nature methods

Article

https://doi.org/10.1038/s41592-024-02226-5

RoboEM: automated 3D flight tracing for synaptic-resolution connectomics

Received: 8 September 2022

Accepted: 26 February 2024

Published online: 21 March 2024

Check for updates

Martin Schmidt[1] ✉, Alessandro Motta[1], Meike Sievers[1,2] & Moritz Helmstaedter[1] ✉

Mapping neuronal networks from three-dimensional electron microscopy (3D-EM) data still poses substantial reconstruction challenges, in particular for thin axons. Currently available automated image segmentation methods require manual proofreading for many types of connectomic analysis. Here we introduce RoboEM, an artificial intelligence-based self-steering 3D 'flight' system trained to navigate along neurites using only 3D-EM data as input. Applied to 3D-EM data from mouse and human cortex, RoboEM substantially improves automated state-of-the-art segmentations and can replace manual proofreading for more complex connectomic analysis problems, yielding computational annotation cost for cortical connectomes about 400-fold lower than the cost of manual error correction.

Extracting the dense neuronal connectivity from three-dimensional electron microscopy (3D-EM) data of brain tissues poses major computational challenges[1–6]. Substantial progress in the field has allowed us to move from fully manual skeleton reconstructions of neurites[3,6–11] via combinations of skeleton reconstruction and automated segmentations[5,12] to proofreading of automated image segmentation (Fig. 1a). This proofreading was initially as laborious as fully manual skeleton reconstructions[4,13–15] but has recently been made more efficient by focused human intervention based on improved automated segmentations[1,2,12,16]. Yet, even for automated methods claiming super-human performance[17] or full automation[12], when applied to large-scale EM datasets, massive manual-annotation efforts are required for the intended connectomic analyses[2,18–20].

Connectomic analyses differ widely in their reconstruction difficulty: for the analysis of pairs of neighboring synapses along a neurite for extraction of learned synaptic configurations[1,12], for example, axon reconstructions of about 10 µm in length are sufficient, and therefore these analyses can already now be fully automated[1,12]. Obtaining neuron-to-neuron connectomes from cortical tissue, however, requires faithful axon reconstruction for at least an additional order of magnitude of axonal length, and has so far not been possible fully automatically. Intermediate-scale connectomic analyses, aimed at axonal synaptic properties for example, demand error-free axonal reconstruction in the range of 30–50 µm (corresponding to about 4–10 synapses per axon stretch, depending on species and tolerable error rates of the intended analysis[1,21]) (Supplementary Note and Supplementary Fig. 1). Such error-free lengths are not yet fully automatically accessible.

Connectomic image analyses have in common that EM data are processed by artificial neural networks to yield voxel-based maps reporting plasma membranes[17,22–28], or the similarity between pairs of image voxels[29], or the association of image voxels to the same foreground object[30–32]. Then, segmentations are computed, and automated methods for the joining or splitting of these initial segmentation objects are currently the main focus of computational improvements[17,24,30–36]. When automated approaches become insufficient, human annotation is used to solve the most challenging of these joining or splitting operations, ideally directed to difficult locations by computational means[1,2,12,37–41], or by human inspection[16,19]. The FocusEM[1] toolset is one such focused reconstruction approach, which automatically detects reconstruction errors and asks human annotators to fly along axons[42] until another segmentation object is reached and thus the problematic location resolved (Fig. 1b). Notably, the usage of image sequences along neurites for alleviating manual reconstruction was already considered in the early days of EM-based neurite tracing[43–47]. More recently, iterative contour tracking methods have been proposed to automate the neurite-following process[48–51], even if these were not yet applied at scale and not learned. Here, by contrast, we wondered whether an artificial neural network could learn to directly control the steering along neurites in 3D and in an end-to-end fashion.

[1]Department of Connectomics, Max Planck Institute for Brain Research, Frankfurt, Germany. [2]Faculty of Science, Radboud University, Nijmegen, the Netherlands. ✉e-mail: martin.schmidt@brain.mpg.de; mh@brain.mpg.de

