

CRYOSWD: SLICED WASSERSTEIN DISTANCE MINIMIZATION FOR 3D RECONSTRUCTION IN CRYO-ELECTRON MICROSCOPY

Mona Zehni, Zhizhen Zhao

Department of ECE and CSL, University of Illinois at Urbana-Champaign

ABSTRACT

Single particle reconstruction (SPR) in cryo-electron microscopy (cryo-EM) is a prominent imaging method that recovers the 3D shape of a biomolecule, given a large number of its noisy projections from random and unknown views. Recently, CryoGAN [1] cast SPR as an unsupervised distribution matching problem and solved it via a Wasserstein generative adversarial network (WGAN) framework. The approach bypasses the estimation of the projection parameters. The reconstruction criterion in CryoGAN is Wasserstein-1 distance. Despite the desirable properties of Wasserstein distances (WD) such as continuity and almost everywhere differentiability, they are difficult to compute and require careful tuning for a stable training. Sliced Wasserstein distance (SWD), on the other hand, has shown desirable training stability and ease to compute. Therefore, we propose to replace Wasserstein-1 distance with SWD in the CryoGAN framework, hence the name *CryoSWD*. In low noise regimes, we show how CryoSWD eliminates the need to have a discriminator which is crucial in CryoGAN. However, coupling CryoSWD with a discriminator boosts its performance, especially in high noise settings. While performing as good as CryoGAN, CryoSWD does not require a gradient penalty term for stabilizing the training and imposing Lipschitz continuity of the discriminator.

Index Terms—Cryo-electron microscopy, 3D ab-initio reconstruction, sliced Wasserstein distance, CryoGAN.

1. INTRODUCTION

Studying the 3D structure of biomolecules is the key to understand their function and enables better design of medical and therapeutic strategies. Recently, cryo-electron microscopy (cryo-EM) has become a prevalent tool to image biomolecules in their native states with no need for crystallization [2]. Specifically, in cryo-EM single particle reconstruction (SPR), samples of the same molecule are frozen in cryogenic temperatures and imaged via an electron microscope [3]. As the relative orientation of the frozen particles with respect to the electron beam is random, each projection image corresponds to a random and unknown view of the 3D molecule. Furthermore, to avoid radiation damage, the electron dose is kept low, thus leading to noisy projection images with low signal-to-noise ratios (SNR). The ultimate goal in

SPR is to recover the 3D shape of the underlying biomolecule given its noisy and random projection image set. We specifically focus on 3D ab-initio SPR which unlike refinement-based solutions, does not require an initial 3D map [4, 5].

SPR approaches either (1) incorporate the estimation of the projection parameters in the reconstruction pipeline or (2) fully bypass it. In one example of the former, the 3D views corresponding to each projection is estimated via template matching or common-line based methods [6]. Next, the 3D map is reconstructed via direct Fourier methods [7] or iterative regularized optimizations [8]. Note that, these methods rely on the quality of the template and the estimated projection parameters degrade with lower SNRs [9]. Other examples of (1) are alternating optimization [10] and maximum likelihood based [11, 12] methods. These methods primarily rely on iterative updates of the 3D map and the projection parameters. As the size of the projection dataset can be very large ($10^4 - 10^6$), updating the projection parameters can become prohibitively expensive.

On the other hand, moment-based ab-initio SPR methods detour the estimation of the projection parameters by first summarizing the large projection dataset into a set of rotation-invariant features [13, 4, 14] and the 3D map is recovered after solving a non-convex optimization. While these approaches no longer require multiple passes through the whole projection dataset, they are often used to find a low-resolution ab-initio model and are sensitive to initialization.

Similarly, CryoGAN [1] circumvents the estimation of the projection parameters while solving SPR through a distribution matching lens. In other words, CryoGAN reframes SPR as finding a 3D map such that the distribution of the projection dataset generated from the estimated 3D map matches the given projection image dataset. To achieve this, Wasserstein-1 distance is chosen as the reconstruction criterion and minimized in a Wasserstein generative adversarial network (WGAN) framework [15].

