Synapse-Aware Skeleton Generation for Neural Circuits

Brian Matejek[1], Donglai Wei[1], Xueying Wang[2], Jinglin Zhao[2], Kálmán Palágyi[3], and Hanspeter Pfister[1]

[1] John A. Paulson School of Engineering and Applied Sciences
Harvard University, Cambridge, MA, USA
[2] Center for Brain Science, Department of Molecular and Cellular Biology
Harvard University, Cambridge, MA, USA
[3] Department of Image Processing and Computer Graphics
University of Szeged, Szeged, Hungary
bmatejek@seas.harvard.edu

Abstract. Reconstructed terabyte and petabyte electron microscopy image volumes contain fully-segmented neurons at resolutions fine enough to identify every synaptic connection. After manual or automatic reconstruction, neuroscientists want to extract wiring diagrams and connectivity information to analyze the data at a higher level. Despite significant advances in image acquisition, neuron segmentation, and synapse detection techniques, the extracted wiring diagrams are still quite coarse, and often do not take into account the wealth of information in the densely reconstructed volumes. We propose a synapse-aware skeleton generation strategy to transform the reconstructed volumes into an information-rich yet abstract format on which neuroscientists can perform biological analysis and run simulations. Our method extends existing topological thinning strategies and guarantees a one-to-one correspondence between skeleton endpoints and synapses while simultaneously generating vital geometric statistics on the neuronal processes. We demonstrate our results on three large-scale connectomic datasets and compare against current state-of-the-art skeletonization algorithms.

Keywords: neural circuits · connectomics · skeleton generation.

1 Introduction

Acquisition techniques [17], automatic segmentation methods [5], and synapse detection strategies [3] in connectomics have all progressed rapidly in the last decade, yielding densely reconstructed volumes at nanometer resolution. These terabyte and petabyte volumes contain hundreds of thousands of interconnected neurons and millions of synaptic connections. Despite the rich detail in the reconstructed 3D volumes, most analysis of this data occurs at a very coarse level [4].

Little research has focused on generating accurate wiring diagrams from the raw reconstructions. Current approaches [5] directly use an off-the-shelf skeletonization method to reduce these volumes into a series of nodes (neurons) and

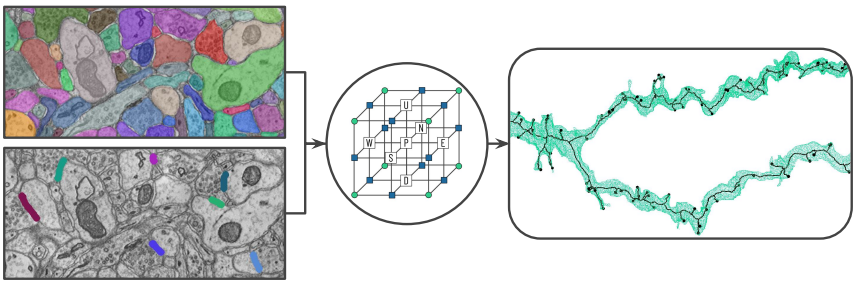

Fig. 1. Our method takes as input a segmentation dataset with cell body information and synapse locations (left). We generate a synapse-aware skeleton using a topological thinning strategy that relies only on local context (center) to produce accurate centerlines connecting all synapses to the cell body (right).

weighted edges (synapses). Unfortunately, such simplification eliminates valuable information needed for accurate biological analysis, such as geodesic distance from synapse to cell body and width of the neuron along the path. Concurrently, theoretical neuroscientists create biophysics-based simulations of simple neuronal circuits by numerically solving a series of differential equations that estimate the voltage change in a dynamic system [6]. These simulations often model neurites as a series of cylinders each with capacitance and resistances given a lack of accurate fine-resolution data, such as neurite width, length, and the accurate enumeration of connecting synapses. In addition, accurate skeleton representations of neuronal processes are increasingly important in the field of connectomics for biologically-constrained reconstruction [11], error correction [2], evaluation [5], and visualization [12].

