doi:10.1017/S1431927622011436
Microsc. Microanal. 28 (Suppl 1), 2022

A Comparison of 3D and 2D U-Net Convolutional Networks for Segmentation in FIB-SEM Imagery

Rengarajan Pelapur[1]*

[1.] Thermo Fisher Scientific, Materials and Structural Analysis, Houston, TX, USA.
*Corresponding author: rengarajan.pelapur@thermofisher.com

In this work, we are demonstrating the differences in 3D [1] and 2D convolutional U-Net [2] networks by showing results of training models on a publicly available dataset [3]. For specific segmentation challenges the 3D image information or morphology of target structures can be captured with 3D convolutions. This could be a powerful discriminative method that would have advantages over performing 2D convolutions. The dataset [3] chosen is a 5x5x5 (x, y and z axis) micrometer sized section of a CA1 hippocampus region of a mouse brain. The resolution of acquisition of this sample was 5x5x5 (x, y and z axis) nanometer using a Focused Ion Beam Scanning Electron Microscope (FIB-SEM). The target problem is to segment mitochondrial structures which are typically 0.5-10 microns in diameter. The image processing challenge with such images is usually the lack of discriminative features such as shape, texture, edges or grayscale intensities that are singularly insufficient to describe difficult structures such as mitochondria. Each slice of image data contains ultra-structures that are complex in terms of grayscale variations, edge features and overall appearance. These ultra-structures would be considered as ‘distractors’ and for traditional machine learning approaches a feature selection with carefully tuned and picked multi-scale feature set is required. This requires expertise in computer vision and image processing that not many domain experts possess. Manually marked data is hard to collect, share, time consuming and laborious [4]. Our work builds up on the frameworks introduced to perform visualization, interactive segmentation and machine learning tools [5, 6]. The goal of the article is to share preliminary results of using 3D U-Net convolutional networks and comparing the results that one can get with 2D U-Net convolutional networks. We have used Amira 3D application for 3D visualization, 2D U-Net with resnet18 [7] backbone. The 3D U-Net convolutional network has been customized and implemented with a resnet18 [7] backbone. Furthermore, we have used python scripts available at https://xtras.amira-avizo.com/xtras/image-segmentation-evaluation-using-standard-metrics-python for the evaluation metrics and quantification of segmentation quality in terms of Jaccard Index, F1 score, Precision and Recall.

The authors [3] provide annotated mitochondria in two sub-volumes: one for testing and another for training. Each sub-volume consists of the first 165 slices of the 1065x2048x1536 pixels of the CA1 hippocampus region of the brain at a spatial resolution of 5x5x5 nm. The two sub-volumes are different regions of interest within the larger volume in order to test true performance for unseen data. We train the 2D convolutional U-Net with 1000 epochs while the 3D convolutional U-Net with only 25 epochs. 3D convolutional U-Net networks are computationally more expensive and per epoch the time taken on similar hardware is about 8-10 times of that of one epoch of the 2D convolutional U-Net network training. We chose similar parameters for both batches of training except for the number of epochs in order to show that in a short number of epochs we can achieve similar results using the 3D convolutional networks in comparison to the 2D implementation.

Preliminary quantitative results are being shared to show the difference between the 2D convolutional U-Net and 3D convolutional U-Net implementation. We have done our best to minimize the difference in parameters for practical use but do acknowledge that the implementations are different. The number of layers, backbone, loss functions and patch sizes have been frozen across the two implementations. Both training and testing for the two implementations were run on a CUDA enabled NVIDIA GPU Quadro P2000 with 4GB of GPU memory. Overall, the 3D U-Net was trained in 31 minutes (75 seconds per epoch) for 25 epochs while the 2D U-Net was trained in 383 minutes (23 seconds per epoch) for 1000 epochs. Our initial results show lower overall numbers but Accuracy and Recall numbers were within 6% for a savings of 92% of runtime. More results will be generated for a more comprehensive overview and shared with the community.

Metrics	2D U-Net (1000 epochs)	3D U-Net (25 epochs)
Precision	0.5062	0.3146
Recall	0.9734	0.9224
F1 Score	0.6660	0.4692
Jaccard Index/Accuracy	0.9481	0.8892

Table 1. Precision, Recall, F1 Score and Jaccard Index for 2D U-Net and 3D U-Net preliminary results. The resnet18 backbone and similar parameters were used except for the maximum number of epochs. Numbers show that the 3D U-Net can generate very similar levels of accuracy and recall with only a fraction of runtime during training.