https://doi.org/10.1038/s41592-024-02226-5

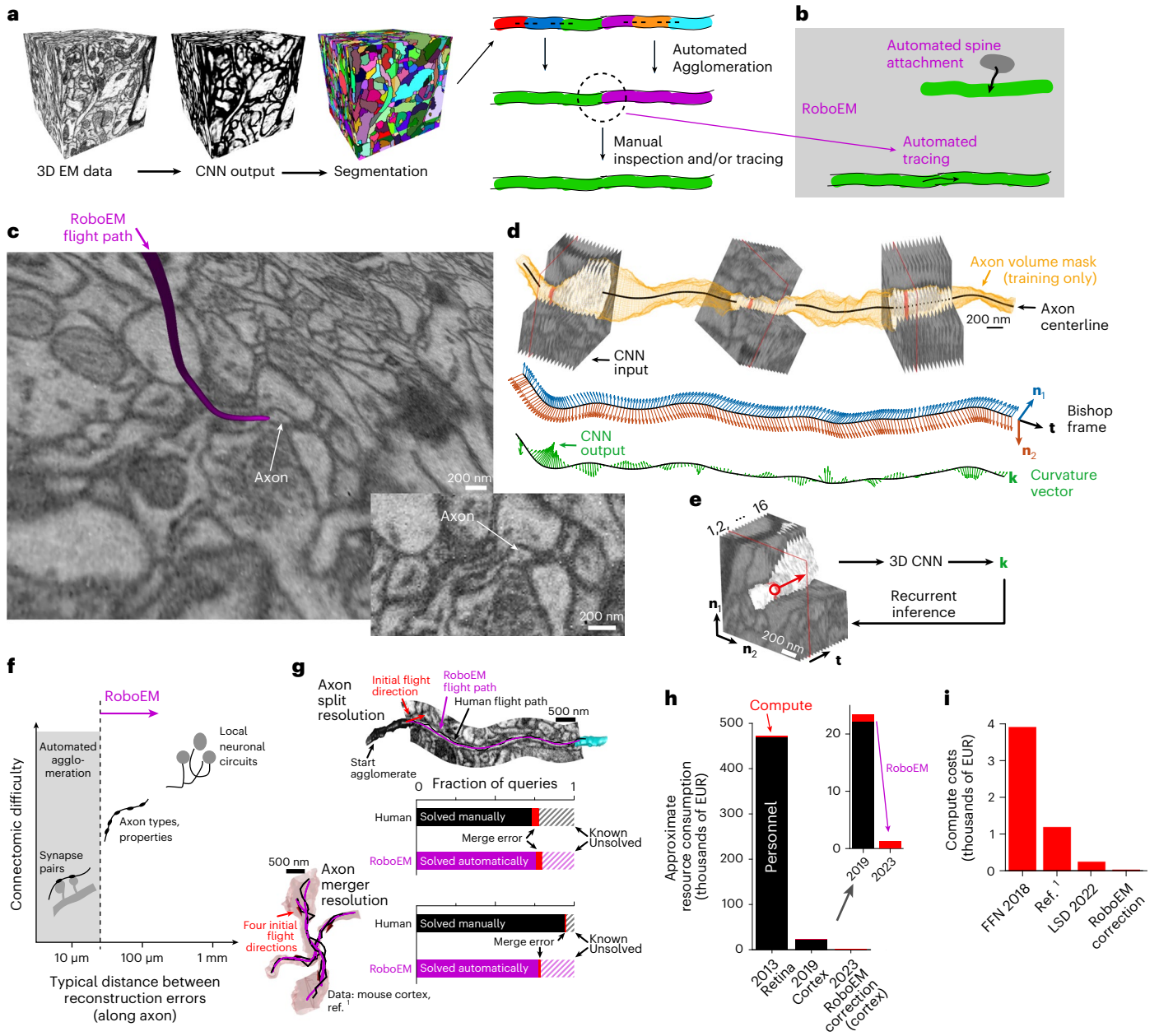


Fig. 1 | Automated neurite tracing for substitution of human annotation needs in neuronal network reconstruction. a, Standard workflow for connectomic analyses from 3D-EM data: initial volume segmentation (two-step process of classification and watershed[17,22,24–28], or directly as foreground classification[30–32]); automated agglomeration (based on interfaces between segments); manual inspection to resolve remaining errors and reach reconstruction quality usable for meaningful connectomic analysis[1,2,12,16,19,37–41]. Data cubes with 10 µm edge lengths are shown. **b**, RoboEM replaces human inspection and correction step by automated connection and/or validation flights solving split and merge errors and attaching remaining spine heads[1]. **c**, Example of the RoboEM flight path along a thin axon in SBEM data[1]. **d**, Design of RoboEM: volumetric EM data as input for prediction of a steering vector that determines the subsequent input. Yellow denotes segmentation mask for 'teaching' corrective steering signals from off-center locations (during training, only). **e**, Detailed sketch of RoboEM inference setup. **f**, Calibration of reconstruction automation by the difficulty of automatable connectomic analyses comes from synaptic pairs-based analyses[1,12,37] via extraction of axonal properties[1,21] (Fig. 2b and Supplementary Fig. 1) to local neuronal circuits. **g**, RoboEM performance in direct comparison to human annotators on axon ending ($n = 90$) and chiasma queries ($n = 100$) for split and merge error resolution[1]. **h**, Effect on resource consumption for connectomic dense reconstructions. **i**, Computing costs for state-of-the-art segmentation and agglomeration: FFNs[32] and local shape descriptors (LSD)[25] are compared against a dense connectomic reconstruction[1] (additionally including costs for synapse and type predictions and processing of human and/or RoboEM skeletons). EUR, euros.

Results

The process of flight tracing along elongated (often very thin) neurites via steering commands has an analogy to image-based road following in autonomous driving[52,53]. Based on this analogy, we developed a convolutional neural network (CNN) architecture to output 3D steering signals directly from neurite-aligned 3D-EM image volumes. We (1) defined a continuous 3D steering framework; (2) defined a membrane-avoiding flight policy to recover from off-centerline positions and noncenterline-aligned orientations; (3) trained a CNN on image-steering pairs from on- and off-centerline positions and orientations in a supervised manner allowing for stable path following during recurrent inference; and (4) defined

https://doi.org/10.1038/s41592-024-02226-5

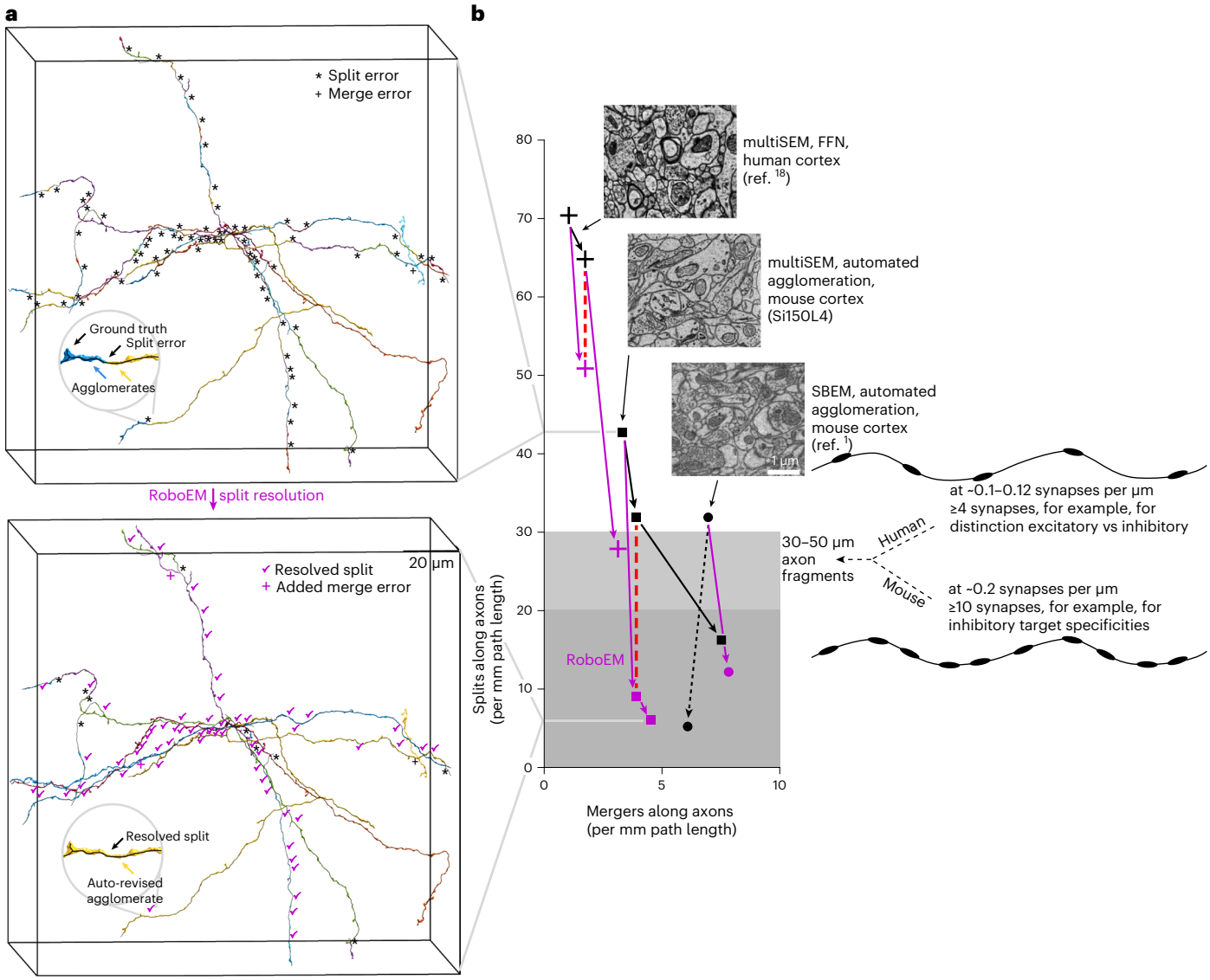


Fig. 2 | RoboEM improves state-of-the-art connectomic reconstruction results in mammalian cortex. a, Example of dense axon sets used for unbiased calibration of reconstruction success. Agglomerates before (top) and after RoboEM correction (bottom). **b**, Quantification of axon split and merge error rate for randomly seeded axons, evaluated on various 3D-EM data: human cortex multiSEM dataset[18] (crosses) (two agglomeration states analyzed (c2,c3) with 8% split resolution from c3 to c2); mouse cortex multiSEM dataset (Si150L4, https://wklink.org/7122); mouse cortex SBEM dataset[1]. Purple symbols show the effect of applying RoboEM to previous reconstruction. The red dashed line denotes the same merge-rate comparison for human data (3.5-fold improved split resolution) and mouse data (3.1-fold improved split resolution). Purple lines denote RoboEM split-rate improvement with modest merge rate increases yielding error rates tolerable for more complex connectomic analyses (Fig. 1f). The black dashed line denotes a human annotator-based error resolution (as published in ref. 1) shown for comparison.

a validation strategy for automated error detection (Fig. 1c–e and Extended Data Fig. 1a,b).

In particular, we used the Bishop frame[54] along interpolated neurite centerline reconstructions to describe the neurites' local directionality, defined the projection for the neurite-aligned 3D-EM input and then used the corresponding Bishop curvatures as the target output steering signals that had to be predicted (Fig. 1d). During training only, a neurite volume mask (obtained by segment pick-up from an oversegmentation) was used to generate off-centerline inputs and corresponding target output steering signals back toward the centerline (Extended Data Fig. 1b). During inference, predicted Bishop curvatures were integrated to yield the next position and orientation (Extended Data Fig. 1c). This approach yielded an automated neurite path annotation agent that mimicked the process of human flight-mode annotation in 3D (ref. 42) (Supplementary Video 1), called RoboEM.