In the context of GANs, properties such as continuity alongside almost everywhere differentiability make Wasserstein distances (WD) a desirable metric to use [15, 16]. However, training stability concerns [17] and intractability of computing WD especially for high-dimensional distributions has led to the use of sliced Wasserstein distance (SWD) [18, 17, 19, 20] as an alternative. SWD computes the

Wasserstein distance between the probability distributions that are projected onto random 1D subspaces. While previous works using SWD mainly focus on learning to generate samples from a target data distribution [18, 17], here our goal is to solve an inverse problem (i.e. cryo-EM SPR).

Motivated by the advantages of SWD in generative modeling, in this work we adopt SWD criterion to solve SPR. Following CryoGAN, we solve SPR via a distribution matching formulation. However, motivated by the aforementioned advantages of SWD, we choose SWD as our distribution matching loss. We show that our method, namely *CryoSWD*, can recover the overall shape of a 3D map without using a discriminator in high SNR regimes. However, jointly training a discriminator improves the quality of the reconstructed 3D map. We hope this paper opens future research directions on the adoption of SWD criterion in various inverse problems.

This paper is organized as follows. We define the SPR forward model in Section 2. We describe our method and results in Sections 3-4 and conclude the paper in Section 5.

2. FORWARD MODEL

Let $V : \mathcal{B}(\mathbb{R}^3) \rightarrow \mathbb{R}^+ \cup \{0\}$ denote the 3D density map compactly supported in a ball $\mathcal{B}(\mathbb{R}^3)$. In cryo-EM, the measurements are noisy random projections of V , modeled as:

$$\zeta_\ell = C_\ell * \mathcal{P}_{R_\ell} V + \varepsilon_\ell, \quad \{\mathcal{P}_{R_\ell} V\}(x, y) = \int V(R_\ell^T \mathbf{x}) dz \quad (1)$$

where $\zeta_\ell : \mathbb{R}^2 \rightarrow \mathbb{R}$, $\ell \in \{1, \dots, L\}$ is the ℓ -th projection image. The X-ray transform $\mathcal{P}_{R_\ell} : \mathcal{B}(\mathbb{R}^3) \rightarrow \mathcal{B}(\mathbb{R}^2)$ takes the line integral along the z -direction of its input rotated by 3×3 rotation matrix $R_\ell \in SO(3)$ in special orthogonal group $SO(3)$ and $\mathbf{x}$ is the 3D coordinate, $\mathbf{x} = [x, y, z]^T \in \mathbb{R}^3$. Note that, R_ℓ defines the projection view corresponding to ζ_ℓ . We assume the projection views are unknown and $\{R_\ell\}_{\ell=1}^L$ are uniformly distributed in $SO(3)$.

In addition, $C_\ell : \mathbb{R}^2 \rightarrow \mathbb{R}$ is the contrast transfer function (CTF) of the microscope affecting ζ_ℓ . Finally, each projection image is corrupted by additive Gaussian noise $\varepsilon_\ell(x, y) \sim \mathcal{N}(0, \sigma^2)$. In practice, the projection images ζ_ℓ have finite resolution and are discretized to $m \times m$ pixels. In addition, we also consider V to be discretized in voxels on its finite compact support. In 3D SPR, the goal is to find V given its large set of random noisy projections $\{\zeta_\ell\}_{\ell=1}^L$.

3. METHOD

3.1. Background on CryoGAN [1]

Let $\mathcal{D}_{\text{real}} = \{\zeta_\ell\}_{\ell=1}^L$ be the real projection dataset with $\mathbb{P}_{\text{real}}$ distribution. Also, $\mathbb{P}_{\text{syn}}(v)$ is the distribution of the synthesized projection dataset $\mathcal{D}_{\text{syn}}$, generated following $\tilde{\zeta} = C * \mathcal{P}_R v + \varepsilon$, with R uniformly distributed in $SO(3)$ and $\varepsilon(x, y) \sim \mathcal{N}(0, \sigma^2)$. CryoGAN [1] defines SPR as:

$$\hat{V} = \min_v W_1(\mathbb{P}_{\text{real}}, \mathbb{P}_{\text{syn}}(v)), \quad (2)$$

where W_1 is the Wasserstein-1 distance [20]. As computing W_1 between high-dimensional distributions is intractable, often Kantorovich-Rubinstein duality [15] is invoked and thus

Algorithm 1 CryoSWD

Require: α, β : learning rates for the discriminator and the 3D map. n_{disc} : the number of update iterations of the discriminator per 3D map update iteration.

