McGill

Topaz for Helical Processing: a Machine Learning Approach for Filament Tracing for High Resolution 3D Reconstruction of Microtubules

Zhe Fan[1], Khanh Huy Bui[1,2]

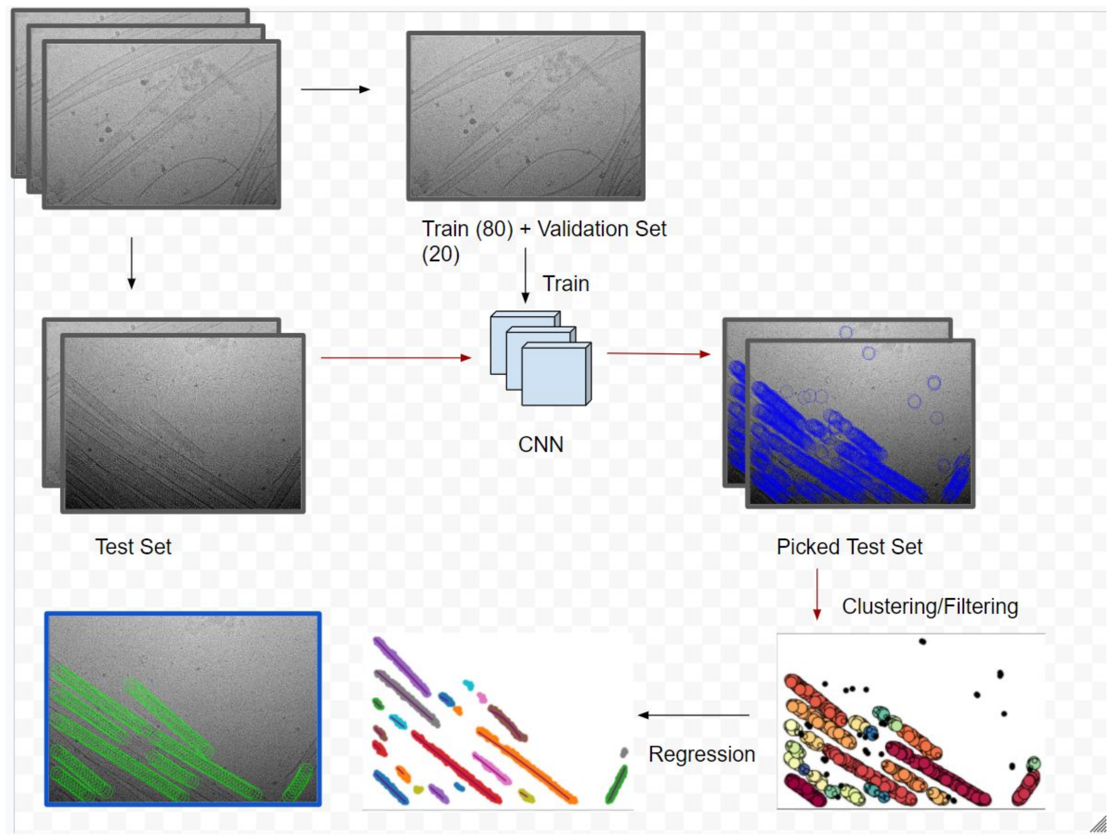
1 - McGill University, 2 - Department of Anatomy and Cell Biology, McGill University

Abstract

Technological advances in single-particle cryo-electron microscopy have led to significant progress in data collection and processing. However, for certain types of samples such as irregularly shaped particles or helical filament, picking of raw particles remains a challenge. Manual selection of those particles is tedious and time-consuming process, typically lasting weeks to complete with a typical dataset. Recent technologies such as SPHIRE-crYOLO[2] and topaz[1] have emerged to improve the efficiency and accuracy of the particle selection process using machine-learning algorithms to perform automatic selection. There aren't many extensive studies involving the use of such technologies for the analysis of filamentous molecules such as microtubules, which requires the information of filament for orientation estimation and alignment. In this study, we implemented a microtubule picking workflow incorporating the ML-based particle picker Topaz, and an additional filament tracing tool we designed to extract equidistant particle locations along the lengths of the microtubules, which cleans up many extraneous particles and optimizes the coordinates for helical processing. We demonstrate the greatly improved efficiency and acceptable accuracy of the helical processing jobs conducted via this workflow on Microtubule micrographs collected for MAP9 analysis.