Next, we investigated to what degree RoboEM could in fact replace human annotation, and which kinds of connectomic analysis would thus become fully automatable (Fig. 1f). For this, we started with a set of dense connectomic analyses of a piece of mammalian neocortex in which, based on automated segmentations, human annotators had been asked to resolve an automatically identified set of problem locations consuming a total of 4,000 work hours[1]. To resolve split errors, endings of axons and spine necks had been queried, asking the human annotators to continue if possible until the task was automatically stopped when another reconstructed object of sufficient size had been reached (Fig. 1g). To resolve merger errors, chiasmatic configurations of axons had been detected and queried, asking the human annotators for proper continuation from one chiasmatic exit into one of the other exits (Fig. 1g). We used RoboEM to replace these human annotations. For this, we issued RoboEM queries analogous to human annotator

queries and required that, additionally, a forward continuation along a neurite should be confirmed by tracing the same neurite location backward (automatic validation). We found that 76% of the ending queries and 78% of the chiasma queries were traced and validated by RoboEM without errors (as judged by manual inspection, compared to fully manual queries yielding 74 and 94%, respectively). Restricting to RoboEM annotations in which forward and backward tracings agreed allowed avoiding most merge errors, yielding only 4% of queries with RoboEM-introduced tracing errors for ending tasks and 1% for chiasma tasks, similar to human annotations.

While this performance indicated that RoboEM could accurately replace human annotation for a range of connectomic analyses, we wanted to quantify this conclusion explicitly by using connectomic analysis itself as the metric for reconstruction success. As described above, some connectomic analyses require more reconstruction accuracy than others. In particular, we considered: A1, paired same-axon same-dendrite synapse analyses aimed at measuring the learned fraction of a connectome[1]; A2, spine rate analyses for identification of interneuron dendrites and A3, axonal type analysis based on the synaptic target distribution of axons. We then used three types of connectome for comparison: CI, the connectome obtained from the fully automated reconstruction, before any focused human annotation[1]; CII, the connectome including 4,000 work hours of human annotation[1] and CIII, the connectome obtained from combining the fully automated reconstruction CI with RoboEM, yielding a fully automated and automatically proofread connectome. We then performed the three types of connectomic analysis, which we expected to increase in connectomic difficulty.

When applying the analyses A1–A3 to the three stages of connectomes CI–CIII, we found that the analysis of paired synapses for quantifying potentially learned synapses in the connectome A1 was already possible with the automated connectome state before any RoboEM-based corrections (CI, upper bound of fraction of paired connections consistent with long term potentiation: CI 11–20%; CII 16–20%; CIII 13–19%, Extended Data Fig. 2a). However, for obtaining correct spine rates for apical dendrites (A2, CI 0.9 ± 0.4; CII 1.3 ± 0.6; CIII 1.2 ± 0.5 spines per µm, mean ± s.d., Extended Data Fig. 2b) and the true fraction of excitatory axons defined by their spine head preference (excitatory axon fractions, A3, CI 75%; CII 87%; CIII 84%, Extended Data Fig. 2c), manual or RoboEM-based corrections were required. In addition, RoboEM-based corrections recovered the axonal target specificities of inhibitory axons onto apical dendrites and smooth dendrites, whereas in the automated state before any annotation this specificity was not detectable (A3, one-sided Kolmogorov–Smirnov test, CI apical dendrites $P = 1.0$%, smooth dendrites $P = 3.6$%; CII apical dendrites $P = 2.7 \times 10^{-4}$, smooth dendrites $P = 1.8 \times 10^{-3}$ and CIII apical dendrites $P = 1.9 \times 10^{-5}$, smooth dendrites $P = 7.1 \times 10^{-4}$, Extended Data Fig. 2d). Thus, using the difficulty of connectomic analyses that can be fully automated as the criterion for evaluating RoboEM-based error correction, we found a shift of automated analysis performance from simpler to more complex connectomic problems (Fig. 1f), yielding 400-fold reduced annotation costs compared to manual error correction (Fig. 1h). The computational cost was 5- to 80-fold lower than other approaches (Fig. 1i and Supplementary Table 4), rendering RoboEM a suitable candidate to run automated error correction as a postprocessing step. These results were obtained on 3D-EM data imaged with serial block-face scanning electron microscopy (SBEM)[55].

Next, we applied RoboEM to data obtained using a state-of-the-art high-throughput 3D-EM imaging approach for mm^3-scale volumes from mammalian brains[18,21] (ATUM-multiSEM, where ATUM stands for automated tape-collecting ultramicrotome[56] followed by multibeam scanning electron microscope (multiSEM)[57]). We used a (150 µm)^3 subvolume centered on cortical layer 4, subsequently referred to as 'Si150L4', cropped from a larger 1.3 × 1.3 × 0.25 mm^3 sized 3D-EM dataset that had been imaged at 4 × 4 × 35 nm^3 voxel size spanning all cortical layers of mouse primary somatosensory cortex (S1). In this subvolume, we quantified RoboEM performance on axons sampled uniformly from the volume ('densely seeded axons', Fig. 2a,b). After automated agglomeration, we seeded RoboEM at automatically detected endings of axon agglomerates, and used it to connect possible missing axonal agglomerates (Fig. 2a). As a result, split rates of axons were reduced sevenfold (42.7 to 6.0 per mm axon path length) while only modestly increasing merger rates (3.3 to 4.5 per mm, Fig. 2b).

To titrate the effect of RoboEM separately for split and merge errors, we used the fact that automated agglomerations (before RoboEM application) typically have a split-versus-merger parameter that can be adjusted independently. We then introduced a similar parameter for RoboEM, and could therefore directly compare the split rates achieved by automated agglomeration and RoboEM at equal merge error rates. We found (Fig. 2b) that the agglomeration solved 25% of split errors yielding 31.9 splits per mm, while RoboEM solved 79% yielding 9.0 splits per mm, compared at 3.9 merge errors per mm axon path length. Hence, in this split and/or merge error regime RoboEM improves split resolution 3.1-fold over state-of-the-art agglomeration (Fig. 2b). The resulting axon reconstructions were comparable to those achieved in the local cortical SBEM dataset[1] (5.2 split and 6.1 merge errors per mm, Fig. 2b).

Similar to thin axons, thin spine necks can pose difficulties for automated reconstruction pipelines of 3D-EM data. To assign synapses onto spine heads to the correct postsynaptic neuron, spine necks need to be reconstructed with high accuracy. We used RoboEM to follow automatically detected spine heads along the spine neck back to the dendritic shaft of origin. Evaluated on a subvolume (Si11L3) in cortical layer 3 sized (11 µm)^3 from the above ATUM-multiSEM dataset of mouse S1 cortex, we found that spine head attachment recall increased from 70 to 94%, while retaining high precision of 97% (85 out of 91 test spine heads attached to the correct dendrite, 3 out of 91 spine heads attached to the wrong dendrite and 3 out of 91 spine heads were unattached, of which one could also not be attached manually and was therefore counted as true negative).

