

4D >Crystal: Deep Learning Crystallographic Information From Electron Diffraction Images

Joydeep Munshi¹, Alexander M Rakowski², Benjamin Savitzky³, Colin Ophus³, Matthew L Henderson², Shreyas Cholia² and Maria KY Chan¹

¹ANL, United States, ²LBNL, United States, ³Lawrence Berkeley National Laboratory, California, United States

Ongoing development in electron detector technology, in the recent years, has enabled a new age of electron microscopy which facilitated easy accessibility of multi-dimensional data. The 4D Camera at Lawrence Berkeley National Laboratory [1], for example, is capable of generating enormous amounts of data (estimated ~ 5 PBs of data each year) due to the development of high-speed electron detector technology. To make the data analysis of 4D Camera easy and accessible, tools such as the robust open-source python-based 4D-STEM (4D- Scanning Transmission Electron Microscope) analysis package (py4DSTEM) have been recently developed [2]. Despite the comprehensive and analysis strategy in many instances, there is a need for image analysis using cutting-edge machine learning (AI/ML) techniques to complement the existing relatively slow, complex and hyperparameter-dependent pipeline. To this end, we propose the design of a fully automated AI/ML python-based pipeline, which is fast, automated, and robust against experimental error and background noise.

In this work, we present one of the targeted analysis pipelines, *4D >Crystal*, an effort to generate a series of reduced crystalline diffraction vectors (Bragg vector map) from a given combination of STEM probe in vacuum with typical 4D-STEM diffraction dataset. Figure 1 presents the workflow of the analysis pipeline for 4D >Crystal. The 4D >Crystal pipeline is composed of three main sub-tasks – 1) crystal selection and data scraping, 2) training data generation and 3) deep neural networks (DNNs) training and evaluation. To efficiently select the crystal structures for the training dataset, we use the python-based pymatgen package [3, 4] and the Materials Project (MP) API to fetch crystal structures from more than 120,000 materials present in the MP database. Additionally, in order to systematically select a subset of the fetched crystal structures, we extract ~ 260 distinct structure prototypes from the AFLOW library of crystallographic prototype [5, 6], representing different types of crystals based on their structural similarity. We select a fraction of structures from the MP database, resembling each structure prototype, following the square-root distribution law. Following the crystal structure selection, we employ a supercell manipulation algorithm (Manipulatt [7]) to pre-process crystal unit cells to an oriented and tilted supercell. Finally, convergent beam electron diffraction (CBED) patterns and the measured potential for each of the manipulated supercells are simulated using the Multislice algorithm [8, 9]. Figure 2 presents a typical example of simulated data input and output labels for the deep learning network training. From a subset of 3082 selected crystals, we are able to generate a training dataset of ~6 million image quartets (inputs: probe + cbed, and outputs: potential (V_g) + out-of-plane tilt (q_z)). To predict Bragg peak locations and their intensities, we implement a modified U-Net [10] type encoder-decoder based deep learning network which employs a cross-correlation Fast Fourier Transform (FFT) layer before the U-net encoder stage. As a test case, we show the performance of the network trained on a small subset of the training dataset (~50,000 training images and 5,000 validation and test images). Use of a cross-correlation layer and the non-linear activation function (Relu) are expected to be effective as they are able to capture the highly non-linear mapping between the input scattering (cbed) intensity and the output potentials (V_g). The ability to quantitatively predict the output potentials (V_g) offers accurate in-plane orientation, robust strain mapping for thick samples and prediction of accurate out-of-plane tilt (q_z) would be capable of estimating 3-dimensional orientation and strain mapping for unknown polycrystalline samples in a robust and, critically, automated manner.

