Learning to Correct Sloppy Annotations in Electron Microscopy Volumes Supplementary

Minghao Chen[1], Mukesh Bangalore Renuka[2], Lu Mi[1], Jeff Lichtman[2], Nir Shavit[1], Yaron Meirovitch[2]
MIT[1], Harvard University[2]

Abstract

Here we provide the supplementary material for the paper "Learning to Correct Sloppy Annotations in Electron Microscopy Volumes." In the paper we propose a new method for the seeded segmentation problem and demonstrate its effectiveness on challenging tasks in electron microscopy connectomics. Our algorithm, EMProof, is based on a neural network which simultaneously expands all seeds, skeletons or imperfect segmented objects within the image, and produces an improved instance segmentation. We provide more details on the experiments conducted in this work and display additional image results. We also provide additional experiments which show that EMProof can be used to learn to track and densely segment objects across sections which demonstrates its ability to learn features of neurons morphology.

1. Overview

Our paper presents an end-to-end pipeline, EMProof, for seeded image segmentation. Unlike previous methods which combine affinity map computation with graph optimization, EMProof operates in one step. Our method leverages the idea of *Cross Classification Clustering* (3C) [6], which solves an instance segmentation problem with traditional classifications. EMProof build on the logarithmic running time of 3C to provide a scalable solution for seeded instance segmentation which is demonstrated on difficult segmentation problems from connectomics.

In section 2, we provide details on the usage of 3C in EMProof. In section 3, we provide the effectiveness of EMProof in its ability to track objects across sections of the dataset (not shown in the main paper). In section 4, we demonstrate the effectiveness of EMProof on the seeded segmentation problem by presenting few examples on the challenging CREMI dataset [1], corroborating the results from the main paper. We compare our pipeline to highly accurate methods such as Learned Random Walker [2] and Learned Watershed [8]. In section 5, we present further examples on the AC3 dataset to display the super-resolution capability of EMProof. In section 6, We demonstrate scalability of EMProof and its performance on a large-scale somatosensory dataset [4] which contains $\sim$ 80 GB of raw images.

2. Cross-Classification Clustering (3C)

The highlight of Cross-Classification-Clustering (3C) is its effective and simple extension of previous single-object classification methods for connectomics, like FFN [3] and MaskExtend [5]. This strategy largely reduces run-time from linear time of the original solutions to logarithmic time in the number of objects. This is a non-trivial task for a neural network as there is quite little semantic meaning in neuronal cross-sections. In effect, as we explain below, sequences of labels given for each object are required to provide sufficient guidance and supervision for the the network to "know" how to map the labels of objects from the source image to their similar and respective neurons in the target image.

As shown in fig. 1, we use k different colors to label different neuron skeletons. This is done arbitrarily and uniquely at random. We repeat this procedure l times, where l is not larger than the logarithm of the number of skeletons in the input image (determined per image in during inference). We then have l different skeleton images, for each of the l iterations. This can be thought of as a repeated colorization procedure. The collection of l images defines the neurons in a unique way, in exact correspondence to their original IDs (which are realized as sequences of labels). Then the skeleton images are fed to the neural network with their respective EM images. The iterated calls to the network will generate a collection of expanded skeleton images, for each of the l images. Each expanded image will use k colors, but together those images define as a sequence (per pixel) the original IDs. This is the aggregation step as defined in fig. 1. We set k=4 for our experiments. If we consider n to be the number of distinct neuron IDs in the stack, then we only need to call the network $\log_4(n)$ times to get the final instance segmentation. Therefore, this strategy largely reduces the time when compared to single object classification methods, like FFN [3] and MaskExtend