Finally, we wondered whether RoboEM could also be applied for improving existing state-of-the-art automated segmentations that are currently being manually proofread and had been generated by flood-filling networks (FFN)[18,32]. We evaluated RoboEM on a subvolume of size (150 µm)^3 from a recently published mm^3-scale ATUM-multiSEM dataset from human cortex[18] and found that application of RoboEM on axonal endings obtained from the dense reconstructions of an FFN segmentation[32] solved 57% of the splits on a random set of axons traced throughout the subvolume. This reduced the split rate from 65 to 28 splits per mm, reaching split lengths in the range required for automated connectomic analyses, while increasing merge rate from 1.7 to 3.1 per mm (in the range of merge rates for connectomic analyses as in ref. 1). When comparing the reduction of split errors in FFN versus RoboEM at an identical merge error rate (c2, 1.7 mergers per mm), we find that the FFN-based agglomeration from c3 to c2 only solves 8% of the split errors of c3. By contrast, RoboEM-based agglomeration solves 28% of split errors of c3 at identical merge error rate as c2 is thereby 3.5-fold more effective in split resolution than FFN (Fig. 2b). A detailed analysis of remaining axon merge errors for the two multiSEM reconstructions with RoboEM corrections furthermore showed that 86–92% of these merge errors are caused by the preceding segmentation or agglomeration, not by RoboEM itself. Further, more than or equal to 83% of major merge errors found were in close proximity to 3D data misalignment, artifacts and/or missing sections, suggesting that greater robustness to these 3D-EM data related issues could help further reduce merge errors (Supplementary Table 3 and Supplementary Note).

These results are important for two reasons. First, evaluation of automated reconstruction performance on random ('densely seeded') axons that are not necessarily connected to a soma within the mm^3-scale volume provides a representative quantification of

reconstruction quality for dense connectomic reconstruction. Restriction to soma-proximal axons quantifies results for soma-based proximal connectomic data, but underestimates reconstruction errors for dense axons. For example, the evaluation of FFN in human cortex, restricted to axons connected to a soma within the volume underestimates split and merge errors of random axons four- to fivefold (Extended Data Fig. 3). When using soma-based reconstructions in small brain volumes[32], error rate quantification is highly biased to the much easier-to-reconstruct dendrites and therefore cannot be interpreted for axonal reconstructions. Second, for the connectomic analysis of axons in cortical neuropil, it is essential to obtain an interpretable number of output synapses along axons automatically. Given a certain rate of synapses along axons, this means that there is a minimum intersplit distance that needs to be achieved to obtain an interpretable number of synapses per axonal segment (Fig. 2b). The rate of synapses per axon can vary strongly between neuronal tissue types and species, such that axonal reconstructions in human cortex, for example, require at least twofold longer intersplit distances (that is, twofold lower split error rates) than in mouse cortex to achieve similar synaptic statistics per axon (Fig. 2b).

Discussion

In summary, we find that our automated proofreading using RoboEM can fully replace human annotation for connectomic analyses of increasing difficulty and improve currently available state-of-the-art automated segmentations. This is particularly notable for segmentations obtained using FNNs[18,32], since FFNs have a certain similarity in their design to RoboEM by recursively reconstructing individual neurites through prediction of neurite continuation from 3D-EM subvolumes. RoboEM, on the other hand, by focusing on centerline neurite tracing, may learn some aspects of axonal morphology directly and adds a notion of growth inertia to the axonal reconstruction. This enabled reconstruction problems to be solved that were not addressed by volume-based methods.

In mm^3-scale 3D-EM datasets, the reconstruction of millimeter- to centimeter-long axons for obtaining neuron-to-neuron connectomes currently demands substantial human proofreading[18,19]. RoboEM, showing 3.5-fold better split reduction compared to state-of-the-art agglomeration by FFN[18], allows for immediate reduction of manual-annotation needs. In addition, while we have calibrated RoboEM performance on smaller datasets (around 100–150 µm on a side), the obtained results allow us to estimate performance of RoboEM for iterative soma-seeded axonal reconstructions in larger datasets. By limiting RoboEM error correction to iterative soma-seeded axon reconstruction, and taking into account the higher accuracy on soma-proximal axons, we estimate RoboEM to extend fully automated axon reconstruction by multiple millimeters per neuron (Supplementary Note), making fully automated connectomic reconstruction at the scale of 'peta-scale' volumes containing thousands to ten thousand neurons in mammalian cortex and other tissue plausible.

Since RoboEM directly transforms EM image volumes into centerline tracings of neurites, it allows end-to-end learning of the key connectomic challenge: to follow axons over very long distances, also at locally thin stretches. As automated error correction framework, RoboEM allows for substantially improved reconstructions while increasing compute costs by less than 20% even compared to the recently proposed local shape descriptor framework[25], which itself is reported to offer FFN-scale accuracy at a much reduced computational cost. With the direct end-to-end strategy, RoboEM may allow optimization for both additional accuracy and computational efficiency, which will be the next key challenge in connectomics for exabyte-scale datasets to come[58,59].

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41592-024-02226-5.

References

1. Motta, A. et al. Dense connectomic reconstruction in layer 4 of the somatosensory cortex. *Science* **366**, eaay3134 (2019).
2. Scheffer, L. K. et al. A connectome and analysis of the adult *Drosophila* central brain. *eLife* **9**, 57443 (2020).
3. Eichler, K. et al. The complete connectome of a learning and memory centre in an insect brain. *Nature* **548**, 175–182 (2017).
4. Kasthuri, N. et al. Saturated reconstruction of a volume of neocortex. *Cell* **162**, 648–661 (2015).
5. Helmstaedter, M. et al. Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature* **500**, 168–174 (2013).
6. Wanner, A. A., Genoud, C. & Friedrich, R. W. 3-dimensional electron microscopic imaging of the zebrafish olfactory bulb and dense reconstruction of neurons. *Sci. Data* **3**, 160100 (2016).
7. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **314**, 1–340 (1986).
8. Bock, D. D. et al. Network anatomy and in vivo physiology of visual cortical neurons. *Nature* **471**, 177–182 (2011).
9. Lee, W. C. A. et al. Anatomy and function of an excitatory network in the visual cortex. *Nature* **532**, 370–374 (2016).
10. Helmstaedter, M., Briggman, K. L. & Denk, W. High-accuracy neurite reconstruction for high-throughput neuroanatomy. *Nat. Neurosci.* **14**, 1081–1088 (2011).
11. Schmidt, H. et al. Axonal synapse sorting in medial entorhinal cortex. *Nature* **549**, 6–9 (2017).
12. Kornfeld, J. M. et al. An anatomical substrate of credit assignment in reinforcement learning. Preprint at *bioRxiv* https://doi.org/10.1101/2020.02.18.954354 (2020).
13. Takemura, S.-Y. et al. A visual motion detection circuit suggested by *Drosophila* connectomics. *Nature* **500**, 175–181 (2013).
14. Kim, J. S. et al. Space-time wiring specificity supports direction selectivity in the retina. *Nature* **509**, 331–336 (2014).
15. Mishchenko, Y. et al. Ultrastructural analysis of hippocampal neuropil from the connectomics perspective. *Neuron* **67**, 1009–1020 (2010).
16. Dorkenwald, S. et al. FlyWire: online community for whole-brain connectomics. *Nat. Methods* **19**, 119–128 (2022).
17. Lee, K., Zung, J., Li, P., Jain, V. & Seung, H. S. Superhuman accuracy on the SNEMI3D connectomics challenge. Preprint at https://doi.org/10.48550/arXiv.1706.00120 (2017).
18. Shapson-Coe, A. et al. A connectomic study of a petascale fragment of human cerebral cortex. Preprint at *bioRxiv* https://doi.org/10.1101/2021.05.29.446289 (2021).
19. MICrONS Consortium et al. Functional connectomics spanning multiple areas of mouse visual cortex. Preprint at *bioRxiv* https://doi.org/10.1101/2021.07.28.454025 (2021).
20. Zheng, Z. et al. A complete electron microscopy volume of the brain of adult *Drosophila melanogaster*. *Cell* **174**, 730–743 (2018).
21. Loomba, S. et al. Connectomic comparison of mouse and human cortex. *Science* **377**, eabo0924 (2022).
22. Berning, M., Boergens, K. M. & Helmstaedter, M. SegEM: efficient image analysis for high-resolution connectomics. *Neuron* **87**, 1193–1206 (2015).
23. Zeng, T., Wu, B. & Ji, S. DeepEM3D: approaching human-level performance on 3D anisotropic EM image segmentation. *Bioinforma.* **33**, 2555–2562 (2017).
24. Funke, J. et al. Large scale image segmentation with structured loss based deep learning for connectome reconstruction. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**, 1669–1680 (2019).

