doi:10.1017/S1431927620020450
Microsc. Microanal. 26 (Suppl 2), 2020

Particle Picking in Cryo-TEM Images Using Machine Learning

Jovany Merham, Alexander Rakowski and Joesph Patterson

University of California-Irvine, Irvine, California, United States

Cryogenic Transmission Electron Microscopy (Cryo-TEM) is a vital tool for many chemists, material scientists, and structural biologists. Regardless of the use case, the first step in Cryo-TEM analysis is image segmentation and particle picking[1]. The current method most widely used by researchers is manual segmentation. The goal during the process is to maximize the number of particles identified in a given image, while minimizing incorrect particles picked and the time spent by researchers manually picking those particles.

Image segmentation takes valuable time away from researchers which could be spent in more productive ways. In addition, it increases the risk of human bias added to the sample of particles collected. For example, one of the most challenging aspects of image segmentation is the often low contrast between the particle and the amorphous ice background, making the particles hard to distinguish by eye. Additionally, researchers may unknowingly preferentially select particles with a size or shape which fits their current hypothesis. Leading to a bias in the size and topology distribution. With increasing use of Cryo-TEM techniques including high throughput and automated experiments, more robust and consistent analysis techniques are needed for ease of production and comparison of results.

By utilizing the EMPIAR[2] database, convolutional neural networks and various statistical learning methods, we will show the development of a new algorithm to segment Cryo-TEM images. The algorithm will use various image segmentation[3,4] techniques and not just simple hierarchical neural networks, promising much faster training and more accurate predictions. This work builds on previous work done by Groschner, et.al., who used a U-Net architecture, and other advanced segmentation techniques[5]. The EMPIAR database contains a large set of Cryo-TEM images with labeled particles, providing a great training and validation dataset. And despite the bias in the available data with all the particles being only biological samples, we will present a robust enough algorithm which generalizes with ease. Because rather than training the algorithm to recognize particles, the main focus is to recognize the amorphous ice layer. This provides a universal attribute that all Cryo-TEM images have regardless of the specimen. Making the algorithm of wide applicability to chemists, material scientists, and structural biologists among many others. This new algorithm hopes to provide a new exciting tool for researchers to help accelerate and standardize data generation and processing.

References

1. Sigworth, F. J. (n.d.). Principles of cryo-EM single-particle image processing. 65(1), 57–67. https://doi.org/10.1093/jmicro/dfv370

2. A. Iudin, P.K. Korir, J. Salavert-Torres, G.J. Kleywegt & A. Patwardhan. "EMPIAR: A public archive for raw electron microscopy image data." Nature Methods 13 (2016). https://dx.doi.org/10.1038/nmeth.3806

3. Image segmentation with Mask R-CNN - Jonathan Hui - Medium. (n.d.). Retrieved February 18, 2020, from https://medium.com/@jonathan_hui/image-segmentation-with-mask-rcnnebe6d793272

4. Kayalibay, B., Jensen, G., & van der Smagt, P. (2017). CNN-based Segmentation of Medical Imaging Data. Retrieved from http://arxiv.org/abs/1701.03056

5. Groschner, C., Choi, C., Nguyen, D., Ophus, C., & Scott, M. (2019). Machine Learning for High Throughput HRTEM Analysis. Microscopy and Microanalysis, 25(S2), 150–151. https://doi.org/10.1017/s143192761900148x