Single-Shot Phase Retrieval in Cryo–EM via Latent Conditional Diffusion Model

Xiaodong Yang [1 2] **Yixiao Yang** [2] **Joel Yeo** [1 3] **N. Duane Loh** [1 3 4]
[1]*NUS Centre for Bio-imaging Sciences, National University of Singapore, Singapore 117557, Singapore* [2]*School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China* [3]*Department of Physics, National University of Singapore, Singapore 117551, Singapore* [4]*Department of Biological Sciences, National University of Singapore, Singapore 117551, Singapore*. Correspondence to: Yixiao Yang yangyixiao@bit.edu.cn, N. Duane Loh duaneloh@nus.edu.sg.

1. Introduction

Cryo-electron microscopy (cryo-EM) has emerged as a powerful imaging modality for resolving the structures of biological macromolecules at near-atomic resolution [1]. By rapidly freezing specimens in vitreous ice, cryo-EM enables imaging of biological samples in a close-to-native state while avoiding crystallization. Despite its remarkable success, cryo-EM imaging remains fundamentally constrained by extremely low electron dose requirements, which are necessary to mitigate radiation damage to sensitive biological structures [2].

One of the central challenges in cryo-EM image processing is phase retrieval. Due to the nature of electron imaging systems, the recorded measurements correspond to intensity-only observations in the detector plane, while the phase information—critical for accurate structural reconstruction—is not directly accessible [3]. Recovering phase information from noisy intensity measurements constitutes a severely ill-posed inverse problem, particularly under the stringent dose limitations inherent to cryo-EM [4].

A variety of techniques have been developed for phase retrieval in cryo-electron microscopy. In practice, the most widely adopted approach is based on the contrast transfer function (CTF), which provides a linearized approximation under the weak-phase object assumption and forms the backbone of modern cryo-EM reconstruction pipelines [5]. However, this linear assumption fundamentally limits high-resolution recovery and is particularly susceptible to the zero-crossing problem, where essential spatial frequency components are completely suppressed [6]. To overcome these limitations, iterative phase retrieval methods [7] have been explored by enforcing measurement consistency and prior constraints through repeated optimization. While capable of modeling nonlinear effects beyond CTF correction, these methods often exhibit slow convergence, strong sensitivity to initialization, and degraded robustness under low-SNR conditions.

In recent years, learning-based approaches have been increasingly investigated for cryo-EM inverse problems, including denoising, CTF correction, and phase retrieval [8, 9, 10, 11]. Despite their promising performance, these methods typically suffer from limited generalization and strong dependence on training distributions. More recently, diffusion probabilistic models have achieved state-of-the-art performance in image generation and restoration by learning a stable denoising process from a Gaussian prior [12, 13, 14]. Beyond unconditional generation, diffusion models have been extended to conditional and physics-guided inference for inverse problems [15], demonstrating strong performance in severely ill-posed imaging tasks such as deblurring, super-resolution, and tomographic reconstruction.

Applying diffusion models directly in pixel space for cryo-EM phase retrieval is challenging due to the extremely low signal-to-noise ratio of the measurements and the high spatial resolution of cryo-EM images, which together lead to unstable conditional guidance and prohibitive computational cost. In the proposed framework, a variational autoencoder serves as a structure-aware dimensionality reduction module that enhances the effective signal-to-noise ratio by suppressing noise-dominated components while preserving essential structural features from iterative phase retrieval results. Diffusion inference is therefore performed in this compact latent space, enabling stable and efficient conditional phase retrieval. As illustrated in Fig. 1, the measured intensity is first computationally propagated back to the exit-wave plane using an iterative phase retrieval method, encoded into a low-dimensional latent representation, and subsequently refined and decoded back to the clean high-resolution phase image domain after diffusion model and decoder, respectively.

In this framework, the variational autoencoder (VAE) [16] learns a compact latent representation that preserves essential structural information while suppressing high-frequency noise. Operating diffusion inference in this lower-dimensional and structured latent manifold improves stability and contrast under extremely low-SNR conditions. By integrating measurement consistency with learned generative priors in latent space, the proposed method enables robust phase retrieval without performing diffusion directly in pixel space, while substantially reducing computational cost and accelerating both training and sampling.

