Physics-informed neural network for single-shot phase retrieval in cryo-EM

Yixiao Yang [1 2] **Joel Yeo** [1 3] **Yong Zi Tan** [1 2] **N. Duane Loh** [1 3]

[1] *Centre for Bioimaging Sciences, National University of Singapore, Singapore* [2] *Department of Biological Sciences, National University of Singapore, Singapore* [3] *Department of Physics, National University of Singapore, Singapore*. Correspondence to: N. Duane Loh duaneloh@nus.edu.sg.

1. Introduction

Cryo-electron microscopy (cryo-EM) has emerged as a transformative imaging technique for structural biology, enabling the visualization of macromolecular structures at near-atomic resolution [1]. However, several challenges remain, particularly in the case of small biomolecules such as biomolecular fragments. Small particles impart only subtle phase shifts in the incident electron wave, leading to low contrast in the resulting images. This makes it challenging to extract meaningful structural information, especially for particles that exhibit structural flexibility.

To enhance contrast, physical phase plates (PPP) have been introduced in cryo-EM to modulate the phase of the transmitted electron wave. Devices such as the Zernike phase plate (ZPP), Volta phase plate (VPP), and laser phase plate (LPP) provide enhanced contrast by shifting phase information, thereby improving the detectability of weakly scattering biological specimens [2]. However, despite their advantages, PPPs face several practical challenges, including fabrication complexity, beam-induced contamination, and phase drift over time.

In recent years, computational phase retrieval methods have become popular as an alternative, aiming to estimate the underlying phase directly from the detected observations via reconstruction algorithms [3, 4]. Traditional phase retrieval methods, such as iterative optimization algorithms, often suffer from slow convergence, sensitivity to noise, and reliance on multiple frames or additional prior knowledge. These limitations hinder the efficiency and accuracy of structure determination, especially in single-particle analysis.

In this work, we propose a physics-informed neural network (PINN) for single-shot phase retrieval in cryo-EM, which embeds the fundamental physical constraints of the imaging process within the neural network architecture to achieve accurate phase retrieval. Through end-to-end training on a simulated dataset, the proposed method is demonstrated to have the potential to bridge the gap between theory and experiments.

2. Method

2.1 The phase retrieval problem in cryo-EM

In cryo-EM, the detector cannot measure the complex-valued field and only obtains an intensity image [1], which can be expressed as:

$$g(\mathbf{x}) = |\psi_{\text{exit}}(\mathbf{x}) \otimes h(\mathbf{x})|^2 + n, \tag{1}$$

[1] See Appendix A for more details.

where $\psi_{\text{exit}}(\mathbf{x})$ is the exit wave, $h(\mathbf{x})$ represents the complex-valued point-spread function of the microscope, $\otimes$ denotes the convolution operation, and n is the observed noise.

The phase retrieval problem aims to estimate the exit wave from the intensity measurement and usually involves solving an optimization problem:

$$\hat{\psi}_{\text{exit}}(\mathbf{x}) = \underset{\psi_{\text{exit}}(\mathbf{x})}{\operatorname{argmin}} \, ||g(\mathbf{x}) - |\psi_{\text{exit}}(\mathbf{x}) \otimes h(\mathbf{x})|^2||_2^2. \tag{2}$$

2.2 Proposed physics-informed neural network

To solve this ill-posed problem (2), we propose a physics-informed neural network that leverages the cryo-EM image formation model and data-driven statistical priors. The overall architecture diagram of the network is shown in Fig. 1.

The proposed network adopts a K-stage progressive reconstruction framework, and each stage contains two modules, i.e., physics informed module and denoising & deicing module. The input of proposed network is the defocused intensity measurement collected by the detector and corresponding cryo-EM system parameters, and the output is the estimated phase of the exit wave. Specifically, the mathematical expression of PINN is as follows:

$$\begin{aligned} \hat{\psi}^K_{\text{exit}}(\mathbf{x}) &= f_{\text{PINN}}\left(g(\mathbf{x}), h(\mathbf{x})\right) \\ &= \left[f_{\text{UNet}} \circ f_{\text{Physics}}\right]^K \left(g(\mathbf{x}), h(\mathbf{x})\right), \end{aligned} \tag{3}$$

where f_{Physics} and f_{UNet} denote the operators of the physics informed module and denoising & deicing module, respectively.

Physics-informed module. This module aims to fully exploit the acquired defocused intensity image in conjunction with the corresponding cryo-EM image formation model to iteratively refine the estimated exit wave. Specifically, the current estimate of the exit wave is forward-propagated to the detection plane, where its amplitude is replaced with the observed intensity-derived amplitude. Subsequently, the updated wavefront is backpropagated to the exit wave plane, mathematically expressed as:

$$f_{\text{Physics}}(\psi(\mathbf{x})) \triangleq \left(\frac{\psi(\mathbf{x}) \otimes h(\mathbf{x})}{|\psi(\mathbf{x}) \otimes h(\mathbf{x})|}\sqrt{g(\mathbf{x})}\right) \otimes h^{-1}(\mathbf{x}). \tag{4}$$

Denoising & deicing module. This module is designed to mitigate the theory-experiment gap caused by the ice layer and observed noise. To this end, we employ a neural network (classical UNet architecture) to perform both denoising and deicing on the currently estimated real-valued projected potential.