The most commonly used skeletonization methods in the connectomic literature are the Tree-structure Extraction Algorithm for Accurate and Robust Skeletons (TEASER) [16] and its variants [19]. Alternative skeletonization approaches in the volume processing and graphics communities extract the medial axis from 3D volumes through the gradual erosion of their surfaces [9,13]. These methods rely only on local context to eliminate voxels while simultaneously preserving the topology of the original volume [13], but they do not maintain biologically relevant details. We present a novel synapse-aware skeleton generation strategy to transform volumetric connectomic data into a format for detailed analysis of the wiring diagram, accurate simulations, and improved reconstructions.

Our method builds on the class of topological thinning algorithms [9,10,13] and simplifies the input volume while still maintaining essential geometric attributes (Figure 1). Our thinning based approach produces a center-line for the neuron with accompanying estimated neurite widths along the skeleton. We guarantee that the skeleton connects all detected synapses along the neuron. We further refine the skeleton to enforce specific topological properties based on the underlying biology and efficiently calculate the geodesic distance from each synapse

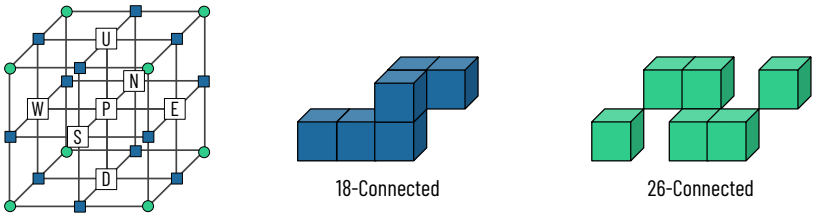

Fig. 2. Following existing work, we use only the local neighborhood around a point p to determine deletion (left). $N_6(p)$ contains the point p and the six points labeled **U**, **D**, **N**, **S**, **E**, and **W**. $N_{18}(p)$ includes the points in $N_6(p)$ and the twelve ■ points. $N_{26}(p)$ is the set of $N_{18}(p)$ and the eight points marked by ●. Two points x_0 and x_n are 18- (center) and 26-connected (right) if each point x_i along the path is in $N_j(x_{i-1})$.

to the cell body, i.e., the soma. Our parameterless algorithm requires no training data. We evaluate our method on three large-scale connectomic datasets from three different species and compare our results to state-of-the-art skeletonization algorithms. Our algorithm generalizes to other medical image domains such as skeleton generation of the cardiovascular or peripheral nervous system where the endpoints correspond to capillaries or nerve endings, respectively.

2 Methodology

Our synapse-aware skeleton generation approach adapts previous topological thinning algorithms [9,13] to guarantee a one-to-one correspondence between endpoints and synapses. We efficiently estimate the width of the neuronal processes during the thinning process. Subsequently, we calculate the geodesic distances from synapses to soma and refine the skeleton to enforce certain biological constraints—namely that neurons are acyclic.

Notation. Our method takes two inputs: a point cloud representing a neuronal process and a list of synapse locations. Here we use the following notations: $\mathbb{Z}^3$ is the set of points, B is the set of object points, and $\mathbb{Z}^3 \setminus B$ is the set of background points.

We define three local neighborhoods of different scope around a pixel p which we call $N_6(p)$, $N_{18}(p)$, and $N_{26}(p)$ (Figure 2, left). Two points x_0 and x_n are *j-connected* (for $j = 6, 18, 26$) if there exists a path $< x_0, x_1, ..., x_n >$ where each x_i is in $N_j(x_{i-1})$ for $i = 1, ..., n$ (Figure 2, center, right). An object is a maximally 26-connected set of object points. Conversely, a background component is a maximally 6-connected set of background points. An *endpoint* has only one object point in $N_{26}(p)$. Hence we consider digital pictures specified by the quadruple ($\mathbb{Z}^3$, 26, 6, B) [7].