The main contributions of this work are as follows: we propose a latent conditional diffusion framework for cryo-EM phase retrieval, in which diffusion inference is performed in a variationally learned latent space. By integrating measurement consistency with learned generative priors, the proposed method enables robust phase retrieval. Experiments on both simulated and real cryo-EM datasets demonstrate improved phase reconstruction fidelity and significantly reduced computational cost, making the approach practical for large-scale cryo-EM data processing.

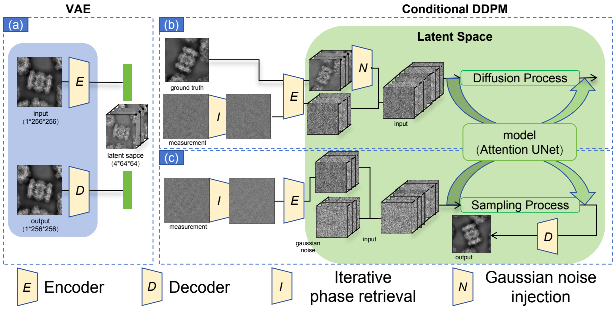

Fig. 1: The proposed network framework diagram. (a) represents the training process of the variational autoencoder, (b) represents the training process of the conditional diffusion model, and (c) represents the generation process of the entire method.

2. Method

As illustrated in Fig. 1, the proposed framework consists of two main components: a variational autoencoder (VAE) and a conditional diffusion model, which are trained separately and jointly used during inference to perform cryo-EM phase retrieval from real measurements through an encoder–diffusion–decoder pipeline.

Fig. 1 (a) depicts the training process of the VAE. We employ a dataset of 10,000 high-resolution cryo-EM images of size 256 × 256, which are compressed by the encoder into a 64 × 64 latent representation and subsequently reconstructed by the decoder back to the original resolution with minimal information loss. This learned latent space provides a compact and structured domain in which diffusion-based inference can be conducted more efficiently and stably than in the original pixel space.

Fig. 1 (b) illustrates the training procedure of the latent conditional diffusion model for phase retrieval. A dataset of 10,000 clean high-resolution phase images and their corresponding measurements is used for training. The measurements are processed by an iterative phase retrieval method and encoded into the latent space using a pretrained encoder. In the latent domain, Gaussian noise is progressively added to clean phase representations, while the encoded phase-consistent measurements serve as conditional inputs. The diffusion model is trained to perform conditional denoising in latent space, learning a generative prior consistent with both cryo-EM imaging physics and the underlying phase distribution.

Fig. 1 (c) depicts the inference pipeline for single-shot cryo-EM phase retrieval. A measured observation is first processed by an iterative phase retrieval method and encoded into latent space using the trained encoder. The latent representation is combined with Gaussian noise and refined through an iterative reverse diffusion process to recover a latent phase estimate, which is finally decoded to produce the high-resolution phase image.

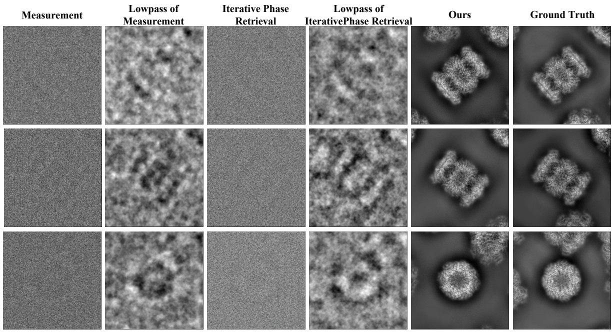

Fig. 2: The reconstruction result of the phase retrieval algorithm based on simulation data.

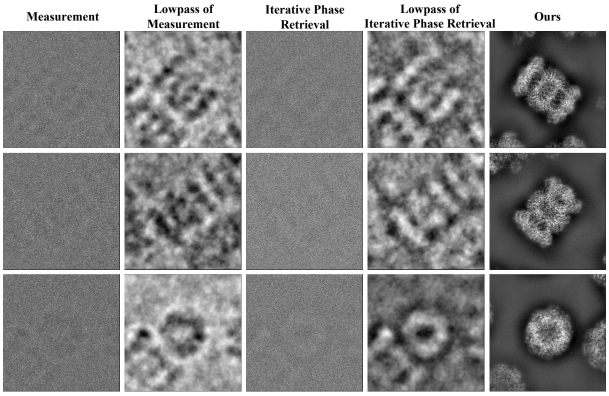

Fig. 3: The reconstruction result of the phase retrieval algorithm based on real experimental data.

