# CryoSphere: SO(3)-equivariant method for cryo-EM pose estimation

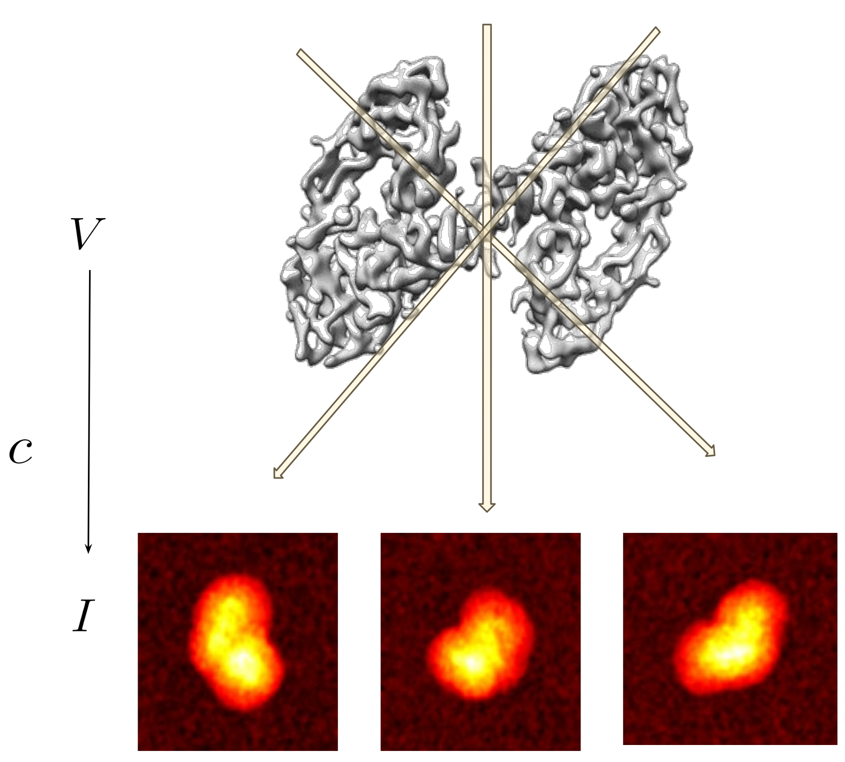
Dev Chheda, Divya Nori
dchheda@mit.edu, divnor80@mit.edu

## Introduction

- Cryo-EM is the state-of-the-art technique to determine the 3D structure of proteins at high resolution
- The output of a single cryo-EM run is $10^4$ to $10^7$ noisy projection images, all taken from unknown random orientations
- Here, we focus on the task of **3D pose estimation of protein electron density volume from projection images** to investigate the following:

*Can we reconstruct protein structure using fewer projection images by exploiting symmetries?*

- The homogenous, single-particle setting treats a protein as volumetric map $V$, and projection operator $c$ maps $V$ to image $I$




$$I(r_x, r_y) = g * \int_{\mathbb{R}} V(R\mathbf{r})\, dr_z + \epsilon$$

$$\mathbf{r} = \begin{bmatrix} r_x \\ r_y \\ r_z \end{bmatrix} \quad \begin{aligned} V &: \mathbb{R}^3 \to \mathbb{R} = \text{volumetric map} \\ I &: \mathbb{R}^2 \to \mathbb{R} = \text{image} \\ c &: V \mapsto I = \text{projector} \end{aligned}$$