Malandain and Bertrand [10] prove the following theorem to determine if an object point p is simple (i.e., removal from the set B does not alter the topology):

Theorem 1. *An object point p is simple in a picture ($\mathbb{Z}^3$, 26, 6, B) if and only if all of the following conditions hold:*

1. *The set $N_{26}(p) \cap (B \setminus p)$ contains exactly one 26-component.*
2. *The set $N_6(p) \setminus B$ is not empty.*
3. *Any two points in $N_6(p) \setminus B$ are 6-connected in the set $N_{18}(p) \setminus B$.*

All simple points are surface points by Condition 2 of Theorem 1.

Topological Thinning. Every endpoint is a simple point by Theorem 1. Therefore to generate skeletons, specific points are preserved to avoid a complete reduction to a single point in the case of an acyclic input. Some strategies include preserving endpoints [1] or defining a second class of geometric constraints as non-simple isthmuses [13]. We differ from previous approaches by designating synapse locations as always non-simple and thus non-deletable. Any other point in the volume can be removed if it adheres to the requirements of Theorem 1. This synapse-aware endpoint strategy produces skeletons that are better suited for higher-level analysis.

We employ a sequential thinning procedure to erode the surface uniformly in all directions [13]. Each iteration consists of six sub-iterations where we consider surface points whose corresponding neighbor at location **U**, **N**, **E**, **S**, **W**, and **D** is a background point (Figure 2, left). For each of these six sub-iterations, we identify the simple points that are potentially deletable. After collecting all the simple points, we reiterate through the list and delete any that are still simple. This dual-pass approach is necessary since a point may lose its simple designation based on neighboring deletions. After deletion, we add any neighboring points that are now on the surface to the list of surface voxels.

Width Estimation. For each object point, we store an estimate for the distance from the point to the surface. We initialize these estimates to 0 for all surface voxels and to ∞ for all internal voxels. When a surface point p is deleted during a thinning iteration, we look at the neighborhood $N_{26}(p)$. We update the distance-to-surface estimate for a neighboring voxel only if its distance to p plus the distance estimate at p is less than its current value. As the surface erodes, our internal distance estimates better approximate the actual distance to the original surface. The distance estimates at the skeleton points correspond to half of the width of the neuronal process at that cross-section.

Geodesic Distance Calculation. Neuronal processes are acyclic, i.e., the genus of the cell membrane surface is zero. However, errors in the input volumes can produce bubbles and tunnels in the segmentation which our topological thinning strategy would preserve (Figure 3). Furthermore, although we can quickly generate the neuron width at a given location with limited overhead, we have not yet determined the geodesic distance from each synapse to the cell body. We simultaneously enforce the acyclic constraint and produce geodesic distances with the following procedure. First, we run Dijkstra's shortest path algorithm on the skeleton with the soma as the source. We only keep skeleton

points that belong on the shortest path from a synapse to the soma. This process removes any loops in the skeleton since any shortest path cannot contain loops. This additional step produces the geodesic distance from every skeleton point (including synapses) to the cell body and enforces the acyclic property of neurons on the skeleton.

Fig. 3. *Loop Removal.* Since the thinning step preserves topology, any holes in the input segmentation cause loops in the skeleton. On the left, we see an example skeleton after thinning with several loops (the black spheres represent synapses). After refinement, these loops disappear, leaving a cleaner skeleton (right).

3 Experiments

We evaluate our methods on three large-scale connectomic datasets from three different species: rat, fruit fly, and zebra finch. JWR (rat) contains six fully-reconstructed neurons with manual segmentation and synapse identification. For FIB-25 (fruit fly), human proofreaders refined an initial segmentation produced by a context-aware two-stage agglomeration framework [14]. Synapses underwent a similar process of automatic detection and human refinement. Fully automatic techniques segmented neurons [5] and identified synapses [3] for the J0126 (zebra finch) dataset. The JWR, FIB-25, and J0126 datasets have resolutions of $32 \times 32 \times 30\,\mathrm{nm}^3$, $10 \times 10 \times 10\,\mathrm{nm}^3$, and $18 \times 18 \times 20\,\mathrm{nm}^3$, respectively.