The Work-Flow (Methods)

1. We collected a dataset consisting of 3588 micrographs of singlet microtubules (Assembled in Taxotere + DMSO) recorded via cryogenic electron microscopy.
2. Performed manual picking on a training set of 100 micrographs, randomly selected from the batch.
3. Using the topaz pipeline[1] (*Bepler et Al, 2019*), we trained a Convolutional Neural Network (CNN) for classifying positive particles from the training set.
4. Performed particle coordinate prediction on the unseen micrographs using the CNN model.
5. Performed additional processing on picked coordinates using Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to extract particle clusters which resemble filaments and filtering out noisy coordinates.
6. Random Sample Consensus (RANSAC) regression performed on particle clusters to linearly fit filaments and specify equidistant (the "inter-box" distance[3], *Scheres et Al*) neighbouring particles along the filament, as a standard step of helical processing in Relion.
7. Relion Processing of all samples using the newly fitted, automatically picked particles separated by the inter-box distance.



Picking Results

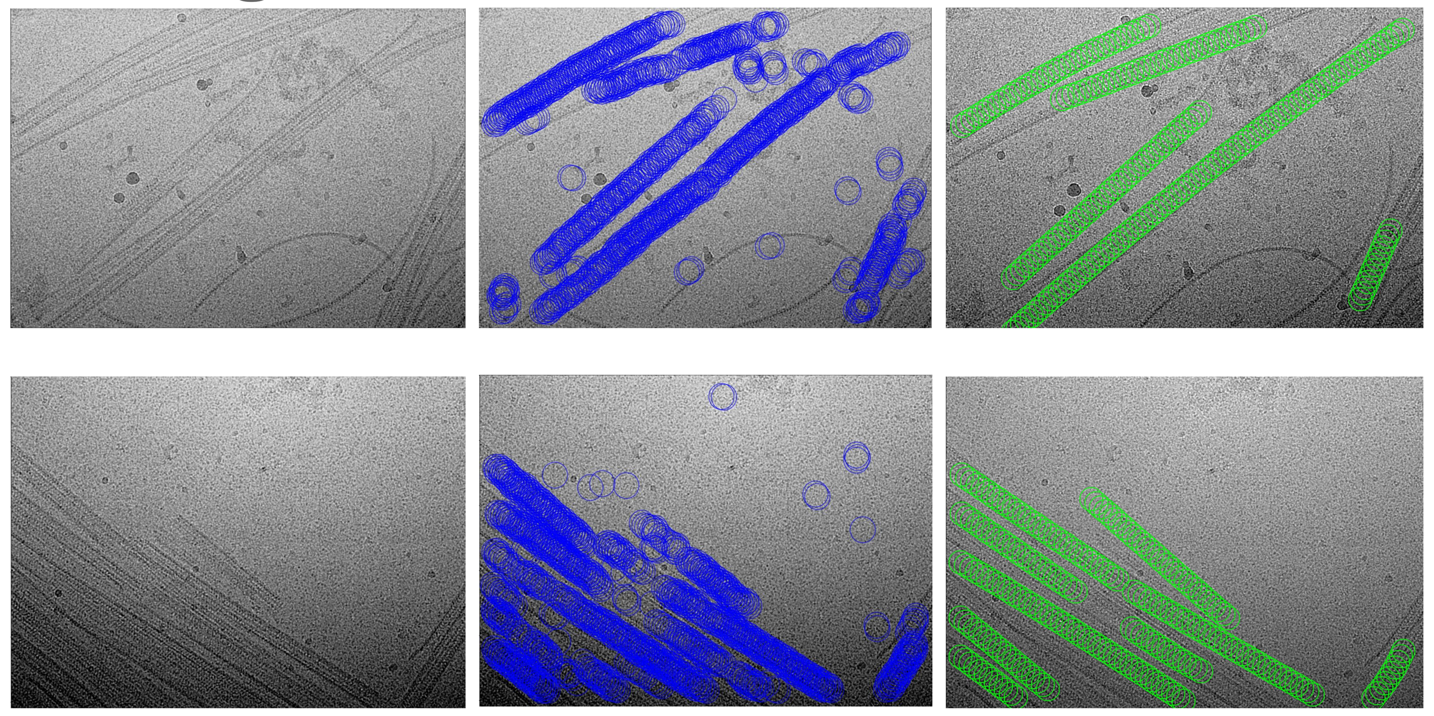


Fig 1. Examples of picking completed by Topaz and the Post-Topaz Processing steps
Left column: Raw micrographs
Middle column: Topaz Picking (CNN)
Right column: Post-Topaz Processing (DBSCAN Clustering and RANSAC regression)

