doi:10.1017/S1431927622011631
Microsc. Microanal. 28 (Suppl 1), 2022

Maximizing Neural Net Generalizability and Transfer Learning Success for Transmission Electron Microscopy Image Analysis in the Face of Small Experimental Datasets

Katherine Sytwu[1], Luis Rangel DaCosta[1,2], Catherine Groschner[2] and Mary C. Scott[1,2*]

[1.] Molecular Foundry, Lawrence Berkeley National Lab, Berkeley, CA, USA
[2.] Department of Materials Science and Engineering, University of California, Berkeley, Berkeley, CA, USA
* Corresponding author: mary.scott@berkeley.edu

The increasing ability to perform high throughput electron microscopy has created a need for robust, automated analysis that appears addressable by machine learning (ML) tools. Indeed, approaches such as convolutional neural nets (CNNs) are finding increased application in scientific data analysis tasks, including analysis of electron microscopy imaging data [1-3]. While fast detectors and automated imaging protocols may create datasets of tens of thousands of images, only a small subset of the experimental data can be feasibly manually labelled to create training data for a CNN. Therefore it is desirable to establish 1. what experimental variables should be included in a training dataset for robust performance and maximum accuracy across workflows, and 2. similarly, what are the best practices for establishing transfer learning workflows, where initial training is performed on large libraries of simulated data.

Here, we take the example of pixelwise segmentation of high-resolution transmission electron microscopy (HRTEM) data to evaluate best practices for data curation and to test the performance of trained CNNs on data with characteristics not found in the training data. Pixelwise image segmentation is an important example of an image analysis task for electron microscopy data, where each pixel is classified according to the structures it contains. Segmentation effectively divides images into regions of interest that can be further analyzed or classified, and is therefore a critical first step to many image processing pipelines [4]. Convolutional neural nets have shown high performance in segmentation tasks, but their behavior is difficult to interpret owing to their complex and nonlinear behavior. Specifically, it is difficult to predict where CNNs will fail to generalize, that is, where they fail to segment properly when shown data types not included in their training data. In electron microscopy, this could mean varying experimental parameters, such as radiation dose, sample geometry, or nanomaterial structural features. In the case of transfer learning from simulated data, simulation parameters must also be considered.

To understand how CNNs trained on experimental data generalize, we have performed a systematic study of CNN performance when trained and tested on data with varying experimental conditions and sample characteristics. We have explored the role of varying magnification, image dose, and defocus value, as well as sample characteristics such as nanomaterial size and substrate (Fig. 1). The results give insight into what types of features neural networks use to classify image pixels under different training strategies. Many trends, such as failure to generalize across magnifications, are not surprising, however other results illustrate the importance of specific types of dataset diversity and potential sources of bias in data analysis pipelines.

Curating experimental datasets of tens of thousands of labelled images is not possible for a typical researcher. However, fast and accurate electron microscopy image simulation protocols [5] enable generation of large simulated training datasets. These can be employed in a "transfer learning" strategy, that is, early training is performed on the simulated data and later training is performed on experimental data. Again taking the example of pixelwise segmentation of HRTEM images, we test generalizability and transfer learning strategies using simulated datasets (Fig. 2). We find that the segmentation accuracy of models trained on simulated datasets including variations in sample geometry also depends on the characteristics of the experimental test dataset. For example, it is intuitive that including smaller nanoparticles in the training data boosts performance on datasets containing small nanoparticles. However, while transfer learning can overcome poor initial performance, network training dynamics can vary significantly depending on the training and test data characteristics as well as the transfer learning strategy. By systematically investigating training dataset diversity and the success of various transfer learning approaches, we gain insight into reliable transfer learning strategies. We also monitor how initial learning is related to network generalizability. Taken together, our work has resulted in best practices to maximize CNN accuracy and generalizability in the limit of small experimental datasets, and has established how network performance dynamics during training relate to CNN accuracy and generalizability [6].

