

Automatic Pipeline for Cryo-EM Data Preprocessing

Yilai Li and Michael Cianfrocco

University of Michigan, Ann Arbor, Michigan, United States

Cryo-electron microscopy (cryo-EM) is arguably the fastest growing technique in structural biology (Callaway, 2020). Since 2012, the number of deposited structures determined with cryo-EM has been increasing at rates unmatched by the other techniques, namely X-ray crystallography and nuclear magnetic resonance (NMR). This has led to the huge investments in the required infrastructure, made by universities and government-funding agencies worldwide. Moreover, with the recent improvement in sample preparation and data collection (Jain et al., 2012; Arnold et al., 2017; Cheng et al., 2018; Zivanov et al., 2018; Darrow et al., 2019; Ravelli et al., 2019), a single cryo-EM instrument can easily generate more than one dataset (5,000 - 8,000 movies) per day. Data quality assessment and user-free preprocessing will soon become the bottleneck in the “high-throughput” era of cryo-EM.

Although there has been a lot of development in software and algorithms for cryo-EM data processing in the past few years (Punjani et al., 2017; Zivanov et al., 2018; Tegunov and Cramer, 2019), manual and subjective decisions are still heavily involved in solving a structure (Fig. 1, left). As experts represent a continually shrinking fraction of cryo-EM users, the complexity in data processing frustrates many new users. On the other hand, even for cryo-EM experts, manual data assessment and supervision through all the preprocessing steps are still tedious and time consuming.

To improve the automation of cryo-EM data processing, we have developed an automatic pipeline for cryo-EM data preprocessing and assessment (Fig. 1, right) (Li et al., 2019). Specifically, we used a combination of deep learning and image processing to transform the expert knowledge to data assessment models that can replace subjective decisions in the preprocessing steps. Based on the first version of our automatic pipeline, we made a number of improvements which further improves the robustness and reduces the computational cost. Moreover, we are exploring new algorithms for particle sorting to provide a quick readout of particle quality to enable faster preprocessing routines. We believe that our automatic pipeline for cryo-EM data preprocessing will benefit the users from a range of backgrounds.

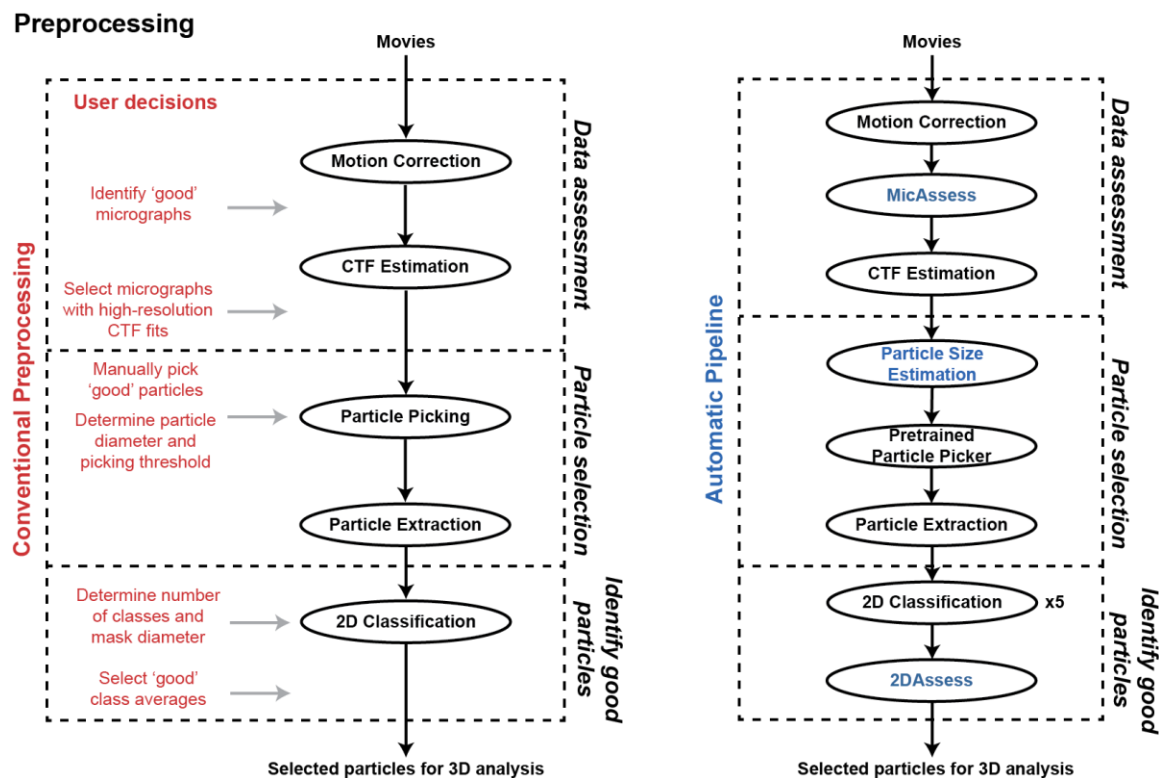


Figure 1. Conventional cryo-EM preprocessing vs. automatic preprocessing pipeline.

References

- Arnold, S. A. et al. (2017) 'Blotting-free and lossless cryo-electron microscopy grid preparation from nanoliter-sized protein samples and single-cell extracts', *Journal of structural biology*, 197(3), pp. 220–226.
- Callaway, E. (2020) 'Revolutionary cryo-EM is taking over structural biology', *Nature*, 578(7794), p.201.
- Cheng, A. et al. (2018) 'High resolution single particle cryo-electron microscopy using beam-image shift', *Journal of structural biology*, 204(2), pp. 270–275.
- Darrow, M. C. et al. (2019) 'Chameleon: Next Generation Sample Preparation for CryoEM based on Spotiton', *Microscopy and microanalysis: the official journal of Microscopy Society of America, Microbeam Analysis Society, Microscopical Society of Canada*. Cambridge University Press, 25(S2), pp. 994–995.
- Jain, T. et al. (2012) 'Spotiton: a prototype for an integrated inkjet dispense and vitrification system for cryo-TEM', *Journal of structural biology*, 179(1), pp. 68–75.
- Li, Y. et al. (2019) 'High-throughput cryo-EM enabled by user-free preprocessing routines', *bioRxiv*. <https://www.biorxiv.org/content/10.1101/2019.12.20.885541v1.abstract>.
- Punjani, A. et al. (2017) 'cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination', *Nature methods*, 14(3), pp. 290–296.
- Ravelli, R. B. G. et al. (2019) 'Automated cryo-EM sample preparation by pin-printing and jet vitrification', *bioRxiv*. doi: 10.1101/651208.
- Tegunov, D. and Cramer, P. (2019) 'Real-time cryo-electron microscopy data preprocessing with Warp', *Nature methods*. doi: 10.1038/s41592-019-0580-y.
- Zivanov, J. et al. (2018) 'New tools for automated high-resolution cryo-EM structure determination in RELION-3', *eLife*, 7. doi: 10.7554/eLife.42166.