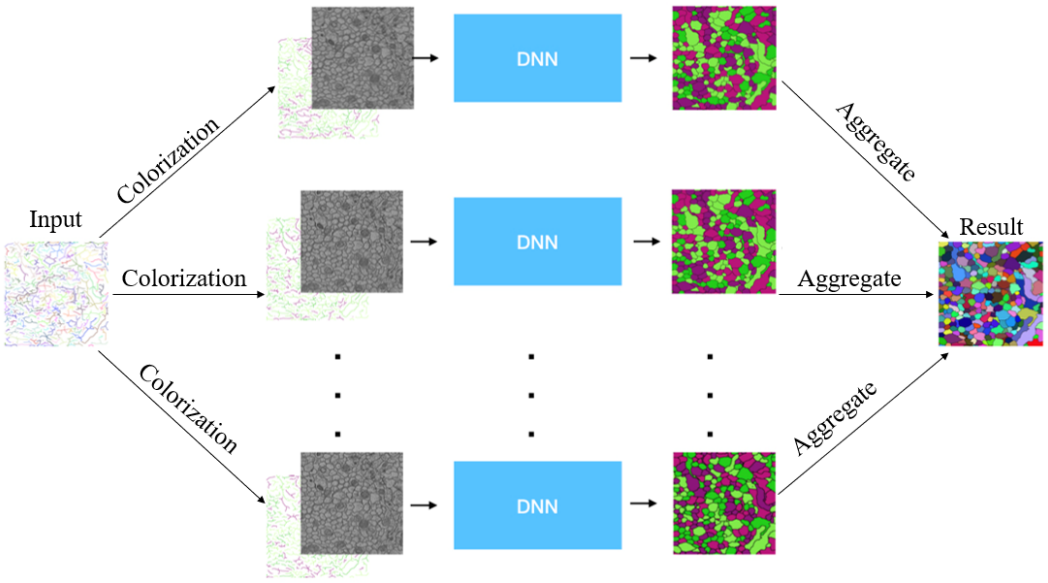

Figure 1. Overview of the internal workings of 3C [7] within EMProof.

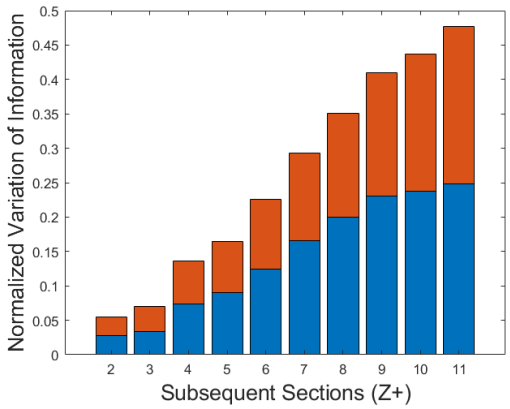

Figure 2. Tracking capability of EMProof. An instance segmentation of the source image (z=1) is propagated to subsequent sections and the Variation of Information is computed. EMProof can be used to propagate segmentation across three adjacent sections with marginal loss in accuracy, offering additional 4x speedup in reconstruction time.

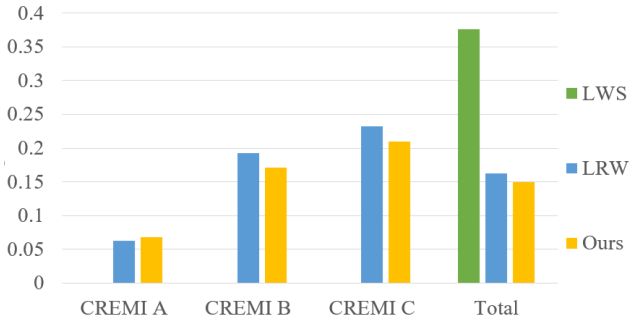

Figure 3. Comparison between the accuracy (VI) of EMProof and the previous state-of-the-art seeded segmentation algorithms Learned Random Walker [2] and Learned Watershed [8].

[5], while preserving their accuracy [6]. In turn, as we show in this paper, 3C allows also to improve the state of the art of the seeded segmentation problem.