25. Sheridan, A. et al. Local shape descriptors for neuron segmentation. *Nat. Methods* **20**, 295–303 (2022).
26. Jain, V. et al. Supervised learning of image restoration with convolutional networks. In *Proc. IEEE 11th International Conference on Computer Vision (2007)* 1–8 (IEEE, 2007).
27. Turaga, S. C. et al. Maximin affinity learning of image segmentation. In *Advances in Neural Information Processing Systems (NIPS 2009)* (eds. Bengio et al.) Vol. 22, 1865–1873 (2009).
28. Turaga, S. C. et al. Convolutional networks can learn to generate affinity graphs for image segmentation. *Neural Comput.* **22**, 511–538 (2010).
29. Lee, K., Lu, R., Luther, K. & Seung, H. S. Learning and segmenting dense voxel embeddings for 3D neuron reconstruction. *IEEE Trans. Med. Imaging* https://doi.org/10.1109/TMI.2021.3097826 (2021).
30. Meirovitch, Y. et al. A multi-pass approach to large-scale connectomics. Preprint at https://doi.org/10.48550/arXiv.1612.02120 (2016).
31. Meirovitch, Y. et al. Cross-classification clustering: an efficient multi-object tracking technique for 3-D instance segmentation in connectomics. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* Vol. 2019, 8425–8435 (IEEE, 2019).
32. Januszewski, M. et al. High-precision automated reconstruction of neurons with flood-filling networks. *Nat. Methods* **15**, 605–610 (2018).
33. Li, H., Januszewski, M., Jain, V. & Li, P. H. Neuronal subcompartment classification and merge error correction. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)* Vol. 12265 (eds Martel, A. L. et al.) 88–98 (Springer, 2020).
34. Zung, J., Tartavull, I., Lee, K. & Seung, H. S. An error detection and correction framework for connectomics. In *Advances in Neural Information Processing Systems (NIPS 2017)* Vol. 30 (eds. Guyon et al.) 6818–6829 (2017).
35. Dmitriev, K., Parag, T., Matejek, B., Kaufman, A. & Pfister, H. Efficient correction for EM connectomics with skeletal representation. In *Proc. British Machine Vision Conference (BMVC)* (2018).
36. Nguyen, K. T., Jang, G. & Jeong, W.-K. RLCorrector: reinforced proofreading for connectomics image segmentation. Preprint at https://doi.org/10.48550/arXiv.2106.05487 (2021).
37. Dorkenwald, S. et al. Binary and analog variation of synapses between cortical pyramidal neurons. *eLife* **11**, e76120 (2022).
38. Zhao, T., Olbris, D. J., Yu, Y. & Plaza, S. M. NeuTu: software for collaborative, large-scale, segmentation-based connectome reconstruction. *Front. Neural Circuits* **12**, 101 (2018).
39. Hubbard, P. M. et al. Accelerated EM connectome reconstruction using 3D visualization and segmentation graphs. Preprint at *bioRxiv* https://doi.org/10.1101/2020.01.17.909572 (2020).
40. Haehn, D., Kaynig, V., Tompkin, J., Lichtman, J. W. & Pfister, H. Guided proofreading of automatic segmentations for connectomics. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 9319–9328 (IEEE, 2018).
41. Plaza, S. M. Focused proofreading to reconstruct neural connectomes from EM images at scale. In *Proc. Deep Learning and Data Labeling for Medical Applications* (Carneiro, G. et al.) 249–258 (Springer, 2016).
42. Boergens, K. M. et al. webKnossos: efficient online 3D data annotation for connectomics. *Nat. Methods* **14**, 691–694 (2017).
43. Macagno, E. R., Levinthal, C. & Sobel, I. Three-dimensional computer reconstruction of neurons and neuronal assemblies. *Annu. Rev. Biophys. Bio.* **8**, 323–351 (1979).
44. Stevens, J. K., Davis, T. L., Friedman, N. & Sterling, P. A systematic approach to reconstructing microcircuitry by electron microscopy of serial sections. *Brain Res. Rev.* **2**, 265–293 (1980).
45. Levinthal, C. & Ware, R. Three dimensional reconstruction from serial sections. *Nature* **236**, 207–210 (1972).
46. Fiala, J. C. Reconstruct: a free editor for serial section microscopy. *J. Microsc.* **218**, 52–61 (2005).
47. Fiala, J. C. & Harris, K. M. Computer-based alignment and reconstruction of serial sections. *Microscopy Anal.* **52**, 5–7 (2002).
48. Jurrus, E. et al. Axon tracking in serial block-face scanning electron microscopy. *Med. Image Anal.* **13**, 180–188 (2009).
49. Macke, J. H. et al. Contour-propagation algorithms for semi-automated reconstruction of neural processes. *J. Neurosci. Methods* **167**, 349–357 (2008).
50. Vazquez-Reina, A., Miller, E. & Pfister, H. Multiphase geometric couplings for the segmentation of neural processes. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 2020–2027 (IEEE, 2009).
51. Jeong, W. K. et al. Scalable and interactive segmentation and visualization of neural processes in EM datasets. *IEEE Trans. Vis. Comput. Graph.* **15**, 1505–1514 (2009).
52. Pomerleau, D. A. Alvinn: an autonomous land vehicle in a neural network. *Adv. Neural Inf. Process Syst.* **1**, 305–313 (1989).
53. Bojarski, M. et al. End to end learning for self-driving cars. Preperint at https://doi.org/10.48550/arXiv.1604.07316 (2016).
54. Bishop, R. L. There is more than one way to frame a curve. *Am. Math. Monthly* **82**, 246–251 (1975).
55. Denk, W. & Horstmann, H. Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS Biol.* **2**, e329 (2004).
56. Hayworth, K. J., Kasthuri, N., Schalek, R. & Lichtman, J. W. Automating the collection of ultrathin serial sections for large volume TEM reconstructions. *Microsc. Microanal.* **12**, 86–87 (2006).
57. Eberle, A. L. et al. High-resolution, high-throughput imaging with a multibeam scanning electron microscope. *J. Microsc.* **259**, 114–120 (2015).
58. Motta, A., Schurr, M., Staffler, B. & Helmstaedter, M. Big data in nanoscale connectomics, and the greed for training labels. *Curr. Opin. Neurobiol.* **55**, 180–187 (2019).
59. Abbott, L. F. et al. The mind of a mouse. *Cell* **182**, 1372–1376 (2020).

Methods

Animal experiments

This section applies to the acquisition of the barrel cortex multiSEM dataset of a 28-day-old male mouse (species, *Mus musculus*, strain C57BL/6-J, one animal). The mouse cortex SBEM dataset[1] and the human cortex multiSEM dataset[18] analyzed in this study were taken from the respective publications.

All animal-related experimental procedures were performed according to the law of animal experimentation issued by the German Federal Government under the supervision of local ethics committees and according to the guidelines of the Max Planck Society. Experimental procedures were approved by Regierungspräsidium Darmstadt, file number V54-19c20/15-F126/1002.

The animals were bred in captivity at the animal husbandry department of the Max Planck Institute for Brain research. The room temperature was 22 °C, relative humidity 55% (±10%) and the light cycle was 12 h light/12 h dark. Autoclaved water and feed Sniff standard mouse extruded breeding or mouse/rat husbandry both ad libitum was provided. Animals were kept in breeding in type 2 long IVC greenline cages with red house and nesting material, bedding Lignocel BK8-15, under specific pathogen-free conditions.