Acknowledgement:

1. Use of the Center for Nanoscale Materials, an Office of Science user facility, was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357.
1. This research used resources of the National Energy Research Scientific Computing Center; a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.
1. Work at the Molecular Foundry was supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

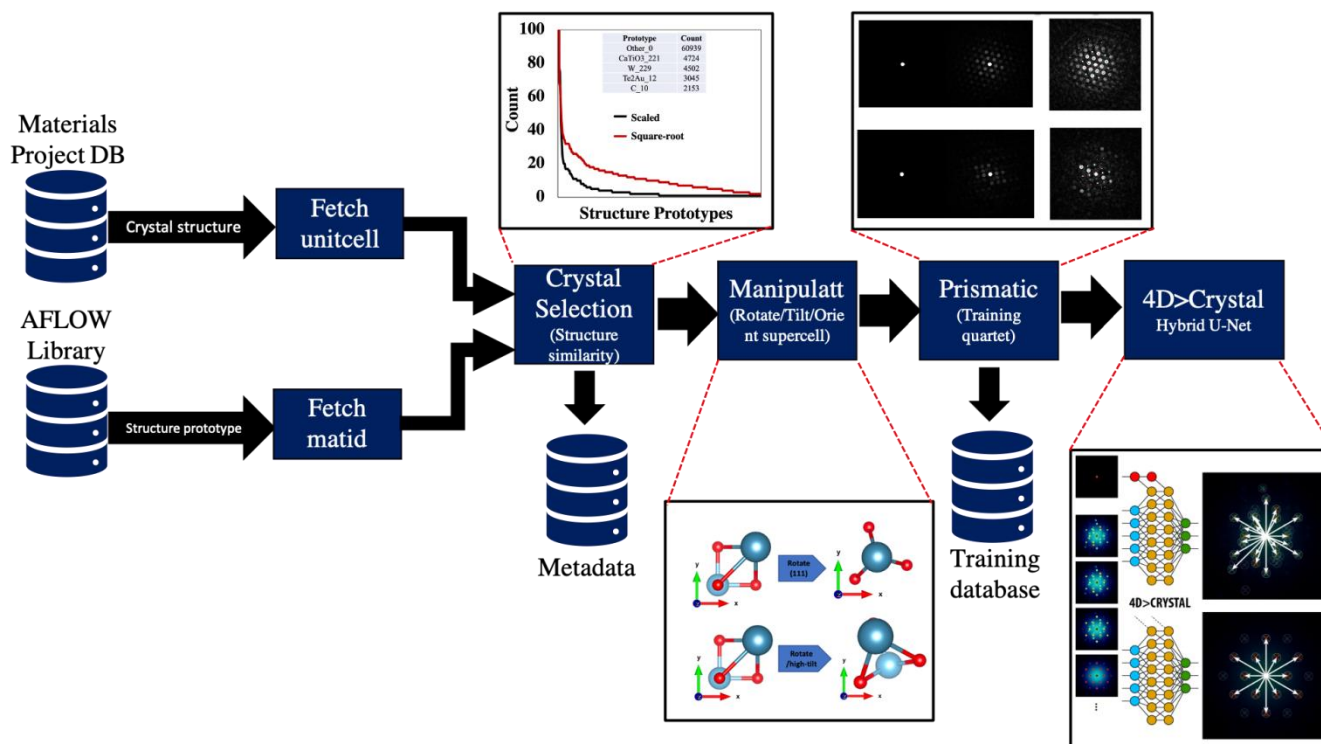


Figure 1. Fully automated AI/ML workflow for robust and accurate Bragg peak position and intensity detection from 4D STEM electron diffraction dataset.

