Simulation-to-Real Alignment for Cryo-EM Image Modeling via Conditional GAN-Based Refinement

Ziyue Jiang [1] **Yixiao Yang** [2] **Joel Yeo** [3] **N. Duane Loh** [3 4]

[1]*National University of Singapore, Singapore 117551, 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 117557, Singapore.*
Correspondence to: N. Duane Loh duaneloh@nus.edu.sg.

1. Introduction

Cryo-electron microscopy (Cryo-EM) is now a standard tool for protein structure determination, enabling high-quality reconstructions for many macromolecular complexes [1, 2]. However, several regimes remain challenging in practice—notably small proteins and highly heterogeneous or flexible targets—where reconstruction and classification can still be brittle even in widely used pipelines such as RELION and CryoSPARC [2, 3, 4]. A key factor is the extremely low signal-to-noise ratio (SNR) of experimental particle images, which makes reconstruction and data-driven models sensitive to small mismatches in image statistics.

Deep learning has improved multiple stages of the Cryo-EM workflow, including particle picking and heterogeneity modeling [2]. Representative examples include Topaz [5], CryoDRGN [6], and AlphaCryo4D [7]. Many methods rely on weak or limited supervision, and learning signals can become unreliable at very low SNR. Fully supervised training is more stable when reliable targets (e.g., cleaner or less-degraded images) exist, yet such paired "ground-truth" data are intrinsically difficult to obtain at scale [2].

Supervised learning typically requires paired data linking noisy observations to reliable targets. A common workaround is to synthesize such pairs using physics-based forward models with known parameters. However, simulators only approximate real image formation [8], leading to a persistent simulation–experiment gap [9]. This gap often appears as mismatches in appearance and frequency statistics, which degrades the transfer from simulation-trained models to experimental data. In practice, models trained on simulated particles often fail to generalize because simulated images do not faithfully match the frequency statistics of real observations.

To address this problem, we study particle-level alignment: given paired simulated and experimental particle crops, we refine simulated images to better match experimental observations while preserving pose and contrast transfer function (CTF)-conditioned content, where CTF models the microscope's modulation of spatial frequencies.

Recent work such as CryoGEM explores physics-informed generative modeling to bridge simulation and measurement statistics [10]. In contrast, we focus on a paired, lightweight refinement module that can be inserted into simulation-driven pipelines. We train a conditional GAN (pix2pix-style) to map simulated crops (`sim`) to experimental observations (`raw`). Because PatchGAN-based objectives emphasize local texture statistics while weakly constraining global frequency structure, we incorporate the radially averaged log power spectrum (RPSD) to encourage frequency-domain consistency, providing a practical way to reduce the simulation–experiment gap with quantitative evidence beyond visual inspection.

2. Methods

We study paired Cryo-EM particle translation from simulated images to experimental observations. Paired samples are constructed on a per-particle basis by simulating each experimental particle with CryoSIM [11], using its pose estimated from CryoSPARC homogeneous refinement and the corresponding CTF parameters, yielding one-to-one aligned (`sim`, `raw`) pairs. Each training pair is (x, y) where $x = \texttt{sim} \in \mathbb{R}^{256\times256}$ and $y = \texttt{raw} \in \mathbb{R}^{256\times256}$. Our goal is to generate $\hat{y} = G(x)$ (denoted as `fake`) whose *visual statistics* match `raw`.

2.1 Conditional GAN backbone

Generator. G is a lightweight U-Net for image-to-image translation [12, 13]: $\texttt{fake} = G(\texttt{sim})$.

Discriminator. We use a conditional PatchGAN D [13] that operates on local patches, taking the channel-wise concatenation $[\texttt{sim}, z]$ where $z \in \{\texttt{raw}, \texttt{fake}\}$, and outputs a patch-level realism map. The adversarial learning objective follows the standard GAN formulation [14].

2.2 Training objective

We follow the standard pix2pix training scheme with alternating updates of G and D [13]. The generator is optimized with (i) an adversarial term to match the `raw` domain and (ii) a paired L_1 term to stabilize learning. To better match global frequency statistics under extremely low SNR, we add a frequency-domain constraint based on the radially averaged log power spectrum (RPSD), motivated by common Fourier and amplitude-spectrum diagnostics in Cryo-EM and CTF assessment [15, 16]:

$$\mathcal{L}_G = \lambda_{\text{GAN}}\mathcal{L}_{\text{GAN}} + \lambda_1\|G(x) - y\|_1 + \lambda_{\text{RPSD}}\mathcal{L}_{\text{RPSD}} \quad (1)$$

where λ_{GAN}, λ_1, and λ_{RPSD} are scalar weights for the adversarial, pixel-domain L_1, and frequency-domain terms, respectively, and $\mathcal{L}_{\text{RPSD}}$ is defined in Eq. (2).

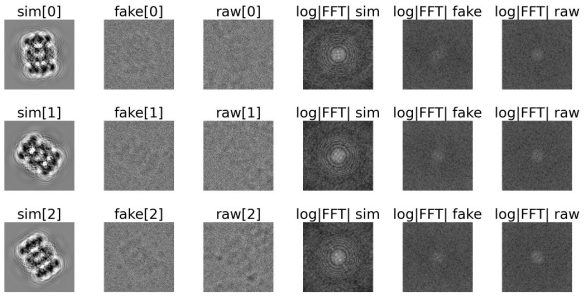

Fig. 1: Qualitative and FFT diagnostics at step 3000 (fixed visualization batch). Shown columns: `sim`, `fake`, `raw`, and $\log|\mathcal{F}(\cdot)|$ for `sim`/`fake`/`raw`.

