

CryoDRGN: Deep Generative Models for Reconstructing Heterogeneous 3D Structures from Cryo-electron Micrographs

Ellen Zhong, Tristan Bepler, Bonnie Berger and Joseph Davis

Massachusetts Institute of Technology, Cambridge, Massachusetts, United States

Single particle cryo-electron microscopy (cryo-EM) has been successfully used to determine the structures of large multidomain complexes [1]. In this technique, a purified solution of a target macromolecular complex is fixed in a thin layer of vitreous ice and imaged with a transmission electron microscope. As each image contains a snapshot of a unique molecule, the technique has the ability to uncover the equilibrium conformational ensemble relevant to the molecule of interest. However, structural heterogeneity complicates 3D reconstruction, a computational task wherein 10^{4-7} noisy, randomly oriented 2D images are combined into a 3D structure. A standard technique for heterogeneous reconstruction is 3D classification, where heterogeneity is modeled as originating from a discrete mixture model of independent structures [2]. However, this discrete approach can fail under more complex mixtures ($> > 10$ states) or continuous forms of structural heterogeneity.

Here, we introduce a novel machine learning based reconstruction algorithm for heterogeneous cryo-EM reconstruction [3]. We develop a deep generative model for 3D structure from a continuous low dimension latent space, which is trained using an encoder-decoder architecture. Our model for 3D structure predicts electron density as a function of Cartesian coordinates and the latent variable. Once trained, the encoding neural network maps individual particles to a low-dimensional latent space, which can provide insight into the degree of structural heterogeneity in the dataset. Empirically, we find that the latent space encoding can also be used to identify and filter out-of-distribution junk particles from the dataset. The decoding neural network can then be used to reconstruct 3D density maps from the learned latent space and thus generate an ensemble of structures, including generation of continuous trajectories.

This approach, termed cryoDRGN (Deep Reconstructing Generative Networks) represents the first application of deep neural networks to the cryo-EM reconstruction task. We demonstrate the efficacy of this tool in reconstructing both discretely and continuously heterogeneous structures from simulated datasets. On publicly available datasets, we find that this tool is able to reveal residual heterogeneity in high resolution datasets, uncover compositionally diverse structures without a priori specification of the number of states, and model continuous conformational changes of large macromolecular machines.

References

- [1] Cheng, Y. (2018). Single-particle cryo-EM—How did it get here and where will it go. *Science*, 361(6405), 876–880. <http://doi.org/10.1126/science.aat4346>
- [2] Scheres, S. H. W., Valle, M., Nuñez, R., Sorzano, C. O. S., Marabini, R., Herman, G. T., & Carazo, J.-M. (2005). Maximum-likelihood Multi-reference Refinement for Electron Microscopy Images. *Journal of Molecular Biology*, 348(1), 139–149. <http://doi.org/10.1016/j.jmb.2005.02.031>
- [3] Zhong, E. D., Bepler, T., Davis, J. H., & Berger, B. (2019, September 11). Reconstructing continuous distributions of 3D protein structure from cryo-EM images. *arXiv.org*.