EM image datasets, segmentations, RoboEM training, validation and test sets

RoboEM was developed and tested on a 92.6 × 61.8 × 94.8 μm^3 SBEM dataset from layer 4 primary somatosensory cortex of a 28-day-old mouse previously densely reconstructed[1]. The tissue was conventionally en bloc stained[60] and imaged at 11.24 × 11.24 nm^2 and nominal cutting thickness of 28 nm. Training and validation axons were sampled from a set of axons seeded by means of presynaptically classified segments obtained using SynEM[61] and skeleton traced by student annotators. To also acquire a volume reconstruction of training axons, segments of the oversegmentation obtained using SegEM with parameters as set for the whole-cell segmentation of the cortex dataset[22] were picked up and combined. The resulting volume mask was also used to iteratively optimize the interpolated skeleton tracing to yield a better centerline approximation. A set of 14 axons with 1.2-mm path lengths were used for training, and up to 0.7 million weight updates corresponding to ~260 epochs were run. The validation set consisted of 13 axons with 1.4-mm path length, where branches with less than 5 µm were excluded for better heuristic error detection. Results on the validation sets over the course of training are depicted in Extended Data Fig. 1b. A third set of ten axons with 1.7 mm path length seeded from a $(2.5\ \mu m)^3$ bounding box was used as a test set (Fig. 2b and Extended Data Fig. 1c). These are the same axons as previously used to evaluate human and semiautomated segmentation[1,42]. For automated spine head attachment on this dataset, the axon-trained RoboEM was evaluated on a random subset of 50 spine heads previously attached by human annotators, as well as on the set of spine heads previously used as a test set for semiautomated segmentation[1].

Next, RoboEM was evaluated on a $(150\ \mu m)^3$ subvolume (dataset ID Si150L4, https://wklink.org/7122) from a 1.3 × 1.3 × 0.25 mm^3 dataset from the barrel cortex of a 28-day-old male mouse ('Animal experiments' section) stained following the protocol by Hua et al.[62] with small modifications, sectioned at 35 nm using ATUM[56] and imaged at 4 × 4 nm^2 using multiSEM[57]. The segmentation and agglomeration applied to the dataset at a voxel size of 8 × 8 × 35 nm^3 (downsampled in *xy* by a factor of 2) were developed in collaboration with scalable minds GmbH, partly based on published approaches. In brief, a 3D U-Net[63] was used to predict per cardinal axis voxel affinities[17] from which a watershed-based oversegmentation was generated. The segments from the oversegmentation were then combined using hierarchical agglomeration[24]. In addition, neurite type predictions, blood vessel and nuclei detection were incorporated into the agglomeration to further reduce merge errors. Neurite type predictions were also used for spine head detection, which was the basis for RoboEM-based spine head attachment. For RoboEM training on axons, a training set was acquired based on a set of ten soma-seeded axons from layer 4, for which student annotators picked up segments from the oversegmentation to acquire a volume reconstruction. Here, we then used Kimimaro[64] followed by subsampling and B-spline interpolation to extract centerline skeletons from the volume reconstruction. This yielded a training set with a total of 21-mm axon path length. As axon validation set a random ten out of 20 axons seeded from a layer 4 bounding box of size $(2\ \mu m)^3$ were traced within a $(50\ \mu m)^3$ bounding box yielding 1-mm path length. The error rate on the validation set was minimal after 3.35 million gradient updates. For the axon test set, another $(1.5\ \mu m)^3$ bounding box within layer 4 was densely annotated and a random subset of five axons were traced within the $(150\ \mu m)^3$ subvolume Si150L4 yielding 1.7 mm path length. A separate RoboEM model was trained on spine head attachment. This training and validation set was generated from a set of 20 $(5\ \mu m)^3$ bounding boxes sampled within layer 4 and annotated for spine heads. A subset of around 1,000 spine heads were volume annotated at 4 × 4 × 35 nm^3 by student annotators from the spine head through the spine neck up to the dendritic trunk. Here, we again used Kimimaro[64] to extract centerline skeletons from the volumetric masks for training yielding 2 mm of spine neck tracings. Additionally, a random subset of 76 spine heads with 0.2-mm path length from the 20 bounding boxes was skeleton traced to serve as validation set on which the error was minimal after 0.9 million gradient updates. The evaluation was done on another randomly selected $(5\ \mu m)^3$ bounding box (Si11L3 https://wklink.org/2458) containing 91 densely annotated spine head to dendritic trunk skeleton tracings.

For the evaluation of RoboEM-based error correction of existing state-of-the-art segmentation, we applied RoboEM to a published mm^3-scale multiSEM dataset[18] from human cortex with voxel size 4 × 4 × 33 nm^3 segmented and agglomerated using FFNs[32]. Specifically, we focused on a $(150\ \mu m)^3$ bounding box containing 6.5 mm of axon path length of the provided ground truth skeleton tracings. The published ground truth skeleton tracings in this box were then used to evaluate FFN on soma-seeded axons[18] (Extended Data Fig. 3). To fine-tune RoboEM on this dataset, we generated ground truth skeleton tracings of dense seeded axons by sampling a bounding box of size $(2.5\ \mu m)^3$ within a centered $(15\ \mu m)^3$ bounding box, annotating all processes in this bounding box and then sampling a random subset of five axons, which were traced throughout the $(150\ \mu m)^3$ bounding box yielding 1.25 mm path length. Here, skeleton annotations were already done with high precision along the centerline such that no postprocessing was necessary. The volumetric neurite mask needed for training was generated by means of segment pick-up from the c3 FFN segmentation[18]. The best RoboEM model checkpoint, based on the validation set from the axon training in the mouse cortex ATUM-multiSEM dataset (subvolume Si150L4), was then used as initialization and RoboEM was trained for another 1.95 million gradient updates until converging in terms of reset-based error rates on the training axons. For the evaluation of FFN and RoboEM another bounding box of size $(1.5\ \mu m)^3$ was annotated for a random subset of five axons traced throughout the $(150\ \mu m)^3$ bounding box of the human cortex multiSEM dataset[18], yielding 1.4 mm path length.

Neurite flight reconstruction

We phrase the problem of neurite reconstruction from a 3D-EM volume as a centerline reconstruction task, in which a neurite is represented by a sequence of visited points. A CNN[65] is trained on the task of predicting the local neurite continuation from a neurite-centered and -aligned 3D-EM subvolume (Fig. 1d,e and Extended Data Fig. 1a) similar to human annotation flight mode[42]. Integration of the predicted neurite continuation yields a new position and orientation, which is used to generate the subsequent CNN input. Iterative application of this procedure turns a starting location and orientation into a neurite skeleton

reconstruction using only 3D-EM data and without intermediary steps, such as volume segmentation.

The input to the CNN consists of a 96 × 96 × 16 voxel neurite-centered and -aligned 3D-EM subvolume that covers, for example for axons, a field of view of ~1 × 1 × 0.7 µm^3 (Fig. 1e and Extended Data Fig. 1a). The third (z) dimension corresponds to the current flight direction, while the current position is made the center of the fourth z plane. The field of view is thus asymmetric along the z direction with more contextual information available in forward than in reverse flight direction. Empirically, this allows for better steering toward the 'exit' within axonal varicosities. The EM data is projected onto neurite-aligned planes by means of trilinear interpolation.

For the CNN architecture, we use seven 3D strided convolutional layers, followed by a dropout layer (dropout rate 0.5), three fully connected layers and a last linear layer that estimates the two steering commands and the distance to membrane along the flight direction (Extended Data Fig. 1a). The two steering commands are the Bishop curvatures[54] described in more detail below. A two-dimensional CNN architecture previously proposed for image-based road following[53] served as a starting point for the architecture in this work. As nonlinearity we tested rectified linear units[66] (ReLU) and exponential linear units[67] (ELU), and found ELU to work better (Extended Data Fig. 1b).

