



THE UNIVERSITY OF
CHICAGO

DEPARTMENT OF STATISTICS

Master's Thesis Presentation

Ke Shao

Department of Statistics
The University of Chicago

“Deep Generative Modeling for Cryo-EM”

Thursday, February 15, 2024, at 2:00 PM
Jones 111, 5747 S. Ellis Avenue

Abstract

Cryo-electron microscopy (Cryo-EM) has become a crucial instrument in structural biology for understanding protein structures. This method enables the imaging of proteins in a solution by directing an electron beam through the protein molecule onto a detector. The challenge lies in reconstructing the volume from an assortment of molecular “shadows” captured under various unknown rotations. A range of deep learning algorithms tailored to Cryo-EM images, such as CryoDRGN, CryoAI, CryoFIRE, have significantly advanced our capability to reconstruct 3D volumes from these images. Concurrently, other techniques based on different data types, like SMILES and sequences, have shown promising results in predicting protein folding. However, there are limited attempts to leverage multiple data types together. In this presentation, we’ll start by examining the latest advancements in volume reconstruction from CryoEM images. We’ll then delve into the challenge of integrating different types of data. First, we’ll explore how recent research by Ryan Tibshirani can be adapted to merge multiple data types for our specific problem. Following this, we’ll look at another strategy that incorporates AlphaFold predictions to enhance protein reconstruction.