DeepCryoPicker

An unsupervised learning approach for fully automated single particle picking in Cryo-EM images

Zachary M. Lipperd, Jianlin Cheng

Project Overview

The advancement of cryogenic electron microscopy (cryo-EM) technology has stimulated a revolution in structural biology of studying large protein complexes and assemblies that were unable to be well studied before. However, Computational reconstruction of these protein structures from cryo-EM image data remains a time-consuming, labor-intensive, error-prone, and often inaccurate process.

Difficulties in creating these computational models come from three major areas:

- Bottleneck in picking protein particles in cryo-EM images
- Substantial noise in 3D cryo-EM density maps generated from particle images,
- lack of automated and accurate methods to build protein structures from density maps.

To address the issue of picking protein particles in cryo-EM images, the goal of DeepCryoPicker is to develop 2D transformer networks built on top of the attention mechanism that perform better than traditional convolutional and recurrent neural networks in image processing to pick single protein particles accurately and automatically in cryo-EM image data via a novel combination of unsupervised and supervised learning.

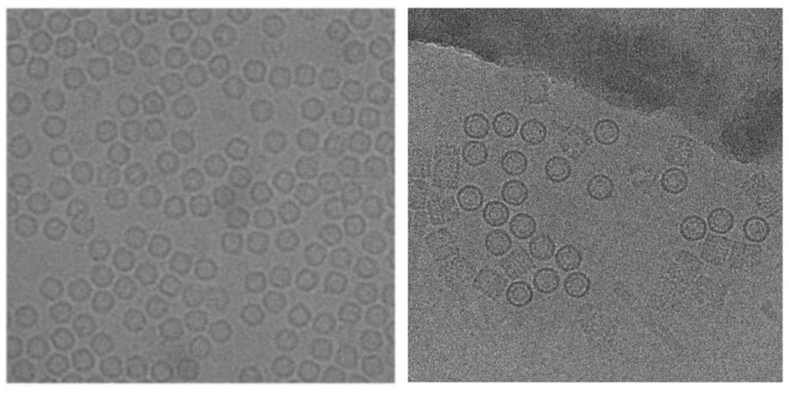
Background

Protein complexes are macromolecules essential to the functioning and well-being of all living organisms. The structure of a protein complex, in particular the region of interaction between multiple protein subunits (i.e., chains), has a notable influence on the biological function of the complex. Fundamental research in fields such as drug discovery and materials science can benefit from an enhanced understanding of these protein complexes.

We plan to develop advanced deep learning methods to reconstruct protein structures automatically and accurately from cryo-EM data, leveraging the large amount of high-resolution cryo-EM data accumulated in the field and the latest advances in the deep learning technology. The end goal is to build high-resolution full-atom structures of any protein.

Cryo-EM

Cryogenic electron microscopy (cryo-EM) has emerged as a major experimental technology to determine protein structures as it reached atomic resolution (1.2-4Å). Compared to traditional techniques (i.e., X- ray crystallography and nuclear magnetic resonance), cryo-EM has the unique capability of determining the quaternary structures of large protein complexes and assemblies that are difficult or impossible for traditional techniques to handle.



Cryo-EM micrographs contains two-dimensional projections of the particles in different orientations. Generally, cryo-EM images have low contrast, due to the similarity of the electron density of the protein to that of the surrounding solution, as well as the limited electron dose used in data collection.

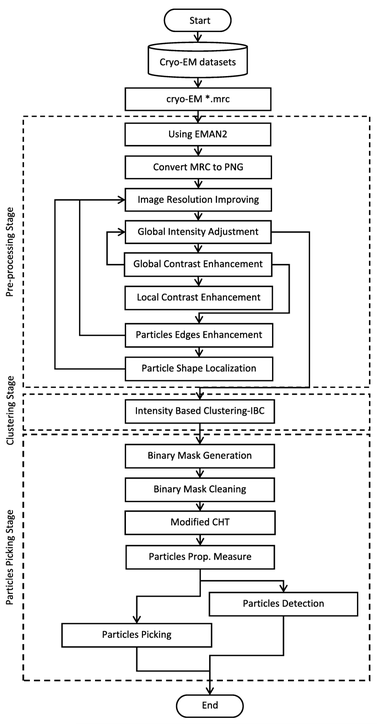
In addition, the micrographs may contain sections of ice, deformed particles, protein aggregates, etc., which can complicate particle picking. Because a large number of single-particle images must be extracted from cryo-EM micrographs to form a reliable 3D reconstruction of the underlying structure, particle recognition, represents a significant bottleneck in cryo-EM structure determination.

Particle Picking

Particle picking from 2D micrographs is a challenge due to the diversity of particle shapes and the extremely low signal-to-noise ratio of the micrographs. Human involvement is required to create a high-quality set of particles for input to the downstream structure determination steps. Previous supervised machine learning methods for particle picking require large training dataset, which requires extensive manual annotation.

Goal is to create a fully automated, unsupervised approach for single particle picking in cryo-EM micrographs.