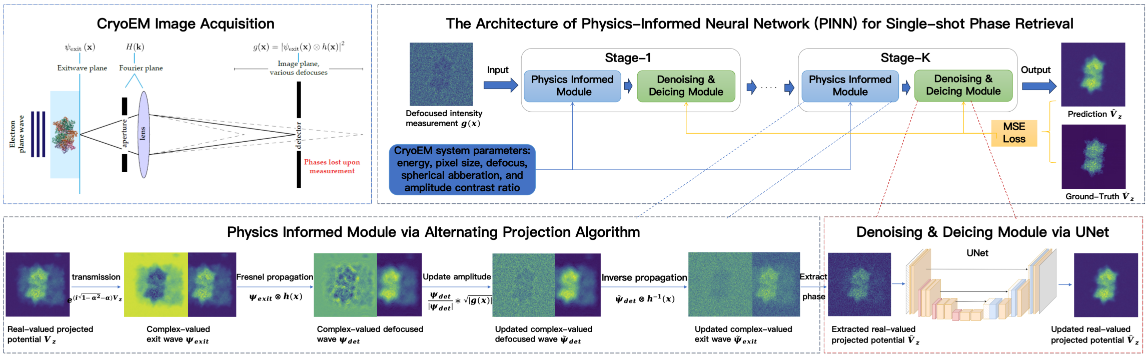

Fig. 1: Schematic diagram of the cryo-EM image acquisition and the proposed neural network architecture.

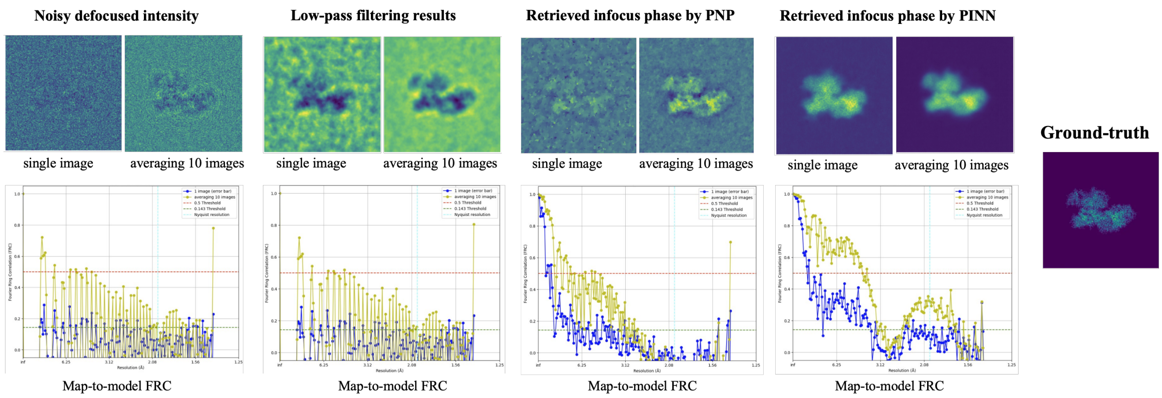

Fig. 2: Comparisons of reconstruction results using different methods.

Given an input phase profile (i.e., $\psi'(\mathbf{x})$), we mathematically define the operation of this neural network as $f_{\text{UNet}}(\psi'(\mathbf{x}))$.

To train the proposed network, we simulated and constructed a paired dataset consisting of defocused observations and their corresponding exit wave phases. The network parameters are updated by minimizing the mean square error (MSE) loss function between the predicted phases and the ground truth phases.

3. Results and Discussions

3.1 Implementation details

We generated simulated defocused intensity images of the particle '5MAC' [5] from various orientations, along with their corresponding exit wave phases, resulting in a dataset of 500 paired training samples. The imaging parameters were configured as follows: electron beam energy of 200 kV, amplitude contrast ratio of 0.1, defocus distance of 1000 nm, spherical aberration coefficient of 2 mm, and pixel size of 1 Å. And we implemented the proposed network on pytorch and empirically chose the number of stages to be 10. The neural network was trained for 500 epochs using a learning rate of 0.0001.

3.2 Compare with the existing methods

To assess the efficacy of the proposed method, we conducted experiments on the particle '8HLP' [6], which is entirely distinct from the training dataset. Fig. 2 presents a comparative analysis of the simulated observations, low-pass filtering results, reconstruction results using the iterative plug-and-play (PNP) phase retrieval algorithm with the total-variation denoiser [7], and reconstructions generated by the proposed PINN. For each method, we display both the reconstruction from a single observation and the averaging results derived from 10 observations captured at the same projection angle.