3. Tracking Experiments

In this section we test the tracking capabilities of EMProof. Tracking capability refers to the ability of a pipeline to track objects from one part of the dataset to a different part of the dataset that is farther away from it in the image stack. A capacity of the network to track objects would result in improvements in reconstruction time as it allows users skip some sections during the annotation process. Moreover, an ability to segment an image based on skeletons of another image requires the network to learn semantic and morphological features of the neuronal objects. In this experiment, Tracking is always performed from the first image and is not done recursively.

As demonstrated by fig. 2, the Variation of Information of reconstructed images increases proportionally with the distance from the source image. Fig 4 shows a few of the results obtained by our tracking experiment. We see that the predictions deteriorate after section 4. We observe that qualitatively plausible results are obtained for three consecutive images. The skipping of three images would yield a speedup of 4x in reconstruction time. Combined with the 8x improvement in running time of EMProof compared with a human annotator (without tracking), we anticipate that a sloppy annotator can produce reconstructions that are com-

parable to an accurate annotator but up to 30x faster.

4. Seeded segmentation comparisons

In this section we present some further examples of the seeded segmentation capabilities of EMProof. Fig. 3 compares the Variation of Information scores between EMProof and current state of the art techniques. We also present few examples of EMProof in action in fig. 5 where we show some results of EMProof on the CREMI [1] dataset both on the easier section (section A) as well as the much harder section (section C) of the dataset. The examples shown and comparisons use round seeds as shown in figs.5 and 6 to ensure fair and equal comparisons between techniques. The performance of our pipeline on the harder image set of the dataset (Section C) in fig. 6 demonstrates the robustness of our pipeline. EM Images from section C are considerably more blurry and are often comprised of thin objects as demonstrated in the paper's main text.

5. Result on Super-resolution

We provide more results for the super-resolution experiment used on the public electron microscopy volume AC3 of [4]. Traditional SR tasks often deals with predicting a higher resolution version of an image, utilizing a model which captures both global and local image features. Our SR is defined by the requirement to enhance the resolution of an instance segmentation. We learn to map the pair of EM image (a raw image) and its instance segmentation, to a higher quality segmentation. We test the capability of EMProof on SR, by artificially downsampling true instance segmentation images, computing their skeletons, and then requiring EMProof to reverse those skeletons and produce the the high quality objects. This is a useful approach for connectomics because in many cases computing the high accuracy segmentation is delayed due to costs associated with running at high resolution. In those cases, EMProof can learn to improve the geometry of the existing results without needing to re-invoke the original pipeline, which could be highly expensive. Furthermore, this demonstrates that a human annotator can segment the image at low resolution and EMProof will take care of the task of "snapping" the human segmentation to the high resolution images.

Fig. 7 depicts the results of the super-resolved images on AC3. The results are nearly indistinguishable from the GT, where in many cases the network produces neater boundaries compared with the manual segmentation.

6. Skeletons to Volumes: EMProof performance tests using S1 dataset of [4]

EMProof was tested on the Somatosensory cortex dataset of [4] (*S1*) which consists of 1840 images each being $16{,}000 \times 16{,}000$ pixels. The first experiment was described in the main text where EMProof was required to improve the resolution of a single object mask given as binary image-stack (MIP-level 4, i.e. downsampled by a factor of $2^4{=}16$ in each X,Y dimension). EMProof was trained to improve the resolution by using respective raw images and object's 2-D skeletons. The prediction for this task took around 20 minutes. We now report on a second experiment on that dataset in which EMProof was required to reverse the skeletonization of all the objects in the dataset. Fig. 8 depicts the results of the reconstruction of intricate objects from the dataset, showing dendrites and their spines. It takes around 8 hours to predict the whole dataset. EMProof was able to recover the morphology of the objects in 3-D by accurately expanding skeletons, even in cases where objects branch are extremely thin.