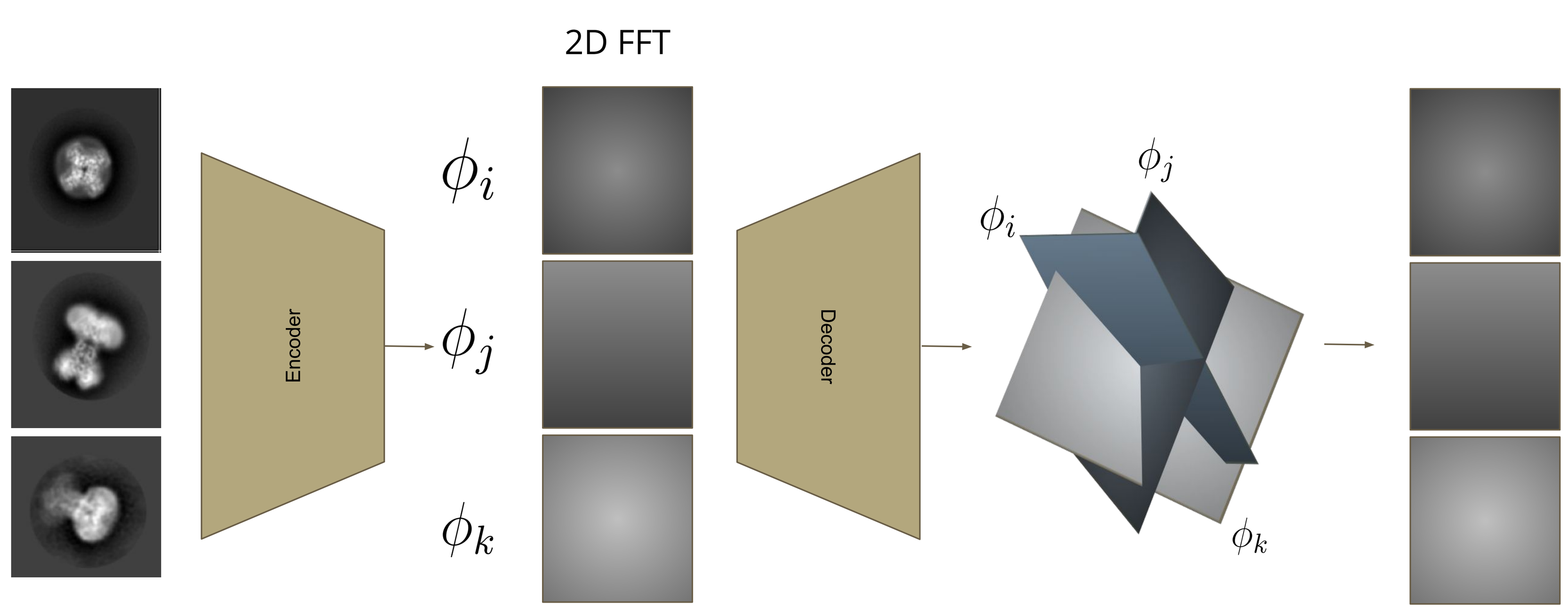
$$R \in SO(3) = \text{unknown pose}$$

$$g = \text{microscope blurring}$$

$$\epsilon = \text{Gaussian noise}$$

## Related Work

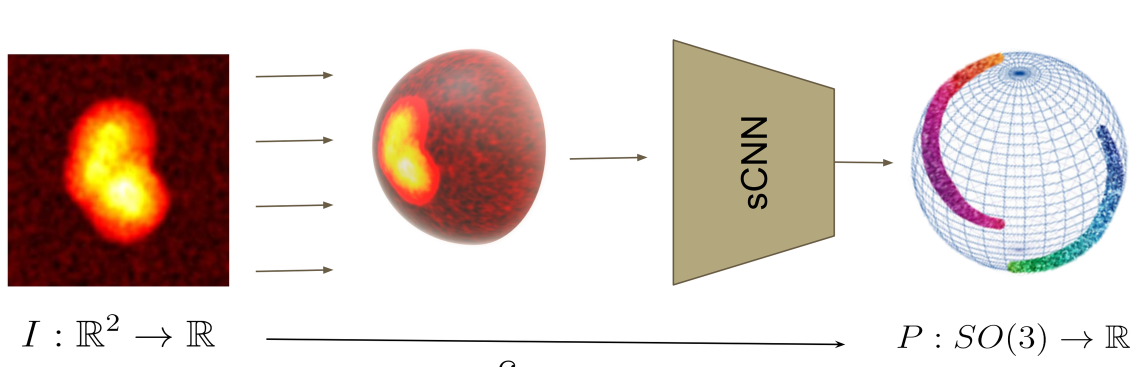
- Current state-of-the-art is **CryoFIRE** which employs an image-to-pose encoder and pose-to-slice decoder, reconstructing the 3D volume using Fourier Slice Theorem
- The model is trained using a symmetric loss between projections of the 3D model at predicted orientations and ground truth projections




- Notably, CryoFIRE predicts a single orientation for every projection image, but we aim to instead learn a more complex representation of pose from each image
- The representation we learn is a probability distribution over orientations in SO(3), as done in prior work **Image2Sphere**
- Image2Sphere exploits 3D object symmetry for pose prediction by projecting images onto a half-sphere and performing spherical convolutions
- The model is trained in a supervised manner using cross entropy loss, inspiring our training protocol