Furthermore, we compute the Fourier Ring Correlation (FRC) [8] curve between each reconstruction and the ground-truth result as a quantitative evaluation metric. The results indicate that, for single-image reconstructions, the proposed method not only enhances image contrast substantially but also yields a marked improvement in quantitative FRC measurements, as assessed by the 0.5 threshold criterion. Furthermore, in the subsequent 2D classification task, averaging 10 images processed with our method leads to a further enhancement in the attainable resolution, again confirmed by the 0.5 FRC criterion. We believe that the proposed PINN method has the potential to bridge the gap between theory and experiment and enhance single-particle reconstruction in cryoEM.

Acknowledgments

Y.Y. acknowledges the support of the Singapore Ministry of Education, AcRF Tier 1 grant, as well as the China Scholarship Council (202306030172).

References

[1] Werner Kühlbrandt. The resolution revolution. *Science*, 343(6178):1443–1444, 2014.

[2] Hong-Wei Wang and Xiao Fan. Challenges and opportunities in cryo-em with phase plate. *Current Opinion in Structural Biology*, 58:175–182, 2019.

[3] Yoav Shechtman, Yonina C Eldar, Oren Cohen, Henry Nicholas Chapman, Jianwei Miao, and Mordechai Segev. Phase retrieval with application to optical imaging: a contemporary overview. *IEEE signal processing magazine*, 32(3):87–109, 2015.

[4] Kaiqiang Wang, Li Song, Chutian Wang, Zhenbo Ren, Guangyuan Zhao, Jiazhen Dou, Jianglei Di, George Barbastathis, Renjie Zhou, Jianlin Zhao, et al. On the use of deep learning for phase recovery. *Light: Science & Applications*, 13(1):4, 2024.

[5] Laura H Gunn, Karin Valegård, and Inger Andersson. A unique structural domain in methanococcoides burtonii ribulose-1, 5-bisphosphate carboxylase/oxygenase (rubisco) acts as a small subunit mimic. *Journal of Biological Chemistry*, 292(16):6838–6850, 2017.

[6] Yiqing Wei, Zhuoya Yu, Lili Wang, Xiaojing Li, Na Li, Qinru Bai, Yuhang Wang, Renjie Li, Yufei Meng, Hao Xu, et al. Structural bases of inhibitory mechanism of cav1. 2 channel inhibitors. *Nature Communications*, 15(1):2772, 2024.

[7] Yixiao Yang and Ran Tao. Single-shot fractional fourier phase retrieval. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2023.

[8] Marin Van Heel and Michael Schatz. Fourier shell correlation threshold criteria. *Journal of structural biology*, 151(3):250–262, 2005.

[9] Cr J Humphreys and PB Hirsch. Absorption parameters in electron diffraction theory. *The Philosophical Magazine: A Journal of Theoretical Experimental and Applied Physics*, 18(151):115–122, 1968.

[10] Miloš Vulović, Lenard M Voortman, Lucas J van Vliet, and Bernd Rieger. When to use the projection assumption and the weak-phase object approximation in phase contrast cryo-em. *Ultramicroscopy*, 136:61–66, 2014.

Appendix A. Image formation model in cryo-EM

In cryo-EM, the biological specimens usually assume Humphrey's prior [9], which can be defined as

$$\tilde{V}(x,y,z) = \sqrt{1-\alpha^2}V(x,y,z) + \mathrm{i}\alpha V(x,y,z), \quad \text{(A1)}$$

where V is the real-valued scattering potential and α denotes the amplitude contrast ratio.

By applying projection approximation [10], the 3D complex-valued scattering potential of the sample at a given orientation is projected along the z-axis:

$$\tilde{V}_z(\mathbf{x}) = \int \tilde{V}(x,y,z)dz = \sqrt{1-\alpha^2}V_z(\mathbf{x}) + \mathrm{i}\alpha V_z(\mathbf{x}). \quad \text{(A2)}$$

The influence of the sample on the incident beam is expressed by the exit wave function as:

$$\psi_{\text{exit}}(\mathbf{x}) \approx \exp[\mathrm{i}\sigma\tilde{V}_z(\mathbf{x})] = \exp[(\mathrm{i}\sqrt{1-\alpha^2}-\alpha)\sigma V_z(\mathbf{x})], \quad \text{(A3)}$$

where σ is the interaction parameter.

Through the imaging system, the final image captured by the detector is expressed as:

$$g(\mathbf{x}) = |\psi_{\text{exit}}(\mathbf{x}) \otimes h(\mathbf{x})|^2 + n, \quad \text{(A4)}$$

where $\otimes$ denotes the convolution operation, n is the observed noise, and $h(\mathbf{x})$ represents the complex-valued point-spread function of the microscope, which is usually described by its Fourier transform:

$$H(\mathbf{k}) = \exp[-\mathrm{i}2\pi(-\frac{1}{2}\Delta f\lambda k^2 + \frac{1}{4}C_s\lambda^3 k^4)], \quad \text{(A5)}$$

where $k = |\mathbf{k}|$, Δf is the defocus distance, λ is the wavelength of the incident electron beam, and C_s is the spherical aberration constant.