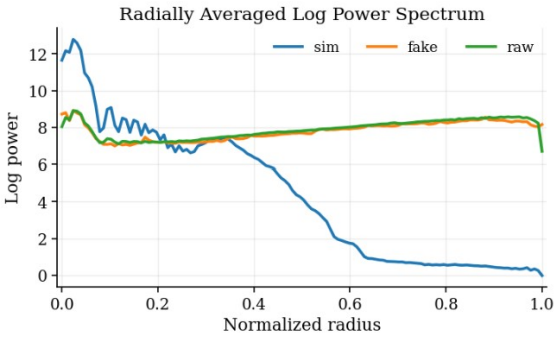

Fig. 2: Radially averaged log power spectrum (RPSD) at step 3000 (fixed visualization batch). `fake` aligns closely with `raw` over most radii, while `sim` shows a clear mismatch.

2.3 Frequency-domain diagnostic (RPSD)

Cryo-EM images exhibit characteristic frequency statistics (e.g., Thon rings in Fourier space that are closely related to CTF behavior and micrograph quality assessment) [15, 16]. To quantify frequency-domain agreement between `fake` and `raw`, we compute the *radially averaged log power spectrum* (RPSD). Let $P(u) = |\mathcal{F}(u)|^2$ be the 2D power spectrum and $s(u)$ the radial average of $\log(1 + P(u))$ over normalized radius. We define:

$$\mathcal{L}_{\text{RPSD}} = \left\| s(G(x)) - s(y) \right\|_1 \tag{2}$$

and report the same quantity as a diagnostic metric during training and validation.

2.4 Implementation and outputs

We record losses and RPSD on a fixed visualization batch, save checkpoints, and export test predictions (`sim`, `fake`, `raw`) for downstream analysis, including spatial comparisons and FFT/RPSD plots.

3. Results

We evaluate the paired translation `sim`→`fake` primarily in terms of frequency-domain consistency.

3.1 Qualitative and frequency-domain evidence

Fig. 1 shows representative examples from the fixed visualization batch. Compared to the structured appearance of `sim`, the generated `fake` more closely matches the experimental `raw` in both visual texture statistics and Fourier-domain patterns, indicating improved alignment of acquisition-related frequency characteristics under extremely low SNR.

To summarize frequency alignment more compactly, Fig. 2 reports the radially averaged log power spectrum (RPSD) curves. Across most radii, the `fake` curve closely tracks `raw`, while `sim` deviates noticeably, particularly toward higher normalized radii. This indicates that the translation reduces the simulation-to-experiment gap in frequency statistics.In preliminary runs without the RPSD term, frequency alignment is visibly weaker, consistent with PatchGAN's limited global frequency constraints.

Table 1: Metrics at step 3000 (from the training log).

Metric	Train (fixed vis batch)	Validation
L_D	0.248796	–
L_G	0.367816	–
GAN loss	0.126427	–
Paired L1 loss	0.233720	–
RPSD L1	0.131791	0.130181
RPSD MSE	0.101840	0.097661
Pixel L1	–	0.234102
Pixel MSE	–	0.086429

3.2 Quantitative metrics

Table 1 summarizes the metrics at step 3000. In addition to standard pix2pix losses [13], we report frequency-domain discrepancies between `fake` and `raw` on both the fixed visualization batch and the validation split. Notably, RPSD L1/MSE values on validation are comparable to those on the fixed batch, suggesting that the observed frequency alignment generalizes beyond the visualization examples.

Future work. An important next step is to evaluate whether refinement of simulated particles improves downstream generalization from simulation-trained models to experimental Cryo-EM data, for example the generated particles and real data are used simultaneously for downstream tasks such as 3D reconstruction or phase recovery, and then compared.

Acknowledgments

The authors would like to acknowledge the computational resources from the NUS Centre for Bio-Imaging Sciences, as well as the NUS AI for Science MSc program.

References

[1] Werner Kühlbrandt. Microscopy: Cryo-em enters a new era. *eLife*, 3:e03678, 2014.

[2] Jose Luis Vilas, Jose Maria Carazo, and Carlos Oscar S Sorzano. Emerging themes in cryoem—

single particle analysis image processing. *Chemical Reviews*, 122(17):13915–13951, 2022.

[3] Sjors H. W. Scheres. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, 180(3):519–530, 2012.

[4] Ali Punjani, John L. Rubinstein, David J. Fleet, and Marcus A. Brubaker. cryosparc: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, 14(3):290–296, 2017.

[5] Tristan Bepler, Andrew Morin, 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.

[6] Ellen D. Zhong, Tristan Bepler, Bonnie Berger, and Joseph H. Davis. Cryodrgn: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature Methods*, 18(2):176–185, 2021.

[7] Zhaolong Wu, Enbo Chen, Shuwen Zhang, Yinping Ma, and Youdong Mao. Visualizing conformational space of functional biomolecular complexes by deep manifold learning. *International Journal of Molecular Sciences*, 23(16):8872, 2022.

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

[9] Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017.

[10] Jiakai Zhang, Qihe Chen, Yan Zeng, Wenyuan Gao, Xuming He, Zhijie Liu, and Jingyi Yu. Cryogem: Physics-informed generative cryo-electron microscopy. *Advances in Neural Information Processing Systems*, 38, 2024.

[11] 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, 2023.

[12] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241, 2015.

[13] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976, 2017.

[14] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2014.

[15] Alexis Rohou and Nikolaus Grigorieff. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*, 192:216–221, 2015.

[16] Kai Zhang. Gctf: Real-time CTF determination and correction. *Journal of Structural Biology*, 193(1):1–12, 2016.