CryoAI: Amortized Inference of Poses for *Ab Initio* Reconstruction in Cryo-EM

A. Levy[1,2], F. Poitevin[2], J. Martel[1], Y. Nashed[2], A. Peck[2], N. Miolane[3], D. Ratner[2], M. Dunne[2], G. Wetzstein[1]

1: Stanford University – 2: SLAC National Laboratory – 3: UC Santa Barbara

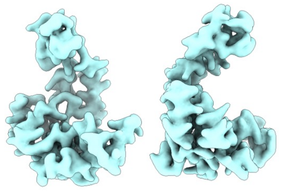
Introduction

Cryogenic Electron Microscopy (cryo-EM) enables biologists to understand the shapes, and thereby the functions, of biomolecules like proteins. In a cryo-EM experiment, the same molecule is imaged from a set of unknow viewing angles (poses). The pose for each image must be jointly inferred with the 3D structure of the molecule.

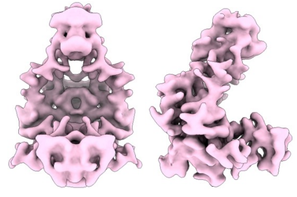
CryoAI learns to map images to their associated pose via a shared encoder, thereby avoiding the expensive "orientation matching step" and amortizing the reconstruction over the size of the dataset. It is the first scalable approach to work with large modern cryo-EM datasets.

Ablation Study

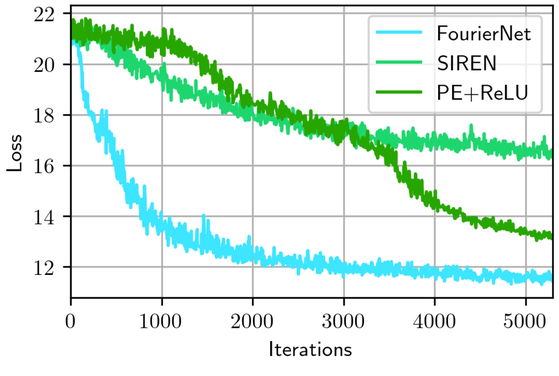
Symmetric Loss



L2 Loss



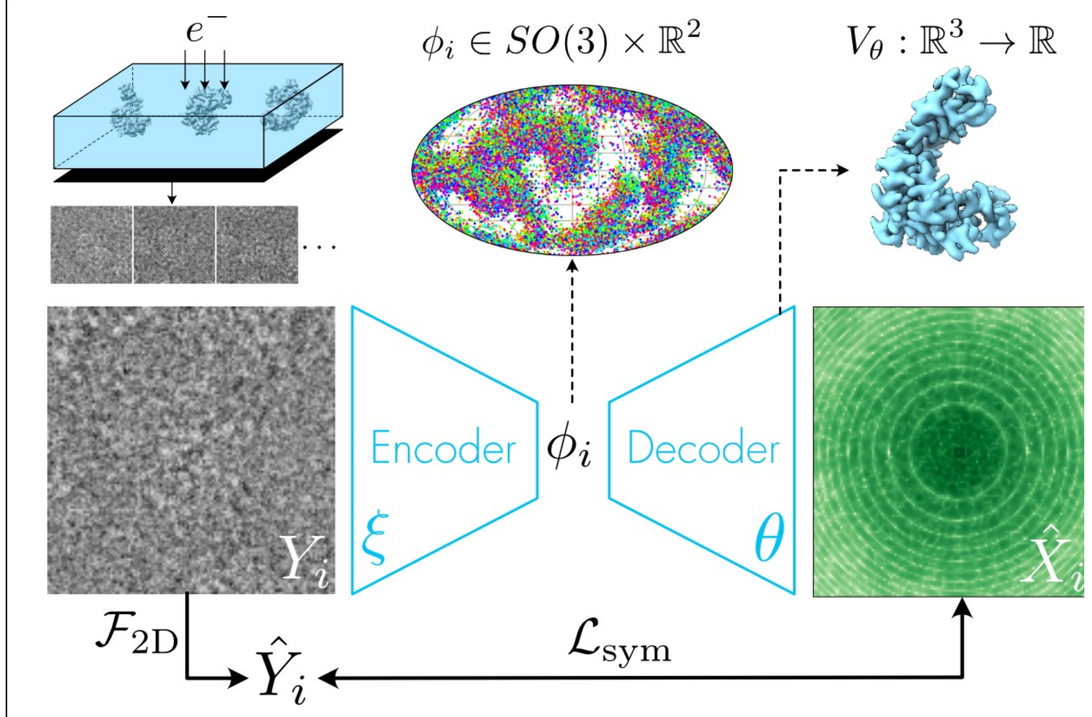
Specialized Neural Representation



The symmetric loss prevents the model from getting stuck in symmetrical equilibrium states, due to the handedness ambiguity.