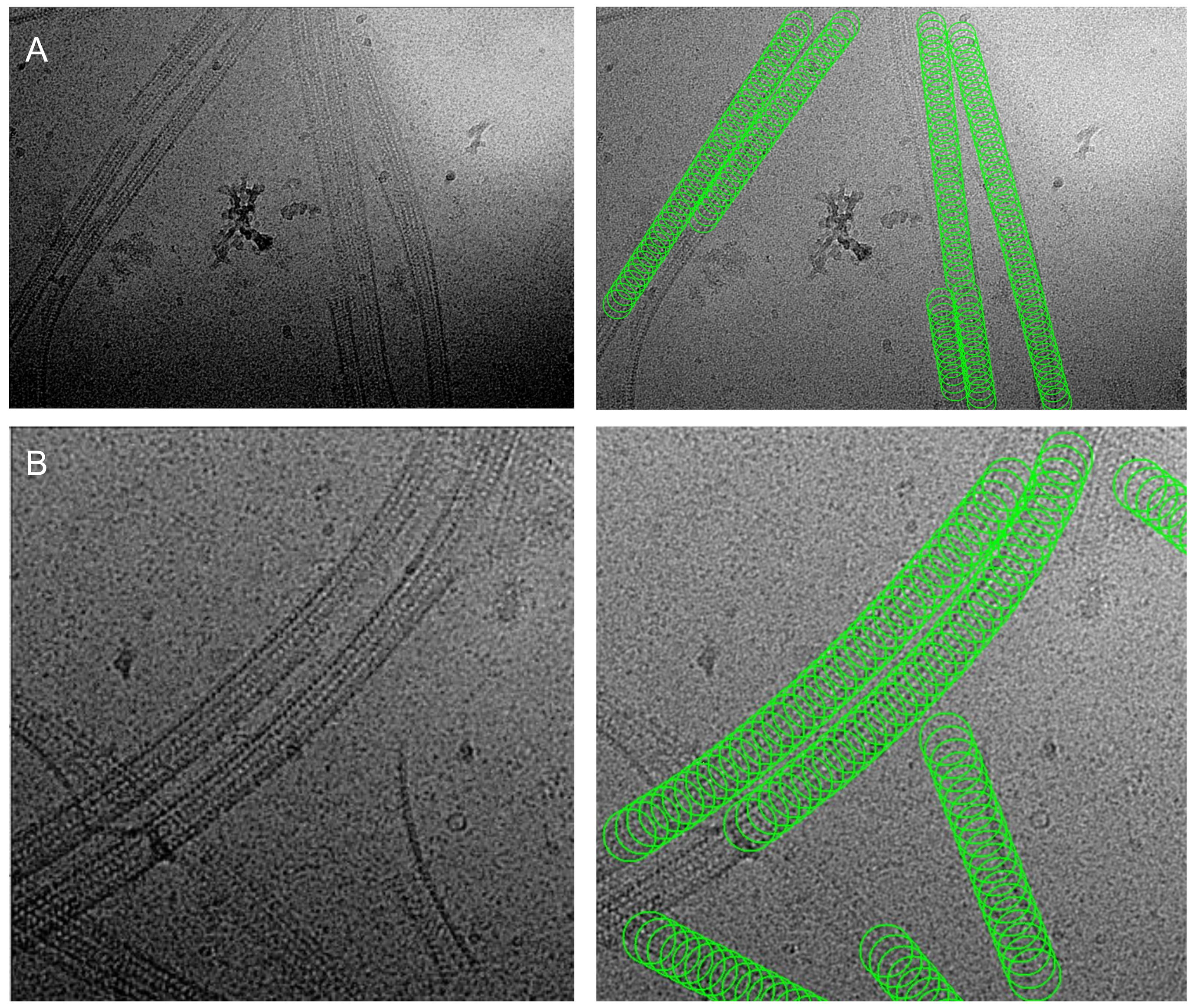
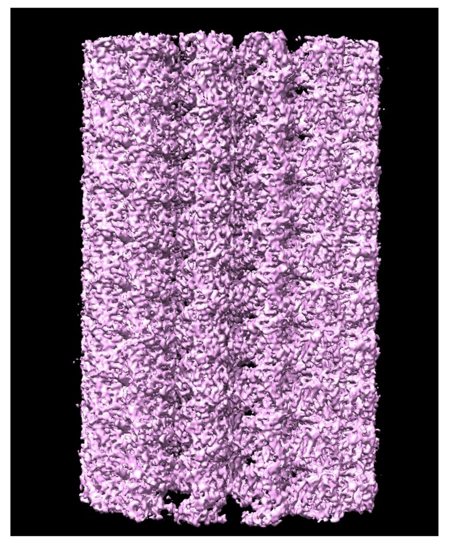


