Self-supervised deep learning method for in-cell cryo-electron tomography

Frosina Stojanovska[1,2], Anna Kreshuk[3], Julia Mahamid[1,3], Judith Zaugg[1,4]
[1]Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany, [2]Collaboration for Joint PhD Degree between EMBL and Heidelberg University, Heidelberg, Germany, [3]Cell Biology and Biophysics Unit, European Molecular Biology Laboratory, Heidelberg, Germany, [4]Genome Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany

Cryo-electron tomography (cryo-ET) is a powerful technique for visualizing and analyzing macromolecular complexes within their native cellular context. However, cryo-ET analysis is hindered by the low signal-to-noise ratio (SNR) inherent to cryo-ET data and the lack of ground truth data, which pose significant challenges for supervised automated mining of molecular patterns within the cellular environment. As a result, the accurate localization and identification of macromolecular structures of interest, particularly those that are small and less abundant, continue to be a major challenge in the analysis of cryo-ET data.

To address these challenges, we developed a new self-supervised deep learning approach tailored for the dense (voxel-wise) representation of in-cell cryo-ET data. This method generates high-resolution representations of cellular information at the voxel level, facilitating precise segmentation of structural details within tomograms, including particles (globular macromolecular complexes) and filaments (for example, DNA). To evaluate the performance of the model, we created an extensive simulated dataset (Purnell, 2023) that closely mimics a crowded cellular environment, featuring membranes, actin and microtubule filaments, and over 100 PDB protein structure entries of varying sizes. Additionally, we enhanced the dataset by simulating the presence of DNA structures.

Experimental results from the simulated data from the 2021 SHREC competition (Gubins, 2020) and our new simulated crowded dataset demonstrate the efficiency of our method in extracting detailed information about membranes, actin, microtubules, particles, and filaments on a voxel level. To achieve further separation of different types of particles, we conducted an experiment in which we extracted subtomograms for each detected particle and generated embedding representations for every subtomogram. This approach resulted in a new representation space, where distinct clusters were formed using unsupervised clustering, effectively separating different types of particles. The ability to distinguish and separate different types of structural information highlights the potential of our

method for advancing the analysis of complex cellular structures in cryo-ET data.

Furthermore, we applied our method to tomograms of Mycoplasma pneumoniae (O'Reilly, 2020), which capture the entire cell in a single tomogram. We successfully extracted structural information of the membrane, particles, and putative DNA filaments, creating a comprehensive 3D structural cell model.

In conclusion, our novel self-supervised deep learning approach demonstrates significant potential in overcoming challenges associated with ground truth generation and accelerating biological discoveries from cryo-ET data. By enabling accurate segmentation and extraction of macromolecular and filament information, even in cases with limited or missing annotations, our method advances the analysis of cryo-ET data. Furthermore, the proposed approach can be utilized to construct a comprehensive 3D Mycoplasma pneumoniae cell model from in situ tomograms, showcasing its potential for diverse applications in the field.

Keywords:

self-supervised deep learning, cryo-ET, segmentation

Reference:

Purnell, C., Heebner, J., Swulius, M. T., Hylton, R., Kabonick, S., Grillo, M., ... & Swulius, M. T. (2023). Rapid synthesis of cryo-ET data for training deep learning models. bioRxiv.
Gubins, I., Chaillet, M. L., van Der Schot, G., Veltkamp, R. C., Förster, F., Hao, Y., ... & Bunyak, F. (2020). SHREC 2020: Classification in cryo-electron tomograms. Computers & Graphics, 91, 279-289.
O'Reilly, F. J., Xue, L., Graziadei, A., Sinn, L., Lenz, S., Tegunov, D., ... & Rappsilber, J. (2020). In-cell architecture of an actively transcribing-translating expressome. Science, 369(6503), 554-557.