## Background & Workflow

- Our general workflow involves projecting input images onto a half-sphere and training a spherical convolutional neural network to predict a probability distribution over SO(3)
- Similar to previous work, our encoder is **approximately SO(3) equivariant**




$$s : I \mapsto P = \text{encoder}$$

$$s \circ c : V \mapsto P$$

**Equivariance condition:**

$$\boxed{(s \circ c)(RV) = R(s \circ c)(V)}$$

- At inference time, the sCNN encoder is used to predict a probability distribution for each input image which can be used for reconstruction with Fourier Slice Theorem
- Fundamentally, cryo-EM images are highly noisy and the proteins we aim to learn symmetries on are complex, requiring specific modifications to the training regime

## Methods

Data & Label Generation → Model Building → Model Training

### Data & Label Generation

*Synthetic Images*

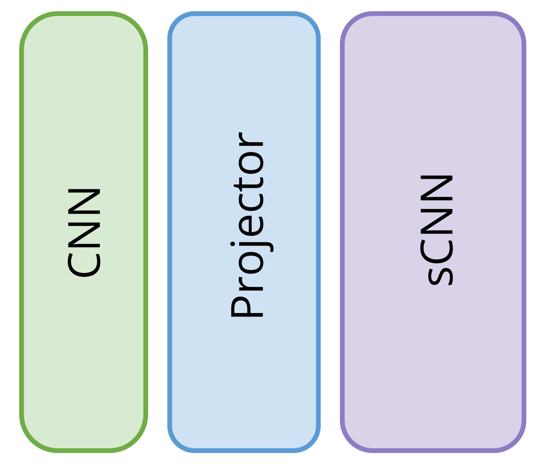
- Using ground truth 3D electron density map (from PDB), performed orthographic projection from *n* orientations
- Added Gaussian noise and blur

*Training Labels*

- For each data point, initialized a grid over SO(3) with bin width 15 degrees
- Each grid cell contains MSE with respect to ground truth image from that pose
- Generated binary mask over spherical grid where 1 represents the largest 5% of 1/MSE values

### Model Building

*Architecture*




- Featurized 2D projection image with 2 Conv2d layers and ReLU activation
- Project featurization onto half-sphere using depthwise convolution
- Learned spherical signal over SO(3) with S2 convolution, SO(3) activation, and SO(3) convolution

### Model Training

*Loss Function*

- Used a Binary Cross Entropy Loss between the predicted probability map and ground truth
- Ground truth probability mask generated by

$$\Pr[R \mid I] = e^{-\lambda \mathrm{MSE}(I, c(RV))^2}$$

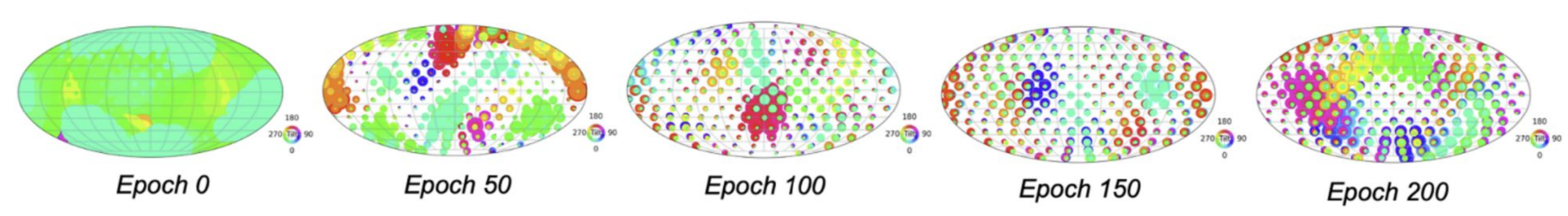
before bitmasking

- $\lambda$ tuned to produce separation between similar and dissimilar poses to encourage learning of symmetries

