

SI100F Project: Bio-image Processing

Date:

TAs responsible for the project:

高若谷 : gaorg@epphysics.cn

慕安 : mua@epphysics.cn

许多 : xuduo2022@shanghaitech.edu.cn

高聿翀 : gaoych@shanghaitech.edu.cn

1. Background

In recent years, electron cryo-microscopy (cryo-EM) has allowed the 3D imaging of an increasing number of biological macromolecules at resolutions sufficient for de novo atomic modelling. Among these high resolution structures most were resolved by single-particle analysis (SPA). However, identifying a sufficient number of high-quality particles from low SNR micrographs sometimes can be very tough, which may find many false positives, especially for unusually shaped and small proteins with very low contrast.

2. Project description :

The SPA dataset with 2D micrographs containing purified proteins was provided from the Electron Microscopy Public Image Archive (EMPIAR, <https://www.ebi.ac.uk/empair/>).

You need to pick particles in 2D micrographs from the provided dataset, output the coordinate of each individual protein particle and visualize the results.

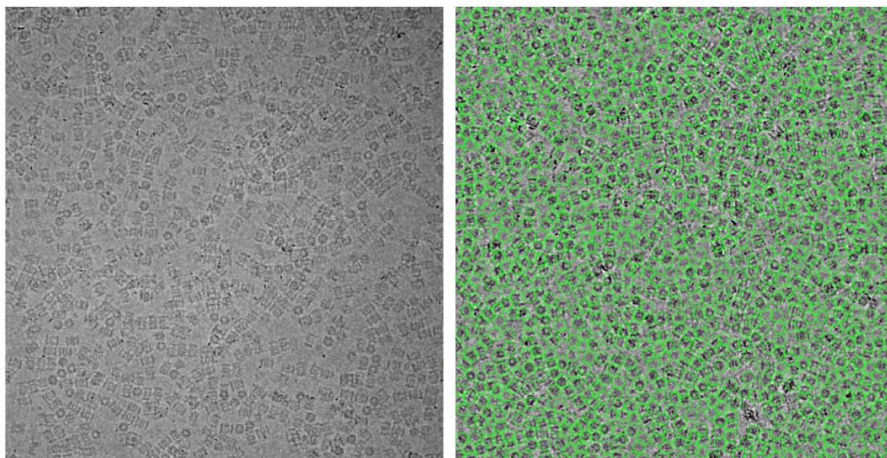


Figure 1. A raw micrograph and another micrograph with selected particle picks

In this project, you need to Finish the following tasks:

Task1: EMAN2 Binary installation.

You need to install EMAN2 Binary in your anaconda environment, either install it into an existing Anaconda setup or create a fresh environment is ok. The detailed instructions are contained in supplement material *01. EMAN2 Binary Installation Instructions.pdf*.

Task2: Manual particle picking.

You need to use GUI of EMAN2 to perform particle picking. See *02. EMAN2 Particle Picking Tutorial for SI100 2023 .pdf* for detail.

Task3: Python code running.

Run given python file. The detailed instructions are contained in supplement material *03. Brief Intro of Code and Assets for SI100 2023 v0.1.pdf*.

Task4: Python implementation of picking function.

Implement your own version of particle picking using python and apply it in real cryo-EM dataset. The detailed instructions are contained in supplement material *03. Brief Intro of Code and Assets for SI100 2023 v0.1.pdf*.

Reference:

Sjors H W S. Semi-automated selection of cryo-EM particles in RELION-1.3[J]. Journal of Structural Biology, 2015, 189(2):114-122.

Bepler T, Morin A, Rapp M, et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs[J]. Nature Methods, 2019, 16(11):1153-116

3. Project schedule :

Week 1 (May 8th - 14th): Task1 + Task2

Week 2 (May 15th - 21th): Run code of Task3

Week 3+4 (May 22th - Jun 4th): Implement picking function in Task4

Until Jun 4th 23:59 : Package submission

Jun 4th : Presentation including result demo and Q&A

4. Project evaluation :

GUI: 10 pts

Source code running: 10 pts

Programming: 10 pts

Result evaluation: 20 pts

Q&A and Demo: 10 pts

Report: 40pts

GUI, source code running and programming are basic tasks you should complete, you should select pickles correctly, successfully run source code and finish programming to obtain these scores.

In result evaluation part: We will check your result (coordinates in json files) with the standard output result. Scores of this part is quantified and the specific criteria will be published later.

In Q&A and Demo part: You should demo your code to TA in presentation and answer questions from TA.

At the last week, you need to follow the following steps to submit project package.

(Note: each person must submit the package into the system, but the member from the same group can submit the same package.)

- Submit the report as the pdf file (eg: SI100F-Group1-project.pdf) to Blackboard
- Package all the files into <eg: SI100F-Group1 -project>zip, also submit the zip file to Blackboard

Please write the report in English carefully and complete the coding tasks with Python.

Failing to submit the project package as the required format (including not writing in English) will deduct 10 points from your score.

Failing to submit the project package before the due date will receive 50% of the score. No submission will be no score for the project.

5. Content of Report

The report should contain 4 parts

Part 1: For Task 2:

1. Briefly describe the process of particle selection using GUI and your method of particle picking. (5')
2. Show the figure after picking (select 1-2 images). (5')

Part 2: For Task 3:

1. Show the figure after picking (select 1-2 images). (5')

Part 3: For Task 4:

1. Describe your implementation of particle picking. (5')
2. Show the figure after picking (select 1-2 images). (5')

Part 4: Discussion

Among all the methods you have applied on these bio-image, discuss the difference of these methods and briefly explain why. Also, any other things you confused about or found interesting can be discussed here. (15')

6. Content of Zip

The Zip should contain:

Part 1: For Task 2&3:

1. json files of the picked coordinates, save them in different folder. e.g. <Taks2><Task3>

Part 2: For Task 4:

1. Your implemented code boxer_gauss.py.
2. json files of the picked coordinates.

Pack all your results and compress them in one zip file.