Require: Initialize $\hat{V}$ with zeros and (ϕ, φ) randomly.

Output: Estimated 3D map $\hat{V}$ given $\mathcal{D}_{\text{real}} = \{\zeta_\ell\}_{\ell=1}^L$.

```

1: while not converged do
2:   Draw  $N_\Omega$  random samples of  $\{\omega_k\}_{k=1}^{N_\Omega}, \omega_k \in \mathbb{S}^{n-1}$ .
3:   for  $t = 0, \dots, n_{\text{disc}} - 1$  do
4:     Sample a random batch from  $\mathcal{D}_{\text{real}}, \{\zeta_b\}_{b=1}^B$ .
5:     Sample a random batch from  $\mathcal{D}_{\text{syn}}, \{\tilde{\zeta}_b\}_{b=1}^B$ .
6:     Update  $(\phi, \varphi)$  with gradient ascent steps using the
       gradients of (9) with respect to  $(\phi, \varphi)$ .
7:   end for
8:   Sample a random batch from  $\mathcal{D}_{\text{real}}$  and  $\mathcal{D}_{\text{syn}}$ .
9:   Compute the embeddings of the real and synthetic
       batches, i.e.  $u_b = \phi(\zeta_b)$  and  $\tilde{u}_b = \phi(\tilde{\zeta}_b)$ .
10:  Compute real and synthetic projected embeddings,
       i.e.  $\{\langle \tilde{u}_b, \omega_k \rangle\}_{k=1}^{N_\Omega}$  and  $\{\langle \tilde{u}_b, \omega_k \rangle\}_{k=1}^{N_\Omega}$ .
11:  For each  $\omega_k$ , sort the batch of projected embeddings.
12:  Update  $\hat{V}$  using gradient descent step with the following
       gradient.
```

$$\nabla_v \left(\frac{1}{N_\Omega B} \sum_{k=1}^{N_\Omega} \sum_{b=1}^B \|\langle u_{\pi_b}, \omega_k \rangle - \langle \tilde{u}_{\pi'_b}, \omega_k \rangle\|^2 \right)$$

```

13: end while
```

solving (2) is equivalent to:

$$\hat{V} = \min_v \sup_{\|f\|_L \leq 1} \mathbb{E}_{\zeta \sim \mathbb{P}_{\text{real}}} [f(\zeta)] - \mathbb{E}_{\tilde{\zeta} \sim \mathbb{P}_{\text{syn}}(v)} [f(\tilde{\zeta})] \quad (3)$$

with f belonging to the family of 1-Lipschitz functions. In WGAN [15] context, f is modeled by a neural network critic and the Lipschitz continuity constraint is enforced through gradient penalty terms [21].

3.2. CryoSWD

In CryoSWD, we replace the W_1 criterion used in CryoGAN with SWD. Specifically, we use the quadratic sliced Wasserstein-2 distance (SW_2^2). Suppose $\mathbb{P}_{\text{real}}$ and $\mathbb{P}_{\text{syn}}$ are probability distributions on $\mathbb{R}^d$, $d = m^2$ (consider flattened versions of the projection images). Then, the quadratic SW_2 between the two distributions is defined as [17]:

$$SW_2^2(\mathbb{P}_{\text{real}}, \mathbb{P}_{\text{syn}}) = \int_{\mathbb{S}^{d-1}} W_2^2(\mathbb{P}_{\text{real}}^\omega, \mathbb{P}_{\text{syn}}^\omega) d\omega \quad (4)$$

$$\approx \frac{1}{|N_\Omega|} \sum_{k=1}^{N_\Omega} W_2^2(\mathcal{D}_{\text{real}}^{\omega_k}, \mathcal{D}_{\text{syn}}^{\omega_k}) \quad (5)$$