We compare our proposed method against two baselines: TEASER [16] and an isthmus-based topological thinning approach [13]. Both baselines are particularly susceptible to surface noise, which creates many spurious endpoints. Additionally, the TEASER strategy requires a significant amount of memory per voxel. Therefore, we downsample the volumes for both of these datasets to a resolution of $100 \times 100 \times 100\,\mathrm{nm}^3$. Our proposed method is robust to surface noise since endpoints only occur at designated synapse locations.

Table 1. We evaluate our method on three connectomic datasets from three different animal species. The FIB-25 and J0126 datasets contain many neuron fragments.

Name	Species	Volume	No. Neurons	No. Synapses
JWR	Rat	$106 \times 106 \times 93\,\mu\mathrm{m}^3$	6	10,203
FIB-25 [18]	Fruit Fly	$36 \times 29 \times 69\,\mu\mathrm{m}^3$	491	63,258
J0126 [8]	Zebra Finch	$96 \times 98 \times 114\,\mu\mathrm{m}^3$	371	84,098

Table 2. We evaluate the proposed method versus two baselines approaches on three metrics: NRI score, estimated neurite width error, and the number of points in the skeleton. The TEASER algorithm calculates the distance transform on the downsampled data, so the width given is the expected difference with upsampling.

Method	JWR			FIB-25			J0126	
	NRI ↑	Width ↓	Points ↓	NRI ↑	Width ↓	Points ↓	Width ↓	Points ↓
Proposed	1.000	25.26 nm	43,088	1.000	13.16 nm	14,529	23.54 nm	24,533
TEASER	0.265	20 nm*	45,890	0.387	25 nm*	13,577	24 nm*	45,946
Isthmus Thinning	0.216	N/A	607,583	0.427	N/A	35,604	N/A	673,256

We evaluate our results using three heuristics. The Neural Reconstruction Integrity (NRI) score indicates how well a given segmentation (or skeleton) maintains the underlying wiring diagram of the brain [15]. An NRI score near 1.0 indicates that most of the intracellular pathways between pairs of synapses are preserved. For our baselines, we link synapses to endpoints that fall within 800 nanometers of each other. Second, we calculate the mean absolute error of our width prediction over a random subset of 20% of skeleton points. Lastly, we evaluate the simplicity of each skeleton with the number of remaining points [13].

4 Results

Table 2 shows how our proposed method performs over three evaluation metrics.

NRI. Our method guarantees a one-to-one correspondence between endpoints and synapses, we produce a perfect NRI score of 1.0 on the JWR and FIB-25 datasets. We cannot evaluate the NRI score on the J0126 dataset since there is no ground truth. With the addition of *merge* and *split* errors in the segmentation, our NRI score would continue to match that of the input data. Both the TEASER and the isthmus thinning strategies have significantly lower NRI scores ranging from 0.216 to 0.427.

Width Estimation. Our method achieves a mean absolute error of 25.26 nm, 13.16 nm, and 23.54 nm on the JWR, FIB-25, and J0126 datasets, respectively. The TEASER algorithm calculates the distance transform on the segmentation data as its first step. The mean absolute error for the TEASER skeleton generation strategy would be zero if the algorithm ran on the high-resolution data. However, since we need to downsample the data, the widths vary from the exact distances. We show the expected absolute error for any point between the downsampled distance transform and the correct surface distance.

Skeleton Simplicity. Our skeletons have the fewest total points on average on the JWR and J0126 datasets and are within 8% of the number of points of the TEASER skeletons on the Fib-25 dataset. The J0126 segmentation has numerous holes, particularly in the cell bodies. The isthmus thinning strategy preserves topology, which leads to a large number of skeleton points surrounding each hole. A refinement step would significantly reduce the number of skeleton points for this method.