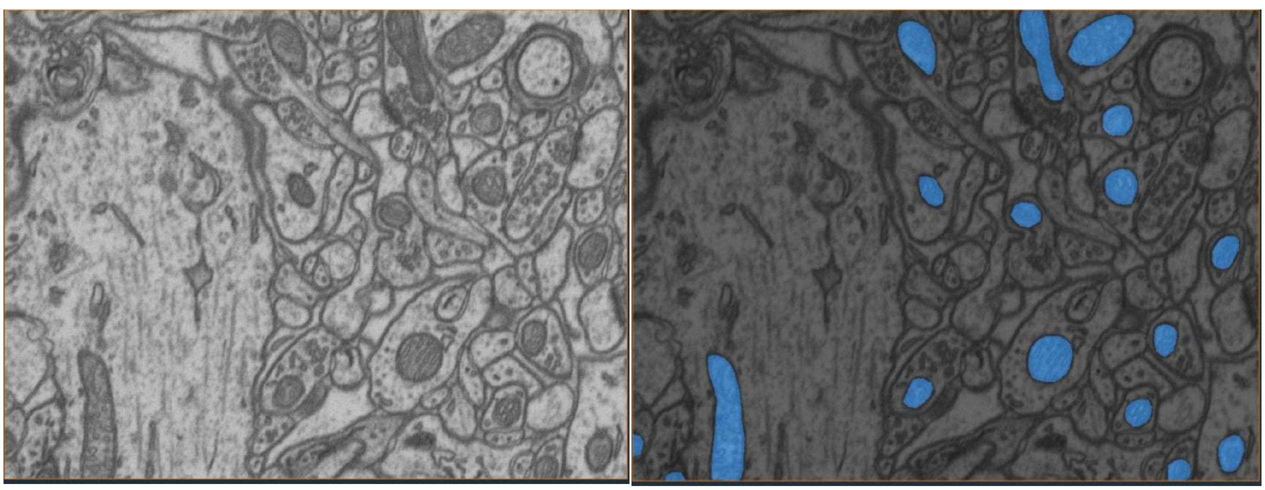

Figure 1. (L) Grayscale data showing training sub-volume from a region of interest from the larger dataset and (R) showing manually annotated ground-truth of mitochondria on the same slice in blue.

https://doi.org/10.1017/S1431927622011436 Published online by Cambridge University Press

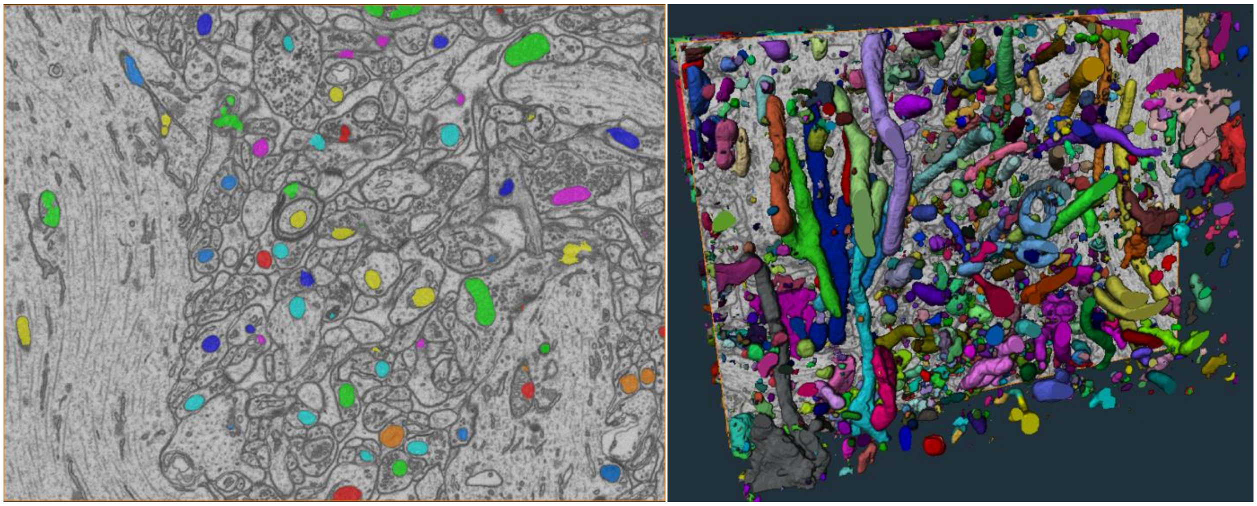

Figure 2. (L) Full resolution grayscale data showing detected mitochondria with singularly repeating colors denoting separated mitochondrial structures (R) showing image with 3D segmented and separated mitochondrial structures in the 1065x2048x1536 pixels image. Detection is a result of binarized output of the U-Net trained model prediction output followed by a 3D connected components algorithm to perform labeling

References:

[1] Ö Çiçek et al., International conference on medical image computing and computer-assisted intervention (Springer, Cham) 2016.
[2] O Ronneberger, P Fischer and T Brox, International Conference on Medical image computing and computer-assisted intervention (Springer, Cham) 2015.
[3] Mitochondria Detection in EM Stacks, cvlab.epfl.ch/data/data-em/
[4] A Lucchi, Y. Li and P Fua, Conference on Computer Vision and Pattern Recognition (CVPR), (Portland, Oregon, USA) 2013. DOI : 10.1109/Cvpr.2013.25
[5] R Pelapur and MM Heck, M&M (2019).
[6] R Pelapur, M&M (2020).

[7] K He, Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.