

Robust Denoising of Cryo-EM Images via β -GAN

Hanlin Gu

Hong Kong University of Science and Technology

Ilona Unarta

Hong Kong University of Science and Technology

Xuhui Huang

Hong Kong University of Science and Technology

Yuan Yao (✉ yuany@ust.hk)

Hong Kong University of Science and Technology

Research Article

Keywords: PSNR, GANs, RNAP, MSE, Cryo-EM

Posted Date: January 6th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-139567/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Robust Denoising of Cryo-EM Images via β -GAN

Hanlin Gu¹, Ilona Christy Unarta², Xuhui Huang^{2,3}, Yuan Yao^{1,*}

¹Department of Mathematics, Hong Kong University of Science and Technology

²Department of Chemistry, Center of System Biology and Human Health, State Key Laboratory of Molecular Neuroscience, Hong Kong University of Science and Technology

³Department of Chemistry, State Key Laboratory of Molecular Neuroscience, Hong Kong University of Science and Technology, Kowloon, Hong Kong

*Corresponding: yuany@ust.hk

ABSTRACT

The cryo-electron microscopy (Cryo-EM) becomes popular for macromolecular structure determination. However, the 2D images which Cryo-EM detects are of high noise and often mixed with multiple heterogeneous conformations and contamination, imposing a challenge for denoising. Traditional image denoising methods and simple Denoising Autoencoder can not remove Cryo-EM image noise well when the signal-noise-ratio (SNR) of images is meager and contamination distribution is complex. Thus it is desired to develop new effective denoising techniques to facilitate further research such as 3D reconstruction, 2D conformation classification, and so on. In this paper, we approach the robust denoising problem for Cryo-EM images by introducing a family of Generative Adversarial Networks (GAN), called β -GAN, which is able to achieve robust estimate of certain distributional parameters under Huber contamination model with statistical optimality. To address the challenge of robust denoising where the traditional image generative model might be contaminated by a small portion of unknown outliers, β -GANs are exploited to enhance the robustness of denoising Autoencoder. The method is evaluated by both a simulated dataset on the *Thermus aquaticus* RNA Polymerase (RNAP) and a real dataset on the *Plasmodium falciparum* 80S ribosome dataset (EMPIRE-10028), in terms of Mean Square Error (MSE), Peak Signal to Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM) and 3D Reconstruction as well. The results show that equipped with some designs of β -GANs and the robust ℓ_1 -Autoencoder, one can stabilize the training of GANs and achieve the state-of-the-art performance of robust denoising with low SNR data and against possible information contamination. Our proposed methodology thus provides an effective tool for robust denoising of Cryo-EM 2D images, which is helpful for 3D structure reconstruction.

Introduction

The cryo-electron microscopy (Cryo-EM) has become one of the most popular techniques to resolve the atomic structure. In the past, Cryo-EM was limited to large complexes or low-resolution models. Recently the development of new detector hardware has dramatically improved the resolution in Cryo-EM¹, which makes Cryo-EM widely used in a variety of research fields. Different from X-ray crystallography, Cryo-EM has the advantage of preventing the recrystallization of inherent water and re-contamination. Also, Cryo-EM is superior to Nuclear Magnetic Resonance spectroscopy (NMR) in solving macromolecules in the native state. In addition, both X-ray crystallography and NMR require large amounts of relatively pure samples, whereas Cryo-EM requires much fewer samples². For this celebrated development of Cryo-EM for the high-resolution structure determination of biomolecules in solution, the Nobel Prize in Chemistry in 2017 was awarded to three pioneers in this field³.

However, it is a computational challenge in processing raw Cryo-EM images, due to heterogeneity in molecular conformations and high noise. Macromolecules in natural conditions are usually heterogeneous, i.e., multiple metastable structures may coexist in the experimental samples^{4,5}. Such conformational heterogeneity adds extra difficulty to the structural reconstruction as we need to assign each 2D image to not only the correct projection angle but also its corresponding conformation. This imposes a computational challenge that one needs to denoise the Cryo-EM images without losing the key features of their corresponding conformations. Moreover, in the process of generating Cryo-EM images, one needs to provide a view using the electron microscope for samples that are in frozen condition. Thus there are two types of noise: one is from ice, and the other is from the electron microscope. Both of them are significant in contributing high noise in Cryo-EM images and leave a difficulty to the detection of particle structures (Figure

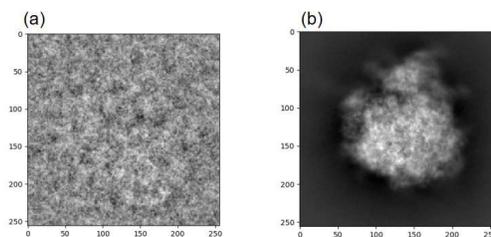


Figure 1. (a) a noisy Cryo-EM image (b) a reference image

Figure 1. (a) a noisy Cryo-EM image (b) a reference image

1 shows a typical noisy Cryo-EM image with its reference image which is totally non-identifiable to human eyes). In extreme cases, some experimental images even do not contain any particles, rendering it difficult for particle picking either manually or automatically⁶. How to achieve robust denoising against such kind of contamination thus becomes a critical problem. So it is a great challenge to develop robust denoising methods for Cryo-EM images to reconstruct heterogeneous biomolecular structures.

There are a plethora of denoising methods developed in applied mathematics and machine learning that could be applied to Cryo-EM image denoising. Most of them in Cryo-EM are based on unsupervised learning, which don't need any reference image data to learn. Wang, J. et al.⁷ proposed a filtering method based on non-local means, which makes use of the rotational symmetry of some biological molecules. Also, Wei, D. et al.⁸ designed the adaptive non-local filter, which takes advantage of a wide range of pixels to estimate the denoised pixel values. Besides, Xian, Y. et al.⁹ compared transform domain filtering method: BM3D¹⁰ and dictionary learning method: KSVD¹¹ in denoising problem in Cryo-EM. However, all of these do not work well in low Signal-Noise-Ratio (SNR) situations like Cryo-EM. In addition, Covariance Wiener Filtering (CWF)¹² is proposed for image denoising. It demonstrates that CWF needs large sample size of data in order to estimate the covariance matrix correctly, although it has an attractive denoising effect.

Deep learning technique has entered the field of image denoising with its rapid progress in image classification. One of the most popular methods is denoising Autoencoder (DA) motivated by Vincent, P. et al.¹³ that needs reference data to learn a compressive representation (encoding) for a set of data. The extension of DA in Xie, J. et al.¹⁴ exploits sparsity regularization in addition to the reconstruction loss in order to avoid overfitting. Other developments such as Zhang, K. et al.¹⁵ takes advantage of the residual network architecture and Agostinell, F. et al.¹⁶ combine several sparse denoising Autoencoder directed to be robust to different noise. However, deep Autoencoders have not been applied to Cryo-EM denoising up to our knowledge.

On the other hand, Generative Adversarial Networks (GAN) recently gains its popularity in machine learning field and provides a promising new approach for Cryo-EM image denoising. The modern version of GAN proposed in¹⁷ is mainly composed of two parts: generator (G : generate the new samples) and discriminator (D : determine whether the samples are real or generated (fake)). In pursuit of a minimax zero-sum game, it aims to minimize the Jensen-Shannon (JS) divergence between distributions of generated samples and true samples, hence called JS-GAN. Various GANs are then studied, and in particular, Arjovsky et al.¹⁸ proposed WGAN, which uses Wasserstein distance in replacement of the JS divergence. Gulrajani, I. et al.¹⁹ further improved WGAN with the addition of the gradient penalty that makes the training more stable. For image denoising, Yang, Q. et al.²⁰ applied GAN to medical image denoising in the low noise situation. In particular, Su, M. et al.²¹ applied JS-GAN to Cryo-EM image denoising in a homogeneous molecular setting. Recently, Gao, C. et al.^{22,23} showed that a general family of GANs (β -GANs, including JS-GAN and TV-GAN, etc.) enjoys robust reconstruction when the data set contain outliers under Huber contamination models. Therefore, such GANs equip us with a natural technique for robust denoising for Cryo-EM images, which becomes the central topic in this paper.

In this paper, we investigate a robust denoising of Cryo-EM images based on joint training of β -GANs and denoising Autoencoders. In particular, our main contributions are as follows.

- In order to better describe the complex generative process in Cryo-EM images, we enhance the traditional image generative model with Huber contamination model, where a small portion of samples allows for unknown contamination. To recover the clean image in this new model, we introduce a family of β -GAN, which is able to achieve the robustness of denoising against partial agnostic contamination of samples (e.g., (.5, .5)-GAN or (1, 1)-GAN in this family work best in this paper, where β -GAN has two parameters: α and β , often written as (α, β) -GAN).
- We exploit a joint training of GANs and denoising Autoencoders toward robust denoising. Both Autoencoder and GANs help each other for Cryo-EM denoising in low signal-noise-ratio scenarios. On the one hand, Autoencoder helps stabilize GANs during training, without which the training processes of GANs are often collapsed due to high noise; on the other hand, GANs help Autoencoder in denoising by sharing information in similar samples via distribution learning and enhancing the robustness against contaminations.
- Numerical experiments and reconstruction are conducted with both simulated dataset on the *Thermus aquaticus* RNA Polymerase (RNAP) and a real dataset on the *Plasmodium falciparum* 80S ribosome dataset (EMPIRE-10028). The experiments on those datasets show the validity of the proposed methodology and suggest that: some designs of β -GANs, such as (.5, .5)-GAN and (1, 1)-GAN, joint with robust ℓ_1 -Autoencoder are among the best choices in robust denoising against unknown contamination; on the other hand, despite that WGANs achieve superb performance in contamination-free scenarios, they deteriorate significantly under contaminated samples.