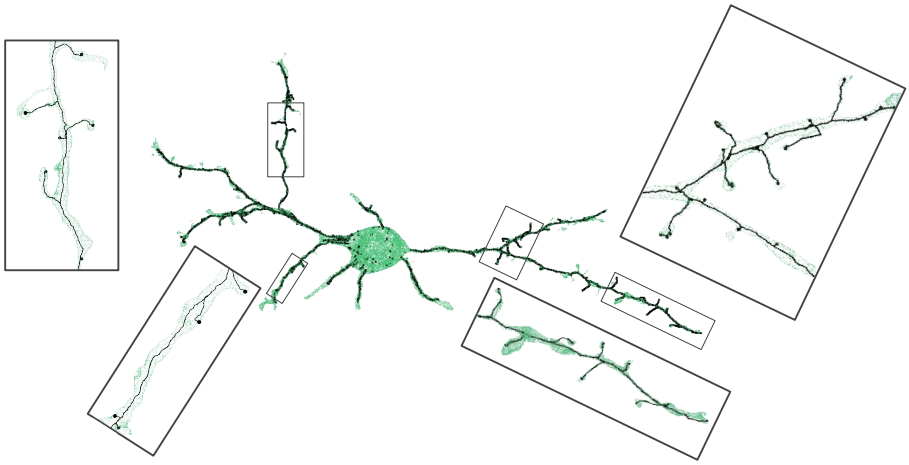

Fig. 4. Our synapse-aware skeleton generation strategy accurately produces centerlines on several neurons over three large-scale representative datasets. We zoom into four locations on one neuron from J0126 dataset. The black spheres indicate synapses.

Geodesic Distance Calculation. Using geodesic distances provides a much more accurate measurement of the path from a synapse to the soma than the frequently used Euclidean distance. The geodesic distance between each synapse and the soma is 47% (12 µm) farther on average over the three datasets. Two of the JWR neurons share a synaptic connection. The geodesic distance from the soma to the presynaptic terminals and from the postsynaptic density to the neighboring soma is 145% farther than the corresponding Euclidean distance. These distances need to be accurate for neuron simulation.

Qualitative Results. Figure 4 shows the generated skeleton for a complete neuron from the J0126 dataset. The black spheres indicate synapse locations. Our skeleton refinement process removes self-loops caused by errors in the input segmentation enforcing the constraint that neurons are acyclic (Figure 3). This refinement process reduces the number of skeleton points by a factor of 3.48× on average on the FIB-25 dataset.

Computational Efficiency. For each point in the neuron, we require 8 bytes for a linear index, 1 byte for the status (surface, inside, or synapse), and 4 bytes for the distance to the surface estimation. We average throughput of over $100,000$ voxels per second on the FIB-25 and JWR datasets. Our skeleton refinement step takes a negligible amount of time for all tested neurons with an average throughput of nearly $65,000$ skeleton points per second. We only consider skeleton points when calculating the geodesic distance, a significant reduction from the overall volume (on average 297× fewer points on the FIB-25 and JWR datasets).

5 Conclusion

Large-scale connectomic datasets contain fully-segmented neurons and synapse locations. Current attempts to generate wiring diagrams from the reconstructed data leave behind a wealth of information such as neurite width and geodesic distance from synapse to soma. We present a novel synapse-aware skeleton generation strategy to transform the reconstructed volumetric data into an abstract yet expressive format for detailed analysis, accurate simulation, and improved reconstruction. We compare our method against state-of-the-art skeletonization methods on 868 neurons and neuron fragments over three different datasets. Our code is freely available at https://www.rhoana.org/synapseaware.

Acknowledgements. H. Pfister is supported in part by NSF grant IIS-1607800. We thank Joergen Kornfeld and Winfried Denk's group for the J0126 data and synapses, and the Connectomics Group at Google led by Viren Jain for the segmentation. For the JWR dataset, we thank Jeff Lichtman's group at Harvard University for image acquisition, alignment, and ground truth labeling.