where W_2 denotes the Wasserstein-2 distance, $\mathbb{S}^{d-1}$ is the d -dimensional unit sphere and $\mathbb{P}_{\text{real}}^\omega$ ($\mathbb{P}_{\text{syn}}^\omega$) is the distribution of the projected $\mathbb{P}_{\text{real}}$ ($\mathbb{P}_{\text{syn}}$) distribution along $\omega \in \mathbb{S}^{d-1}$. In practice, (4) is approximated by (5), using a finite summation with N_Ω SW projection directions randomly sampled from $\mathbb{S}^{d-1}$. Also, $\mathcal{D}_{\text{real}}^\omega = \{\langle \zeta, \omega \rangle | \zeta \in \mathcal{D}_{\text{real}}\}$ and $\mathcal{D}_{\text{syn}}^\omega = \{\langle \tilde{\zeta}, \omega \rangle | \tilde{\zeta} \in \mathcal{D}_{\text{syn}}\}$ denote the projected datasets by ω .

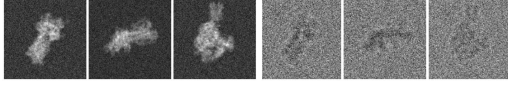


Fig. 1. Examples of projection images of HP1 and HJC maps.

We emphasize the difference between projection operator employed in the computation of SW_2 and the X-ray transform denoted by $\mathcal{P}_R$ in the cryo-EM forward model. While the former operates by taking the inner product of its input with projection directions $\omega \in \mathbb{S}^{d-1}$, the latter takes the line integral of its input along the view specified by $R \in SO(3)$.

As $\mathcal{D}_{\text{real}}^\omega$ and $\mathcal{D}_{\text{syn}}^\omega$ are 1D distributions, the W_2 between the two has a closed form expression based on the solution of the optimal transport map [22, 17, 19] and is written as:

$$W_2^2(\mathcal{D}_{\text{real}}^\omega, \mathcal{D}_{\text{syn}}^\omega) = \frac{1}{L} \sum_{i=1}^L \|\langle \zeta_{\pi_i}, \omega \rangle - \langle \tilde{\zeta}_{\pi'_i}, \omega \rangle\|^2 \quad (6)$$

where π and π' denote the permuted indices sorted such that: $\langle \zeta_{\pi_i}, \omega \rangle \leq \langle \zeta_{\pi_{i+1}}, \omega \rangle$, $\zeta_i \in \mathcal{D}_{\text{real}}$ and $\langle \tilde{\zeta}_{\pi'_i}, \omega \rangle \leq \langle \tilde{\zeta}_{\pi'_{i+1}}, \omega \rangle$, $\tilde{\zeta}_i \in \mathcal{D}_{\text{syn}}$, for all $1 \leq i \leq L$. Finally, plugging (6) in (5) leads to:

$$SW_2^2(\mathbb{P}_{\text{real}}, \mathbb{P}_{\text{syn}}) = \frac{1}{N_\Omega L} \sum_{k=1}^{N_\Omega} \sum_{i=1}^L \|\langle \zeta_{\pi_i}, \omega_k \rangle - \langle \tilde{\zeta}_{\pi'_i}, \omega_k \rangle\|^2. \quad (7)$$

For high-dimensional distributions, to obtain better results, SWD minimization is often accompanied by training a discriminator (parameterized by a neural network) and mapping the dataset to an adversarially learnt embedding space [17, 19]. Thus, our final formulation of CryoSWD becomes:

$$\hat{V} = \arg \min_v \frac{1}{N_\Omega L} \sum_{k=1}^{N_\Omega} \sum_{i=1}^L \|\langle u_{\pi_i}, \omega_k \rangle - \langle \tilde{u}_{\pi'_i}, \omega_k \rangle\|^2, \quad (8)$$

$$\hat{\phi}, \hat{\varphi} = \arg \max_{\phi, \varphi} \sum_{\zeta \in \mathcal{D}_{\text{real}}} \ln[\varphi(\phi(\zeta))] + \sum_{\tilde{\zeta} \in \mathcal{D}_{\text{syn}}} \ln[1 - \varphi(\phi(\tilde{\zeta}))] \quad (9)$$

