ROBUST SINGLE-PARTICLE CRYO-EM IMAGE DENOISING AND RESTORATION

Jing Zhang[1,2,], Tengfei Zhao[1,*], ShiYu Hu[1], Xin Zhao[2,†]*

1. Institute of Automation, Chinese Academy of Sciences, Beijing, China
2. School of Computer and Communication Engineering
University of Science and Technology Beijing, Beijing, China

ABSTRACT

Cryo-electron microscopy (cryo-EM) has achieved *near-atomic level* resolution of biomolecules by reconstructing 2D micrographs. However, the resolution and accuracy of the reconstructed particles are significantly reduced due to the extremely low signal-to-noise ratio (SNR) and complex noise structure of cryo-EM images. In this paper, we introduce a diffusion model with post-processing framework to effectively denoise and restore single particle cryo-EM images. Our method outperforms the state-of-the-art (SOTA) denoising methods by effectively removing structural noise that has not been addressed before. Additionally, more accurate and high-resolution three-dimensional reconstruction structures can be obtained from denoised cryo-EM images.

Index Terms— Singe particle cryo-electron microscopy, image denoising, image restoration, diffusion model

1. INTRODUCTION

Cryo-electron microscopy (cryo-EM), as a crucial research tool in structural biology, enables the determination of molecular structures with high resolution, which relies on the accurate interpretation of particle structures [1]. However, due to the limitations of low electron dose, 2D cryo-EM images always suffer from extremely low signal-to-noise ratio (SNR) (see Fig. 1(a)), and the presence of complex noise can result in inaccurate protein structure analysis. Currently, the advancement of computer vision has infused new vitality into various cryo-EM tasks, including image denoising and restoration [2][3], image clustering[4], particle picking [5], and more.

The basic task of cryo-EM image processing is to minimize the interference of randomly distributed background noise, and restores the frozen state of particles in glassy ice. One representative tool for this purpose is Topaz-Denoise [6], which is publicly available denoising model specifically designed for cryoEM images. It leverages the Noise2Noise framework [7] to directly learn the denoising from the original data, thereby improving the SNR and accelerating particle picking. The following 2-D clustering is also designed to generate high-quality particle class averages through operations such as grouping and alignment. It has a certain guiding
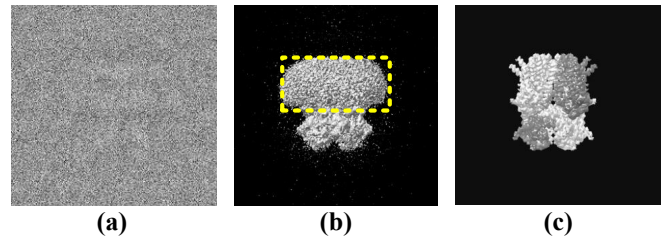
Fig. 1. Noises cryo-EM images of the HCN1 channel[12]. **(a)** Real cryo-EM image (EMPIAR-10081). **(b)** 3D cryo-EM density map (EMD-8511, yellow box represents structural noise). **(c)** Biological structure (PDB: 5u6o).

effect on the evaluation of data quality and subsequent 3D reconstruction [8]. However, cryo-EM images contain not only the aforementioned randomly distributed background noise but also structural noise caused by the presence of carbon and ice in the background, as well as variations in molecular conformation [9]. Traditional image processing methods like Block-matching and 3-D filtering (BM3D) [10] and K-SVD [11] are very difficult to restore clear particles in high background noise, and it is more challenging to achieve structural noise reduction. As demonstrated in Figs. 1(b) and 1(c), where the reconstructed 3D structure of the human HCN1 hyperpolarization-activated cyclic nucleotide-gated ion channel still exhibits indistinguishable structural noise when compared with the original biological structure.