Results

Network Architecture and Hyperparameters

In this paper, we exploit a family of (α, β) -GANs jointly trained with ℓ_p -Autoencoder, shown in Algorithm 1. For (α, β) -GAN, we reports two types of choices: (1) $\alpha = 1, \beta = 1$; (2) $\alpha = 0.5, \beta = 0.5$ since they show the best results in our experiments, while the others are collected in the supplementary. For WGAN, the gradient penalty with parameter $\mu = 10$ is used to accelerate the speed of convergence and hence the algorithm is denoted as WGAN_{gp} below. The trade-off (regularization) parameter of ℓ_1 or ℓ_2 reconstruction loss is set to be $\lambda = 10$ through out this section, while an ablation study on varying λ is also discussed in the supplementary.

For the optimization method, we chose Adam²⁴. The learning rate of the discriminator is $\eta_d = 0.001$, and the learning rate of the generator is $\eta_g = 0.01$. We choose batch size $m = 20$, $k_d = 1$, and $k_g = 2$ in Algorithm 1.

For the choice of network architecture, the best results in the experiments of this paper come from the ResNet architecture²¹ shown in Figure 2, which has been successfully applied to study biological problems such as predicting protein-RNA binding. The generator in such GANs exploits the Autoencoder network architecture, while the discriminator is a binary classification ResNet. In the supplementary, we also discuss a Convolutional Network without residual blocks and the PGGAN²⁵ architecture with their experimental results, respectively. Reproducible codes can be downloaded at: <https://github.com/gh11995/denoise-gan-in-cryo-EM>.

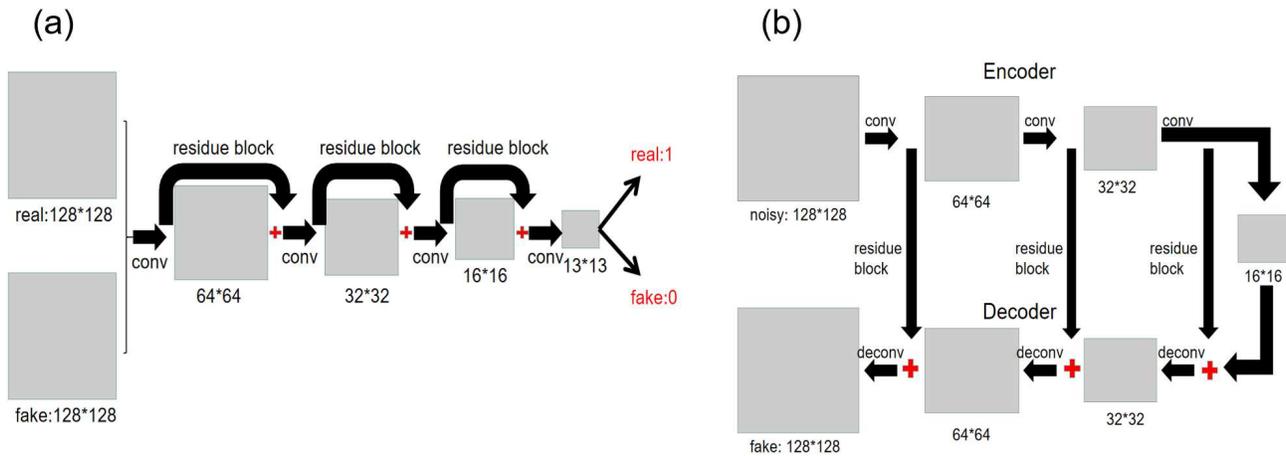


Figure 2. The architectures of (a) discriminator (D) and (b) generator (G) which borrow the residue structure. The input image size (128×128) here is adapted to RNAP dataset, while in EMPIRE-10028 dataset it is 256×256 with a similar architecture.

Results for RNAP

Denoising without contamination

In this part, we attempt to denoise the noisy image without the contamination (i.e., $\varepsilon = 0$ in model (2)). In order to present the advantage of GAN, we compare the denoising result in different methods. Table 1 shows the MSE and PSNR of different methods in SNR 0.05 and 0.1. We recognize the traditional methods such as KSVD, BM3D, Non-local mean, and CWF can remove the noise partially and extract the general outline, but they still leave the unclear piece. However, deep learning methods can perform much better. Specifically, we observe that GAN-based methods, especially WGAN_{gp} + ℓ_1 loss and $(.5, .5)$ -GAN + ℓ_1 loss, perform better than denoising Autoencoder methods, which only optimizes ℓ_1 or ℓ_2 loss (ℓ_1 -Autoencoder represents ℓ_1 loss, ℓ_2 -Autoencoder represents ℓ_2 loss, GAN + ℓ_1 represents adding ℓ_1 regularization in GAN generator loss). The adversarial process inspires the generation process, and the additional ℓ_1 loss optimization speeds up the process of generation towards reference images. Notably, WGAN_{gp} and $(5, .5)$ - or $(1, 1)$ -GANs are among the best methods, where the best mean performance up to one standard deviation are all marked in bold font. Specifically, compared with $(.5, .5)$ -GAN, the WGAN_{gp} get better PSNR and SSIM in SNR 0.1; the $(.5, .5)$ -GAN shows the advantage in PSNR and SSIM in SNR 0.05 while $(1, 1)$ -GAN is competitive within one standard deviation. Also, Figure 3(a) presents the denoised images of denoising methods in SNR 0.05. For the convenience of comparison, we choose a clear open-conformation to present, and the performances show that WGAN_{gp} and (α, β) -GAN can grasp the “open” shape completely and derive the more explicit pictures than other methods.

Table 1. Denoising result without contamination in simulated RNAP dataset: MSE, PSNR and SSIM of different models, such as BM3D¹⁰, KSVD¹¹, Non-local means⁸, CWF¹², DA and GAN-based methods.

Method/SNR	MSE		PSNR		SSIM	
	0.1	0.05	0.1	0.05	0.1	0.05
BM3D	3.52e-2(7.81e-3)	5.87e-2(9.91e-3)	14.54(0.15)	12.13(0.14)	0.20(0.01)	0.08(0.01)
KSVD	1.84e-2(6.58e-3)	3.49e-2(7.62e-3)	17.57(0.16)	14.61(0.14)	0.33(0.01)	0.19(0.01)
Non-local means	5.02e-2(5.51e-3)	5.81e-2(8.94e-3)	13.04(0.50)	12.40(0.65)	0.18(0.01)	0.09(0.01)
CWF	2.53e-2(2.03e-3)	9.28e-3(8.81e-4)	16.06(0.33)	20.31(0.41)	0.25(0.01)	0.08(0.01)
ℓ_2 -Autoencoder	3.13e-3(7.97e-5)	4.02e-3(1.48e-4)	25.10(0.11)	23.67(0.77)	0.79(0.02)	0.79(0.01)
ℓ_1 -Autoencoder	3.16e-3(7.05e-5)	4.23e-3(1.32e-4)	25.05(0.09)	23.80(0.13)	0.77(0.02)	0.76(0.01)
(0,0)-GAN + ℓ_1	3.06e-3(5.76e-5)	4.02e-3(5.67e-4)	25.25(0.04)	24.00(0.06)	0.78(0.03)	0.78(0.03)
WGANgp + ℓ_1	2.95e-3(1.41e-5)	4.00e-3(8.12e-5)	25.42(0.04)	24.06(0.05)	0.83(0.02)	0.80(0.03)
(1,1)-GAN + ℓ_1	2.99e-3(3.51e-5)	4.01e-3(1.54e-4)	25.30(0.05)	24.07(0.16)	0.82(0.03)	0.79(0.03)
(.5,.5)-GAN+ ℓ_1	3.01e-3(2.81e-5)	3.98e-3(4.60e-5)	25.27(0.04)	24.07(0.05)	0.79(0.04)	0.80(0.03)

What’s more, in order to test the denoised results of β -GAN, we reconstruct the 3D volume by RELION in 200000 images of SNR 0.1, which are denoised by (.5, .5)-GAN + ℓ_1 . The value of pixel size, amplitude contrast, spherical aberration and voltage are 1.6, 2.26, 0.1 and 300. For the other terms, retaining the default settings in RELION software. Figure 3(b) and (c) separately show the 3D volume recovered by clean images and denoised images. Also, the related FSC curves are shown in Figure 3(d). Specifically, the blue curve, which represents (.5, .5)-GAN + ℓ_1 denoised images is closed to red curves representing the clean images. We use the 0.143 cutoff criterion in literature (the resolution as Fourier shell correlation reaches 0.143, shown by dash lines in Figure 3(d)) to choose the final resolution: 3.39Å. The structure recovered by our method and FSC curve are as good as the original structure, which illustrates that the denoised result of β -GAN can identify the details of image and be helpful in 3D reconstruction.

In addition, in supplementary we show an example that GAN with ℓ_1 -Autoencoder helps heterogeneous conformation clustering.

Robustness under contamination

In this part, we consider the contamination model $\varepsilon \neq 0$ and Q from purely noisy images. We randomly replace partial samples of our training dataset of RNAP by noise to test the robustness of denoising methods under contamination. There are three ways to test: (A) Only replacing the clean reference images. It implies the reference images are wrong or missing, such that we do not have the reference images to compare. This is the worst contamination case. (B) Only replacing the noisy images. It means the Cryo-EM images the machine produces are broken. (C) Replacing both, which indicates both A and B happen. The latter two are mild contamination cases, especially C that replaces both reference and noisy images by Gaussian noise whose ℓ_1 or ℓ_2 loss is thus well-controlled.

Here we test our robustness of various deep learning based methods using the data of SNR 0.1, and the former three contaminations are applied to randomly replace the samples in the proportion of $\varepsilon \in \{0.1, 0.2, 0.3\}$ of the whole dataset.

Figure 3(e), (f) and (g) compare the robustness of different methods. In all the cases, some β -GANs ((.5, .5)- and (1, 1)-) with ℓ_1 -Autoencoder exhibit relatively universal robustness. Particularly, (1) The MSE with ℓ_1 loss is less than the MSE with ℓ_2 loss, which represents the ℓ_1 loss is more robust than ℓ_2 as desired. (2) The Autoencoder method in ℓ_2 loss and WGANgp show certain robustness in case B and C but are largely influenced by contamination in case A (shown in Figure 3 (e)), indicating the most serious damage arising from type A, merely replacing only the reference image by Gaussian noise. The reason is that the ℓ_2 Autoencoder and WGANgp method are confused by the wrong reference images so that they can not learn the mapping from data distribution to reference distribution accurately. (3) In the type C, the standard deviations of the five best models are larger compared the other two types. The contamination of both noisy y and clean x images influence the the stability of model more than the other two types.

