Convolutional Neural Network as a Solution to Segment and Classify High Resolution TEM Images to Obtain 3D Information

M. Leibovich[1], R. Manzorro[2], Mai Tan[2], S. Mohan[1], Adrià Marcos-Morales[1], C. Fernandez-Granda[1], P. A. Crozier[2,*]

1. Courant Institute of Mathematical Sciences and Center for Data Science, New York University, New York, NY

2. School for the Engineering of Matter, Transport, and Energy, Arizona State University, Tempe, Arizona

* Corresponding author: crozier@asu.edu

Analysis of nanoparticles has enjoyed a continuously increasing interest, with applications in catalysis, medicine or optoelectronics, to name a few. The physical and chemical properties of these particles rely on their exact 3D structure, and multiple approaches have been developed to extract this information. The most accurate and reliable procedure for retrieving the nanoparticle morphology is through scanning transmission electron microscopy (STEM) tomography, which performs well even at the atomic scale [1]. However, tomographic techniques require acquiring at least two different image projections from different zone axes, demanding a stable and stationary system. This limits the temporal resolution necessary to capture dynamics, which are critical to understanding the underlying functionality of interest. Other approaches use the intensity of a single STEM or TEM image, combined with very precise image simulations to extract 3D information [2]. Here too, temporal resolution is limited by the signal-to-noise ratios required in order to reliably measure the intensity of each atomic column. Deep neural networks have shown tremendous results in applications involving processing and analyzing high dimensional data. Specifically, convolutional neural networks (CNNs) were used in applications involving natural image data such as: image classification, segmentation, and object detection [3]. More recently, machine learning based algorithms has shown great success in processing other kinds of images, applied to medical imaging data, and images encountered in biology, chemistry and physics. Pioneering work on 3D shape reconstruction of gold nanoparticles using machine learning was first done on simulated TEM images [4], followed by experimental data [5]. Still, in the presence of low doses, i.e., poor signal-to-noise ratios, the reported results are not accurate.

We propose a semantic segmentation convolutional neural network-based retrieval algorithm for 3D atomic structure, even in the presence of strong noise. At present, we focus our initial efforts on CeO_2 nanoparticles due to their importance in catalytic energy conversion processes. The network can be applied to temporally resolved *in situ* image series providing information on the fluctuations of the structure with time. For example, in CeO_2, this allows us to track both cation migration and the time dependence of oxygen vacancy formation at different locations in a CeO_2 nanoparticle.

The general sketch of the training and evaluation process is described elsewhere [6]. A neural network, as showed in Figure 1, based on U-net architecture is trained using simulated images that captures the heterogeneities of dynamical nanoparticles, including variations in thickness, tilts, point defects, and modifications of electron-optical parameters. Poisson noise, which models the shot noise encountered experimentally on fast direct electron detectors, is added. The output of the network is matched to a mask based on the 3D structure of the nanoparticle, where the mask is optimized to produce stable predictions. Every pixel in the output is classified as either an atom or background, as well as the occupancy of the atomic column. The network parameters are trained by minimizing the output label misclassification error using a spatially weighted cross-entropy loss. The preliminary result on applying this neural network on experiment TEM data is showed in Figure 2.