In recent years, the continuous improvement of computing power and advancements in learning algorithms have made it possible to analyze complex noise. The Noise-Transfer2Clean (NT2C) [13] claims to be the first denoising method to address complex noise for the cryo-EM images. The core idea is using Generative Adversarial Networks (GAN) to learn the noise model of cryo-EM images, and further design a contrast guided noise and signal re-weighted algorithm to solve complex noise. In order to solve the problem that traditional image generation models may be polluted by a small portion of unknown outliers, Hanlin Gu et al. [14] utilized β-GAN achieves robust estimation of certain distribution parameters. It can be noticed that the strong capabilities of generative networks in the denoising and restoration of cryo-EM images due to their unique technical characteristics.

---

*Equal Contribution, †Corresponding Author

At present, the diffusion model [15], as a new branch of the generative networks, has shown brilliant results in natural image processing. Moreover, diffusion model also significantly surpasses GANs in terms of interpretability and stability. Furthermore, the promising development of DeepTracer [16] is encouraging, which currently achieves 3D electron density map denoising by utilizing the patterns of noise and the biological structure. This approach effectively eliminates noise signals, including both background noise and structural noise. Inspired by the achievements of DeepTracer, we aim to explore future research directions in single-particle denoising and restoration. The goal is to directly address complex noise in cryo-EM images and accurately restore the true structure of the particles. By directly improving the image quality, many tedious follow-up works can be circumvented, leading to superior reconstruction outcomes.

In this study, we present a framework that combines the diffusion model with post-processing techniques to address the challenging task of single-particle cryo-EM images denoising and restoring. Our contributions are listed as follows:

- Our method offers a direct approach for denoising and restoring single-particle cryo-EM images in situations of extremely low signal-to-noise ratio (SNR).

- Complex noise in cryo-EM images can be effectively removed by our method, especially structural noise. It is worth noting that no existing method addresses structural noise in real cryo-EM images.

- Our method achieves the best denoising effect on both simulated and real datasets. By leveraging the denoised cryo-EM images, we can attain more accurate and higher-resolution 3D structures.

2. METHOD

2.1. Architecture Design

Single particle cryo-EM images denoising and restoration aims to remove all interference noise and restore the structure of the particle itself. In this work, we not only need to excellently restore the real frozen state of particles under the interference of high background noise, but also collapse the indistinguishable structural noise covering the particle surface. Our proposed method, illustrated in the Fig. 2, involves a primary framework employing a diffusion model to restore the particle's structure, followed by a lightweight post-processing module to expedite model convergence. Below we provide a comprehensive description of our methodology.

2.1.1. Diffusion Model

The initial diffusion model, like the Denoising Diffusion Probabilistic Models (DDPM), is an unconditional generative network that converts the Gaussian distribution into a specific data distribution according to the process of iterative denoising. There are two processes: diffusion and reverse
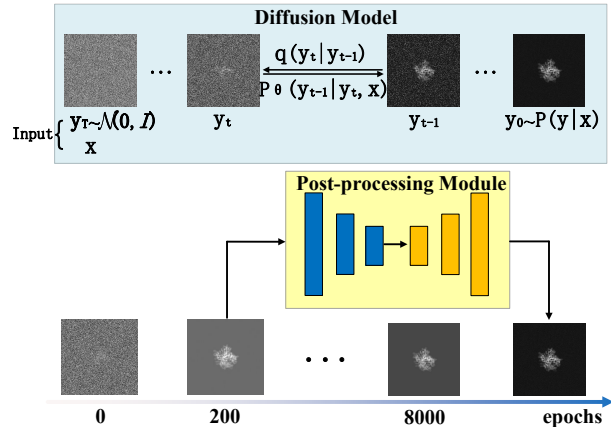
Fig. 2. Framework Overview. It includes two parts: diffusion model and post-processing module. The diffusion model restores the particle structure, and the post-processing accelerates the convergence.

processes. The diffusion process is parameter-free, and the mean and variance are determined at each step. Given the initial data distribution $y_0 \sim q(y)$, the noise is continuously injected until it becomes an independent Gaussian distribution. In the reverse process, a model with parameters $p_\theta(y_0)$ is used to learn to recover the initial data from Gaussian noise.