Furthermore, we take an example in type A contamination with $\varepsilon = 0.1$ for 3D reconstruction. The 3D reconstruction in denoised images with (.5, .5)-GAN + ℓ_1 and ℓ_2 -Autoencoder are shown in Figure 3(h) and (j), and related FSC curve is Figure 3(i). Specifically, on the one hand, the blue FSC curve of ℓ_2 -Autoencoder doesn’t drop, which leads to the worse reconstruction; on the other hand, the red FSC curve of (.5, .5)-GAN + ℓ_1 drops quickly but begins to rise again, whose reason is that some unclear detail of structure mixed angular information in reconstruction. When applying 0.143 cutoff criterion (dashed line in FSC curve), the resolution of (.5, .5)-GAN + ℓ_1 is about 4Å. Although reconstruction of images and final resolution is not better than the clean images, it is much clearer than ℓ_2 -Autoencoder which totally fails in the contamination case. The outcome of the reconstruction demonstrates that (.5, .5)-GAN + ℓ_1 is relatively robust, whose 3D result is consistent

with the clean image reconstruction.

In a summary, some (α, β) -GANs $(.5, .5)$ and $(1, 1)$ here) with ℓ_1 Autoencoder are more resistant to sample contamination, which are better to be applied into the Cryo-EM experimental data.

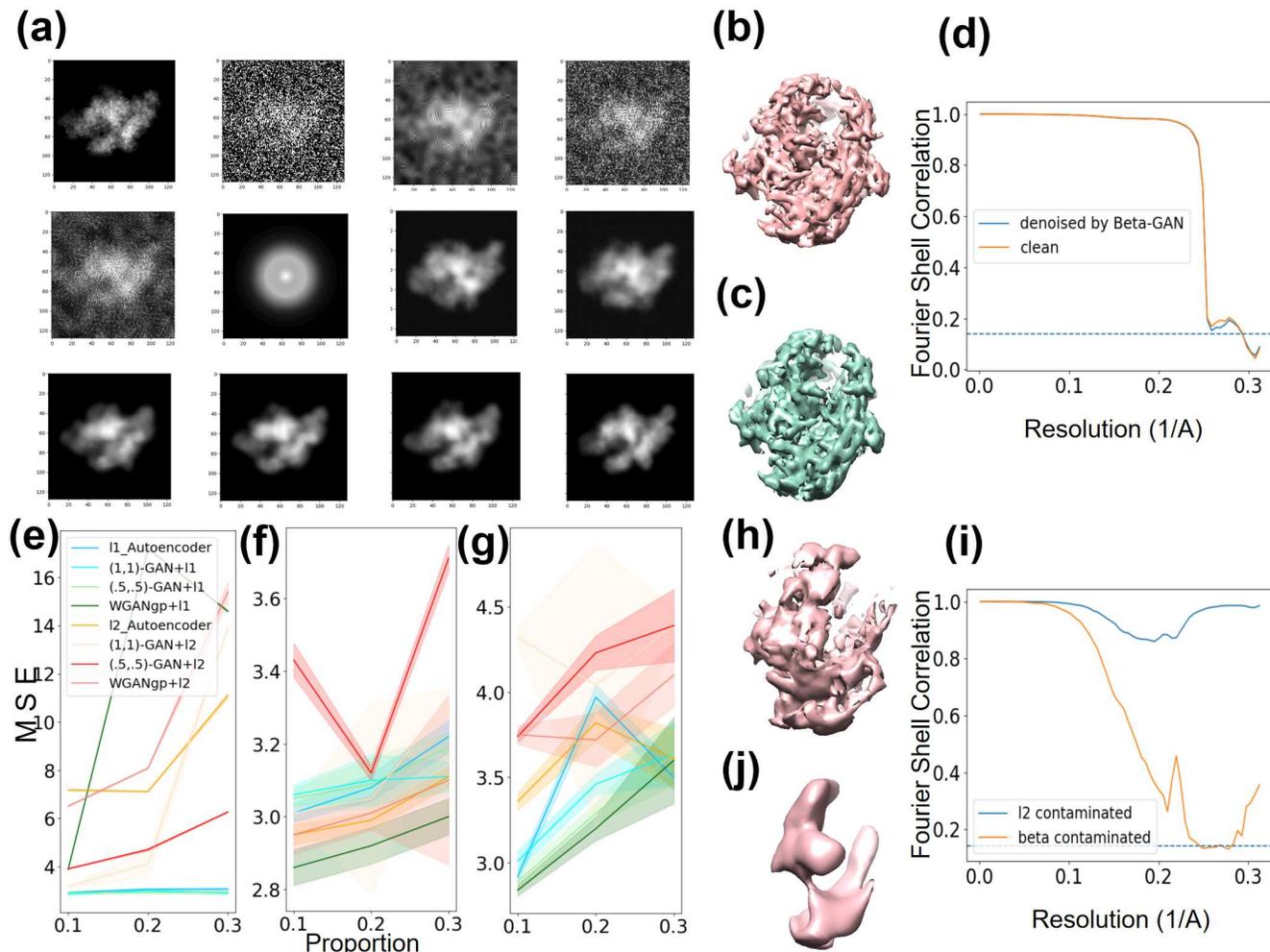


Figure 3. Results for RNAP dataset. (a) is denoised images in different denoised methods (from left to right, top to bottom): Clean, Noisy, BM3D, KSVD, Non-local means, CWF, ℓ_1 -Autoencoder, ℓ_2 -autoencoder, $(1,1)$ -GAN + ℓ_1 , $(0,0)$ -GAN + ℓ_1 , $(.5, .5)$ -GAN + ℓ_1 and WGANgp + ℓ_1 . (b) and (c) are reconstruction of clean images and $(.5, .5)$ -GAN + ℓ_1 denoised images. (d) is FSC curve of (b) and (c). (e), (f) and (g) are robustness tests of various methods under $\epsilon \in \{0.1, 0.2, 0.3\}$ -proportion contamination in three types of contamination: (e) Type A: replacing the reference images with random noise; (f) Type B: replacing the noisy images with random noise; (g) Type C: replacing both with random noise. (h) and (j) are reconstructions of images with $(.5, .5)$ -GAN + ℓ_1 and ℓ_2 -autoencoder under type A contamination, respectively, where ℓ_2 -autoencoder totally fails but $(.5, .5)$ -GAN + ℓ_1 is robust. (i) shows FSC curves of (h) and (j).

Denoising and Reconstruction Results for EMPIRE-10028

The following Figure 4(a) and (b) show the denoising results by different deep learning methods in experimental data: ℓ_1 or ℓ_2 Autoencoders, JS-GAN $((0,0)$ -GAN), WGANgp, and (α, β) -GAN, where we add ℓ_1 loss in all of the GAN-based structures. Although the Autoencoder can grasp the shape of macromolecules, it is a little blur in some parts. What is more, WGANgp and (α, β) -GAN can perform well. According to MSE and PSNR, WGANgp and $(.5, .5)$ -GAN perform better than other deep learning methods, largely consistent with the result of the RNAP dataset. The improvements of such GANs over pure Autoencoders lie in their ability of utilizing structural information among similar images to learn the data distribution better.

Finally, we take reconstruction via RELION of 100000 images, which denoised by $(.5, .5)$ -GAN + ℓ_1 . The parameters are the same as the paper²⁶. The reconstruction results is shown in Figure 4(c). It is demonstrated that final resolution is 3.20\AA ,

