

Master 2 internship proposal

Towards a new approach to deep learning for missing data inference in cryo electron tomography

Supervisor: Slavica Jonic, Research Director CNRS, IMPMC – UMR 7590 CNRS, Sorbonne Université, MNHN, IRD, 4 Place Jussieu, 75005 Paris (slavica.jonic@upmc.fr, <http://www.impmc.upmc.fr/~jonic>)

Single-particle cryo electron microscopy (cryo-EM) and cryo electron tomography (cryo-ET) are complementary image collection techniques that allow multi-scale analysis of biomolecular assemblies in health and disease [1-2]. While single-particle cryo-EM allows determining the 3D structure of purified molecules, cryo-ET allows structural determination of the molecules in their native, cellular environment. Molecules change their conformations to accomplish various biological functions [3]. The conformational heterogeneity in the collected data can be both a huge technical problem and (if the problem is solved) an unparalleled opportunity to describe several relevant conformations at once. Advanced image processing algorithms and software are needed to solve the heterogeneity [4-8].

The analysis of the conformational heterogeneity in cellular cryo-ET data is difficult because of an extremely high noise level (due to a low electron dose to which the sample is exposed in the microscope to avoid structural damage) and a highly crowded cellular environment. An additional problem is that 3D reconstruction is performed using images from a limited range of the tilt angle (images are collected from a sample that is only tilted between -60° and 60° or between -45° and 45°), which induces **distortions in the reconstructed volumes** due to the resulting empty space in the 3D Fourier space (the so-called **missing wedge**). The structural deformations induced by the missing wedge interfere with the structural deformations required for accomplishing biological functions [7-8]. An illustration of the cryo-ET data collection, reconstruction, missing wedge, and conformational heterogeneity is shown in the figure below.

A computational correction of the missing wedge is a difficult and still open problem. The majority of methods constrain the molecular structural data analysis to the Fourier-space region that is outside of the missing wedge region [7] or fill in the missing wedge region by the data in the corresponding region of an average structure obtained by averaging many identical molecules extracted from the reconstructed volume [8]. Very little efforts have been so far invested in addressing this problem by artificial intelligence approaches. IsoNet is the first method that aims at correcting the missing wedge using deep learning and has been shown to be able to correct the reconstructed structure of large assemblies such as viruses or organelles [9].

This Master 2 internship project **aims at establishing the basis of a new deep learning approach for the missing wedge correction**, operating in real space or in Fourier space. This new approach will be compared with IsoNet [9] on synthetic and experimental cryo-ET data. If the time permits, the new missing-wedge correction method will be used in combination with conformational heterogeneity analysis methods [7-8]. The experimental data that will be used for testing the methods will be obtained from the public database EMPIAR (<https://www.ebi.ac.uk/empiar>).

Duration: Up to 6 months.

Gratification: $\approx 600\text{€}$ /month (standard for internships in French public research labs)

Location: The internship student will be located at the IMPMC, Tour 23, Campus Jussieu (Sorbonne Université, 4 Place Jussieu, 75005 Paris).

PhD thesis after the internship: This work could be continued in the framework of a PhD thesis.

Contact for applications and required documents: The applications should be sent to slavica.jonic@upmc.fr. They should include a CV, a motivation letter, and grade transcripts.

References:

1. Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P.M.G.E., Grigoras, I.T., Malinauskaite, L., Malinauskas, T., Miehlung, J., Uchański, T., Yu, L., Karia, D., Pechnikova, E.V., de Jong, E., Keizer, J., Bischoff, M., McCormack, J., Tiemeijer, P., Hardwick, S.W., Chirgadze, D.Y., Murshudov, G., Aricescu, A.R., Scheres, S.H.W. (2020). Single-particle cryo-EM at atomic resolution. *Nature* 587, 152-156. (<https://doi.org/10.1038/s41586-020-2829-0>)
2. Turoňová, B., Sikora, M., Schürmann, C., Hagen, W.J.H., Welsch, S., Blanc, F.E.C., von Bülow, S., Gecht, M., Bagola, K., Hörner, C., van Zandbergen, G., Landry, J., de Azevedo, N.T.D., Mosalaganti, S., Schwarz, A., Covino, R., Mühlebach, M.D., Hummer, G., Krijnse Locker, J., Beck, M. (2020). In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* 370, 203-208. (<https://doi.org/10.1126/science.abd5223>)
3. Dashti, A., Mashayekhi, G., Shekhar, M., Ben Hail, D., Salah, S., Schwander, P., des Georges, A., Singharoy, A., Frank, J., Ourmazd, A. (2020). Retrieving functional pathways of biomolecules from single-particle snapshots. *Nat Commun* 11, 4734. (<https://doi.org/10.1038/s41467-020-18403-x>)
4. **Jonic, S. (2017)**. Computational methods for analyzing conformational variability of macromolecular complexes from cryo-electron microscopy images. *Curr Opin Struct Biol* 43, 114-121. (<https://hal.archives-ouvertes.fr/hal-01435361>)
5. Harastani, M., Vuillemot, R., Hamitouche, I., Moghadam, N.B., and **Jonic, S. (2022)**. ContinuousFlex: Software package for analyzing continuous conformational variability of macromolecules in cryo electron microscopy and tomography data. *J Struct Biol* 214, 107906. (<https://hal.archives-ouvertes.fr/hal-03821777>)
6. Hamitouche, I., and **Jonic, S. (2022)**. DeepHEMNMA: ResNet-based hybrid analysis of continuous conformational heterogeneity in cryo-EM single particle images. *Front Mol Biosci* 9, 965645. (<https://doi.org/10.3389/fmolb.2022.965645>)
7. Harastani, M., Eltsov, M., Leforestier, A., **Jonic, S. (2021)**. HEMNMA-3D: Cryo Electron Tomography Method Based on Normal Mode Analysis to Study Continuous Conformational Variability of Macromolecular Complexes. *Front Mol Biosci* 8, 663121. (<https://doi.org/10.3389/fmolb.2021.663121>)
8. Harastani, M., Eltsov, M., Leforestier, A., and **Jonic, S. (2022)**. TomoFlow: Analysis of Continuous Conformational Variability of Macromolecules in Cryogenic Subtomograms based on 3D Dense Optical Flow. *J Mol Biol* 434, 167381. (<https://doi.org/10.1016/j.jmb.2021.167381>)
9. Liu, Y.-T., Zhang, H., Wang, H., Tao, C.-L., Bi, G.-Q., Zhou, Z.H. (2022). Isotropic reconstruction for electron tomography with deep learning. *Nat Commun* 13, 6482. (<https://doi.org/10.1038/s41467-022-33957-8>)

Cryo electron tomography