References

[1] Cremi dataset. https://cremi.org/data/. 1, 3

[2] Lorenzo Cerrone, Alexander Zeilmann, and Fred A Hamprecht. End-to-end learned random walker for seeded image segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 12559–12568, 2019. 1, 2

[3] Michał Januszewski, Jörgen Kornfeld, Peter H Li, Art Pope, Tim Blakely, Larry Lindsey, Jeremy Maitin-Shepard, Mike Tyka, Winfried Denk, and Viren Jain. High-precision automated reconstruction of neurons with flood-filling networks. *Nature methods*, 15(8):605, 2018. 1

[4] Narayanan Kasthuri, Kenneth Jeffrey Hayworth, Daniel Raimund Berger, Richard Lee Schalek, José Angel Conchello, Seymour Knowles-Barley, Dongil Lee, Amelio Vázquez-Reina, Verena Kaynig, Thouis Raymond Jones, et al. Saturated reconstruction of a volume of neocortex. *Cell*, 162(3):648–661, 2015. 1, 3, 6

[5] Yaron Meirovitch, Alexander Matveev, Hayk Saribekyan, David Budden, David Rolnick, Gergely Odor, Seymour Knowles-Barley Thouis Raymond Jones, Hanspeter Pfister, Jeff William Lichtman, and Nir Shavit. A multi-pass approach to large-scale connectomics. *arXiv preprint arXiv:1612.02120*, 2016. 1, 2

[6] Yaron Meirovitch, Lu Mi, Hayk Saribekyan, Alexander Matveev, David Rolnick, and Nir Shavit. Cross-classification clustering: An efficient multi-object tracking technique for 3-d instance segmentation in connectomics. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8425–8435, 2019. 1, 2

[7] Yaron Meirovitch, Lu Mi, Hayk Saribekyan, Alexander Matveev, David Rolnick, and Nir Shavit. Cross-classification clustering: An efficient multi-object tracking technique for 3-d instance segmentation in connectomics, 2019. 2

[8] Steffen Wolf, Lukas Schott, Ullrich Kothe, and Fred Hamprecht. Learned watershed: End-to-end learning of seeded segmentation. *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct 2017. 1, 2