Due to the single particle restoration requires further denoising on cryo-EM images, unlike DDPM which starts from pure Gaussian noise. The noisy cryo-EM image x can be used as prior conditions to guide model learning. Therefore, in this work we use a conditional diffusion model, Super-Resolution via Repeated Refinement (SR3) [17]. The goal is to learn a conditional model $p_\theta(y_0|x)$. So the reverse process can be expressed as:

$$p_\theta(y_{0:T}|x) = p(y_T) \prod_{t=1}^{T} p_\theta(y_{t-1}|y_t, x) \quad (1)$$

$$p_\theta(y_{t-1}|y_t, x) = \mathcal{N}(y_{t-1}|\mu_\theta(x, y_t, \gamma_t), \sigma_t^2 I) \quad (2)$$

where (x, y_0) are sampled jointly from the data distribution. γ is the variance of the added noise. Through conditioning on γ, the denoising model can understand the denoising level, which is conducive to the selection of the number of diffusion steps and the noise scheduling in the inference process.

2.1.2. Post-processing Module

Diffusion model has shown outstanding performances, but it has a major drawback: it's gradual sampling characteristics will greatly extend the experimental time. In the experiment we observed that for single particle cryo-EM images denoising and restoration, the diffusion model can quickly reveal the particle structure, but it takes longer time to gradually reduce noise. However, the main problem that plagues current methods is the difficulty in identifying particle structures under high-noise conditions. Therefore, we first use

the diffusion model to quickly figure out the particle structure (only trained for 200 epochs), and then we can use a current method to carry out step noise reduction on this basis. In our work, we use a simple U-Net [18] to complete this module. Through experimental observations, under the same hardware configuration, our framework can achieve better results and shorten the training time by nearly 12 times. The training time of our framework is less than 1 day, and the test results are shown in the table 1. When the diffusion model iterates to 5,000 epochs, take 12 days, and the test results of EMD-8511 and EMD-24928 are respectively: MSE (1.028-2), PSNR (21.264), SSIM (0.824) and MSE(1.298e-2) PSNR(20.163) SSIM(0.870).

2.2. Training

The training part of the method is divided into two steps. Firstly, for the diffusion model, the training goal is to optimize the log likelihood function $logp_\theta(y_0|x)$. According to SR3, the optimization objectives are simplified to:

$$\mathbb{E}_{(x,y)}\mathbb{E}_{\epsilon,\gamma} \left\| f_\theta(x, \sqrt{r}y_0 + \sqrt{1-\gamma}\epsilon, \gamma) - \epsilon \right\|_2 \qquad (3)$$

where $\epsilon \sim \mathcal{N}(0, I)$, f_θ is the denoising model trained to predict noise vectors ϵ given any noisy image.

Secondly, for the training of the post-processing module, we use the mean squared error (MSE) loss function between the output images and the ground truth images.

3. EXPERIMENT

3.1. Datasets

In this section, we perform numerical experiments with (i) two simulated datasets and (ii) a real dataset.

The simulated datasets is prepared from the particles of human HCN1 hyperpolarization-activated cyclic nucleotide-gated ion channel and dolphin prestin, which can be obtained in the Electron Microscopy Data Bank (EMDB) with the entries of EMD-8511 [12] and EMD-24928 [19] respectively. We first classify and remove noise in the 3D structure through the DeepTracer [16], and then generate clean 2D projection images in the direction of random sampling of the sphere. The simulated data is generated by adding Gaussian white noise to the clean projected image.

The real dataset is a real-world experimental dataset of the human HCN1 hyperpolarization-activated cyclic nucleotide-gated ion channel. It can be available in the Electron Microscope Pilot Image Archive (EMPIAR) with 10081 entry.

3.2. Experimental Setup

3.2.1. Experimental Scheme

In this paper, all validation experiments are performed on the simulation datasets, and generalization experiments are conducted on the real dataset. Firstly, we compare the effect of denoising complex noise in different method. Then, we construct the 3D particle from the denoised images and compare

Table 1. Denoising results in two simulated datasets: the Mean Square Error (MSE), the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similaity Index Measure (SSIM) of different models. Our method achieves the best results.