The general framework of DeepCryoPicker: Fully Automated Single Particle Picking.



The fully automated approach has three main stages:

- Preprocessing
- Clustering
- Particle Picking

In the preprocessing stage, several image processing methods are applied to enhance the input cryo-EM images such as image normalization, Contrast Enhancement Correction (CEC), etc.

Clustering is done using an intensity-Based Clustering (IBC) that addresses some typical clustering issues such as cluster destabilization due to random initialization of cluster centers.

In the particle picking stage, a final set of particles is selected from clustered particle candidates.

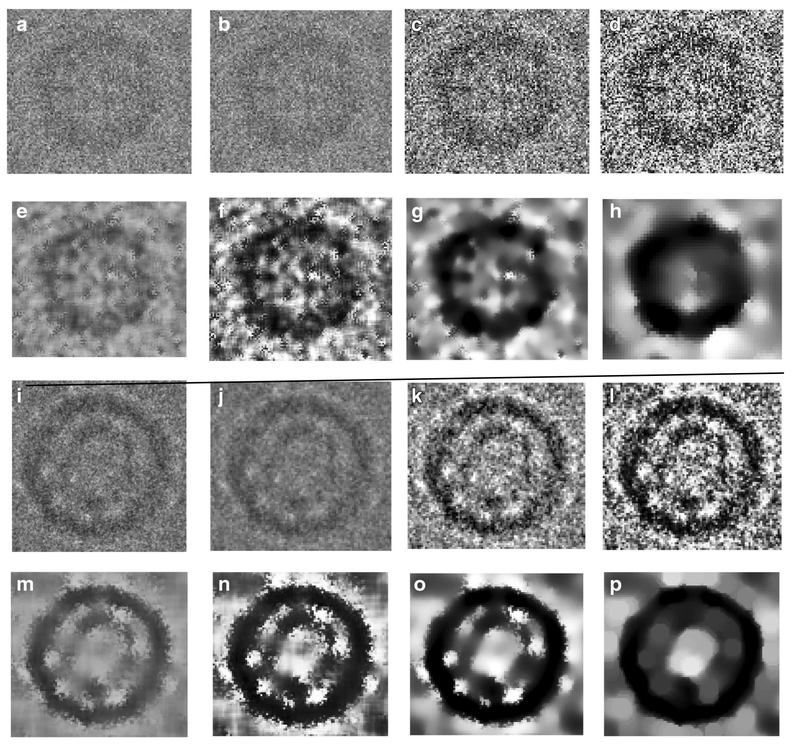
Pre-Processing

In this stage we apply various image preprocessing techniques to improve the quality of noisy cryo-EM images. There are two benefits of using the preprocessing: Firstly, improve the contrast of the cryo-EM images by increasing the particle's intensity. Secondly, pre-grouping the pixels inside each particle makes them easier to be isolated by the clustering algorithm

The preprocessing tools are selected based on three main objectives:

- Enhancing the global contrast of the cryo-EM
- Enhancing the local contrast and increasing the intensity level of each particle
- Enhancing the particle shapes inside the cryo-EM images.

Illustration of effects of the cryo-EM image analysis on a zoom-in selected particle region using two different examples. a and i) zoom-in of selected particle region in the micrograph image. b and j) normalized single particle image region. c and k) contrast enhancement correction (CEC). d and l) after applying the histogram equalization. e and m) image resonation with Wiener filtering. f and n) contrast-limited adaptive histogram equalization. g and o) image guided filtering. h and p) morphological image operation.

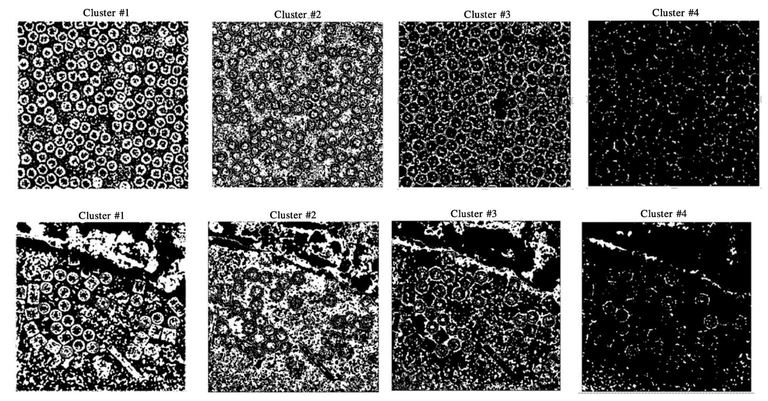


Particle Clustering

The Particle clustering method is based on an intensity distribution model which has been shown to be much faster and more accurate than traditional K-means and Fuzzy C-Means (FCM) algorithms for particle clustering.

This clustering algorithm is based on an intensity distribution model, P(i; d), which relates the intensity difference value d to the signed difference intensity values, i.

