Two-stage Neural Architecture Search for Microscopy Image Segmentation

Matthew Guay[1*], Zeyad Emam[2] and Richard Leapman[1]

[1.] National Institute of Biomedical Imaging and Bioengineering, NIH, Bethesda, USA.
[2.] AMSC Department, University of Maryland, College Park, USA.
* Corresponding author: matthew.guay@nih.gov

Modern microscopy hardware is capable of acquiring gigabytes or terabytes of 2D, 3D, and 4D data in short spans of time [1]. The rapid growth in dataset size necessitates the automation of common image analysis tasks in order to fully leverage the information that the data provides. One fundamental image analysis task is *segmentation* - the grouping of pixels into regions corresponding to image content. In the past decade, automated segmentation algorithms based on deep neural networks have made significant progress on automating difficult microscopy segmentation problems, such as those encountered in large-scale biological electron microscopy using imaging platforms such as *serial block-face scanning electron microscopy* (SBF-SEM) [1]. However, challenges remain for bolstering performance to levels of accuracy required for practical, turn-key solutions for biomedical research. One challenge is identifying neural network architectures best suited for a given segmentation problem. Large neural networks such as those used for image segmentation may require dozens of design choices to fully specify a computational module, and multiple modules may be combined to form a final network architecture. Searching through the combinatorial collection of *hyperparameters* which quantify these design choices is computationally expensive, even with a cluster of GPU-enabled compute nodes available. Naive search strategies such as grid search or random search at the level of individual network layers may fail to discover architectures that are more useful than those designed by hand via human intuition.

In previous work, we determined that randomized search strategies may be modified to discover high-performing segmentation network architectures using modest computational budgets. Encoder-decoder neural networks were recast as a hierarchy of simpler modules, and by modifying architecture hyperparameters at each level of this hierarchy, we were able to discover diverse architectures that met or exceeded the performance of manually-designed networks [2]. This strategy is useful, but further gains can be realized by applying a second search strategy to optimization hyperparameters which have no hierarchical structure. For this we used the Python `blackbox.py` library, which uses efficient high-dimensional parameter sampling and a loss surface approximation algorithm to minimize the hyperparameter optimization objective function using a relatively small number of evaluations [3], performed in parallel across multiple compute nodes on the NIH Biowulf high-performance computing system.

For this paper, the first stage of our neural architecture search used hierarchical random search to generate 2D encoder-decoder segmentation networks, using the same search parameters as in [2]. The best network discovered in Stage 1 achieved a mean intersection-over-union (mIOU) of 0.585, a metric we have found to best correspond to perceptual image segmentation quality. The architecture of this network was used as the basis for Stage 2 of the search, in which 7 of the architecture's optimization hyperparameters (learning rate schedule, momentum weights) and regularization weight hyperparameters were optimized using `blackbox.py`. The best architecture found in Stage 2 of the search achieved a mIOU of 0.609, improving qualitatively and quantitatively over its predecessor in

Stage 1, as indicated in Figure 1. However, one subtle point must be handled when comparing segmentation network architectures - performance is dependent on the random initialization of the trainable weights. To determine whether performance differences were due to hyperparameter choices or chance, we retrained the best Stage 1 network 25 additional times to measure the variability in performance due to initialization. The best retrained Stage 1 network had a mIOU of 0.599, and its segmentation is displayed below in Figure 1. This evidence supports the conclusion that performance improvements in Stage 2 stem at least in part from the enhanced architecture search, despite focusing on a relatively narrow collection of possible hyperparameters. In future work, we will extend this enhanced architecture search to a larger hyperparameter collection, and integrate the technique into the development of ensembles of larger networks that we use for more accurate segmentation for biomedical microscopy.

References:

[1] CR Pfeifer et al., Journal of Structural Biology **189.1** (2015), p. 44.
[2] M Guay et al., 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018).
[3] P Knysh and Y Korkolis in "Blackbox: A Procedure for Parallel Optimization of Expensive Black-Box Functions." arXiv preprint arXiv:1605.00998 (2016).
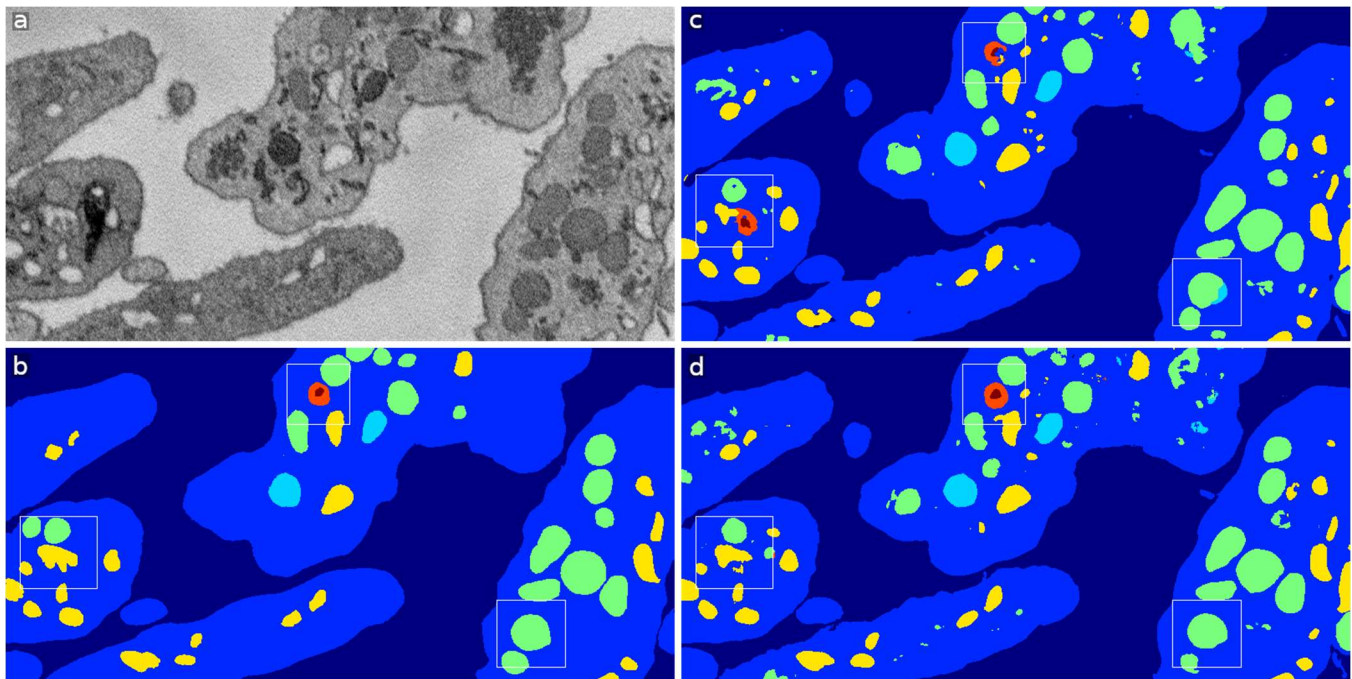
Figure 1. Comparison of the best Stage 1 and Stage 2 networks with image data and ground-truth labeling, with points of comparison highlighting significant differences between the networks. The Stage 2 (hierarchical random search plus blackbox optimization) network does a better job at properly labeling organelles, especially the uncommon red organelle. (**a**) Sample SBF-SEM image of platelet cells. (**b**) Ground-truth segmentation labels. (**c**) Best Stage 1 network after 25 retraining attempts. (**d**) Best Stage 2 network.