Methods	EMD-8511			EMD-24928		
	MSE ↓	PSNR ↑	SSIM ↑	MSE ↓	PSNR ↑	SSIM ↑
BM3D [10]	1.323e-1	4.963	0.425	7.705e-2	7.419	0.635
K-SVD [11]	1.351e-1	4.812	0.271	7.913e-2	7.240	0.404
CWF [2]	1.319e-2	19.071	0.808	2.948e-2	15.354	0.652
DenoiseGAN [21]	3.299e-3	20.693	0.885	4.088e-3	19.405	0.885
RobustGAN [14]	3.239e-3	21.884	0.889	3.484e-3	19.569	0.902
Noise2noise [7]	1.776e-3	24.846	0.913	2.513e-3	22.221	0.909
Ours	**1.449e-3**	**26.775**	**0.932**	**2.141e-3**	**23.356**	**0.923**

them to the groundtruth. Finally, the well-trained model from the denoising model is generalized to the real EMPAIR particle dataset to verify the extensibility of the proposed method.

3.2.2. Evaluation Metrics

As representative metrics in the field of image denoising and restoration, MSE, PSNR and SSIM are used to measure the results of denoising methods. And then we conducted three dimensional reconstruction experiments on the denoising result. The Fourier shell correlation (FSC) [20] measures the normalized crosscorrelation between two 3D maps in Fourier domain at various radial frequency shells. The resolution of the 3D reconstruction results can be defined as the spatial frequency at which the FSC curve is 0.143.

3.3. Experimental results

We make a qualitative and quantitative comparison and analysis of our method with the current mainstream methods, including: (1) BM3D [10], (2) K-SVD [11], (3) Covariance Wiener Filter (CWF) [2], (4) DenoiseGAN [21], (5) Robust-GAN [14], (6) Noise2noise [7].

Table 1 presents a quantitative comparison of the denoised results. Our method outperforms others in terms of all three assessment criteria, demonstrating superior performance. In the first and second rows of Fig. 3, the samples of denoised images by different methods are displayed. We recognize traditional methods such as BM3D, KSVD, and CWF can extract the general outline of the single particle under high noise conditions, especially in CWF results, but they still leave the part of structural noise that is difficult to distinguish. The deep learning methods can perform much better, but they produce ambiguous boundaries. For the sake of fairness here, retraining is done on the datasets whose groundtruth is free of structural noise, which is not operated in their original algorithm.

To verify the generalization and scalability of the proposed method, we conduct comparative experiments with real particle dataset. The specific experimental protocol is to use the optimal network model trained on the simulation dataset to test on the real dataset. As shown in Fig. 3, our method still shows the best denoising effect, while deep learning methods that perform well on simulated data, such as Noise2noise,

