Microsc. Microanal. 28 (Suppl 1), 2022
doi:10.1017/S1431927622002367

Assisting 4D-STEM Data Processing with Unsupervised Machine Learning

Yimo Han[1*], Chuqiao Shi[1], Michael Cao[1] and Yi Jiang[2]

[1.] Department of Materials Science and NanoEngineering, Rice University, Houston, TX, United States.
[2.] Advanced Photon Source, Argonne National Laboratory, Lemont, IL, United States.
* Corresponding author: yh76@rice.edu

From the highest-resolution electron ptychography [1–3] to micrometer-scale strain mapping [4], four-dimensional scanning transmission electron microscopy (4D-STEM) has demonstrated significant advancements in the study of materials. The commercialization of direct electron detectors, such as the electron microscope pixel array detector (EMPAD) [5,6], has made 4D-STEM a more general technique than a few years ago. However, the processing of large 4D data still remains challenging and heavily relies on experts. Here, we present the utilization of unsupervised learning and Bayesian optimization to assist 4D-STEM data processing.

For mapping deformations in materials, we developed a hierarchical unsupervised learning workflow to cluster nanobeam electron diffraction patterns in 4D datasets [7]. This approach provides an initial analysis of the 4D data and uncovers essential features in the sample even without prior knowledge. For example, using this method, we uncovered strain and ripples in two-dimensional (2D) lateral heterojunctions in agreement with previously reported results4 (Figure 1a, b). Applying this approach on a novel 2D ferroelectric SnSe sample, we have identified ferroelectric domains and super domains (Figure 1c-e). Based on the results, a subsequent quantitative analysis of the same 4D data provides details in lattice structure at different domains and domain boundaries.

In addition, we also developed a streamlined data-processing workflow for electron ptychography. Electron ptychography is capable of deep sub-angstrom resolution reconstructions but requires careful selection of multiple reconstruction parameters. We demonstrate a scheme for automatic parameter tuning by implementing Bayesian optimization with Gaussian processes. This approach is able to find high-resolution reconstructions after exploring only 1% of the entire parameter space. With minimal prior knowledge, the approach significantly improves the efficiency of 4D data processing and provides high quality ptychography reconstructions that agree with the experts [8].

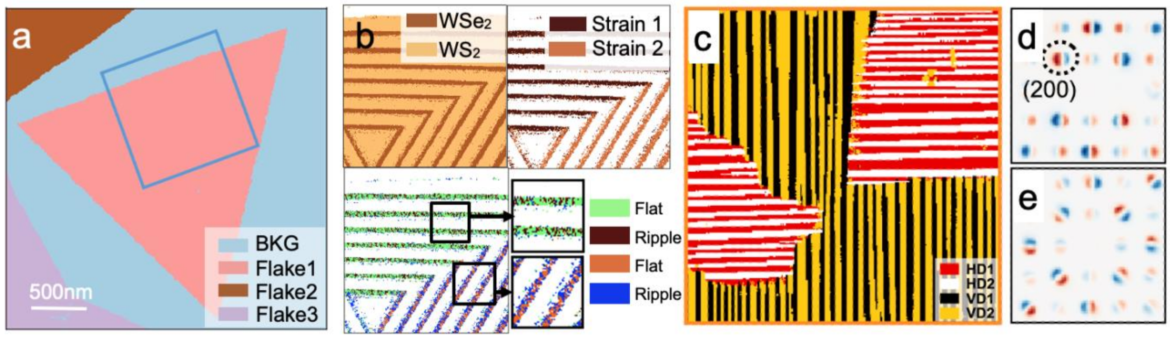

Figure 1. Unsupervised learning to uncover deformations. (a) Clustering results from the 4D dataset of the WS_2-WSe_2 multi-junction sample. (b) Sub-clustering results in WSe_2 with colors indicating strain, flat area, and ripples. (c) Clustering results of SnSe, which contains four different domains. (d) Difference between mean diffraction patterns of yellow and black domains. (e) Difference between mean diffraction patterns of white and black domains.

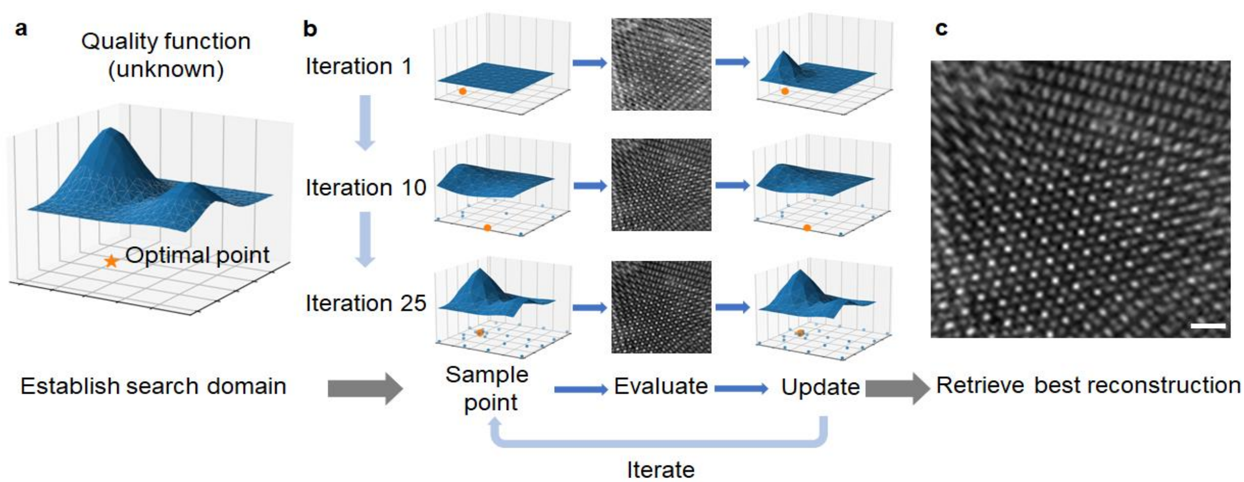

Figure 2. Automatic reconstruction tuning with Bayesian optimization. (a) The workflow aims to find the optimal point that maximizes the unknown quality function. (b) Bayesian optimization iteratively samples more points, performs ptychographic reconstructions, and updates a surrogate model that approximates the quality function. (c) Best reconstruction is retrieved at the end of the process. Scale bar is 5Å.

References:

[1] Y Jiang et al., Nature **559** (2018), p. 343.
[2] Z Chen et al., Nature Communications **11** (2020), p. 1.
[3] Z Chen et al., Science **372** (2021), p. 826.
[4] Y Han et al., Nano Letters **18** (2018), p. 3746.

[5] MW Tate et al., Microscopy and Microanalysis **22** (2016), p. 237.
[6] HT Philipp et al., arXiv preprint arXiv:2111.05889, (2021).
[7] C Shi et al., arXiv preprint arXiv:2111.06496, (2021).
[8] The authors acknowledge funding from the Welch Foundation (C-2065-20210327).

https://doi.org/10.1017/S1431927622002367 Published online by Cambridge University Press