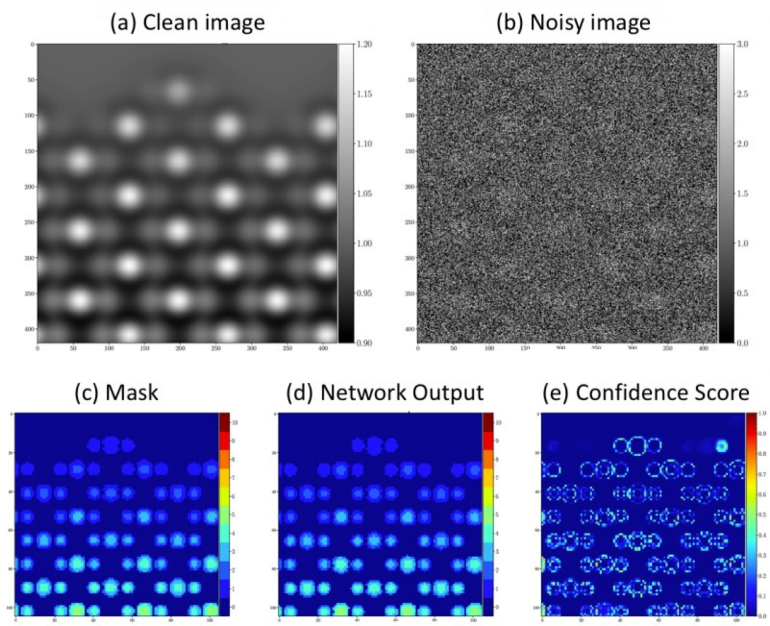

Figure 1. Simulated TEM images: (a) The clean image generated by the solver. (b) The network takes as an input a corrupted version where the value of every pixel $n_i \sim \mathrm{Pois}(c_i)$ is a Poisson random variable. c_i is the corresponding pixel value in the clean image. (c) The target mask. Atomic column depth serves as a label. Smoothing the target labels so that they are gradually changing to the background provides fewer ambiguous predictions. (d) The network's output. While the pixel classification accuracy is not 100% the network predicted all atomic columns values correctly. (e) Pixel valued confidence score. The confidence score is high for the background and column centers, and lower for areas on the boundary, the most susceptible to noise.

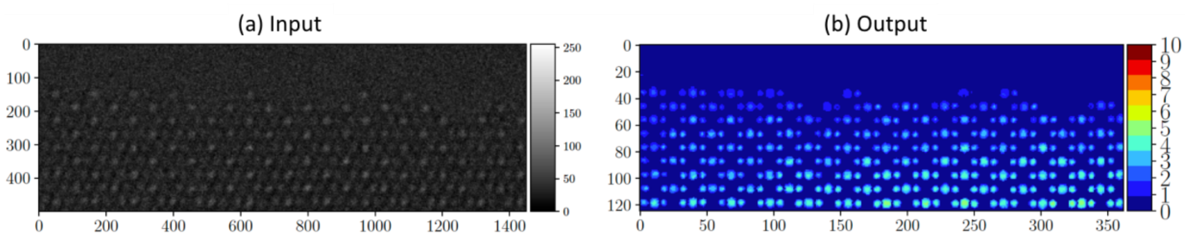

Figure 2. Results on *in situ* experiment data: The network applied to time-resolved TEM images of CeO_2 at the (110) surface in a [110] zone axis (sampled at 7.5 frames per second). (a) The input data with an estimated SNR of 1. (b) The network's output. The atomic column depth decreases as one reaches the surface.

REFERENCES:

[1] Van Aert, S. et al. Nature **470** (2011), p. 374-377. doi: 10.1038/nature09741
[2] Li, Z. Y. et al. Nautre **451** (2008), p. 46-48. doi: 10.1038/nature06470

[3] LeCun, Y., Bengio Y., and Hinton G. Nature **521** (2015), p. 436-444. doi: 10.1038/nature14539
[4] Madsen, J. et al. Adv. Theory and Sim. **1** (2018) p. 1800037. doi: /10.1002/adts.201800037
[5] Ragone, et al. Comp. Mat. Science (2020) p. 109722. doi: 10.1016/j.commatsci.2020.109722
[6] Manzorro, R. et al. Microsc. Microanal. 27 (2021) p. 464-465. doi: 10.1017/S1431927621002154
[7] The authors acknowledge funding from NSF (DMR 1840841, OAC 1940263 HDR 1940097, OAC 2103936, HDR 1922658), the Simons Foundation, and the use of facilities of Eyring Materials Center at Arizona State University.