Fast SPI Images Classification Using Supervised Machine Learning Methods

Yingchen Shi[1], Ke Yin[2], Xuanxuan Li[1], Xuecheng Tai[3*], Haiguang Liu[1*]

[1] Beijing Computational Science Research Center, Beijing, China
[2] Huazhong University of Science and Technology, Wuhan, China
[3] Hong Kong Baptist University, Hong Kong, China

Convolutional neural network (CNN) and Graph-cut method (GC) are used for classifying single-hit patterns with small training set in single particle imaging. We evaluate the calculation time, averaged pattern intensity curves (SAXS curves) and recovered orientations to compare results from three methods: CNN, GC and Diffusion map embedding (DM). The comparison shows great potential for CNN and GC to set up fast classification pipeline, as they use around 20% of the time consumed by DM to finish the work. However, there are still much room for CNN and GC to improve their classification results.

Diffusion Map Embedding program used to test time consumption is wholly based on software provided by Russel Fung and collaborators at the University of Wisconsin-Milwaukee. EMC algorithms implemented in Dragonfly and tools in SASTBX are used for evaluating classification results. The data are from PR772 experiment collected in 2016 in LCLS.