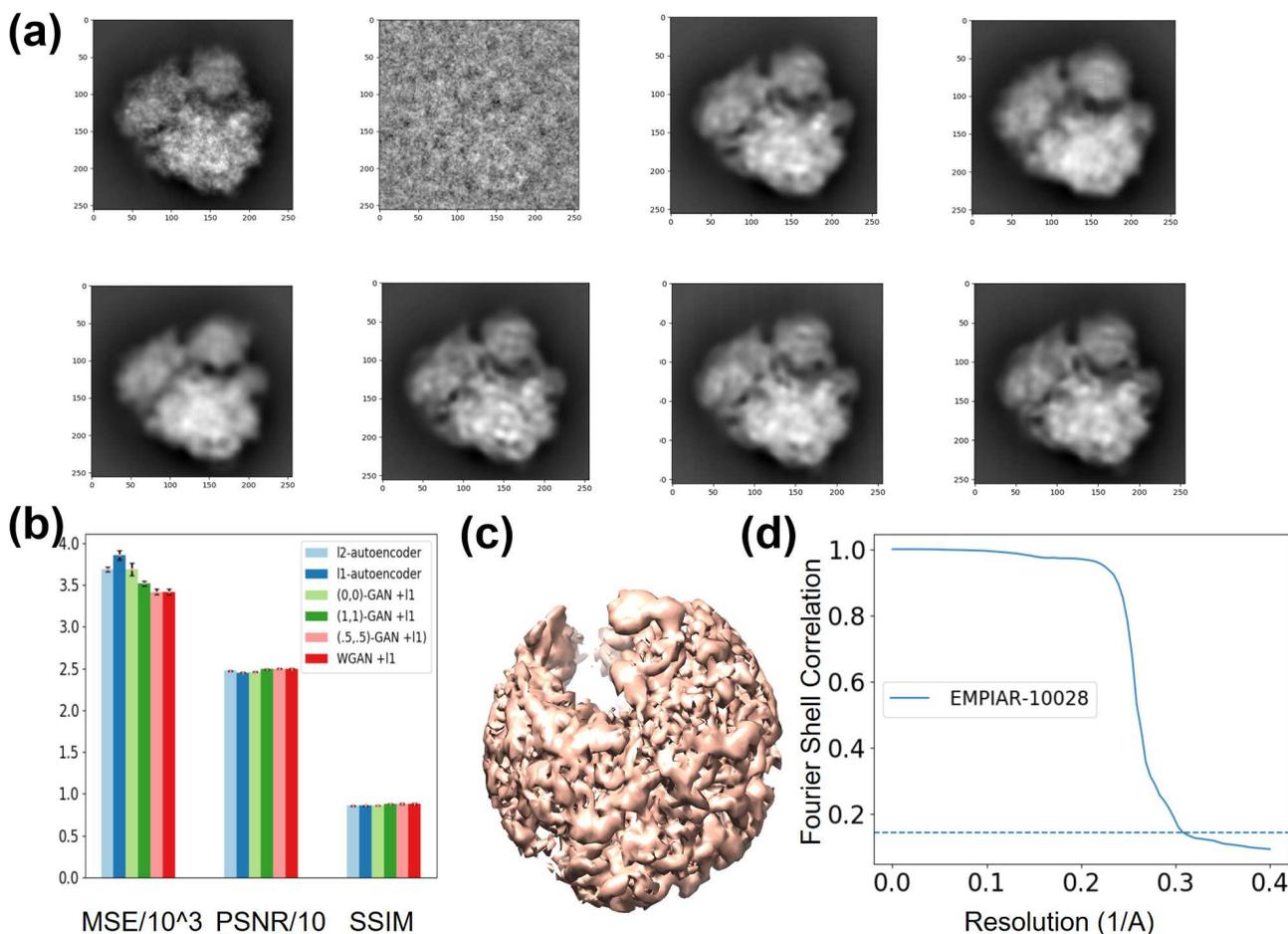


Figure 4. Results for EMPIRE-10028. (a) Comparison in EMPIRE-10028 dataset in different deep learning methods (from left to right, top to bottom): clean image, noisy image, ℓ_1 -Autoencoder, ℓ_2 -Autoencoder, (0,0)-GAN + ℓ_1 , (1,1)-GAN + ℓ_1 , (.5,.5)-GAN + ℓ_1 , WGANgp + ℓ_1 . (b) is the MSE, PSNR and SSIM in different denoised methods. (c) and (d) is the 3D-reconstruction of denoised images by (.5,.5)-GAN + ℓ_1 and the FSC curve, respectively. The resolution of reconstruction from (.5,.5)-GAN + ℓ_1 denoised images is 3.20Å, which is as good as the original resolution.

which is derived by FSC curve in figure 4(d) using the same 0.143 cutoff (dashed line) to choose the final resolution. We note that the final resolution by RELION after denoising is as good as the original resolution 3.20Å reported in²⁶.

Discussion

In this paper, we extend the traditional generative model for Cryo-EM images to a Huber contamination model that includes unknown distributions of contamination. To achieve robust denoising for Cryo-EM images, we propose to exploit β -GANs, a family of Generative Adversarial Networks which is able to achieve robust estimate of distributional parameters with statistical optimality, to enhance the robustness of Denoising Autoencoder and have seen that such a joint training scheme can remarkably improve the performance in Cryo-EM image robust denoising. In this joint training scheme, on the one hand, the reconstruction loss of Autoencoder helps GAN to avoid mode collapse and stabilize training; on the other hand, GAN helps Autoencoder in improving robustness of denoising and utilizing the highly correlated Cryo-EM images since they are 2D projections of one or a few 3D molecular conformations. In experiments of both simulated RNAP data and real EMPIRE-10028 data, joint training of ℓ_1 -Autoencoder combined with (.5,.5)-GAN, (1,1)-GAN, and WGAN with gradient penalty is often among the best performance in terms of MSE, PSNR, and SSIM, when the data is contamination-free. However, when an unknown portion of data is contaminated, especially when the reference data is contaminated, WGAN with ℓ_1 Autoencoder may suffer from the significant deterioration of reconstruction accuracy. Therefore, some β -GANs (e.g. (.5,.5)-GAN and (1,1)-GAN) joint with robust ℓ_1 -Autoencoder are the overall best choices for robust denoising with contaminated and high

noise datasets.

There are also some open problems to pursue in future directions. Most of the deep learning-based techniques in image denoising need reference data, limiting themselves in the application of Cryo-EM denoising. For example, in our experimental dataset EMPIRE-10028, the reference data is generated by the cryoSPARC, which itself becomes problematic in highly heterogeneous conformations. Therefore the reference image we learn may follow a fake distribution. How to denoise without the reference image thus becomes a significant problem. It is still open how to adapt to different experiments and those without reference images. One idea possibly overcoming this hurdle is called "image-blind denoising" proposed by^{27,28}, in which they viewed the noisy image or void image as the reference image to denoise. Moreover, Chen, J. et al.²⁹ tried to extract the noise distribution from the noisy image and gain denoised images through removing the noise for noisy data; Quan, Y. et al.³⁰ augmented the data by Bernoulli sampling and denoise image with dropout. Besides, Bepler, T. et al.^{31,32} applied noise2noise into Cryo-EM image denoising. Nevertheless, all of the methods need noise is independent of the elements themselves. Thus it is hard to remove noise in Cryo-EM because the noise from ice and machine might be correlated to the particles. At last but not the least, for reconstruction problems in Cryo-EM, Zhong, E. D. et al.³³ attempted an end-to-end 3D reconstruction approach based on the network from Cryo-EM images, where they exploited the Variational Autoencoder (VAE) to approximate the forward generative model and recover the 3D structure directly by combining the angle information and image information learned from data. This is one future direction to pursue.

Methods

A Generative Model with Huber Contamination

Let $x \in \mathbb{R}^{d_1 \times d_2}$ be a clean image, often called reference image in the sequel. The generative model of noisy image $y \in \mathbb{R}^{d_1 \times d_2}$ in cryo-EM under the linear, weak phase approximation^{12,34} can be described by

$$y = a * x + \zeta \quad (1)$$

where $*$ denotes the convolution operation, a is the point spread function of the microscope convolving with the clean image and ζ is an additive noise, usually assumed to be Gaussian noise that corrupts the image. In order to remove the noise the microscope brings, traditional Denoising Autoencoder such as³⁵ could be exploited to learn from examples $(y_i, x_i)_{i=1, \dots, n}$ the inverse mapping a^{-1} from the noisy image y to the clean image x .

However, this model is not sufficient in the real case in Cryo-EM. In the experimental data, Cryo-EM images are possibly contaminated by the ice or others, which do not contain any interesting particle information. For example, particles don't exist in all Cryo-EM images, such that even the experimentalists do the manual or automatic particle picking⁶. Such contaminations will significantly affect our denoising efficiency if the denoising methods continuously depend on the sample outliers. Therefore we introduce the following Huber contamination model to extend the image formation model of Equation (1).

Consider that the pair of reference image and experimental image (x, y) is subject to the following mixture distribution P_ε :

$$P_\varepsilon = (1 - \varepsilon)P_0 + \varepsilon Q, \quad \varepsilon \in [0, 1], \quad (2)$$

a mixture of true distribution P_0 of probability $(1 - \varepsilon)$ and arbitrary contamination distribution Q of probability ε . P_0 is characterized by model (1) and Q accounts for the unknown contamination distribution possibly due to ice, broken of data, and so on such that the image sample does not contain any particle information. This is called the Huber contamination model in statistics³⁶. Our purpose is that given n samples $(x_i, y_i) \sim P_\varepsilon$ ($i = 1, \dots, n$), possibly contaminated with unknown Q , learn a robust inverse map $a^{-1}(y)$.

Robust Denoising Method

In this report, we exploit a neural network to approximate the robust inverse mapping $G_\theta : \mathbb{R}^{d_1 \times d_2} \rightarrow \mathbb{R}^{d_1 \times d_2}$, here a neural network parameterized by $\theta \in \Theta$ whose structure will be discussed in the former. Our goal is to ensure that discrepancy between reference image x and reconstructed image $\hat{x} = G_\theta(y)$ is small. Such a discrepancy is usually measured by some non-negative loss function: $\ell(x, \hat{x})$. Therefore, the denoising problem minimizes the following expected loss,

$$\arg \min_{\theta \in \Theta} \mathcal{L}(\theta) := \mathbb{E}_{x, y}[\ell(x, G_\theta(y))] \quad (3)$$

In practice, given a set of training samples $S = \{(x_i, y_i) : i = 1, \dots, n\}$, we aim to solve the following empirical loss minimization problem,

$$\arg \min_{\theta \in \Theta} \widehat{\mathcal{L}}_S(\theta) := \frac{1}{n} \sum_{i=1}^n \ell(x_i, G_\theta(y_i)) \quad (4)$$

For example, the following choices of loss functions will be considered in this paper:

- (**ℓ_2 -Autoencoder**) $\ell(x, \hat{x}) = \frac{1}{2} \|x - \hat{x}\|_2^2 := \frac{1}{2} \sum_{i,j} (x_{ij} - \hat{x}_{ij})^2$, or equivalently $\mathbb{E}\ell(x, \hat{x}) = D_{KL}(p(x) \| q(\hat{x}_\theta))$ where $\hat{x}_\theta \sim \mathcal{N}(x, \sigma^2 I_D)$;
- (**ℓ_1 -Autoencoder**) $\ell(x, \hat{x}) = \|x - \hat{x}\|_1 := \sum_{i,j} |x_{ij} - \hat{x}_{ij}|$, or equivalently $\mathbb{E}\ell(x, \hat{x}) = D_{KL}(p(x) \| q(\hat{x}_\theta))$ where $\hat{x}_\theta \sim \text{Laplace}(x, b)$;
- (**Wasserstein-GAN**) $\ell(x, \hat{x}) = W_1(p(x), q_\theta(\hat{x}))$ where W_1 is the 1-Wasserstein distance between distributions of x and \hat{x} ;
- (**β -GAN**) $\ell(x, \hat{x}) = D(p(x) \| q_\theta(\hat{x}))$ where D is some divergence function to be discussed below between distributions of x and \hat{x} .

