," *Nat. Photonics* **7**(9), 739–745 (2013).
23. Y. Xue et al., "Reliable deep-learning-based phase imaging with uncertainty quantification," *Optica* **6**(5), 618–629 (2019).
24. H. Wang et al., "Fourier ptychographic topography," *Opt. Express* **31**, 11007–11018 (2023).
25. E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: dataset and study," in *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit. Workshops*, pp. 126–135 (2017).
26. Y. Chen, S. Liu, and X. Wang, "Learning continuous image representation with local implicit image function," in *Proc. IEEE/CVF Conf. Comput. Vis. and Pattern Recognit.*, pp. 8628–8638 (2021).
27. L.-H. Yeh et al., "Experimental robustness of Fourier ptychography phase retrieval algorithms," *Opt. Express* **23**(26), 33214–33240 (2015).
28. K. De and V. Masilamani, "Image sharpness measure for blurred images in frequency domain," *Proc. Eng.* **64**, 149–158 (2013).
29. T. Nguyen et al., "Deep learning approach for Fourier ptychography microscopy," *Opt. Express* **26**, 26470–26484 (2018).
30. X. Ou, G. Zheng, and C. Yang, "Embedded pupil function recovery for Fourier ptychographic microscopy," *Opt. Express* **22**(5), 4960–4972 (2014).
31. S. Fu et al., "Field-dependent deep learning enables high-throughput whole-cell 3D super-resolution imaging," *Nat. Methods* **20**(3), 459–468 (2023).
32. M. Chen, Z. F. Phillips, and L. Waller, "Quantitative differential phase contrast (DPC) microscopy with computational aberration correction," *Opt. Express* **26**(25), 32888–32899 (2018).
33. B. Lim et al., "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit. Workshops*, pp. 136–144 (2017).
34. D. C. Ghiglia and L. A. Romero, "Robust two-dimensional weighted and unweighted phase unwrapping that uses fast transforms and iterative methods," *J. Opt. Soc. Amer. A* **11**(1), 107–117 (1994).

Hao Wang is currently a PhD student in the Department of Electrical and Computer Engineering at Boston University. He received his BS degree in optical information science and technology from the University of Science and Technology of China (USTC) and his MS degree in optical engineering from the Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences in 2016 and 2019, respectively. His current research focuses on computational imaging.

Jiabei Zhu is currently a PhD student in the Department of Electrical and Computer Engineering at Boston University. He received his BS degree in optical information science and technology from the USTC in 2019. His current research focuses on computational imaging.

Yunzhe Li is currently a postdoctoral researcher in the Department of Electrical Engineering and Computer Sciences at University of California, Berkeley (UC Berkeley). She received her PhD in electrical engineering from Boston University and her MS degree from Columbia University. Her current research focuses on computational imaging.

Qianwan Yang is currently a PhD student working with Professor Lei Tian in the Computational Imaging Systems Lab at Boston University. She received her BS degree in biomedical engineering in the School of Engineering Sciences from Huazhong University of Science and Technology. Her research interests center around computational imaging, with particular emphasis on utilizing deep learning approaches to expand imaging capabilities.

Lei Tian is currently an associate professor in the Department of Electrical and Computer Engineering and Biomedical Engineering and directs the Computational Imaging Systems Lab at Boston University. He received his PhD from Massachusetts Institute of Technology. He was a postdoctoral associate at UC Berkeley. His research focuses on computational imaging and microscopy.