where $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^n$ is a neural network encoder and $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ is a classifier with a final sigmoid layer. Note that (ϕ, φ) parameterize the discriminator. Also, $u = \phi(\zeta)$, $\zeta \in \mathcal{D}_{\text{real}}$ and $\tilde{u} = \phi(\tilde{\zeta})$, $\tilde{\zeta} \in \mathcal{D}_{\text{syn}}$ are the encoded samples from the real and synthetic datasets. Note that, in (8), as we project the encoded datasets to a n -dimensional space (n can be different from d), $\omega \in \mathbb{S}^{n-1}$. In practice, (8)-(9) is implemented using mini-batches of the real and synthetic datasets. We provide the outline of CryoSWD in Algorithm 1.

4. NUMERICAL RESULTS

In our experiments, we use four different 3D biomolecular maps and generate each from their protein sequences in Chimera [23]. The generated maps are then pre-processed and zero-padded to the size of $128 \times 128 \times 128$. The four maps alongside their protein data bank (PDB) IDs are: Holliday junction complex (5j0n), SARS-CoV-2 RNA-dependent RNA polymerase (7btf), Human patched 1 (6oeu) and pre-catalytic spliceosome (5nrl). Throughout the rest of this draft, we refer to these maps as HJC, Sars, HP1 and Splice. We evaluate CryoSWD on synthetic datasets with $L = 20k$

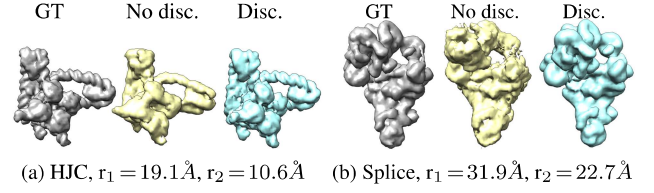


Fig. 2. Effect of the discriminator. r_1 and r_2 are the resolutions of the reconstructed 3D maps, without and with a discriminator, respectively. Here the 3D maps are of dimension $64 \times 64 \times 64$ and the voxel sizes for HJC and Splice are $4\text{\AA}$ and $7.4\text{\AA}$.

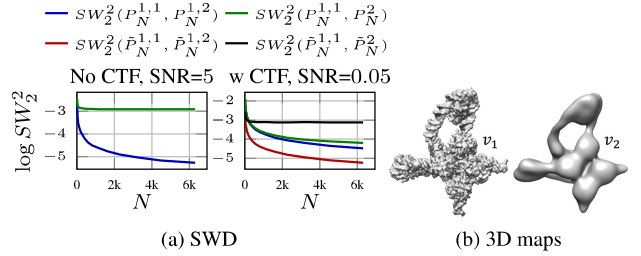


Fig. 3. (a) SWD between the distribution of the synthetic datasets with varying sizes (N) induced by two 3D maps v_1 and its smoothed version v_2 . $\tilde{P}$ denotes the distribution of the encoded dataset by ϕ . (b) v_1 and v_2 maps, used to generate the datasets in (a).

projection images, synthesized following (1). We experiment on datasets with and without the CTF effect. In this draft, we do not consider the effect of image in-plane translations. To generate datasets with CTF, we consider four CTF groups with defocus levels varying in $[1, 2]$ microns range. We specifically study two settings: (1) high SNR ($\text{SNR} = 5$) with no CTF, (2) low SNR ($\text{SNR} = 0.05$) with CTF. In datasets with CTF and without CTF, the size of the projection images are 135×135 and 128×128 , respectively. Examples of projection images under both settings are provided in Fig. 1.

We assess the quality of the estimated maps using Fourier shell correlation (FSC). FSC measures the normalized cross-correlation between two 3D maps in Fourier domain at various radial frequency shells [24]. To determine the resolution of the results, we compute the FSC between the reconstruction and the ground truth (GT) and use $\text{FSC} = 0.5$ criterion.

