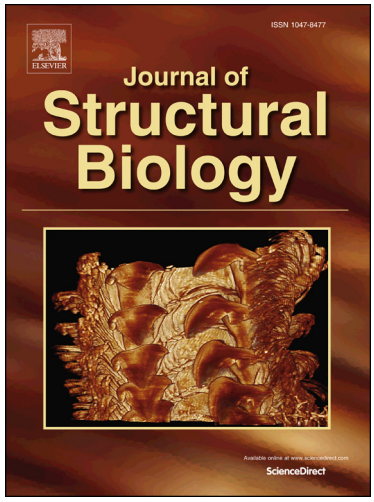

A convolutional autoencoder approach for mining features in cellular electron cryo-tomograms and weakly supervised coarse segmentation

Xiangrui Zeng, Miguel Ricardo Leung, Tzviya Zeev-Ben-Mordehai, Min Xu

PII: S1047-8477(17)30238-1
DOI: https://doi.org/10.1016/j.jsb.2017.12.015
Reference: YJSBI 7147

To appear in: *Journal of Structural Biology*

Please cite this article as: Zeng, X., Leung, M.R., Zeev-Ben-Mordehai, T., Xu, M., A convolutional autoencoder approach for mining features in cellular electron cryo-tomograms and weakly supervised coarse segmentation, *Journal of Structural Biology* (2017), doi: https://doi.org/10.1016/j.jsb.2017.12.015

A convolutional autoencoder approach for mining features in cellular electron cryo-tomograms and weakly supervised coarse segmentation

Xiangrui Zeng[1], Miguel Ricardo Leung[2,3], Tzviya Zeev-Ben-Mordehai[2,3], and Min Xu*[1]

[1]Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, 15213, USA
[2]Division of Structural Biology, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK
[3]Cryo-electron Microscopy, Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, Netherlands

Abstract

Cellular electron cryo-tomography enables the 3D visualization of cellular organization in the near-native state and at submolecular resolution. However, the contents of cellular tomograms are often complex, making it difficult to automatically isolate different *in situ* cellular components. In this paper, we propose a convolutional autoencoder-based unsupervised approach to provide a coarse grouping of 3D small subvolumes extracted from tomograms. We demonstrate that the autoencoder can be used for efficient and coarse characterization of features of macromolecular complexes and surfaces, such as membranes. In addition, the autoencoder can be used to detect non-cellular features related to sample preparation and data collection, such as carbon edges from the grid and tomogram boundaries. The autoencoder is also able to detect patterns that may indicate spatial interactions between cellular components. Furthermore, we demonstrate that our autoencoder can be used for weakly supervised semantic segmentation of cellular components, requiring a very small amount of manual annotation.

Keywords: Cellular Electron CryoTomography; Macromolecular Complex; Subtomogram Classification; Visual Proteomics; Particle Picking; Structural Pattern Mining; Deep Learning; Convolutional Neural Network; Convolutional Autoencoder; Image Semantic Segmentation; Machine Learning; Unsupervised Learning; Weakly Supervised Learning; Pose Normalization;

1 Introduction

Recent developments in cellular electron cryo-tomography (CECT) have enabled the three-dimensional visualization of cellular organization in the near-native state and at submolecular resolution. Subcellular components can be systematically analyzed at unprecedented levels of detail. This *in situ* 3D visualization has made possible the discovery of numerous important structural features in both prokaryotic and eukaryotic cells as well as in viruses [17, 7, 14, 18]. As the approach develops, high quality CECT data will continue to yield valuable insights into the structural organization of the cell. In principle, a tomogram of a cell contains structural information of all cellular components within the field of view. However, cellular structures are densely packed within a small volume, which makes it challenging to systemically extract cellular structural information from tomograms. Imaging limitations, such as low signal-to-noise ratio (SNR) and missing wedge effects, further complicate the systematic recovery of such information. Currently, many CECT structural identification, characterization and segmentation tasks are performed by visual inspection and manual annotation, which can be very laborious. Consequently, the labor-intensive nature of these analyses has become a major bottleneck in CECT studies.

*Corresponding author email: mxu1@cs.cmu.edu