For the mathematical formulation of the network's input and output, the Bishop frame, a local orthonormal coordinate system spanned by a tangential vector $\mathbf{t}$ and two normal vectors $\mathbf{n}_1$ and $\mathbf{n}_2$ and obeying a rotation minimizing property, and the associated Bishop curvatures k_1 and k_2, were found to be particularly suitable[54] (Fig. 1d and Extended Data Fig. 1a). Here, we define the neurite-aligned projection planes for the network's input by means of the normal Bishop vectors, while the network's flight direction corresponds to the tangential vector. The evolution of the Bishop frame unit vectors is coupled via the Bishop curvatures k_1 and k_2, one for each normal vector direction and corresponding to signed curvatures for steering toward left to right and up and down, respectively. Note that the signed curvature has previously been used as steering command for left to right steering in image-based road following limited to planar curves[53]. The task of predicting the neurite continuation in the form of Bishop curvatures can be interpreted as fitting a parabola to the neurite's centerline, where the curvature vector $\mathbf{k} = k_1\mathbf{n}_1 + k_2\mathbf{n}_2$, determines direction and magnitude of bending of this parabola (Fig. 1d and Extended Data Fig. 1a).

Training and inference

For training the above-described CNN architecture on the neurite-following task, we implemented the CNN in TensorFlow[68] and trained with a mini-batch size of 128 using RMSProp[69] with momentum[70,71], for the SBEM dataset[1] and Adam[72], for the multiSEM datasets, on minimizing the mean-squared error plus a L2 regularization loss term on the weights. Weights for layers with ReLU or ELU activations were initialized following He et al.[73], and the final layer was initialized following Glorot et al.[74].

Similar to a finding by Pomerleau[52] for image-based road following, we also found that training on the neurite centerline alone does not yield good generalization performance during inference, as current position and orientation depend on past network decisions and errors can accumulate. To achieve stable path following, we trained the CNN on off-centerline positions and off directions with correspondingly adjusted steering leading back to the neurite centerline, where we refer to the mapping from a particular off position, off-direction state to the adapted steering as the flight policy.

We derived a greedy flight policy based only on local information available to the CNN within the finite field of view (details below). We induced the notion of obstacle (that is, the membrane) avoidance through the usage of a dynamic convergence distance set to the distance to the plasma membrane along the flight direction, which empirically allowed for better performance than a constant value. Additionally, we used the distance to plasma membrane as an auxiliary loss term during training.

To reconstruct neurites during inference, the steering predictions of the CNN for a position and orientation are integrated to a new position and orientation used to generate the subsequent input (Extended Data Fig. 1a, normal inference). For the random rotation inference mode (Extended Data Fig. 1a) we additionally perform a random rotation around the tangential after the integration step, which decorrelates consecutive inputs and comes at negligible computational cost. In both inference modes, given a start position and orientation, the recurrent application of the CNN yields a trace of visited points corresponding to the network's prediction of the neurite centerline.

Whereas for an error correction framework, such as FocusEM[1], both start position and orientation can be provided, for some use cases the orientation might not be available. Specifically, for the task of attaching spine heads to their corresponding dendrites, it can be less obvious how to compute the start orientation. To apply RoboEM on spine neck tracing tasks given only the start positions, we run stochastic forward passes through the dropout layer, known as Monte Carlo dropout[75,76], yielding a sampled distribution of predictions, from which we estimate prediction uncertainty. The uncertainty can then be used to select a start orientation from a list of candidate orientations or rank tracings thereby allowing for a trading off split against merge errors. As there is only a single dropout layer after the convolutional layers (Extended Data Fig. 1a), the computation for different dropout masks is shared up to this point and the overhead of Monte Carlo dropout is negligible (<2% in terms of floating point operations for 128 Monte Carlo samples).

B-spline interpolation and RoboEM step size

To get a continuous representation of neurite branches from sparsely placed nodes from human skeleton reconstructions, we use degree 4 B-spline interpolation[77] yielding a curve $\boldsymbol{\gamma} \in C^3$. We reparametrize the curve to get a curvature adaptive step size $\|\dot{\boldsymbol{\gamma}}\,\Delta t\|$ of

$$\|\dot{\boldsymbol{\gamma}}\|\Delta t = f\frac{d}{1+\frac{p}{2}\kappa},$$

where $\kappa = \|\mathbf{k}\|$ is the curvature, $p = 1$ µm the physical size of the projection plane and f is a step size factor set to $f = 1$ for training that can be adjusted to higher values for inference. While for most evaluations in this work, we also used $f = 1$ during inference, step sizes can be increased up to $f = 5$ yielding correspondingly increased throughput without sacrificing accuracy (Extended Data Fig. 1c). A default step size d matching the smallest dimension of a voxel, such as $d = 11.24$ nm in case of the L4 SBEM dataset[1] at $f = 1$ ensures that no voxels up to a radius of half the projection plane size in the projection plane at the current position of the CNN are skipped. To keep the notation uncluttered, subsequent formulas are expressed in terms of the arc length parametrized curve $\boldsymbol{\gamma}(s)$ indicated by the parameter s.

Bishop frame

The Bishop frame[54] consists of three orthonormal vectors, namely the tangential vector $\mathbf{t}$, in this work also termed flight direction, and two normal vectors $\mathbf{n}_1, \mathbf{n}_2$. The evolution equations of the Bishop frame and the curve $\boldsymbol{\gamma}$ read as follows:

$$\frac{\mathrm{d}}{\mathrm{d}s}\boldsymbol{\gamma} = \mathbf{t}$$

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathbf{t} = k_1\mathbf{n}_1 + k_2\mathbf{n}_2 \equiv \mathbf{k}$$

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathbf{n}_1 = -k_1\mathbf{t}$$

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathbf{n}_2 = -k_2\mathbf{t}.$$

Here, k_1 and k_2 are the Bishop curvatures associated to the normal vector $\mathbf{n}_1$ and $\mathbf{n}_2$, respectively, and $\mathbf{k}$ is the curvature vector (Fig. 1d and Extended Data Fig. 1a). Flips and rotations around the tangential vector applied to both normal vectors and Bishop curvatures are invariance transformations that leave $\mathbf{k}$ and thereby the evolution of $\boldsymbol{\gamma}$ and $\mathbf{t}$ unchanged and are used for data augmentation during training.

The rotation minimizing Bishop frame has weaker requirements for a parameterized curve $\boldsymbol{\gamma}$ than the Frenet–Serret frame often used as local coordinate system in differential geometry of parameterized curves. Specifically, the Bishop frame of a parametrized curve in 3D Euclidean space does not require $\kappa \neq 0$, and only requires $\boldsymbol{\gamma} \in C^2$ instead of $\boldsymbol{\gamma} \in C^3$: for details, see ref. 54. Initial conditions in form of start position and Bishop frame orientation together with the Bishop curvatures uniquely define the centerline curve $\boldsymbol{\gamma}$. The integration of above Bishop equations during inference was performed with either of two methods: (1) forward Euler method followed by Gram–Schmidt orthonormalization to maintain an orthonormal basis for the Bishop frame, or (2) analytically derived evolution equations for a Bishop frame along a parabola. Empirically, we found the latter to work better especially for larger step size factors f (Extended Data Fig. 1c).