We compare CryoSWD against CryoGAN¹ [1]. We fix the same critic/discriminator architecture and initialize the 3D maps with zeros for both methods. The number of discriminator updates per 3D map updates is $n_{\text{disc}} = 4$ and we train both models for 80 epochs. We choose a batch size of $B = 32$ and use Adam optimizer. We keep identical configurations between the CryoGAN and CryoSWD experiments except for a few parameters. In all CryoGAN experiments, following [1] default setting, the discriminator and the 3D maps learning rate is $\alpha = 0.001$, however in CryoSWD, we select the learning rate from $\alpha \in \{5 \times 10^{-5}, 10^{-4}\}$. Also, we

¹Our implementation of CryoSWD and CryoGAN experiments are based on the public CryoGAN [1] repository.

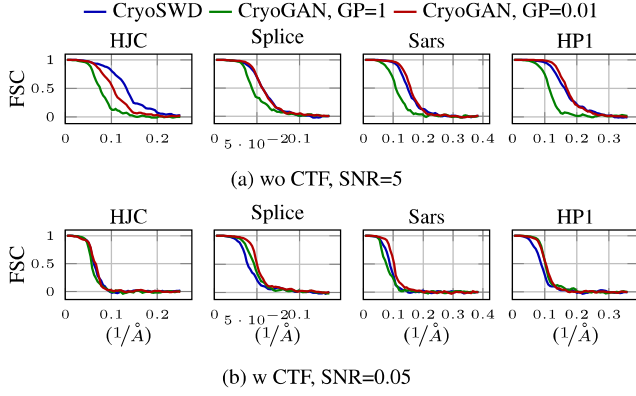


Fig. 4. FSC between the GT and reconstructed 3D maps. The horizontal axis is spatial resolution ($1/\text{\AA}$).

use $N_\Omega = 3 \times 10^4$ and generate $\{\omega_k\}_{k=1}^{N_\Omega}$ by drawing standard normal samples and normalizing them to unit norm.

CryoSWD without a Discriminator: In Fig. 2, we visualize the results of CryoSWD for HJC and Splice datasets, with and without the use of the discriminator. With no discriminator, we optimize the loss function (7) to recover the 3D map and set $B = 500$ and $\alpha = 0.05$. For this experiment, we set $\text{SNR} = 1000$ with no CTF. We observe that, even without the use of the discriminator, minimizing SW_2 in (7), leads to a reconstruction that is close to the overall shape of the GT. However, using a discriminator and an encoder that maps the projection images to an adversarially learnt embedding space helps in reconstructing finer details of the GT.

We noticed that, without a discriminator the reconstruction fails in higher noise regimes. To investigate this, we consider two distributions, $P_1 = P_{\text{syn}}(v_1)$ and $P_2 = P_{\text{syn}}(v_2)$, where v_2 is a smoother version of v_1 . We now define three sets of samples: $\mathcal{S}_N^{1,1}$, $\mathcal{S}_N^{1,2}$ and $\mathcal{S}_N^{2,2}$ each containing N independent samples drawn from P_1 , P_1 and P_2 . In Fig. 3 (green and blue curves), we plot SW_2^2 computed from (7), between the distributions of these sets denoted by $P_N^{1,1}$, $P_N^{1,2}$ and $P_N^{2,2}$. As $\mathcal{S}_N^{1,1}$ and $\mathcal{S}_N^{1,2}$ are sampled from the same distribution, they have a small SWD. Also, in high SNR regime, $P_N^{1,2}$ and $P_N^{2,2}$ are more distant, especially as N grows. However, with increased levels of noise ($\text{SNR} = 0.05$), SWD between the two shrinks and approaches $SW_2(\mathcal{S}_N^{1,1}, \mathcal{S}_N^{1,2})$, despite the differences between v_1 and v_2 . This highlights the difficulty in minimizing SWD in (7) when noise intensifies. On the contrary, if we train a discriminator, the SW_2 distance between the embedded samples of the different datasets (i.e. $\mathcal{S}_N^{1,1}$ and $\mathcal{S}_N^{2,2}$, black and red curves in Fig. 3-(a)) would be significantly larger than the SW_2 distance of the same datasets coming from the same distribution P_1 (blue and green curves in Fig. 3-(a)). This implies that encoding the noisy projections to an adversarially learned embedding space, further separates the two datasets and can lead to better reconstructions.

