A Fast Dimension Reduction Tool to Enhance Performance of 2D Clustering of Cryo-EM Images

Steve S.-C.Chung[1], Hsin-hung Lin[2], Bo-Yao Niu[1], I-Ping Tu[1]*, and Wei-Hau Chang[2]*

[1]*Institute of Statistical Science, Academia Sinica, Taiwan*

[2]*Institute of Chemistry, Academia Sinica, Taiwan*

*Correspondence: iping@stat.sinica.edu.tw; weihau@gate.sincia.edu.tw

Abstract

Thanks to rapid advances in electron microscopy instruments, it is now almost a routine to use cryo-electron microscopy images to reconstruct 3D maps of protein molecules with the details of visualizing the amino acids. However, image processing has become a high dimension problem due to the rapid increase in the number and size of images. Moreover, processing cryo-EM images has encountered unique challenges because of the presence of heavy noise in the images. In the processing pipeline, 2D clustering plays a pivotal role as it groups images of similar view to enhance the signals. In this talk, I will present our recent collaborative work on how a simple, fast and loss-less pre-processing strategy, based on a dimension-reduction method (2SDR), is used to enhance the performance of current 2D clustering algorithms. The benefits observed from tests using various cryo-EM experimental datasets include earlier convergence of a clustering process, increase in the yield of particles, and improvement in the quality of classes. Remarkably, clustering a TRPV1 dataset ion channel with the aid of our pre-processing can sift out homogeneous subset to lift the overall resolution by 0.1 to 0.2 Angstrom. Our findings suggest the 2SDR pre-processing, with light computation, is widely applicable to enhance 2D clustering algorithms of cryo-EM images.