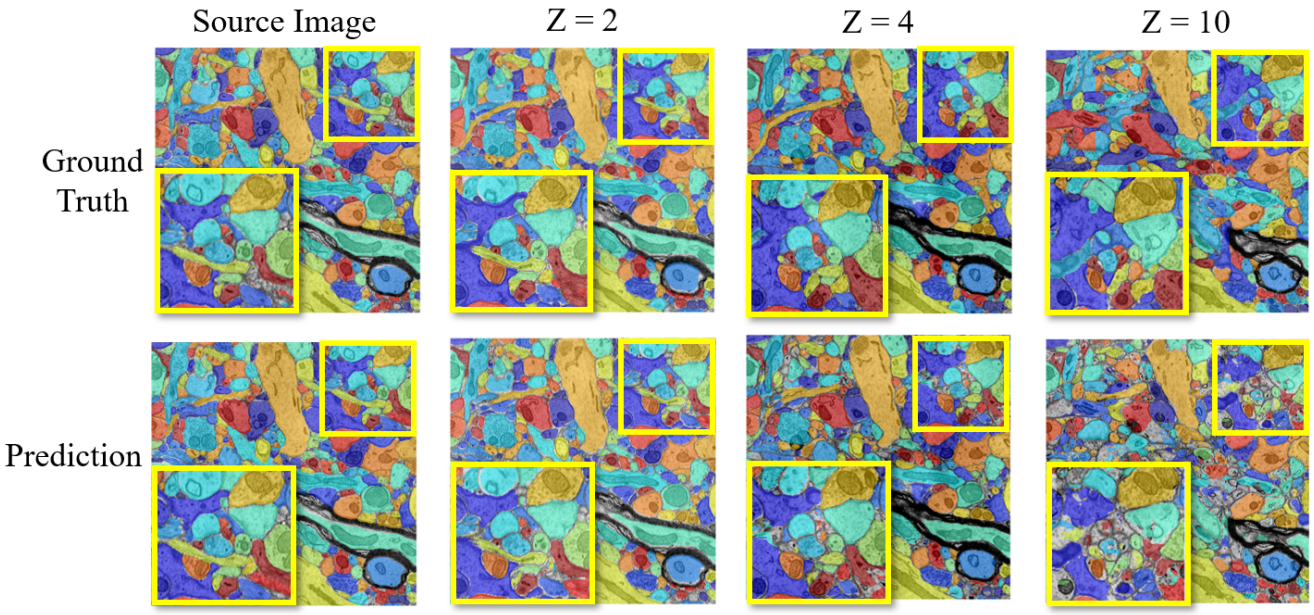

Figure 4. Examples of the tracking ability of EMProof. The source image is represented in the first column. The subsequent columns are predicted segmentation of images $Z{=}\{2, 4, 10\}$ away from the source image in the image stack. We observe that for $Z{=}4$, the segmentation results are qualitatively acceptable (VI are provided in fig. 2.

