Contribution ID: **14** Type: **Contributed Talk (15 min)**

Robust Image Descriptor for Machine Learning based Data Reduction in Crystallography

Monday, 19 September 2022 17:15 (15 minutes)

Recent serial crystallography experiments at FELs produce a large amount of data, where typically the ratio of useful images containing crystal diffraction (hit fraction) is about 5-10% but hit fractions even lower than 0.1 % have been observed in some experiments. Demands on data storage could be greatly reduced by rejecting bad images before saving them to disk, but this requires reliable methods for detecting these images that do not rely on expert tuning or intervention during the experiment. Traditional Non-Machine learning techniques successfully classify good and bad images based on number of detected peak on the images. But those techniques require tuning many parameters that are vary in each experiment. Also traditional machine learning methods like artificial neural networks successfully classify diffraction patterns, but the major challenge is cross-domain performance, in which a classifier trained by a set of local feature vectors extracted on one dataset cannot necessarily be applied to data collected with different samples and experimental settings. Scale-invariance, good localization, and robustness to noise and artifacts are the main properties that a local feature detector should possess. In this paper, we propose a real-time, automatic, and parameter-free image descriptor as a local feature detector for use as input data for ML models. Our method describes each diffraction pattern by a vector, consisting of the number of keypoints (Bragg spots) in different areas of the image. Additionally, we parallelized our novel and parameter-free keypoint detection algorithm to work computationally efficiently compared to the traditional keypoint detection algorithms. A machine learning model trained by this introduced image descriptor(vector) performs well across different experimental settings(cross-domain) to classify diffraction patterns as hit and miss diffraction patterns. Our initial experimental results show a significant improvement in decreasing the domain gap when a ML classifier is trained by new image descriptor and tested by another unseen dataset.

Primary author: Dr RAHMANI, Vahid (DESY)

Co-authors: Dr NAWAZ, Shah (DESY); Mr PALA RAMAKANTHA SETTY, Shabarish (DESY); Mr BHOSALE, Jaisinh (Technical University of Hamburg); Dr PENNICARD, David (DESY); Prof. GRAAFSMA, Heinz (DESY)

Presenter: Dr RAHMANI, Vahid (DESY)

Session Classification: Nano and Material Science

Track Classification: Nano and Material Science