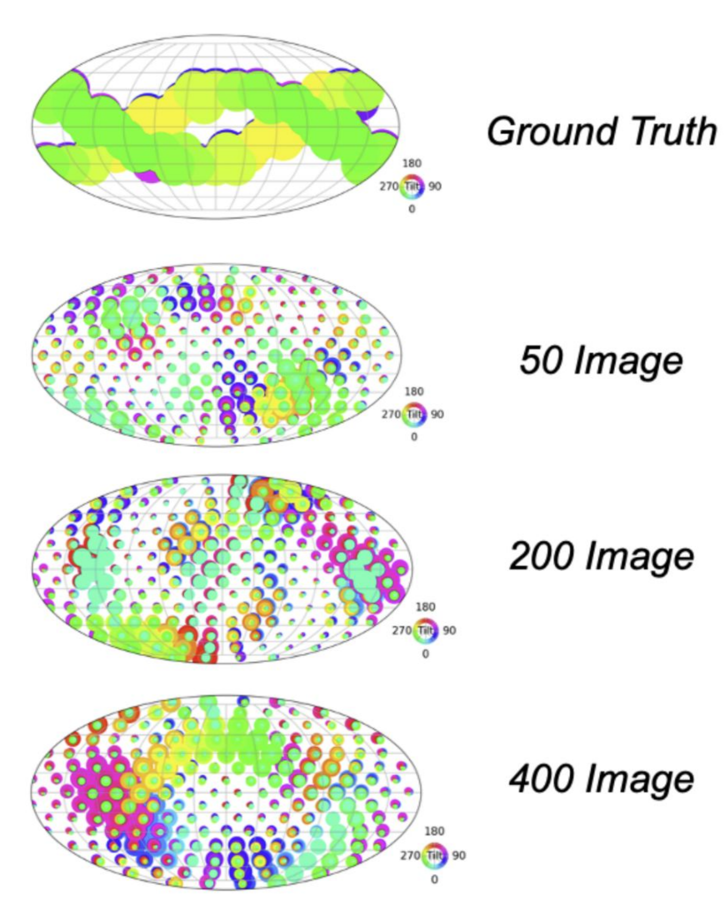
*Hyperparameters*

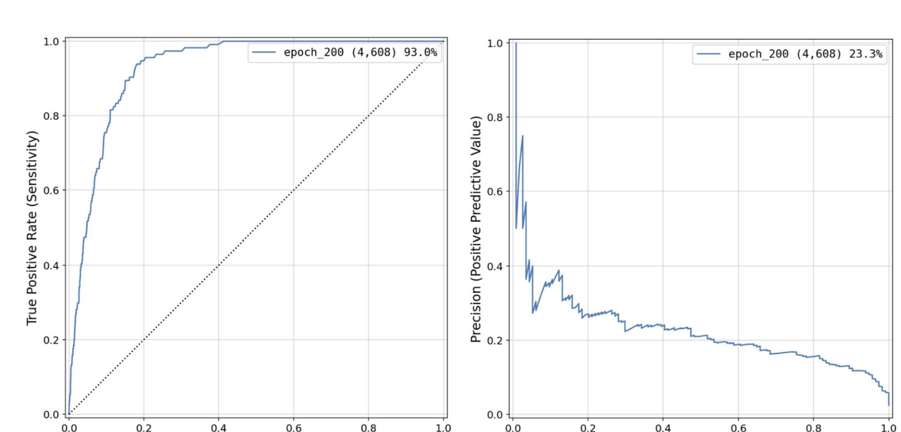
- Model trained for 200 epochs with learning rate of 0.001
- Spherical equivariant layers use lmax = 4

## Experiments




- We trained the model using different numbers of projections (50 to 400)
- The model trained on 400 images for 200 epochs learns the training data and generalizes best
- We quantitatively tested for equivariance, and given that both mean and maximum MSE is considerably lower with our model predictions, our model learns an approximately equivariant function




Training ROC: AUC = 0.930

Training PR: AUC = 0.233

| | Mean | Min | Max |
|---|---|---|---|
| Our Model | **.0050** | **.0024** | **.0075** |
| Random Baseline | .0074 | .0025 | .0144 |

MSE on equivariance tests

## Conclusions

- Equivariant methods can be used to learn protein symmetries from 2D images, as captured by the relation between rotations before imaging and rotations across the probability map
- Although learning against a ground truth probability map is unstable, this can be rectified by using binary masks with carefully chosen thresholds

*Limitations & Future Work*

- Synthetic projection images may provide biased benchmarks; difficult to test on real cryo-EM data (petabytes of storage)
- May not generalize to proteins with symmetries not seen during training
- Integrate pose model with reconstruction using Fourier Slice Theorem
- Increase the resolution used for both images and the spherical CNN to improve the granularity of predicted probability maps and (eventually) reconstruction

## References

[1] **CryoFIRE**: Levy, A. & Wetzstein, G. & Martel, J. & Poitevin, F. & Zhong, E. (2022). Amortized Inference for Heterogeneous Reconstruction in Cryo-EM. *Neural Information Processing Systems*.

[2] **Image2Sphere**: Klee, D. & Biza, O. & Platt, R. & Walters, R. (2023). Image to Sphere: Learning Equivariant Features for Efficient Pose Prediction. *International Conference on Learning Representations*.