Machine Learning Enabled Image Classification for Automated Data Acquisition in the Electron Microscope

Carolin B Wahl, Alexandra Day, Vishu Gupta, Roberto R dos Reis, Wei-keng Liao, Chad A Mirkin, Alok Choudhary, Vinayak P Dravid, Ankit Agrawal

Microscopy and Microanalysis, **29** (Suppl 1), 2023, 1909–1910
https://doi.org/10.1093/micmic/ozad067.986

Microscopy AND Microanalysis

Meeting-report

Machine Learning Enabled Image Classification for Automated Data Acquisition in the Electron Microscope

Carolin B. Wahl[1,2], Alexandra Day[3], Vishu Gupta[3], Roberto R. dos Reis[1,4], Wei-keng Liao[3], Chad A. Mirkin[1,2,5], Alok Choudhary[3], Vinayak P. Dravid[1,2,4,*], and Ankit Agrawal[3,*]

[1]*Department of Materials Science and Engineering, Northwestern University, Evanston, IL, United States*
[2]*International Institute for Nanotechnology, Northwestern University, Evanston, IL, United States*
[3]*Department of Electrical and Computer Engineering, Northwestern University, IL, United States*
[4]*The NUANCE Center, Northwestern University, Evanston, IL, United States*
[5]*Department of Chemistry, Northwestern University, Evanston, IL, United States*

*Corresponding authors: v-dravid@northwestern.edu, ankit-agrawal@northwestern.edu

The development of high-speed detectors and associated data acquisition workflows in electron microscopes has enabled data generation at an unprecedented rate. It is not uncommon to generate terabytes of data per hour which can easily overwhelm computing infrastructure and downstream processing, thereby limiting the full utilization of the microscope hardware [1]. Implementing a selective approach to imaging to only acquire high-fidelity data from areas of interest is common in manual instrument operation, where bright- or dark-field STEM images are acquired first to identify regions of interest for collecting higher-dimensional spectroscopy or diffraction data. In the context of automated acquisitions workflows, this procedure becomes significantly more challenging, as it becomes necessary to quickly make decisions about regions of interest in the absence of a human operator. To that end, we have developed machine learning models working with low-fidelity dark-field STEM images of nanostructure arrays to perform a binary quality assessment to determine the highest priority areas to acquire high-fidelity data.

Our models will be deployed in an automated electron microscopy system for the high-throughput analysis of nanomaterial megalibraries. Megalibraries contain arrays of 10^8 nanoparticles with spatially encoded size and composition gradients on a single chip, these libraries are uniquely suited for the discovery of new materials at high throughput [2]. Given the volume of materials in a megalibrary, we have designed our models to be highly effective at binary classification on unseen images while minimizing the single-inference times and the number of false positive results which would result in long latencies and high downstream processing cost. In this presentation, we will describe the computational challenges and solutions implemented in this design (e.g., memory constraints, training time, Neural Architecture Search (NAS) tools). We present the standard neural network evaluation metrics, training time, and single-inference time for the most accurate model; our best-performing model achieves a precision of >95%. We further report a weighted F-score of >90% on our test data set as a balancing metric of precision and recall for a given application. Given that false positives would incur a high downstream cost in the analysis pipeline, we chose to prioritize precision over recall by a factor of 9-to-1.

This binary classification of low-fidelity data represents an important step towards fully automated electron microscopy workflows, as it introduces a methodology for autonomous decision-making *in situ*. The acceleration resulting from this automation effectively increases the experimental throughput, enabling the analysis of materials at an unprecedented rate [3].

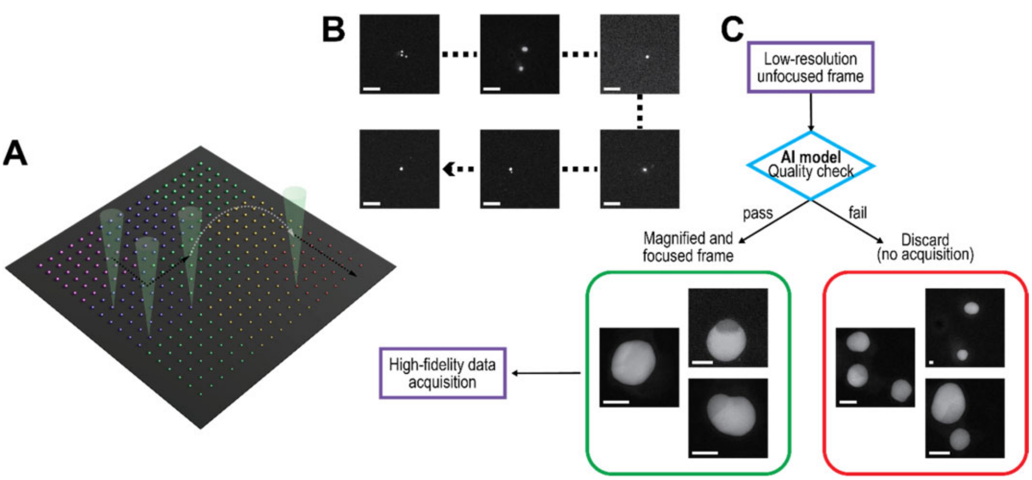

Fig. 1. Automated quality assessment on a nanoparticle megalibrary. (A) Schematic illustration of the screening process, acquiring images at nanoparticle locations. (B) Acquired high-angle annular dark-field (HAADF) images as seen by the model. Scale bars: 200 nm. (C) Illustration of the quality assessment pipeline, where each raw image is fed into the model, returning a binary classifier. If the image passes the test, the microscope can magnify and focus on that region to acquire high-fidelity datasets. Scale bars: 20 nm.

References

1. SR Spurgeon *et al.*, *Nat. Mater.* **20** (2021), p. 274.
2. EJ Kluender *et al.*, *Proc. Nat. Acad. Sci.* **116** (2019), p. 40.
3. The authors acknowledge funding from the Sherman Fairchild Foundation, Inc. and the Toyota Research Institute, Inc. This work is supported in part by the following grants: NIST award 70NANB19H005, DOE awards DE-SC0019358 and DE-SC0021399, and NSF award CMMI-2053929. This work made use of the EPIC facility of Northwestern University's NUANCE Center, which has received support from the Soft and Hybrid Nanotechnology Experimental (SHyNE) Resource (NSF ECCS-1542205); the MRSEC program (NSF DMR-1720139) at the Materials Research Center; the International Institute for Nanotechnology (IIN); the Keck Foundation; and the State of Illinois, through the IIN.