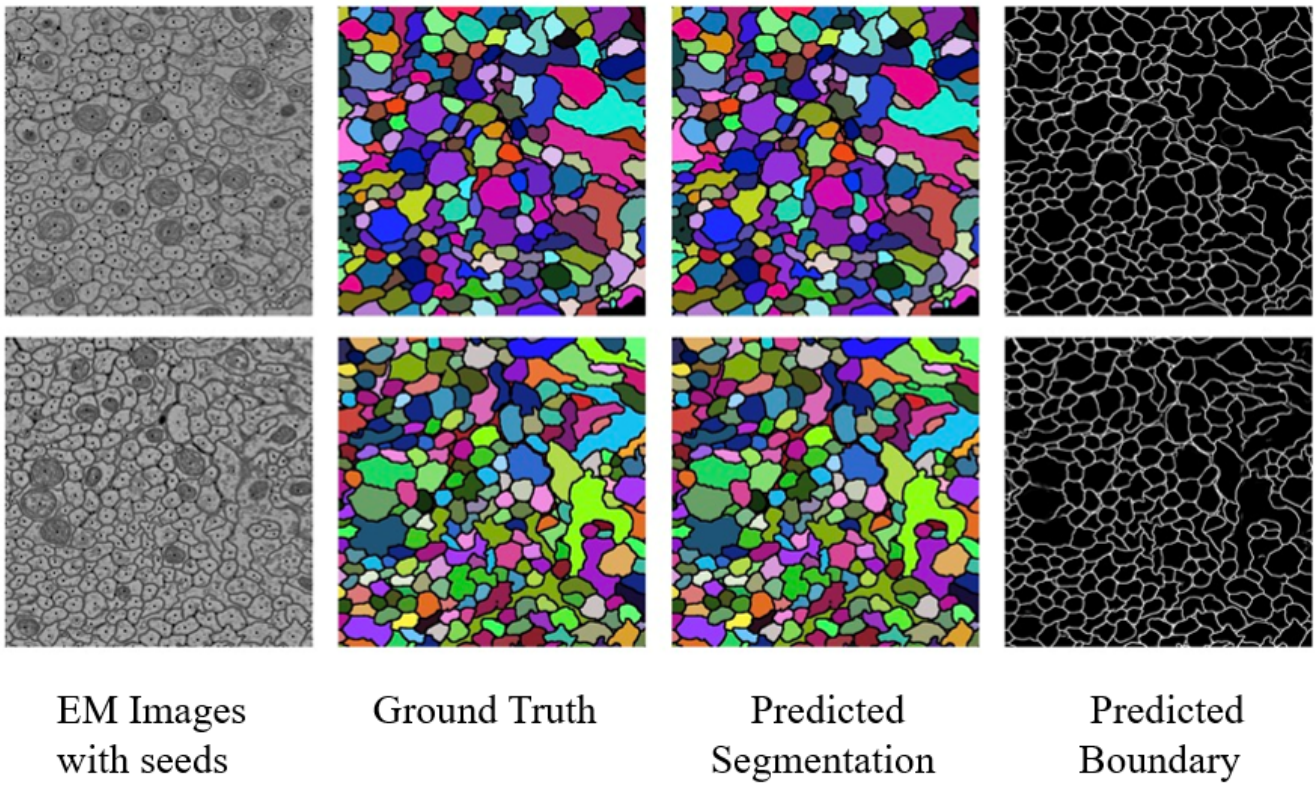

Figure 5. Example result on the CREMI dataset A. Left column shows EM images with super-imposed round seeds. EMProof expands the small seeds to meet the true neuronal boundaries, qualitatively matching the ground truth.

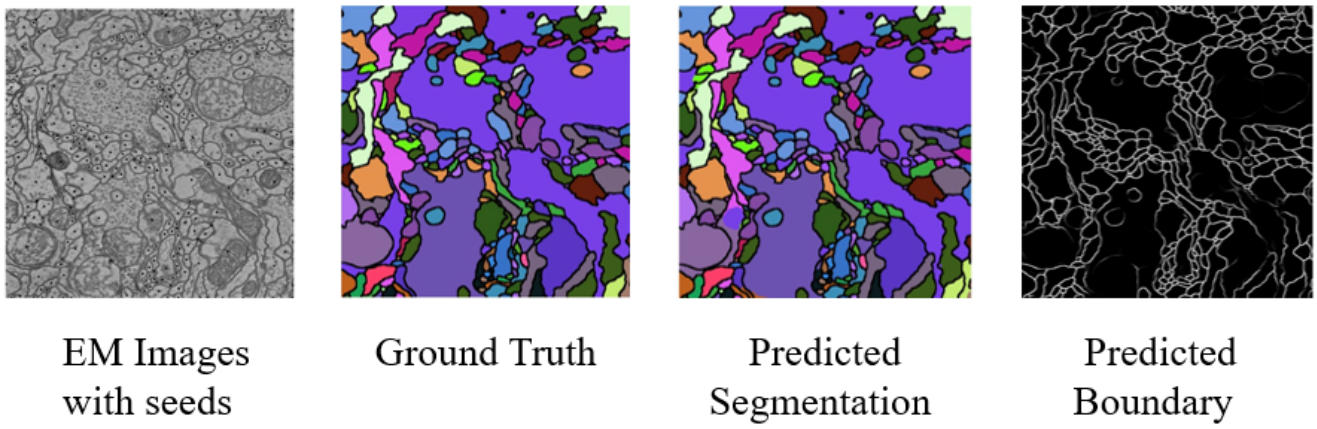

Figure 6. Example result on the challenging CREMI dataset C. Left column shows EM images with super-imposed round seeds. EMProof provides a qualitatively good seeded segmentation of the image, while producing a small number of errors near to the image boundaries (where image context is limited) as well as on one sub-cellular element (mitochondrion) at the bottom left of the electron microscopy image.