Qualitative and Quantitative Results: In Fig. 4-5, we compare CryoSWD against CryoGAN. We noticed that the performance of CryoGAN, especially in low noise settings, is

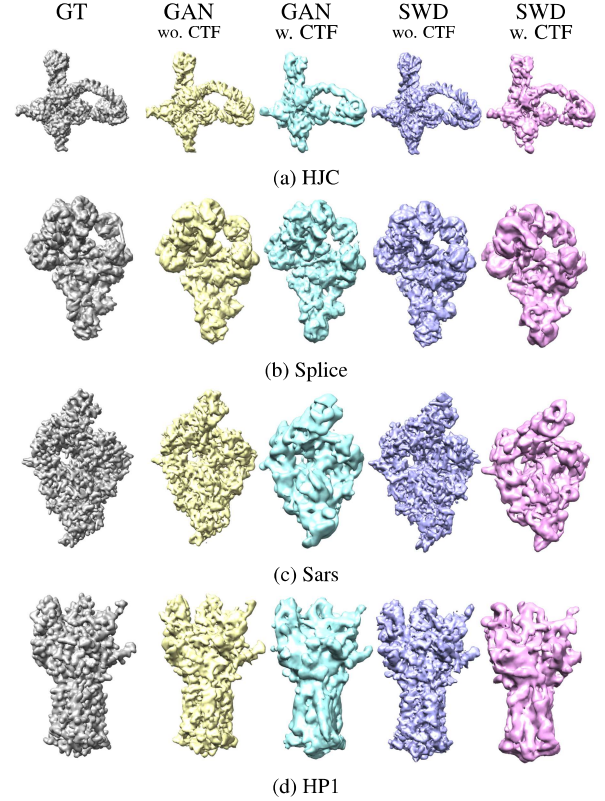


Fig. 5. Visualization of the reconstructed volumes. Gray: GT map, Yellow: CryoGAN [1], no CTF, $\text{SNR}=5$, Cyan: CryoGAN [1], with CTF, $\text{SNR}=0.05$, Purple: CryoSWD, no CTF, $\text{SNR}=5$, Pink: CryoSWD, with CTF, $\text{SNR}=0.05$.

sensitive to the choice of gradient penalty (GP) weight. After tuning GP weight, we observe that the performance of CryoGAN and CryoSWD are close for most datasets. However, unlike CryoGAN, CryoSWD does not need GP term to enforce training stability and in high SNR regimes allows a reasonable reconstruction without training a discriminator.

5. CONCLUSION & FUTURE WORK

In this paper, we proposed to use sliced Wasserstein distance (SWD) as a reconstruction criterion in single-particle cryo-EM. Specifically, we replaced Wassersten-1 with SWD in the CryoGAN pipeline. CryoSWD, similar to CryoGAN, solves SPR in a distribution matching sense. In low noise settings, CryoSWD recovers the 3D maps without the need to train a discriminator. When coupled with a discriminator, CryoSWD results further improve in both low and high noise settings.

In this draft, following CryoGAN, we assumed the distribution of the projection views are known a-priori and uniform. While having a uniform prior empirically works well [1], to better match the projection image dataset distribution, it is important to take into account the recovery of the projection view distribution. Furthermore, we observed that CryoSWD with no discriminator fails with intensified noise. It would be interesting to investigate the effect of multi-scale reconstruction on making CryoSWD more robust to noise.