2997
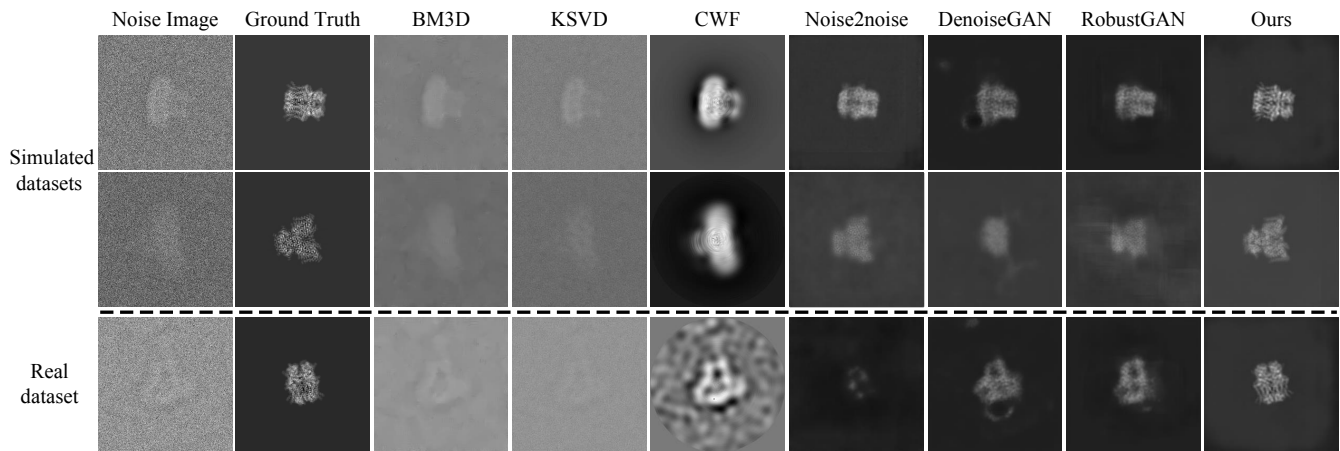
Fig. 3. Comparison of the denoising results from different methods. The first and second rows compare the denoising results of the EMD-8511 and EMD-24928 simulated datasets respectively, and the third row compares the denoising results in the real dataset of the EMPAIR-10081, whose ground truth is the closest projection map. Our method can effectively remove unresolved structural noise on both simulated and real datasets under high noise conditions.
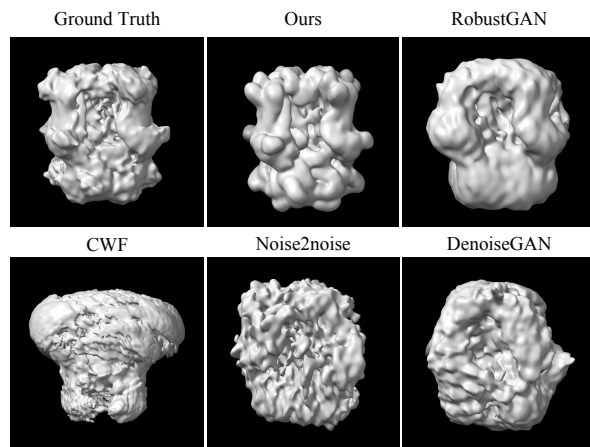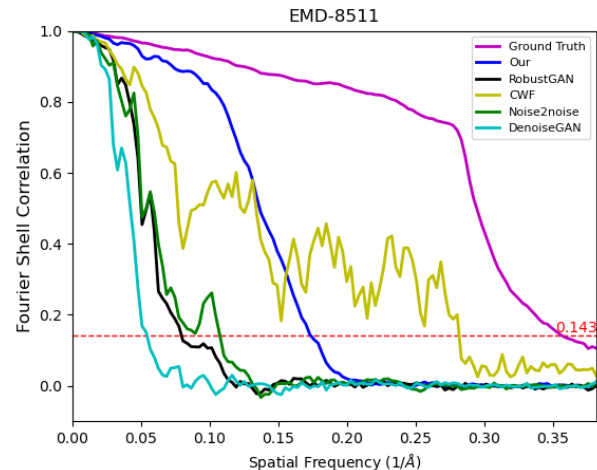

Fig. 4. The 3D reconstruction results of denoised images. Our results are closest to the ground truth.

DenoiseGAN, and RobustGAN, perform very poorly on real data. This also proves the practical value of our method.

In order to further prove the importance of denoising and verify the effect of restoration, we performed 3D reconstruction based on the denoised image using the RELION tool [22]. The Fig. 4 shows the 3D reconstruction results of denoised images based on all algorithms. It can be seen that only our method restores a complete and clear structure (BM3D and K-SVD reconstruction failed).

After reconstruction, we used FSC to determine the resolution and reliability of the reconstructed particle structure. As shown in the Fig. 5 the resolution is determined by the 0.143 cutoff of FSC (Red dashed line). From left to right, the resolution of the particle structure gradually increases. Although our reconstructed result is not as good as the ground truth, it is better than other methods. In addition, a structure with high reliability should decay monotonically toward 0 as

Fig. 5. FSC of the 3D reconstruction results. According to the 0.143 cutoff criterion (Dashed line), as the spatial frequency increases, the resolution of 3D structures gradually increases and the FSC value of highly reliable structures steadily approaches 0. Our results exhibits highly reliability and high-resolution.

the spatial frequency increases (Abscissa). Obviously, the result of CWF is unreliable. Moreover, it can be seen from the Fig. 4 that the reconstructed result still has the structure caused by structural noise.