References

1. Bertrand, G., Aktouf, Z.: Three-dimensional thinning algorithm using subfields. In: Vision Geometry III. vol. 2356, pp. 113–125. International Society for Optics and Photonics (1995)
2. Dmitriev, K., Parag, T., Matejek, B., Kaufman, A., Pfister, H.: Efficient correction for em connectomics with skeletal representation. In: BMVC (2018)
3. Dorkenwald, S., Schubert, P.J., Killinger, M.F., Urban, G., Mikula, S., Svara, F., Kornfeld, J.: Automated synaptic connectivity inference for volume electron microscopy. Nature methods **14**(4), 435 (2017)
4. Fornito, A., Zalesky, A., Breakspear, M.: Graph analysis of the human connectome: promise, progress, and pitfalls. Neuroimage **80**, 426–444 (2013)
5. Januszewski, M., Kornfeld, J., Li, P.H., Pope, A., Blakely, T., Lindsey, L., Maitin-Shepard, J., Tyka, M., Denk, W., Jain, V.: High-precision automated reconstruction of neurons with flood-filling networks. Nature methods **15**(8), 605 (2018)
6. Koch, C.: Biophysics of computation: information processing in single neurons. Oxford university press (2004)
7. Kong, T.Y., Rosenfeld, A.: Digital topology: Introduction and survey. Computer Vision, Graphics, and Image Processing **48**(3), 357–393 (1989)
8. Kornfeld, J., Benezra, S.E., Narayanan, R.T., Svara, F., Egger, R., Oberlaender, M., Denk, W., Long, M.A.: Em connectomics reveals axonal target variation in a sequence-generating network. Elife **6**, e24364 (2017)
9. Lee, T.C., Kashyap, R.L., Chu, C.N.: Building skeleton models via 3-d medial surface axis thinning algorithms. CVGIP: Graphical Models and Image Processing **56**(6), 462–478 (1994)
10. Malandain, G., Bertrand, G.: Fast characterization of 3d simple points. In: Pattern Recognition, 1992. Vol. III. Conference C: Image, Speech and Signal Analysis, Proceedings., 11th IAPR International Conference on. pp. 232–235. IEEE (1992)
11. Matejek, B., Haehn, D., Zhu, H., Wei, D., Parag, T., Pfister, H.: Biologically-constrained graphs for global connectomics reconstruction. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019)

12. Mohammed, H., Al-Awami, A.K., Beyer, J., Cali, C., Magistretti, P., Pfister, H., Hadwiger, M.: Abstractocyte: A visual tool for exploring nanoscale astroglial cells. IEEE transactions on visualization and computer graphics **24**(1), 853–861 (2018)
13. Palágyi, K.: A sequential 3d curve-thinning algorithm based on isthmuses. In: International Symposium on Visual Computing. pp. 406–415. Springer (2014)
14. Parag, T., Chakraborty, A., Plaza, S., Scheffer, L.: A context-aware delayed agglomeration framework for electron microscopy segmentation. PloS one **10**(5), e0125825 (2015)
15. Reilly, E.P., Garretson, J.S., Roncal, W.R.G., Kleissas, D.M., Wester, B.A., Chevillet, M.A., Roos, M.J.: Neural reconstruction integrity: A metric for assessing the connectivity accuracy of reconstructed neural networks. Frontiers in Neuroinformatics **12** (2018)
16. Sato, M., Bitter, I., Bender, M.A., Kaufman, A.E., Nakajima, M.: Teasar: Tree-structure extraction algorithm for accurate and robust skeletons. In: Computer Graphics and Applications, 2000. Proceedings. The Eighth Pacific Conference on. pp. 281–449. IEEE (2000)
17. Suissa-Peleg, A., Haehn, D., Knowles-Barley, S., Kaynig, V., Jones, T.R., Wilson, A., Schalek, R., Lichtman, J.W., Pfister, H.: Automatic neural reconstruction from petavoxel of electron microscopy data. Microscopy and Microanalysis **22**(S3), 536–537 (2016)
18. Takemura, S.y., Xu, C.S., Lu, Z., Rivlin, P.K., Parag, T., Olbris, D.J., Plaza, S., Zhao, T., Katz, W.T., Umayam, L., et al.: Synaptic circuits and their variations within different columns in the visual system of drosophila. Proceedings of the National Academy of Sciences **112**(44), 13711–13716 (2015)
19. Zhao, T., Olbris, D.J., Yu, Y., Plaza, S.M.: Neutu: Software for collaborative, large-scale, segmentation-based connectome reconstruction. Frontiers in Neural Circuits **12** (2018)