Fig 2. Examples of Automatic Picking on Samples containing a. Overlapped filaments and b. Curved Filaments

In dealing with overlapped filaments, Topaz does a good job at eliminating extraneous overlapped segments when predicting. The workflow responds well to variability in the curvature of the filament, and performs linear regression at varying orders as appropriate to trace the filaments.



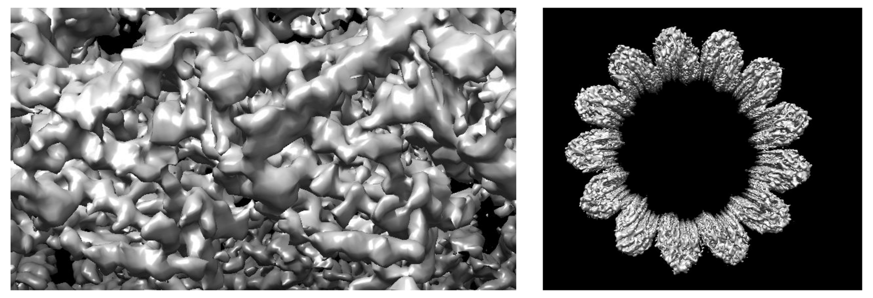


Fig 3. 3D reconstructions
Preliminary 3D reconstructions of 13-PF MT on ReLION using the picked coordinates yielded acceptable resolutions without correcting for the seam. (4.6 Angstroms)
Alpha helical features can be observed clearly.

Classification Results

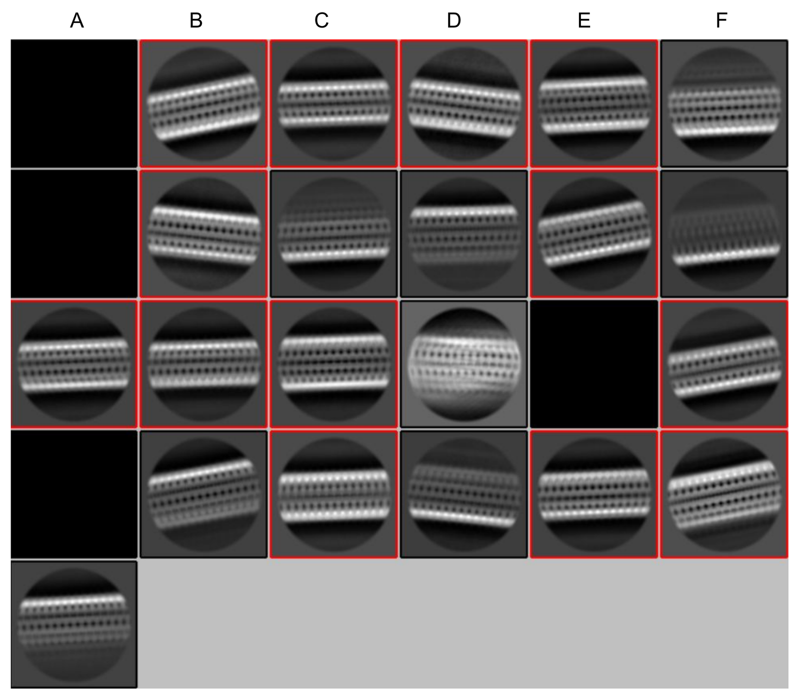


Fig 4. 2D classifications
2D class averages were calculated using the picked particles from the workflow.