4. CONCLUSIONS

In this study, we present a framework based on the diffusion model with post-processing to address the task of single-particle restoration in cryo-EM images. Extensive experimental results substantiate the feasibility and effectiveness of our proposed approach. Our method not only achieves the best denoising results on both simulated and real datasets, but also effectively reduces unresolved structural noise.

5. REFERENCES

[1] J. G. Wu, Y. Yan, D. X. Zhang, et al., "Machine learning for structure determination in single-particle cryo-electron microscopy: A systematic review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 2, pp. 452–472, 2021.

[2] T. Bhamre, T. Zhang, and A. Singer, "Denoising and co-variance estimation of single particle cryo-em images," *Journal of structural biology*, vol. 195, no. 1, pp. 72–81, 2016.

[3] N. F. Marshall, T. Y. Lan, T. Bendory, and A. Singer, "Image recovery from rotational and translational invariants," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 5780–5784.

[4] G. Ji, Y. Yang, and H. B. Shen, "Itervm: an iterative model for single-particle cryo-em image clustering based on variational autoencoder and multi-reference alignment," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2018, pp. 999–1002.

[5] T. Bepler, A. Morin, M. Rapp, et al., "Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs," *Nature methods*, vol. 16, no. 11, pp. 1153–1160, 2019.

[6] T. Bepler, K. Kelley, A. J. Noble, and B. Berger, "Topaz-denoise: general deep denoising models for cryoem and cryoet," *Nature communications*, vol. 11, no. 1, pp. 5208, 2020.

[7] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2noise: Learning image restoration without clean data," *arXiv preprint arXiv:1803.04189*, 2018.

[8] Y. Fan and Z. Zhao, "Cryo-electron microscopy image denoising using multi-frequency vector diffusion maps," in *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2021, pp. 3463–3467.

[9] W. T. Baxter, R. A. Grassucci, H. Gao, and J. Frank, "Determination of signal-to-noise ratios and spectral snrs in cryo-em low-dose imaging of molecules," *Journal of structural biology*, vol. 166, no. 2, pp. 126–132, 2009.

[10] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on image processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[11] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on signal processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[12] C. H. Lee and R. MacKinnon, "Structures of the human hcn1 hyperpolarization-activated channel," *Cell*, vol. 168, no. 1, pp. 111–120, 2017.

[13] H. Li, H. Zhang, et al., "Noise-transfer2clean: denoising cryo-em images based on noise modeling and transfer," *Bioinformatics*, vol. 38, no. 7, pp. 2022–2029, 2022.

[14] H. Gu, Y. Xian, I. C. Unarta, and Y. Yao, "Generative adversarial networks for robust cryo-em image denoising," *arXiv preprint arXiv:2008.07307*, 2020.

[15] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[16] H. Guan and D. Si, "Deeptracer-denoising: Deep learning for 3d electron density map denoising," in *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2022, pp. 2080–2087.

[17] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 4, pp. 4713–4726, 2022.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015, pp. 234–241.

[19] N. Bavi, M. D. Clark, G. F. Contreras, R. Shen, B. G. Reddy, W. Milewski, and E. Perozo, "The conformational cycle of prestin underlies outer-hair cell electromotility," *Nature*, vol. 600, no. 7889, pp. 553–558, 2021.

[20] M. Van Heel and M. Schatz, "Fourier shell correlation threshold criteria," *Journal of structural biology*, vol. 151, no. 3, pp. 250–262, 2005.

[21] M. Su, H. Zhang, K. Schawinski, C. Zhang, and M. A. Cianfrocco, "Generative adversarial networks as a tool to recover structural information from cryo-electron microscopy data," *BioRxiv*, p. 256792, 2018.

[22] S. H. Scheres, "Relion: implementation of a bayesian approach to cryo-em structure determination," *Journal of structural biology*, vol. 180, no. 3, pp. 519–530, 2012.