Architecture Optimization and Interpretability in Neural Networks for HRTEM Segmentation

Catherine Groschner[1] and Mary Scott[2]

[1]University of California Berkeley, Berkeley, California, United States, [2]Lawrence Berkeley National Laboratory, Berkeley, California, United States

Machine learning and deep learning are becoming key elements of Transmission Electron Microscopy (TEM) analysis. Several pipelines have been demonstrated both in the realm of high resolution STEM as well as TEM [3,4]. Most rely on a primary segmentation step with many using the U-Net architecture proposed by Ronneberger and coworkers [2]. However, a plethora of architectures have been proposed for segmentation of real images. There has been an effort in the biomedical community to understand the impact of network architecture on segmentation [5,6]. However, this area is still just beginning to be developed for segmentation of TEM for materials science [1].

Due to the fundamental nature of the segmentation problem in microscopy we chose to analyze network architectures for the segmentation task. We analyze both the classic U-Net architecture as well as the FCN8s network which is based on the standard image classification network VGG-16 [7,8]. Traditional image classification networks are known to have learned features very similar to traditional image processing filters such as filters that respond much like an edge detection filter [9,10]. The FCN8s architecture then builds on these base features to make segmentation maps. We are able to develop this architecture using the pretrained VGG16 from Keras which benefits from prior training on the ImageNet dataset which is much larger than any available dataset for HRTEM [8]. We compare these two in order to understand whether the presence of features focusing on edge detection, and other traditional real image features help in the very noisy HRTEM case and whether we can leverage pre-learned features to push past the performance of U-Nets. We then also analyze the use of convolutional dilation in order to learn the effects of changing the scale over which features can be learned from for the network.

Beyond exploring which architecture yields best results, recent developments also allow us to probe neural networks and make performance more explainable [11,12]. This can not only help us improve network performance but learn other features of the data. In order to gain greater insights we explore using techniques which have been developed for network explainability. We analyze the feature space which leads to accurate predictions. These can be used to generate better training sets and to improve other networks beyond those for segmentation such as structure classification or defect identification.
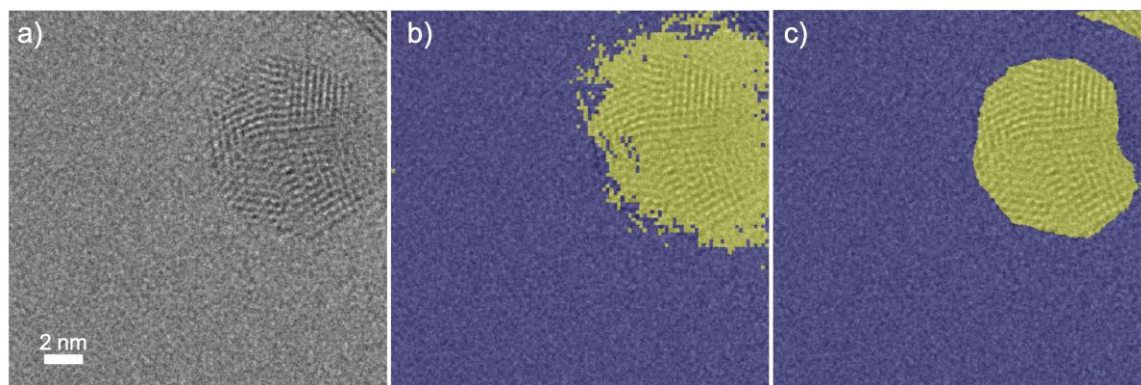
Figure 1. a) A sample micrograph of a gold nanoparticle. b) segmentation of the micrograph by the FCN8s network. c) Segmentation by the U-Net architecture

References

[1] Horwath, J. P. *et al*. Understanding Important Features of Deep Learning Models for Transmission Electron Microscopy Image Segmentation. *ArXiv* (2019).

[2] Ronneberger, O., Fischer, P. & Brox, T. Lecture Notes in Computer Science (2015) p. 234–241.

[3] Madsen, J. *et al.* Adv. Theory Simul. **1800037** (2018) p. 1–12

[4] Ziatdinov, M. *et al.* ACS Nano **11** (2017) p. 12742–12752.

[5] Karim, M. R. *et al*. *Proc. - 2019 IEEE 19th Int. Conf. Bioinforma. Bioeng. BIBE 2019* (2019). p. 415–422

[6] Drozdzal M. *et al*. Carneiro G. et al. (eds) Deep Learning and Data Labeling for Medical Applications. DLMIA 2016, LABELS 2016. Lecture Notes in Computer Science, **10008**. (2016).

[7] Shelhamer, E., Long, J. & Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *ArXiv* (2016).

[8] Simonyan, K. & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv* (2014).

[9] Chollet, Francois. *Deep Learning with Python*. Manning Publications, Shelter Island. (2017).

[10] Gonzalez-Garcia, A., Modolo, D. & Ferrari, V. *Int. J. Comput. Vis.* **126,** 476–494 (2018).

[11] Montavon, G., Samek, W. & Müller, K. R. *Digital Signal Processing* **73,** 1–15 (2018).

[12] Qin, Z., Yu, F., Liu, C. & Chen, X. How convolutional neural network see the world - A survey of convolutional neural network visualization methods. *ArXiv* (2018).