Highlighted classes reflect true positives of the picking.

1F, 2E - 2F, 4B, 4D, 5A indicates presence of MTs that have been split, and such are false positives.

3D indicates the presence of overlapped MTs, and thus are false positives.

The black squares are false positives.

Total Particles: 401979
True Positives: 217166
Precision : 54%

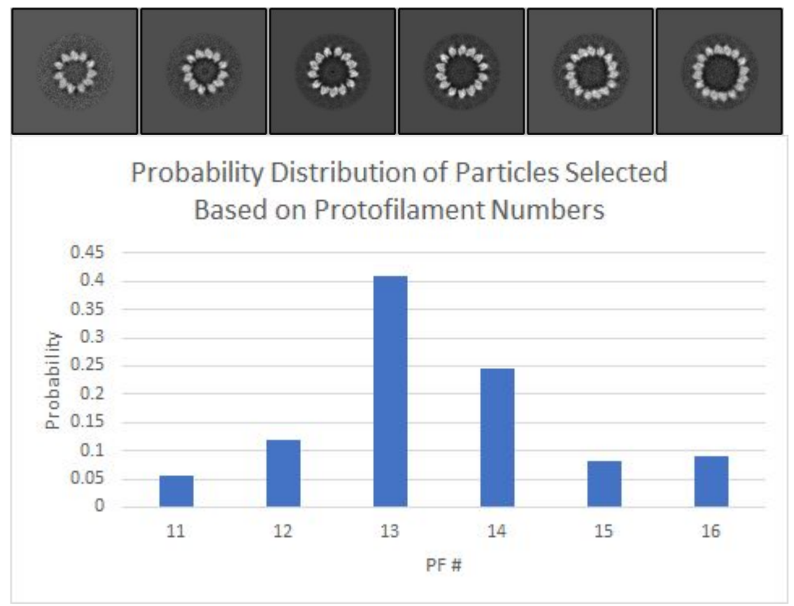


Fig 5. 3D classifications

From PF-11 to PF-16, six classes were generated using the true positive particles.

Distribution of true positive particles, based on the protofilament numbering indicates high concentration of PF-13 and PF-14 Microtubules assembled under Taxotere + DMSO conditions. Consistent with previous research results.[4] (*Sui and Downing, 2010*)

Discussion

Overall, automatic particle picking and helical processing using the proposed workflow was very efficient. The manual picking of 3588 micrographs usually takes weeks to perform by a single person. The creation of the training model and the prediction of particles on all micrographs via Topaz took about 30 minutes in duration. Further processing, such as clustering and regression, takes less than 10 minutes on average when these procedures are implemented with multiprocessing. The precision of the procedure leaves room for improvement, as the automatic picker tends to label split/overlapped microtubules as positive. This may be resolved by expanding the training set for the Neural Network, and by **incrementing the probability threshold for positive particles** in topaz to filter out more false positives, although this may increase the number of false negatives as well. Overall, with the demonstrated efficiency and overall results of this workflow, it remains an optimistic option for the CryoEM processing of Microtubules and many other helical structures.

References

1.Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A. J., & Berger, B. (2019). Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature Methods, 16*(11), 1153-1160. doi:10.1038/s41592-019-0575-8

2.Wagner, T., Merino, F., Stabrin, M., Moriya, T., Antoni, C., Apelbaum, A., . . . Raunser, S. (2019). SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Communications Biology, 2*(1). doi:10.1038/s42003-019-0437-z

3. He, S., & Scheres, S. H. (2017). Helical reconstruction in RELION. *Journal of Structural Biology, 198*(3), 163-176. doi:10.1016/j.jsb.2017.02.003

4.Sui, H., & Downing, K. H. (2010). Structural basis of interprotofilament interaction and lateral deformation of microtubules. *Structure (London, England : 1993), 18*(8), 1022–1031. https://doi.org/10.1016/j.str.2010.05.010