Both the ℓ_2 and ℓ_1 losses consider the reconstruction error of G_θ . The ℓ_2 -loss above is equivalent to assume that $G_\theta(y|x)$ follows a Gaussian distribution $\mathcal{N}(x, \sigma^2 I_D)$, and the ℓ_1 -loss instead assumes a Laplacian distribution centered at x . As a result, the ℓ_2 -loss pushes the reconstructed image \hat{x} toward mean by averaging out the details and thus blurs the image. On the other hand, the ℓ_1 -loss pushes \hat{x} toward the coordinate-wise median, keeping the majority of details while ignoring some large deviations, thus increases the contrast of the reconstructed image and becomes more robust than the ℓ_2 loss against large outliers. Although ℓ_1 -Autoencoder has a more robust loss than ℓ_2 , both of them are not sufficient to handle the contamination in (2). To deal with the Huber contamination model (2), β -GAN will be introduced below.

β -GAN

Recently Gao et al.^{22,23} showed that some types of GANs might achieve robustness for Huber contamination models, playing a similar role as Tukey's median³⁷ in terms of statistical optimality. Therefore it is natural to bring such robust GANs into our considerations. In particular, the following so called β -GAN is shown in²³ to achieve statistically optimal robust estimates.

Adapted to the setting in this paper, β -GAN aims to solve the following minimax optimization problem to find the G_θ ,

$$\min_{G_\theta} \max_D \mathbb{E}[S(D(x), 1) + S(D(G_\theta(y)), 0)] \quad (5)$$

where $S(t, 1) = -\int_t^1 c^{\alpha-1} (1-c)^\beta dc$, $S(t, 0) = -\int_0^t c^\alpha (1-c)^{\beta-1} dc$, $\alpha, \beta \in [-1, 1]$, and D is another neural network called discriminator whose architecture will be discussed on Results. For simplicity, we denote this family with parameters α, β by (α, β) -GAN in this paper.

The family of (α, β) -GAN includes many popular members. For example, when $\alpha = 0, \beta = 0$, it becomes the JS-GAN¹⁷ which aims to solve the following minimax problem whose loss is the Jensen-Shannon divergence,

$$\min_{G_\theta} \max_D \mathbb{E}_{(x,y) \sim P(X,Y)} \{\log(D(x)) + \log(1 - D(G_\theta(y)))\}. \quad (6)$$

When $\alpha = 1, \beta = 1$ the loss is a simple mean square loss; when $\alpha = -0.5, \beta = -0.5$, the loss is boost score. In particular, it is shown in²³ that for all $|\alpha - \beta| < 1$, (α, β) -GAN family is robust in the sense that one can learn an elliptical distribution P_0 from contaminated distributions P_ϵ under the strong contamination model:

$$\{P_\epsilon \in \mathcal{P}(X, Y) : TV(P_\epsilon, P_0) \leq \epsilon\}. \quad (7)$$

More details can be seen in the supplementary. In this report, we are going to see such β -GANs can also help enhance the robustness of Cryo-EM image denoising against contamination.

Yet, we note that Wasserstein GAN (WGAN) is not a member of this family. By formally taking $S(t, 1) = t$ and $S(t, 0) = -t$, we have the following WGAN where an additional gradient penalty is added here (WGANgp)^{18,19}.

$$\min_G \max_D \mathbb{E}_{(x,y) \sim P(X,Y)} \{D(x) - D(G(y)) + \mu \mathbb{E}_{\tilde{x}} (\|\nabla_{\tilde{x}} D(\tilde{x})\|_2 - 1)^2\} \quad (8)$$

where \tilde{x} is uniformly sampled along straight lines connecting pairs of generated and real samples; and μ is a weighting parameter. In WGANgp, the last layer of the sigmoid function in the discriminator network is removed. Thus D 's output range is the whole real \mathbb{R} , but its gradient is close to 1 to achieve Lipschitz-1 functions. Gradient penalty may help stabilize the training of WGAN. Compared to JS-GAN, WGAN aims to minimize the Wasserstein distance between the sample distribution and the generator distribution. Therefore, WGAN is not robust in the sense of contamination models above as arbitrary ϵ portion of outliers can be far away from the main distribution P_0 such that the Wasserstein distance is arbitrarily large.

Joint Autoencoder-GAN and Main Algorithm

Cryo-EM images consist of 2D-projections of the same molecular conformation in different viewing angles, and the reconstruction losses (ℓ_1 or ℓ_2) do not explicitly take into account similar images of similar conformational projections. In addition to possible robustness, GANs can further help denoising by exploiting common information in similar samples during distribution learning; for example, they minimize some divergence or Wasserstein distance between reference image set and denoised image set where similar images can help boost signals for each other.

On the other hand, Autoencoder can help stabilize GANs during training, without which the training processes of GANs are often oscillating and sometimes collapsed due to the presence of high noise (see the supplementary).

For these considerations, in this paper, we propose a combined loss with both GAN and Autoencoder reconstruction loss,

$$\widehat{\mathcal{L}}_{GAN}(x, \widehat{x}) + \lambda \|x - \widehat{x}\|_p^p$$

where $p \in \{1, 2\}$ and $\lambda \geq 0$ is a trade-off parameter for ℓ_p reconstruction loss. Algorithm 1 summarizes the procedure of joint training of Autoencoder and GAN, which will be denoted as ‘‘GAN+ ℓ_p ’’ in the experimental section depending on the proper choice of GAN and p .

Algorithm 1 Joint training of (α, β) -GAN and ℓ_p -Autoencoder.

Input:

1. (α, β) for $S(t, 1) = -\int_t^1 c^{\alpha-1}(1-c)^\beta dc$, $S(t, 0) = -\int_0^t c^\alpha(1-c)^{\beta-1} dc$
or $S(t, 1) = t$, $S(t, 0) = -t$ for WGAN
2. λ regularization parameter of the ℓ_p -Autoencoder
3. k_d number of iterations for discriminator, k_g number of iterations for generator
4. η_d learning rate of discriminator, η_g learning rate of generator
5. ω weights of discriminator, θ weights of generator
- 1: **for** number of training iterations **do**
- 2: • Sample minibatch of m examples $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$ from reference-noisy image pairs.
- 3: **for** $k = 1, 2, \dots, k_d$ **do**
- 4: • Update the discriminator by gradient ascent:
- 5: $g_\omega \leftarrow \frac{1}{m} \sum_{i=1}^m \nabla_\omega [S(D_\omega(x_i), 1) + S(D_\omega(G_\theta(y_i)), 0) + \mu (\|\nabla_{\widehat{x}} D_\omega(\widehat{x}_i)\|_2 - 1)^2]$
 where $\mu > 0$ for WGANgp only;
- 6: $\omega \leftarrow \omega + \eta_d g_\omega$
- 7: **end for**
- 8: **for** $k = 1, 2, \dots, k_g$ **do**
- 9: • Update the generator by gradient descent:
- 10: $g_\theta \leftarrow \frac{1}{m} \sum_{i=1}^m \nabla_\theta [S(D_\omega(G_\theta(y_i)), 0) + \lambda |G_\theta(y_i) - x_i|^p]$, $p \in \{1, 2\}$;
- 11: $\theta \leftarrow \theta - \eta_g g_\theta$
- 12: **end for**
- 13: **end for**

Return: Denoised image: $\widehat{x}_i = G_\theta(y_i)$

Evaluation Method

We exploit the following three metrics to determine whether the denoising result is good or not. They are the Mean Square Error (MSE), the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM).

- (MSE) For images of size $d_1 \times d_2$, the Mean Square Error (MSE) between reference image x and denoised image \widehat{x} is defined as, $MSE := \frac{1}{d_1 d_2} \sum_{i=1}^{d_1} \sum_{j=1}^{d_2} (x(i, j) - \widehat{x}(i, j))^2$. The smaller is the MSE, the better the denoising result is.
- (PSNR) Similarly, the Peak Signal-to-Noise Ratio (PSNR) between reference image x and denoised image \widehat{x} whose pixel value range is $[0, t]$ (1 by default), is defined by $PSNR := 10 \log_{10} \frac{t^2}{\frac{1}{d_1 d_2} \sum_{i=1}^{d_1} \sum_{j=1}^{d_2} (x(i, j) - \widehat{x}(i, j))^2}$. The larger is the PSNR, the better the denoising result is.
- (SSIM) The third criterion is the Structural Similarity Index Measure (SSIM) between reference image x and denoised image \widehat{x} is defined in³⁸, $SSIM = \frac{(2\mu_x \mu_{\widehat{x}} + c_1)(2\sigma_x \sigma_{\widehat{x}} + c_2)(\sigma_{x\widehat{x}} + c_3)}{(\mu_x^2 + \mu_{\widehat{x}}^2 + c_1)(\sigma_x^2 + \sigma_{\widehat{x}}^2 + c_2)(\sigma_x \sigma_{\widehat{x}} + c_3)}$. where μ_x ($\mu_{\widehat{x}}$) and σ_x ($\sigma_{\widehat{x}}$) are the mean and variance of x (\widehat{x}), respectively, $\sigma_{x\widehat{x}}$ is covariance of x and \widehat{x} , $c_1 = K_1 L^2$, $c_2 = K_2 L^2$, $c_3 = \frac{c_2}{2}$ three variables to stabilize the division with weak denominator ($K_1 = 0.01$, $K_2 = 0.03$ by default), L is the dynamic range of the pixel-value (1 by default). The value SSIM of lies in $[0, 1]$, where the closer it is to 1, the better the result is.

Although these metrics are widely used in image denoising, we note that they might not be the best metrics for Cryo-EM images. For example, it shows an example that the best-reconstructed images perhaps do not meet the best MSE/PSNR/SSIM metrics in the supplementary.

In addition to these metrics, we consider the 3D reconstruction based on denoised images. Particularly, we take the 3D reconstruction by RELION³⁹ to validate the denoised result. The procedure of our RELION reconstruction is as follows: firstly creating the 3D initial model, then doing 3D classification, followed by operating 3D auto-refine. Moreover, for heterogeneous conformations in simulation data, we further turn the denoising results into a clustering problem to measure the efficacy of denoising methods, whose details will be discussed in the supplementary.

