**Real-time cryo-electron microscopy data preprocessing with Warp**

**Background:**

Cryogenic electron microscopy (cryo-EM) is an electron microscopy (EM) technique applied on samples cooled to cryogenic temperatures and embedded in an environment of vitreous water. The standard cryo-EM processing pipeline begins with the alignment of dose-fractionated image frames to cancel out sample motion, followed by an estimation of the contrast transfer function. Next, particles are selected and extracted from the images. These steps are referred to as ‘data preprocessing’. Performing ‘data preprocessing’ is aimed to better subsequent 2D and 3D classification and model.

**Methods:**

Warp is a software that automates all preprocessing steps of cryo-EM data acquisition and enables real-time evaluation. At the beginning of the pipeline, Warp reads new data saved by the acquisition software. Warp then estimates and corrects the object motion across the micrograph movie both globally and locally. Next, Warp fits a contrast transfer function (CTF) model, enabling the assignment of local defocus values to any particles extracted from the micrograph later. Warp then uses a neural network-based approach to automatically pick particles from the data with very high accuracy. Finally, Warp exports the resulting dose-weighted particle images to a downstream structure determination program. Coupled with automated acquisition software2,8, Warp provides a continuous low-latency stream of reliably picked and corrected particle images that can be seamlessly fed into 2D classification, ab initio reconstruction and 3D refinement tools using other packages. Reliably performing preprocessing of cryo-EM raw data at the speed of data collection remains a struggle.

**Results:**

To test the performance of Warp, the author reprocessed a published single-particle cryo-EM dataset. For the original set, the best class containing 57,346 particles reached a global resolution of 3.9 Å with a B-factor of −200 Å2. The same particles, updated with the defocus information from Warp, reached a notably higher resolution of 3.5 Å with a B-factor of −170 Å2. This suggests that Warp’s local CTF model is more accurate than the per-particle CTF fitting in gCTF14 used in the original study.

To estimate the frame alignment accuracy independently of 3D refinement, we calculated the average CTF fit quality for aligned movie averages processed with MotionCor2 or Warp. Warp’s averages could be fitted to 2.6 Å, and those of MotionCor2 to 2.7, indicating slightly better frame alignment in Warp. After refinement of particles from the full, completely automated Warp preprocessing pipeline, the best class containing 127,000 particles reached a global resolution of 2.09 Å with a B-factor of −35 Å2.

The preprocessing of 668 movies in Warp required approximately 3 h using a system with 4 Nvidia Titan X GPUs, thus averaging to roughly 220 movies per hour.

**Discussion:**

To further improve cryo-EM data collection efficiency in the future, the author envision an automated feedback loop between the acquisition software and Warp. Such feedback would help to maximize the number of high-quality images collected by quickly skipping areas with low estimated resolution or particle count, adjusting the stage settling time dynamically based on Warp’s motion estimate for the previous image.