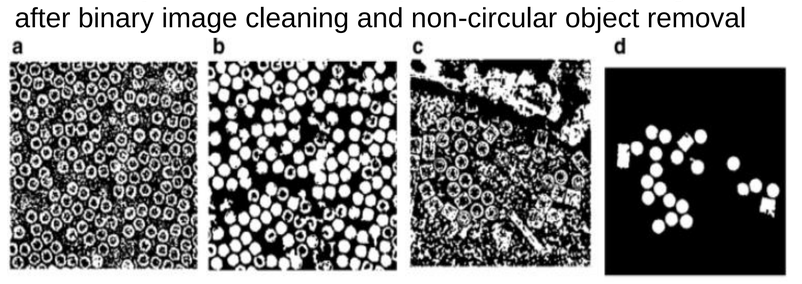
Different cryo-EM image clustering results using an Intensity-Based Clustering Algorithm (ICB)



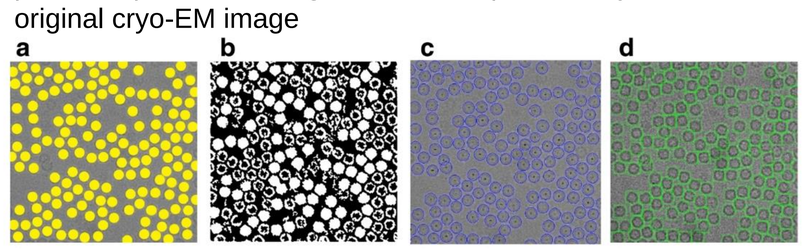
Particle Picking

Particle picking, is based on image cleaning and shape detection using a modified Circular Hough Transform algorithm to effectively detect the shape and center of each particle and create a bounding box to encapsulate the particles.

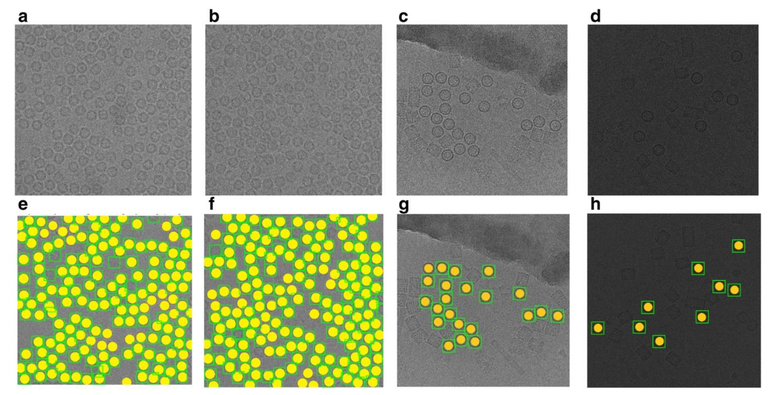
A binary mask of each cryo-EM cluster image is cleaned based on removal of the small and non-circular objects via size filtering and roundness filtering. a and c) particle clustering image before binary image cleaning and non-circular object removal. b and c) article clustering image after binary image cleaning and non-circular object removal



Particles Detection and Picking Results using Modified Circular Hough Transform (CHT). a) particles manually labelled for the cryo-EM image. b) clustering results after the binary image cleaning and non-circular objects removal. c) The center of each particle illustrated by the '+' sign and the radius of each particle by the blue circle around each particle. d) The bounding box for each particle object in the original cryo-EM image



Automated particle picking results. a) A cryo-EM image with a high identical particle density and a lack low-frequency. b) A low SNR cryo-EM image. c) A micrograph image that includes excessively overlapped particles due to confounding artifacts such as ice contamination, degraded particles, and particle aggregates. d) A micrograph that has a very low spatial density and different intensity levels. e, f, g and h) Particle picking results. Yellow is manually labelled particles. Green bounding box is result of particle picking



Electrical Engineering & Computer Science
University of Missouri

Bioinformatics and Machine Learning Lab (BML) focuses on developing machine learning, deep learning, and artificial intelligence (AI) methods to analyze big biological data and address fundamental problems in biological and medical sciences.
BML is directed by Jianlin (Jack) Cheng, PhD, William and Nancy Thompson Professor, Department of Electrical Engineering and Computer Science
Currently working on developing bioinformatics algorithms and tools for protein structure, interaction and function prediction, 3D genomics, biological network modeling, and omics data analysis.

Check out the BML's other papers on picking protein particles from cryo-EM images (which were referenced for the creation this poster)

A. Al-Azzawi, A. Ouadou, M. Highsmith, Y. Duan, J.J. Tanner, and J. Cheng. **DeepCryoPicker: Fully Automated Deep Neural Network for Single Protein Particle Picking in cryo-EM**. *BMC Bioinformatics*

A. Al-Azzawi, A. Quadou, J.J. Tanner, J. Cheng**. A super-clustering approach for fully automated single particle picking in Cryo-EM.** Genes, accepted, 2019

A. Al-Azzawi, A. Quadou, J.J. Tanner, J. Cheng. **AutoCryoPicker: an unsupervised learning approach for fully automated single particle picking in cryo-EM images.** BMC Bioinformatics, accepted, 2019.