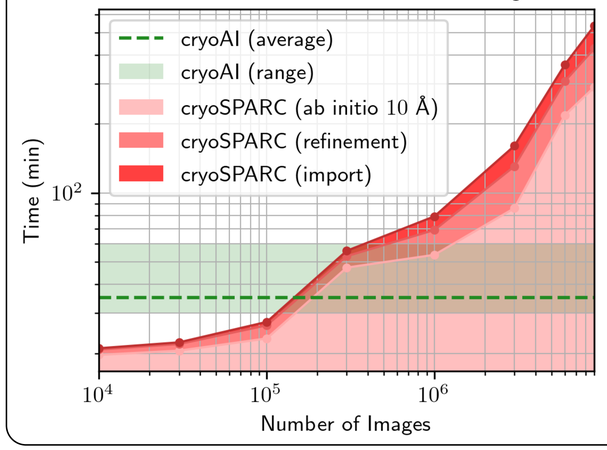
The FourierNet is tailored to approximate functions in Fourier space and speeds up convergence.

Overview

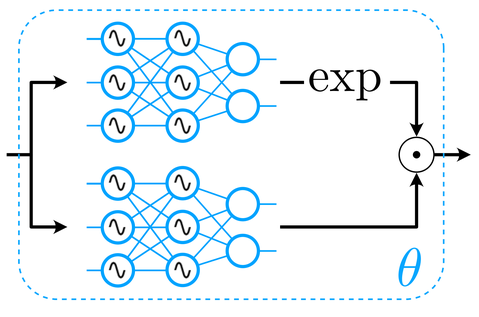


Overview of our method. CryoAI performs volume reconstruction and pose estimation with an auto-encoder architecture. It takes noisy images Y_i as input and outputs a noise-free version $\hat{X}_i$ of the same image, in Fourier space. The encoder ξ maps images to a low dimensional feature ϕ_i, interpreted as a pose by the decoder. The decoder contains a neural representation V_θ (FourieNet) of the volume and uses the image formation model of cryo-EM to predict $\hat{X}_i$, which is then compared with $\hat{Y}_i$ using the "symmetric" loss.

Runtime vs. Number of Images



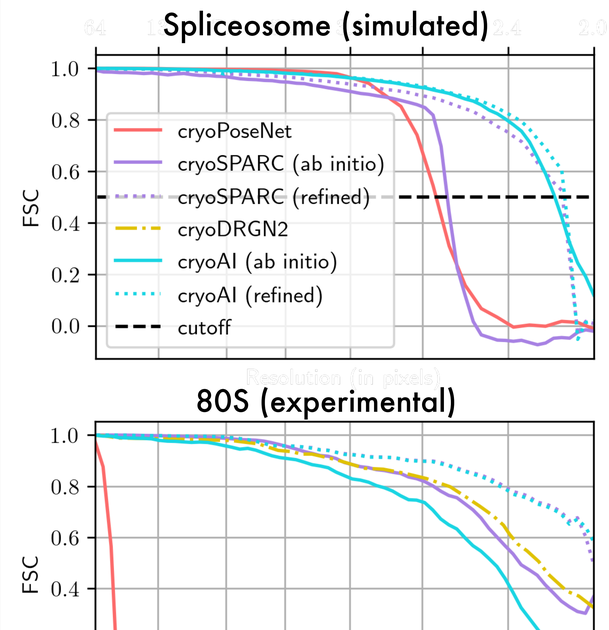
Architecture of a FourierNet



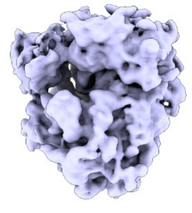
The neural network followed by an **exp** function accounts for the large amplitude variations in Fourier space.

Real and Simulated Datasets

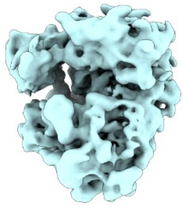
Spliceosome (simulated)



80S (experimental)



CryoSPARC (*ab initio*)



CryoAI

More results here

- Punjani et al. (2017) doi.org/10.1038/nmeth.4169
- Zhong et al. (2021) *CryoDRGN2*, ICCV
- Nashed et al. (2021) *CryoPoseNet*, ICCV Workshops

This work was supported by the U.S. Department of Energy, under DOE Contract No. DE-AC02-76SF00515. We acknowledge the use of the computational resources at the SLAC Shared Scientific Data Facility (SDF).