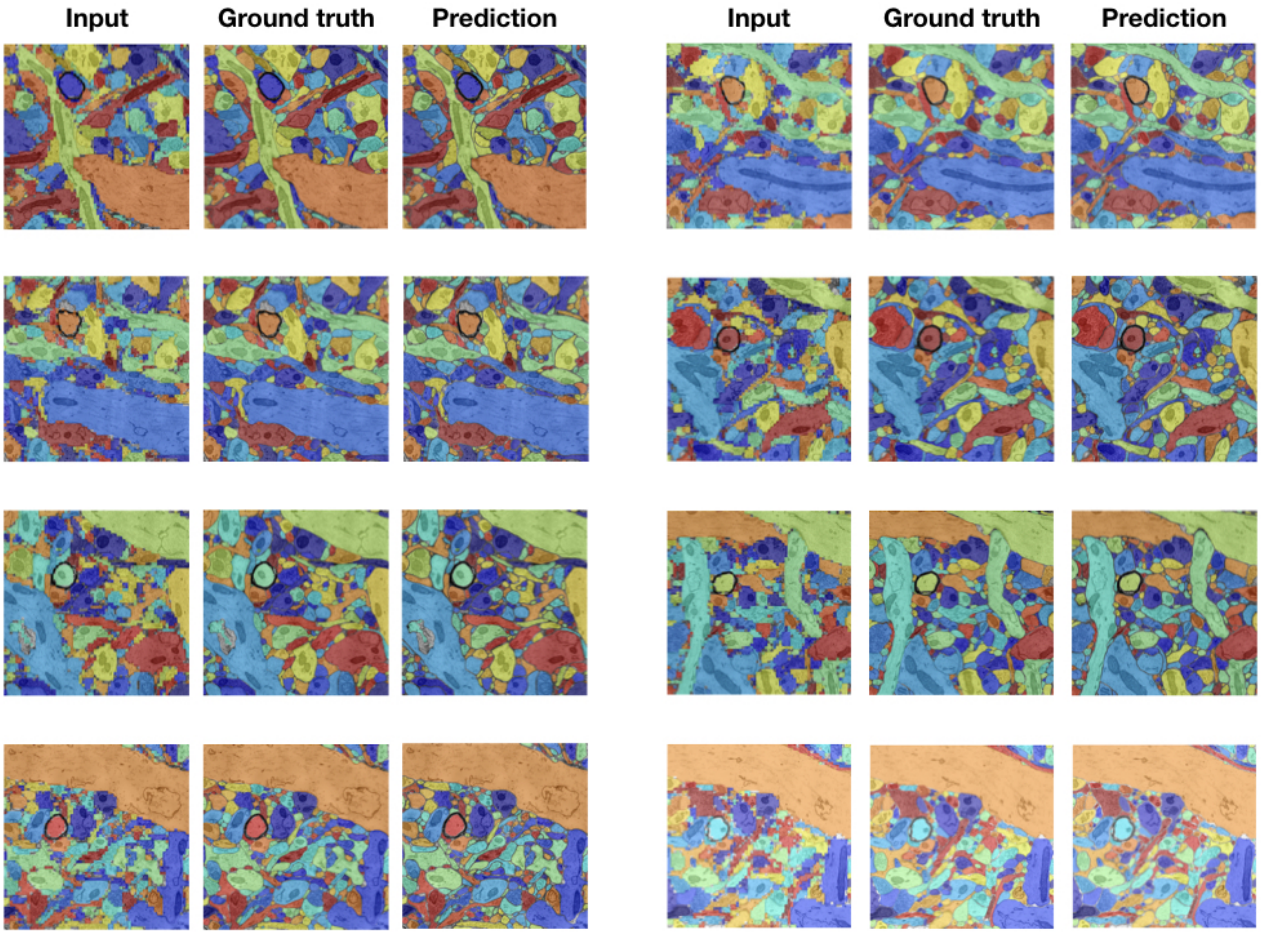

Figure 7. Super-resolution task. The first and fourth columns represent the low-resolution input by nearest-neighbor up-sampling of the AC3 dataset at MIP level 4 (downsampling of 2^4). Our EMProof method takes the low-resolution labeled images and skeletonize them first. Then, EMProof predicts the high-accuracy and high-resolution version image (third and sixth columns). The second and fifth columns show ground truth. Other examples and accuracy measurements (VI) are provided in the main paper.