RNAP: Simulation Dataset

We design a conformational heterogeneous dataset obtained by simulations. We use *Thermus aquaticus* RNA Polymerase (RNAP) in complex with σ^A factor (*Taq* holoenzyme) for our dataset. RNAP is the enzyme that transcribes RNA from DNA (transcription) in the cell. During the initiation of transcription, the holoenzyme must bind to the DNA, then separate the double-stranded DNA into single-stranded⁴⁰. *Taq* holoenzyme has a crab-claw like structure, with two flexible domains, the clamp and β pincers. The clamp, especially, has been suggested to play an important role in the initiation, as it has been captured in various conformations by CryoEM during initiation⁴¹. Thus, we focus on the movement of the clamp in this study. To generate the heterogeneous dataset, we start with two crystal structures of *Taq* holoenzyme, which vary in their clamp conformation, open (PDB ID: 1L9U⁴²) and closed (PDB ID: 4XLN⁴³) clamp. For the closed clamp structure, we remove the DNA and RNA in the crystal structure, leaving only the RNAP and σ^A for our dataset. The *Taq* holoenzyme has about 370 kDa molecular weight. We then generate the clamp intermediate structures between the open and closed clamp using multiple-basin coarse-grained (CG) molecular dynamic (MD) simulations^(44,45). CG-MD simulations simplify the system such that the atoms in each amino acid are represented by one particle. The structures from CG-MD simulations are refined back to all-atom or atomic structures using PD2 ca2main⁽⁴⁶⁾ and SCRWL4⁽⁴⁷⁾. Five structures with equally-spaced clamp opening angle are chosen for our heterogeneous dataset (shown in Figure 5). Then, we convert the atomic structures to $128 \times 128 \times 128$ volumes using Xmipp package⁴⁸ and generate the 2D projections with an image size of 128×128 pixels. We further contaminate those clean images with additive Gaussian noise at different signal noise ratios (SNR): $SNR = 0.05$. The SNR is defined by “ $SNR = \text{Var}(\text{Signal})/\text{Var}(\text{Noise})$ ” in the real space. For simplicity, we did not apply the contrast transfer function (CTF) to the datasets, and all the images are centered. Figure 5 shows the five conformations pictures.

Training data size is 25000 paired images (noisy and reference images), Test data to calculate the MSE, PSNR and SSIM is another 1500 paired images.

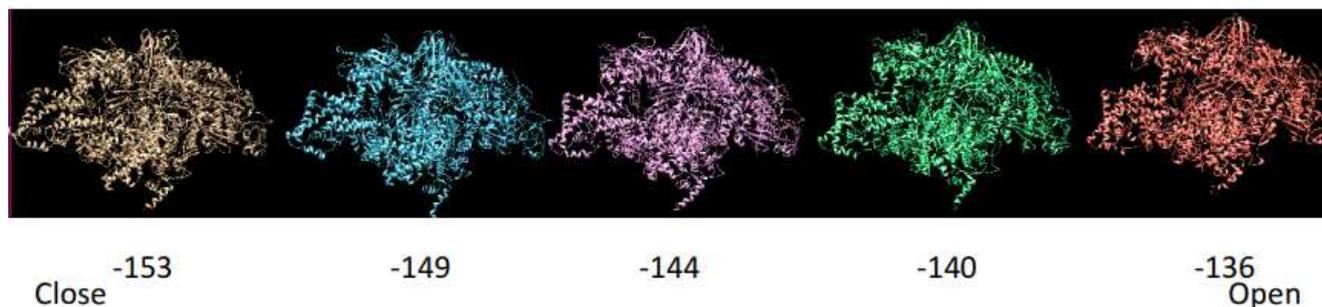


Figure 5. Five conformations in RNAP heterogeneous dataset, from left to right are close conformation to open conformation of different angles.

EMPIAR-10028: Real Dataset

This is a real-world experimental dataset that was firstly studied in²⁶: the *Plasmodium falciparum* 80S ribosome dataset (EMPIAR-10028). They recover the Cryo-EM structure of the cytoplasmic ribosome from the human malaria parasite, *Plasmodium falciparum*, in complex with emetine, an anti-protozoan drug, at 3.2\AA resolution. Ribosome is the essential enzyme that translates RNA to protein molecules, the second step of central dogma. The inhibition of ribosome activity of *Plasmodium falciparum* would effectively kill the parasite²⁶. We can regard this dataset to have homogeneous property. This dataset contains 105247 noisy particles with an image size of 360×360 pixels. In order to decrease the complexity of the computing, we pick up the center square of each image with a size of 256×256 , since the surrounding area of the image is entirely useless that does not lose information in such a preprocessing. Then the 256×256 images are fed as the input of the G_θ -network (Figure 2). Since the GAN-based method needs clean images as reference, we prepare their clean counterparts in the following way: we first use cryoSPARC1.0⁴⁹ to build a 3.2\AA resolution volume and then rotate the 3D-volume by the Euler angles obtained by cryoSPARC to get projected 2D-images. The training data size we pick is 19500, and the test data size is 500.

References

1. Kühlbrandt, W. The resolution revolution. *Science* **343**, 1443–1444 (2014).
2. Bai, X.-C., McMullan, G. & Scheres, S. H. How cryo-em is revolutionizing structural biology. *Trends biochemical sciences* **40**, 49–57 (2015).
3. Shen, P. S. The 2017 nobel prize in chemistry: cryo-em comes of age. *Anal. bioanalytical chemistry* **410**, 2053–2057 (2018).
4. Frank, J. *Three-dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state* (Oxford University Press, 2006).
5. Scheres, S. H. Processing of structurally heterogeneous cryo-em data in relion. In *Methods in enzymology*, vol. 579, 125–157 (Elsevier, 2016).
6. Wang, F. *et al.* Deeppicker: A deep learning approach for fully automated particle picking in cryo-em. *J. structural biology* **195**, 325–336 (2016).
7. Wang, J. & Yin, C. A zernike-moment-based non-local denoising filter for cryo-em images. *Sci. China Life Sci.* **56**, 384–390 (2013).
8. Wei, D.-Y. & Yin, C.-C. An optimized locally adaptive non-local means denoising filter for cryo-electron microscopy data. *J. structural biology* **172**, 211–218 (2010).
9. Xian, Y. *et al.* Data-driven tight frame for cryo-em image denoising and conformational classification. In *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 544–548 (IEEE, 2018).
10. Dabov, K., Foi, A., Katkovnik, V. & Egiazarian, K. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing* **16**, 2080–2095 (2007).
11. Aharon, M., Elad, M. & Bruckstein, A. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing* **54**, 4311–4322 (2006).
12. Bhamre, T., Zhang, T. & Singer, A. Denoising and covariance estimation of single particle cryo-em images. *J. structural biology* **195**, 72–81 (2016).
13. Vincent, P., Larochelle, H., Bengio, Y. & Manzagol, P.-A. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning*, 1096–1103 (ACM, 2008).
14. Xie, J., Xu, L. & Chen, E. Image denoising and inpainting with deep neural networks. In *Advances in neural information processing systems*, 341–349 (2012).
15. Zhang, K., Zuo, W., Chen, Y., Meng, D. & Zhang, L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Process.* **26**, 3142–3155 (2017).
16. Agostinelli, F., Anderson, M. R. & Lee, H. Adaptive multi-column deep neural networks with application to robust image denoising. In *Advances in Neural Information Processing Systems*, 1493–1501 (2013).
17. Goodfellow, I. *et al.* Generative adversarial nets. In *Advances in neural information processing systems*, 2672–2680 (2014).
18. Arjovsky, M., Chintala, S. & Bottou, L. Wasserstein gan. *arXiv preprint arXiv:1701.07875* (2017).
19. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V. & Courville, A. C. Improved training of wasserstein gans. In *Advances in neural information processing systems*, 5767–5777 (2017).
20. Yang, Q. *et al.* Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE transactions on medical imaging* **37**, 1348–1357 (2018).
21. Su, M., Zhang, H., Schawinski, K., Zhang, C. & Cianfrocco, M. A. Generative adversarial networks as a tool to recover structural information from cryo-electron microscopy data. *BioRxiv* 256792 (2018).
22. Gao, C., Liu, J., Yao, Y. & Zhu, W. Robust estimation and generative adversarial nets. In *International Conference on Learning Representations (ICLR), New Orleans, Louisiana, United States* (2019). ArXiv preprint arXiv:1810.02030.
23. Gao, C., Yao, Y. & Zhu, W. Generative adversarial nets for robust scatter estimation: A proper scoring rule perspective. *J. Mach. Learn. Res. accepted* (2019). ArXiv preprint arXiv:1903.01944.
24. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
25. Karras, T., Aila, T., Laine, S. & Lehtinen, J. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196* (2017).

