doi:10.1017/S143192761900148X
Microsc. Microanal. 25 (Suppl 2), 2019
© Microscopy Society of America 2019

Machine Learning for High Throughput HRTEM Analysis

Catherine Groschner[1*], Christina Choi[1], Dat Nguyen[1], Colin Ophus[2], and Mary Scott[1,2]

[1.] Department of Materials Science and Engineering, UC Berkeley, Berkeley, CA
[2.] Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, Berkeley, CA
[*] Corresponding author: cgroschner@berkeley.edu

The surge in interest in nanomaterials in the past decade is ascribable, in large part, to the specialized properties due to size, shape, and structure at the nanoscale. Catalysts are a good example of this almost atom-by-atom dependence, with the number and coordination of each atom within a particle impacting the performance. A recent study even has shown that the simple rearrangement of 25 gold atoms from a spherical to cylindrical shape can significantly impact the efficacy of that catalyst [1]. The length scale of such materials makes characterization of these materials more difficult. Most studies must rely on ensemble measurements such as powder x-ray diffraction (XRD) or small angle x-ray scattering (SAXS) and assume a homogenous population such that the ensemble measurement is representative of each particle. This means that information on structural heterogeneity in samples is completely lost. The only technique which can give local structural information on a nanoparticle by nanoparticle basis is high-resolution transmission electron microscopy (HRTEM) or high resolution scanning transmission electron microscopy (HR-STEM). HRTEM is generally extremely low throughput due to constraints on data analysis. How to make HRTEM a more high-throughput process has become a goal for the TEM community and has large implications for the understanding of structure-property relationships in nanomaterials.

Currently, no method exists to analyze HRTEM data in a high throughput manner across a large range of samples. The recognition of defects and shape require semantic segmentation (partitioning of the image on a pixel by pixel basis) which can be extremely difficult due to the low signal to noise ratio in most HRTEM images. Recent advances in convolutional neural networks (CNNs) for computer vision have made advanced segmentation problems solvable [2-5]. However, only in the last year have CNNs started to be applied to electron microscopy. Both Ziatdinov *et al* and Madsen *et al* have reported using a convolutional neural network to identify atoms in electron micrographs [6,7]. But neither of these methods looks at classifying structural defects or segmenting out entire nanoparticles, instead only being able to identify individual atomic columns. Therefore, a method is needed which focuses on local structural features, such as stacking faults and other two dimensional defects in a manner which is high throughput and generalizable so that it can be applied across many samples. To date this has not been done in atomic resolution electron micrographs. In this talk we introduce machine learning techniques for high throughput analysis of local structural features in nanomaterials and apply these techniques to real micrographs.

As a proof of concept, our technique focuses on two tasks: semantic segmentation of the nanoparticle region from the micrograph and then a basic classification as to whether the projection contains a stacking fault. For the segmentation task we implement a CNN with a U-Net architecture [2]. This CNN we trained both with synthetic and real micrographs of CdSe quantum dots. Synthetic images are generated using the multislice method outlined by Kirkland [8]. Using F1-score to measure accuracy on the segmentation, we achieve an accuracy of 0.9 out of 1 on segmentation of synthetic images. However, we have found that neural networks trained only on this synthetic data fails when applied to real data. To

https://doi.org/10.1017/S143192761900148X Published online by Cambridge University Press

address this we have found that image preprocessing of real micrographs can dramatically improve results. Figure 1a and 1b show samples of successful segmentation on both synthetic and real data.

Once images are segmented they are then fed into a random forest classifier which is trained to determine whether a stacking fault is visible in the nanoparticle. The random forest was trained only on features generated from real images. Features included portions of the Fourier transform and Sobel edges. Using these features we achieved a cross validation score of 0.8 for determination of four classes: no particle (empty), no atomic column resolution (null), no stacking fault seen (no), and stacking fault present (yes). Figure 1c presents the confusion matrix for the random forest classifier showing the rates of success for each class. The combination of CNN segmentation and random forest classification demonstrate well the possibility of high throughput HRTEM micrograph analysis. The final goal of this work will be to use neural networks trained on simulated data on real images in order to gain statistical information on nanoparticle structure. In this talk we will discuss the two machine learning techniques previously discussed and their dependencies, as well as provide a demonstration of a high throughput analysis pipeline for HRTEM micrographs.

References:

[1] Zhao, S. *et al.* ACS Catal. **8** (2018) p. 4996–5001.
[2] Ronneberger, O., Fischer, P. & Brox, T. Lecture Notes in Computer Science (2015) p. 234–241.
[3] Shin, H. C. *et al.* IEEE Trans. Med. Imaging **35** (2016) p. 1285–1298.
[4] Krizhevsky, A., Sutskever, I. & Hinton, G. E. Adv. Neural Inf. Process. Syst. (2012) p.1–9.
[5] Szegedy, C. Nips (2013) p. 1–9 .
[6] Ziatdinov, M. *et al.* ACS Nano **11** (2017) p. 12742–12752.
[7] Madsen, J. *et al.* Adv. Theory Simul. **1800037** (2018) p. 1–12 .
[8] Kirkland, E. J. Advanced Computing in Electron Microscopy, (Springer, New York).
[9] 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. This work is supported by the NSF GRFP under Grant No. 1752814 and by STROBE: NSF-STC (DMR-1548294).

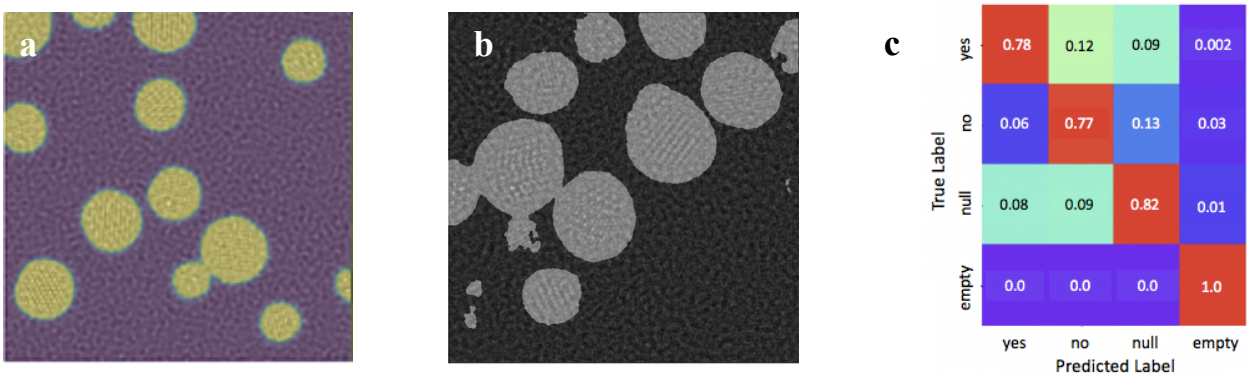

Figure 1. Results for segmentation and structure classification. a) Segmentation of a simulated HRTEM image. b) Segmentation of real image after preprocessing. c) Confusion matrix for random forest classification of stacking fault content, with the diagonal from top left to bottom right indicating percent correctly classified.

https://doi.org/10.1017/S143192761900148X Published online by Cambridge University Press