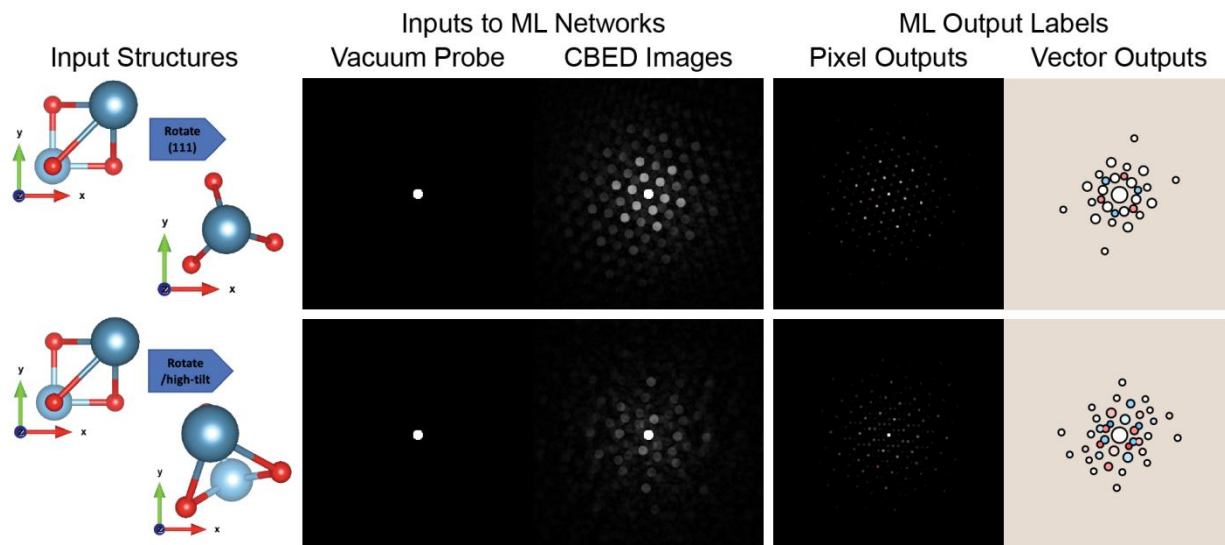


Figure 2. Example manipulation of perovskite (CaTiO₃) unit cell (a) rotated along (111) zone axis, (b) rotated and high-tilted and their training quartets (input probe, diffraction image and output potential and out-of-plane tilt, qz)

References

1. Ciston, J., et al., *The 4D Camera: Very High Speed Electron Counting for 4D-STEM*. *Microscopy and Microanalysis*, 2019. **25**(S2): p. 1930-1931.
2. Savitzky, B.H., et al., *py4DSTEM: A software package for multimodal analysis of four-dimensional scanning transmission electron microscopy datasets*. arXiv preprint arXiv:2003.09523, 2020.
3. Ong, S.P., et al., *Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis*. *Computational Materials Science*, 2013. **68**: p. 314-319.
4. Jain, A., et al., *Commentary: The Materials Project: A materials genome approach to accelerating materials innovation*. *APL Materials*, 2013. **1**(1): p. 011002.
5. Hicks, D., et al., *The AFLOW Library of Crystallographic Prototypes: Part 2*. *Computational Materials Science*, 2019. **161**: p. S1-S1011.
6. Mehl, M.J., et al., *The AFLOW Library of Crystallographic Prototypes: Part 1*. *Computational Materials Science*, 2017. **136**: p. S1-S828.
7. Munshi, J., *manipulatt v2021.01.14, Tool to manipulate primitive cell to rotate, tilt, tile to a supercell*. Github repository, 2021 (<https://github.com/MaterialEyes/manipulatt>).
8. Ophus, C., *A fast image simulation algorithm for scanning transmission electron microscopy*. *Advanced Structural and Chemical Imaging*, 2017. **3**(1): p. 13.
9. Pryor, A., C. Ophus, and J. Miao, *A streaming multi-GPU implementation of image simulation algorithms for scanning transmission electron microscopy*. *Advanced Structural and Chemical Imaging*, 2017. **3**(1): p. 15.
10. Ronneberger, O., P. Fischer, and T. Brox. *U-net: Convolutional networks for biomedical image segmentation*. in *International Conference on Medical image computing and computer-assisted intervention*. 2015. Springer.