3. Result

Fig. 2 and Fig. 3 present the validation results on simulated and real experimental cryo-EM datasets, respectively. The experimental cryo-EM data used in this study are obtained from the publicly available EMPIAR database, specifically EMPIAR-10025, which corresponds to the 2.8 Å resolution reconstruction of the Thermoplasma acidophilum 20S proteasome [17]. We compare the original measurements, reconstructions obtained using iterative phase retrieval methods, and the results produced by the proposed latent conditional diffusion framework. For fair comparison, both the raw measurements and the outputs of traditional methods are additionally processed using a low-pass filter. Across both simulated and experimental settings, the proposed method consistently recovers higher-resolution phase information and exhibits superior structural fidelity compared with conventional approaches under identical filtering conditions, yielding enhanced structural detail and more reliable phase reconstructions.

Acknowledgments

The authors would also like to acknowledge the computational resources from NUS Centre for Bio-Imaging Sciences. This work was supported by the

China Scholarship Council (CSC) under Grant No. 202506030120.

References

[1] Eva Nogales. The development of cryo-em into a mainstream structural biology technique. *Nature methods*, 13(1):24–27, 2016.

[2] Richard Henderson. The potential and limitations of neutrons, electrons and x-rays for atomic resolution microscopy of unstained biological molecules. *Quarterly reviews of biophysics*, 28(2):171–193, 1995.

[3] Miloš Vulović, Raimond BG Ravelli, Lucas J van Vliet, Abraham J Koster, Ivan Lazić, Uwe Lücken, Hans Rullgård, Ozan Öktem, and Bernd Rieger. Image formation modeling in cryo-electron microscopy. *Journal of structural biology*, 183(1):19–32, 2013.

[4] Pierre Thibault and Manuel Guizar-Sicairos. Maximum-likelihood refinement for coherent diffractive imaging. *New Journal of Physics*, 14(6):063004, 2012.

[5] Sjors HW Scheres. Relion: implementation of a bayesian approach to cryo-em structure determination. *Journal of structural biology*, 180(3):519–530, 2012.

[6] Joel Yeo, Benedikt J Daurer, Dari Kimanius, Deepan Balakrishnan, Tristan Bepler, Yong Zi Tan, and N Duane Loh. Ghostbuster: A phase retrieval diffraction tomography algorithm for cryo-em. *Ultramicroscopy*, 262:113962, 2024.

[7] Apoorv Pant, Manidipa Banerjee, and Kedar Khare. Quantitative phase imaging of single particles from a cryoem micrograph. *Optics Communications*, 506:127588, 2022.

[8] Dimitry Tegunov and Patrick Cramer. Real-time cryo-electron microscopy data preprocessing with warp. *Nature methods*, 16(11):1146–1152, 2019.

[9] Tristan Bepler, Andrew Morin, Micah Rapp, Julia Brasch, Lawrence Shapiro, Alex J Noble, and Bonnie Berger. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature methods*, 16(11):1153–1160, 2019.

[10] Yixiao Yang, Joel Yeo, Yong Zi Tan, and N Duane Loh. Can computational phase retrieval help 2d classification in cryo-em? In *13th Asia Pacific Microscopy Congress 2025 (APMC13)*, page 390. ScienceOpen, 2025.

[11] Yixiao Yang, Joel Yeo, Yong Zi Tan, and N Duane Loh. Physics-informed neural network for single-shot phase retrieval in cryo-em. In *AI4X 2025 International Conference*.

[12] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[13] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

[14] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

[15] Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*, 2023.

[16] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. *arXiv*, abs/1312.6114, 2013.

[17] Melody G Campbell, David Veesler, Anchi Cheng, Clinton S Potter, and Bridget Carragher. 2.8 å resolution reconstruction of the thermoplasma acidophilum 20s proteasome using cryo-electron microscopy. *Elife*, 4:e06380, 2015.