26. Wong, W. *et al.* Cryo-em structure of the plasmodium falciparum 80s ribosome bound to the anti-protozoan drug emetine. *Elife* **3**, e03080 (2014).
27. Lehtinen, J. *et al.* Noise2noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189* (2018).
28. Krull, A., Buchholz, T.-O. & Jug, F. Noise2void-learning denoising from single noisy images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2129–2137 (2019).
29. Chen, J., Chen, J., Chao, H. & Yang, M. Image blind denoising with generative adversarial network based noise modeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3155–3164 (2018).
30. Quan, Y., Chen, M., Pang, T. & Ji, H. Self2self with dropout: Learning self-supervised denoising from single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1890–1898 (2020).
31. Bepler, T., Kelley, K., Noble, A. J. & Berger, B. Topaz-denoise: general deep denoising models for cryoem and cryoet. *Nat. communications* **11**, 1–12 (2020).
32. Palovcak, E., Asarnow, D., Campbell, M. G., Yu, Z. & Cheng, Y. Enhancing the signal-to-noise ratio and generating contrast for cryo-em images with convolutional neural networks. *IUCrJ* **7** (2020).
33. Zhong, E. D., Bepler, T., Davis, J. H. & Berger, B. Reconstructing continuous distributions of 3d protein structure from cryo-em images. In *ICLR* (2020).
34. Frank, J. Two-dimensional averaging techniques. *Three-Dimensional Electron Microsc. Macromol. Assemblies: Vis. Biol. Mol. Their Nativ. State* 131–144 (1996).
35. Bai, C. *et al.* Robust contrast-transfer-function phase retrieval via flexible deep learning networks. *Opt. letters* **44**, 5141–5144 (2019).
36. Huber, P. J. Robust estimation of a location parameter. In *Breakthroughs in statistics*, 492–518 (Springer, 1992).
37. Tukey, J. W. Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians, Vancouver, 1975*, vol. 2, 523–531 (1975).
38. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* **13**, 600–612 (2004).
39. Scheres, S. H. Relion: implementation of a bayesian approach to cryo-em structure determination. *J. structural biology* **180**, 519–530 (2012).
40. Browning, D. F. & Busby, S. J. The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.* **2**, 57–65 (2004).
41. Chen, J. *et al.* Stepwise promoter melting by bacterial rna polymerase. *Mol. Cell* (2020).
42. Murakami, K. S., Masuda, S. & Darst, S. A. Structural basis of transcription initiation: Rna polymerase holoenzyme at 4 Å resolution. *Science* **296**, 1280–1284 (2002).
43. Bae, B., Feklistov, A., Lass-Napiorkowska, A., Landick, R. & Darst, S. A. Structure of a bacterial rna polymerase holoenzyme open promoter complex. *Elife* **4**, e08504 (2015).
44. Okazaki, K.-i., Koga, N., Takada, S., Onuchic, J. N. & Wolynes, P. G. Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proc. Natl. Acad. Sci.* **103**, 11844–11849 (2006).
45. Kenzaki, H. *et al.* Cafemol: A coarse-grained biomolecular simulator for simulating proteins at work. *J. Chem. Theory Comput.* **7**, 1979–1989 (2011).
46. Moore, B. L., Kelley, L. A., Barber, J., Murray, J. W. & MacDonald, J. T. High-quality protein backbone reconstruction from alpha carbons using gaussian mixture models. *J. computational chemistry* **34**, 1881–1889 (2013).
47. Krivov, G. G., Shapovalov, M. V. & Dunbrack Jr, R. L. Improved prediction of protein side-chain conformations with scwrl4. *Proteins: Struct. Funct. Bioinforma.* **77**, 778–795 (2009).
48. Marabini, R. *et al.* Xmipp: an image processing package for electron microscopy. *J. structural biology* **116**, 237–240 (1996).
49. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryosparc: algorithms for rapid unsupervised cryo-em structure determination. *Nat. methods* **14**, 290 (2017).

Acknowledgements

X.H. acknowledges support from the Hong Kong Research Grant Council (T13-605/18-W, AoE/M-09/12 and AoE/P-705/16) and Innovation and Technology Commission (ITCPD/17-9). The research of Y. Y. was supported in part by Hong Kong Research Grant Council (HKRGC) grant 16303817, ITF UIM/390, as well as awards from Tencent AI Lab, Si Family Foundation, and Microsoft Research-Asia.

Author contributions statement

H.G took the experiments and analysed the result. I.C.U provided the dataset. X.H and Y.Y supervised the project and wrote the manuscript.

Figures

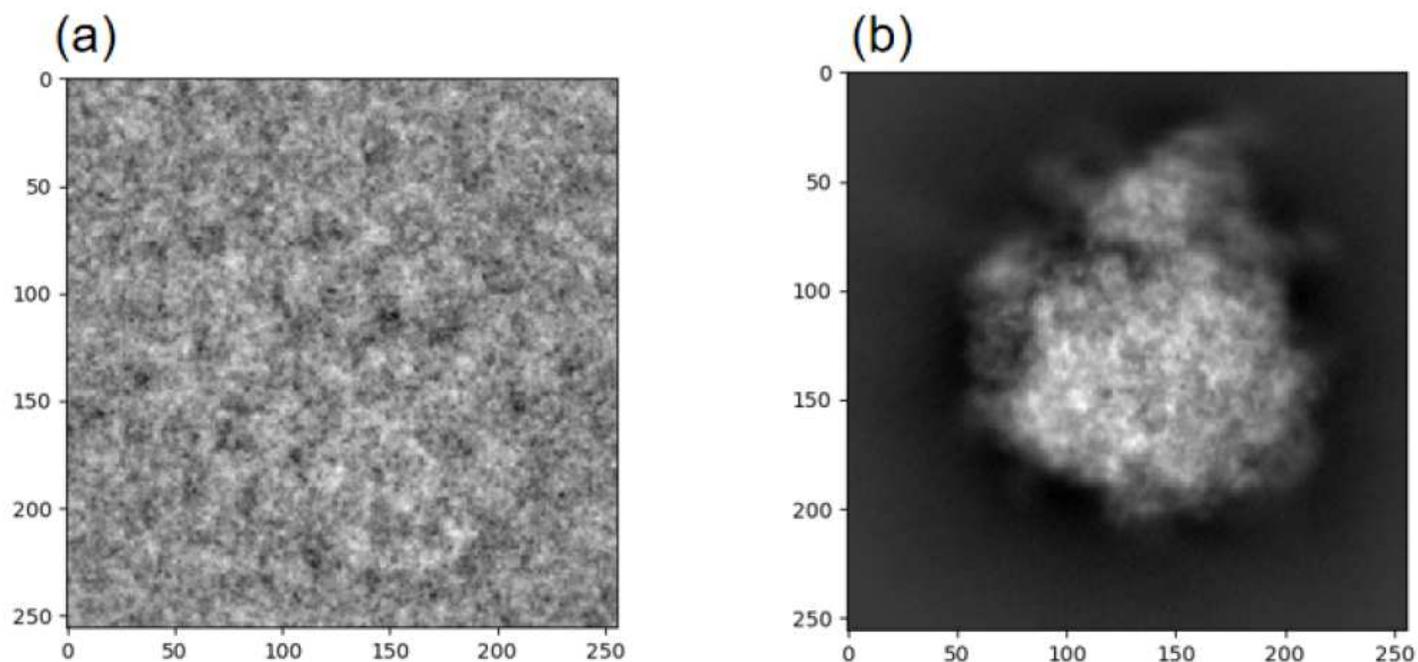


Figure 1

(a) a noisy Cryo-EM image (b) a reference image

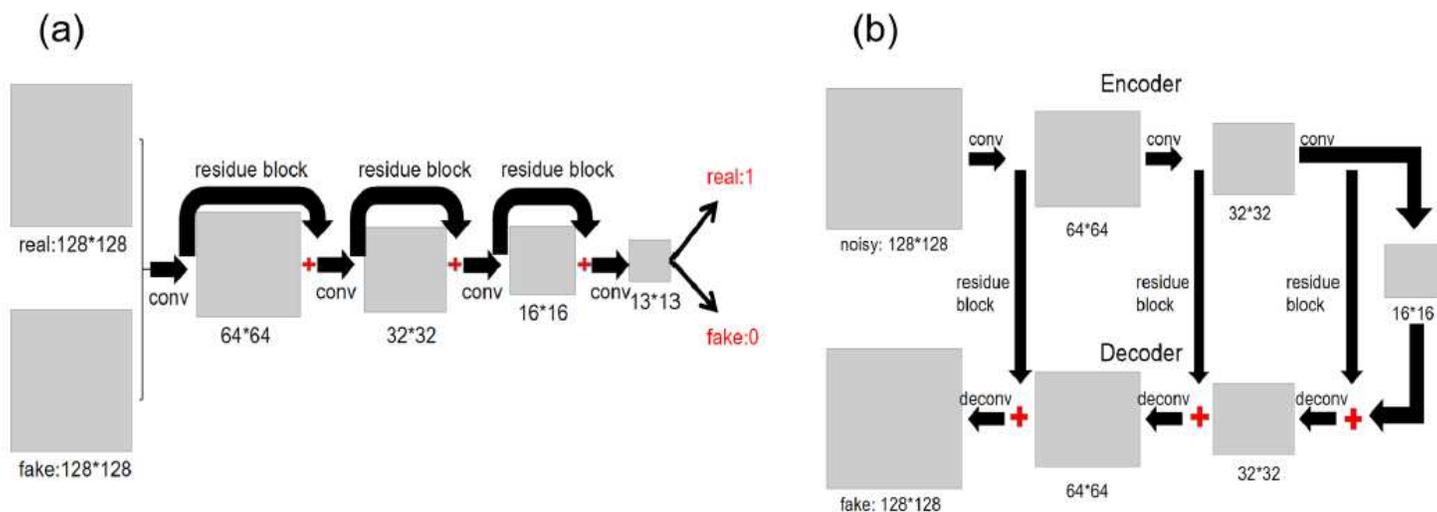


Figure 2

The architectures of (a) discriminator (D) and (b) generator (G) which borrow the residue structure. The input image size (128 X 128) here is adapted to RNAP dataset, while in EMPIRE-10028 dataset it is 256 X 256 with a similar architecture.