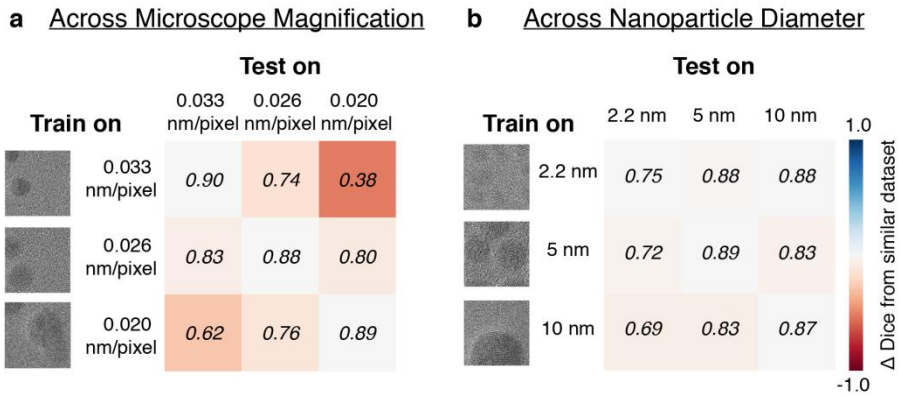

Figure 1: Generalizability of CNNs applied to experimental data. In these confusion matrices, the numerical Dice score is indicated for various combinations of training and test data. The colorbar indicates the deviation from the ideal training case (i.e., training and testing on the same dataset). (a) CNNs are not scale invariant, and it is therefore not surprising that they do not generalize across magnifications. (b) Training on images of differently sized nanoparticles indicates that CNNs trained on data from larger nanoparticles generalize the worst.

https://doi.org/10.1017/S1431927622011631 Published online by Cambridge University Press

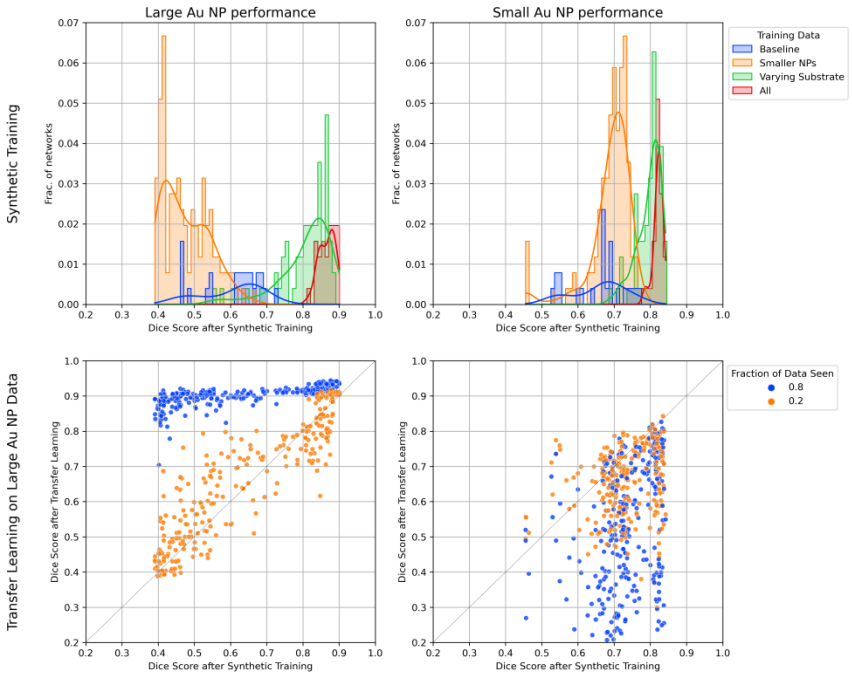

Figure 2: Generalizability and transfer learning performance of CNNs trained on synthetic data. Comparison of performance of CNNs trained purely on synthetic data then immediately tested on experimental data with varying nanoparticle size indicates that varying synthetic parameters such as nanoparticle and substrate geometry results in different training behavior and performance trends on different experimental datasets. Transfer learning results for Au nanoparticle data indicate performance suffers when experimental data contains particle size bias.

[1] L Ruoqian et al., Scientific reports **11**(1) (2021): p. 1.
[2] R Manzorro et al., Microscopy and Microanalysis **27**(S1) (2021): p. 464.
[3] A Khan et al., Bulletin of the American Physical Society **65** (2020).
[4] CK Groschner, C Choi and MC Scott, Microscopy and Microanalysis **27**(3) (2021): p. 549.
[5] L Rangel DaCosta et al, Micron **151** (2021): p. 103141.

[6] 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.

https://doi.org/10.1017/S1431927622011631 Published online by Cambridge University Press