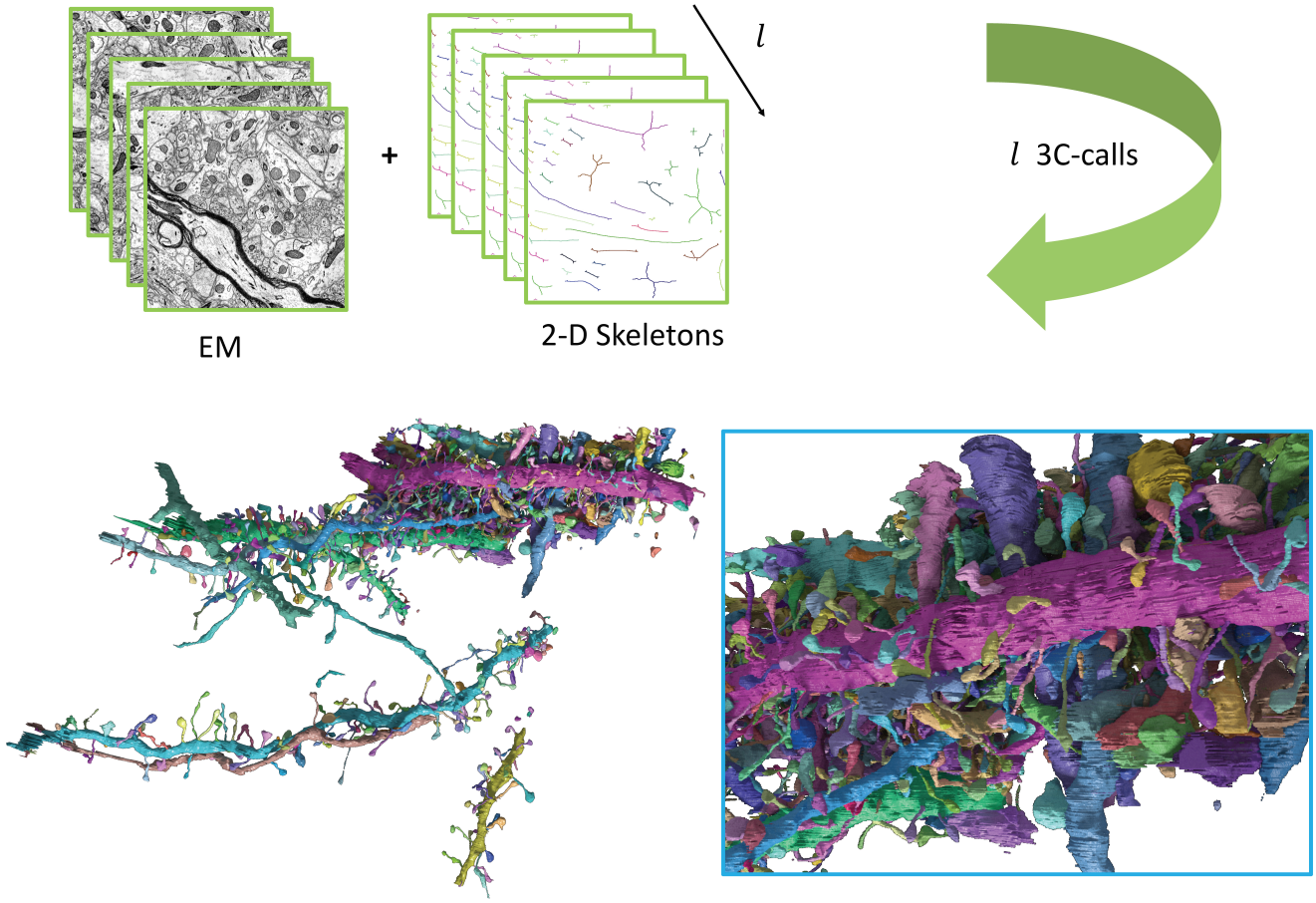

Figure 8. Skeletons to Volumes. The time performance and the quality of EMProof was tested on a large scale dataset [4]. EMProof was required to simultaneously reverse the skeletonization of all the objects in the dataset (one image at a time). EMProof was able to recover the morphology of the objects in 3-D by accurately expanding the skeletons, even in cases where objects branch and/or are extremely thin. Qualitatively, EMProof is able to effectively reverse the 2-D skeletons and reproduce the correct 3-D shapes of the neurons in the dataset. It takes around 8 hours to predict the whole dataset.