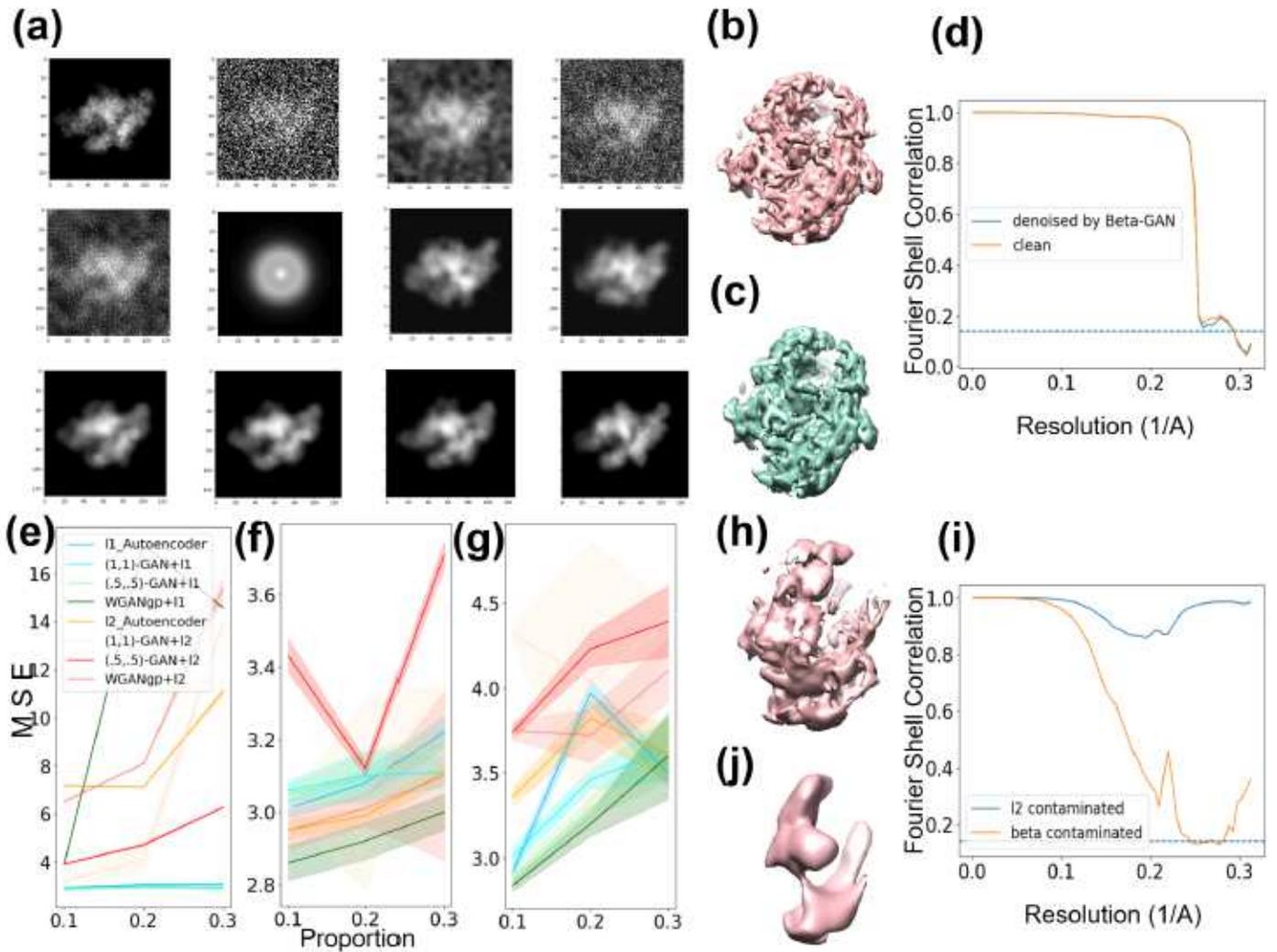


Figure 3

Results for RNAP dataset. (a) is denoised images in different denoised methods (from left to right, top to bottom): Clean, Noisy, BM3D, KSVD, Non-local means, CWF, ℓ_1 -Autoencoder, ℓ_2 -autoencoder, (1,1)-GAN + ℓ_1 , (0,0)-GAN + ℓ_1 , (:5; :5)-GAN + ℓ_1 and WGANgp + ℓ_1 . (b) and (c) are reconstruction of clean images and (:5; :5)-GAN + ℓ_1 denoised images. (d) is FSC curve of (b) and (c). (e), (f) and (g) are robustness tests of various methods under $\epsilon = 2 \times 10^{-2}$ proportion contamination in three types of contamination: (e) Type A: replacing the reference images with random noise; (f) Type B: replacing the noisy images with random noise; (g) Type C: replacing both with random noise. (h) and (j) are reconstructions of images with (:5; :5)-GAN + ℓ_1 and ℓ_2 -autoencoder under type A contamination, respectively, where ℓ_2 -autoencoder totally fails but (:5; :5)-GAN + ℓ_1 is robust. (i) shows FSC curves of (h) and (j).

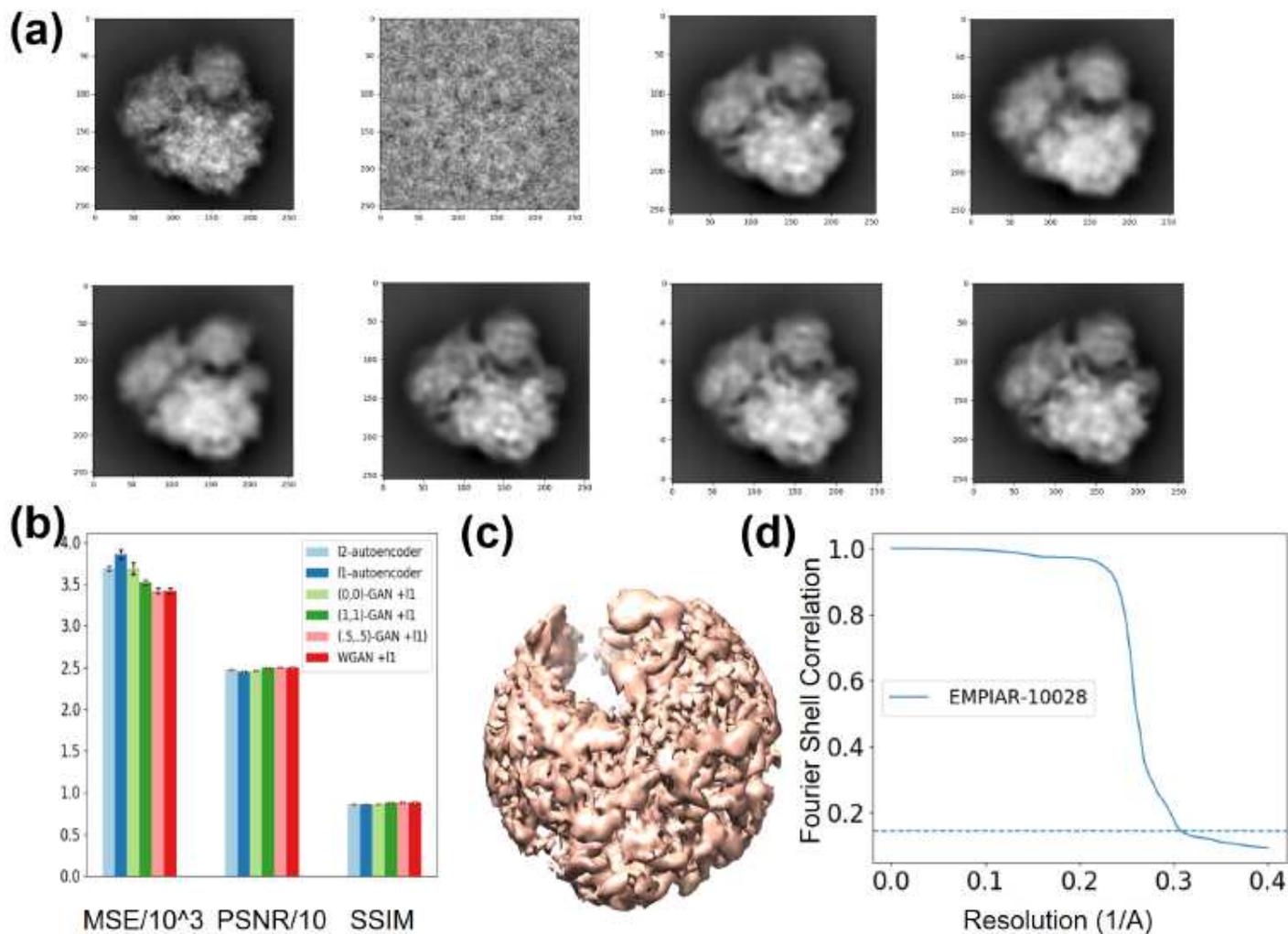


Figure 4

Results for EMPIRE-10028. (a) Comparison in EMPIRE-10028 dataset in different deep learning methods (from left to right, top to bottom): clean image, noisy image, ℓ_1 -Autoencoder, ℓ_2 -Autoencoder, (0,0)-GAN + ℓ_1 , (1,1)-GAN + ℓ_1 , (.5, .5)-GAN + ℓ_1 , WGANgp + ℓ_1 . (b) is the MSE, PSNR and SSIM in different denoised methods. (c) and (d) is the 3D-reconstruction of denoised images by (.5, .5)-GAN + ℓ_1 and the FSC curve, respectively. The resolution of reconstruction from (.5, .5)-GAN + ℓ_1 denoised images is 3.20 $^\circ$ A, which is as good as the original resolution.

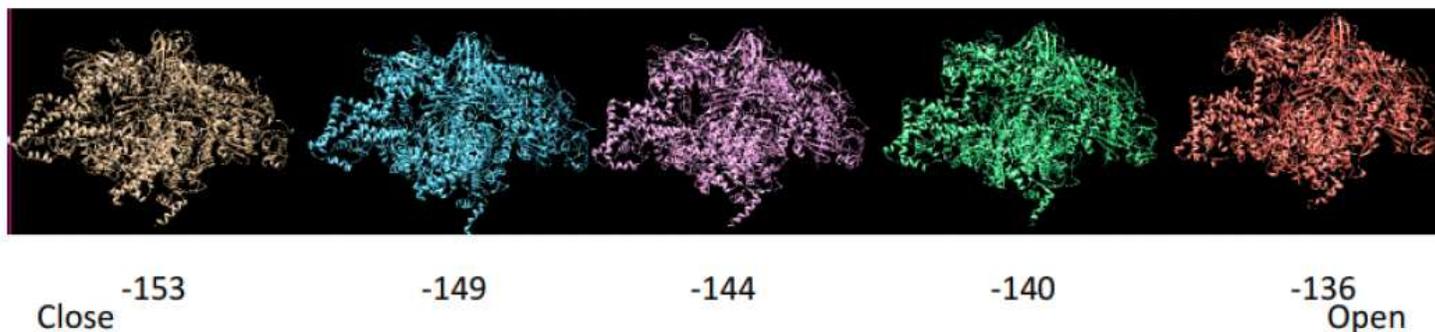


Figure 5

Five conformations in RNAP heterogeneous dataset, from left to right are close conformation to open conformation of different angles.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [robustGANdenosingAPPENDIXd1.pdf](#)