6. REFERENCES

- [1] H. Gupta, M. T. McCann, L. Donati, and M. Unser, "Cryogan: A new reconstruction paradigm for single-particle cryo-em via deep adversarial learning," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 759–774, 2021.
- [2] A. Singer and F. J. Sigworth, "Computational methods for single-particle electron cryomicroscopy," *Annual Review of Biomedical Data Science*, vol. 3, p. 163, 2020.
- [3] J. Frank, *Three-dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state*. Oxford university press, 2006.
- [4] E. Levin, T. Bendory, N. Boumal, J. Kileel, and A. Singer, "3d ab initio modeling in cryo-em by autocorrelation analysis," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 1569–1573, 2018.
- [5] M. Zehni, L. Donati, E. Soubies, Z. Zhao, and M. Unser, "Joint angular refinement and reconstruction for single-particle cryo-em," *IEEE Transactions on Image Processing*, vol. 29, pp. 6151–6163, 2020.
- [6] A. Singer, R. R. Coifman, F. J. Sigworth, D. W. Chester, and Y. Shkolnisky, "Detecting consistent common lines in cryo-em by voting," *Journal of Structural Biology*, vol. 169, no. 3, pp. 312–322, 2010.
- [7] P. A. Penczek, R. Renka, and H. Schomberg, "Gridding-based direct fourier inversion of the three-dimensional ray transform," *Journal of the Optical Society of America A*, vol. 21, pp. 499–509, Apr 2004.
- [8] L. Donati, M. Nilchian, C. O. S. Sorzano, and M. Unser, "Fast multiscale reconstruction for cryo-em," *Journal of Structural Biology*, vol. 204, no. 3, pp. 543–554, 2018.
- [9] F. J. Sigworth, "Principles of cryo-EM single-particle image processing," *Microscopy*, vol. 65, pp. 57–67, 12 2015.
- [10] A. Barnett, L. Greengard, A. Pataki, and M. Spivak, "Rapid solution of the cryo-EM reconstruction problem by frequency marching," *SIAM Journal on Imaging Sciences*, vol. 10, no. 3, pp. 1170–1195, 2017.
- [11] 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.
- [12] A. Punjani, J. L. Rubinstein, D. J. Fleet, and M. A. Brubaker, "cryosparc: algorithms for rapid unsupervised cryo-em structure determination," *Nature methods*, vol. 14, no. 3, p. 290, 2017.
- [13] Z. Kam and I. Gafni, "Three-dimensional reconstruction of the shape of human wart virus using spatial correlations," *Ultramicroscopy*, vol. 17, no. 3, pp. 251 – 262, 1985.
- [14] S. Huang, M. Zehni, I. Dokmanić, and Z. Zhao, "Orthogonal Matrix Retrieval with Spatial Consensus for 3D Unknown-View Tomography," *arXiv e-prints*, July 2022.
- [15] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proceedings of the 34th International Conference on Machine Learning*, vol. 70 of *Proceedings of Machine Learning Research*, pp. 214–223, Aug. 2017.
- [16] V. M. Panaretos and Y. Zemel, "Statistical aspects of wasserstein distances," *Annual Review of Statistics and Its Application*, vol. 6, no. 1, pp. 405–431, 2019.
- [17] I. Deshpande, Z. Zhang, and A. G. Schwing, "Generative modeling using the sliced wasserstein distance," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [18] J. Wu, Z. Huang, D. Acharya, W. Li, J. Thoma, D. P. Paudel, and L. V. Gool, "Sliced wasserstein generative models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3713–3722, 2019.
- [19] I. Deshpande, Y.-T. Hu, R. Sun, A. Pyrros, N. Siddiqui, S. Koyejo, Z. Zhao, D. Forsyth, and A. G. Schwing, "Max-sliced wasserstein distance and its use for gans," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 10648–10656, 2019.
- [20] S. Kolouri, S. R. Park, M. Thorpe, D. Slepcev, and G. K. Rohde, "Optimal mass transport: Signal processing and machine-learning applications," *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 43–59, 2017.
- [21] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," *Advances in neural information processing systems*, vol. 30, 2017.
- [22] C. Villani, *Optimal Transport: Old and New*. Springer Science & Business Media, 2008.
- [23] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, "Ucsf chimera—a visualization system for exploratory research and analysis," *Journal of Computational Chemistry*, vol. 25, no. 13, pp. 1605–1612, 2004.
- [24] G. Harauz and M. van Heel, "Exact filters for general geometry three dimensional reconstruction," *Optik*, vol. 73, pp. 146–156, 1986.