Flight policy

To derive a flight policy for stable path following during inference, the Taylor series expansion up to second order $\boldsymbol{\gamma}_T$ for the known neurite centerline curve $\boldsymbol{\gamma}(s)$ with corresponding Bishop frame $\mathbf{t}$, $\mathbf{n}_1$, $\mathbf{n}_2$ and curvature vector $\mathbf{k}$, and the Taylor series expansion up to second order $\underline{\boldsymbol{\gamma}}_T$ of the unknown off-centerline, off-direction curve $\underline{\boldsymbol{\gamma}}(s)$, with $\mathbf{t}$, $\mathbf{n}_1$, $\mathbf{n}_2$, $\mathbf{k}$:

$$\boldsymbol{\gamma}_T(s) = \boldsymbol{\gamma} + s\,\mathbf{t} + \frac{s^2}{2}\mathbf{k}$$

$$\underline{\boldsymbol{\gamma}}_T(s) = \underline{\boldsymbol{\gamma}} + s\,\underline{\mathbf{t}} + \frac{s^2}{2}\underline{\mathbf{k}}$$

were used.

Knowing the current off position $\underline{\boldsymbol{\gamma}}$ and off-direction $\mathbf{t}$, as well as the current closest position on the neurite centerline $\boldsymbol{\gamma}$ with correct neurite-aligned orientation $\mathbf{t}$ and local shape of the curve in terms of $\mathbf{k}$, that is correct steering for the on centerline case, we derived suitable corrected steering $\mathbf{k}$ converging back to the centerline within some distance s_c, that is, which minimizes the future distance $\|\boldsymbol{\Gamma}_T\|$ defined by:

$$\underbrace{\boldsymbol{\gamma}_T - \underline{\boldsymbol{\gamma}}_T}_{\boldsymbol{\Gamma}_T} = \underbrace{\boldsymbol{\gamma} - \underline{\boldsymbol{\gamma}}}_{\boldsymbol{\Gamma}} + s\underbrace{(\mathbf{t} - \underline{\mathbf{t}})}_{\mathbf{T}} + \frac{s^2}{2}\underbrace{(\mathbf{k} - \underline{\mathbf{k}})}_{\mathbf{K}}.$$

We therefore require:

$$\nabla_{\underline{\mathbf{k}}} \|\boldsymbol{\Gamma}_T(s_c)\|^2 = 0,$$

where $\nabla_{\underline{\mathbf{k}}} = \underline{\mathbf{n}}_1 \partial_{\underline{k}_1} + \underline{\mathbf{n}}_2 \partial_{\underline{k}_2}$. Defining the projection operator $\mathcal{P}_{\underline{\mathbf{n}}_1,\underline{\mathbf{n}}_2} = \underline{\mathbf{n}}_1 \otimes \underline{\mathbf{n}}_1 + \underline{\mathbf{n}}_2 \otimes \underline{\mathbf{n}}_2$, where $\otimes$ denotes an outer product, the solution for $\underline{\mathbf{k}}$ from the above equation reads as follows:

$$\underline{\mathbf{k}} = \mathcal{P}_{\underline{\mathbf{n}}_1,\underline{\mathbf{n}}_2}\left(\frac{2}{s_c^2}[\boldsymbol{\Gamma} + s_c\mathbf{t}] + \mathbf{k}\right).$$

Iterative application with updated values for $\boldsymbol{\gamma}_T$ at the closest point $\boldsymbol{\gamma}$ from the current position $\underline{\boldsymbol{\gamma}}$ yields a stable path following flight policy with a free parameter s_c controlling the convergence speed. Trajectories for different values of s_c are plotted in Extended Data Fig. 1b. For this work, during training, we set s_c to the current distance to membrane along the flight direction and thereby train RoboEM on a membrane avoidance strategy.

Direction prediction with Monte Carlo dropout

For the direction prediction, as for example needed for spine head attachment where only an initial position is given, we sample $I = 256$ roughly equidistant orientations $\mathbf{t}_i$ and run Monte Carlo dropout[75] with $M = 128$ samples. From the sampled Bishop curvature predictions $\mathbf{k}_{im}$, we compute the mean curvatures $\langle\kappa\rangle_i$ and the covariances of the Bishop curvatures cov_i per orientation i over the Monte Carlo samples. The uncertainty estimate u_i is then taken as the square root of the largest eigenvalue of the covariance matrix divided by the mean curvature:

$$u_i = \sqrt{\max(\mathrm{eigvals}(\mathrm{cov}_i))}/\langle\kappa\rangle_i.$$

Using cosine similarity for neighboring angles within 30°, we construct a weighted average of uncertainties $\hat{u}_i$ as

$$\hat{u}_i = \frac{u_i}{2} + \frac{1}{2}\sum_j \hat{w}_{ij} u_i$$

with weights $\hat{w}_{ij}$

$$\hat{w}_{ij} = w_{ij} / \sum_j w_{ij}$$

$$w_{ij} = \begin{cases} \langle \mathbf{t}_i | \mathbf{t}_j \rangle^{16} & \text{if } \arccos(\langle \mathbf{t}_i | \mathbf{t}_j \rangle) < 30° \\ 0 & \text{otherwise} \end{cases}$$

to give more stable predictions. The orientation with minimal averaged uncertainty $\hat{u}_i$ was used as first orientation candidate. For spine head attachment in the mouse cortex ATUM-multiSEM dataset (subvolume Si11L3), we used a second candidate that was >110° from the first and has again minimal averaged uncertainty among the remaining orientations.

Model selection from validation set

To choose a model from a pool of trained models with different learning rates, fields of view and so on, we evaluated RoboEM on a validation set consisting of human skeleton annotations without any segmentation. Specifically, RoboEM was evaluated on linear neurite branches by running recurrent inference starting from both sides. In case RoboEM reaches a first set of thresholds (Supplementary Table 1) concerning the distance or the angle with respect to the ground truth tracing, we consider the tracing to be 'experimental', that is the progress along the ground truth tracing is temporarily not considered until the first set of thresholds is not exceeded anymore. If a second set of thresholds (Supplementary Table 1) is exceeded, the tracing is classified as erroneous and a reset to the closest point on the ground truth before the tracing turned experimental is performed. Tracing is stopped, when the closest point on the ground truth is at the other end of the axonal branch and the tracing is at that time classified as correct.

To relate resets caused by steering errors to merge and split error rates, we consider each reset both a merge error into a wrong process and a split error due to not continuing the neurite of interest. Hence, we count each reset as two errors and report this as a reset-based error rate. While this cannot be used to directly compare with segmentation error rates, it serves as a metric for model selection.

For model selection on the SBEM L4 dataset[1], we find that averaged over all models and training iterations the random inference mode outperforms its normal mode counterpart by 33% (range 12–52%). The best performing model on the validation set uses only EM data as input, has an ELU activation function, was trained for 700,000 training iterations and yielded 17 errors per mm (Extended Data Fig. 1b).

Split and merge error evaluation

For the evaluation of split and merge errors of agglomerations before and after RoboEM-based error correction (Fig. 2 and Extended Data

Fig. 3), we detected merge errors that extended further than 2.2 µm from the ground truth and manually verified that this heuristic accurately detects merge errors. For agglomerations before RoboEM corrections, each merge error was counted as 0.5 and divided by the ground truth path length to yield the merge error rate. This is because each merge error usually connects two neurites and counting them as one error per neurite instead of 0.5 would overestimate the total amount of merge errors. For sparse evaluations of RoboEM-based error corrections limited to agglomerates that overlap with the ground truth, as done in both multiSEM datasets evaluated in this work, additional mergers introduced by RoboEM were counted as one instead of 0.5, which accounts for mergers from agglomerates not overlapping with ground truth and therefore not observable in a sparse evaluation. For split errors, we restricted the set of agglomerates to be evaluated to those agglomerates that overlap more than 2.5 µm with the ground truth. We introduced this overlap threshold to avoid domination of split errors by many small agglomerates or unagglomerated segments along thin stretches of axons. Note that despite this overlap length threshold, around 90% of the ground truth was still covered.

For the multiSEM datasets, RoboEM-based correction was restricted to ending resolution, for which endings were extracted from skeleton representations of those agglomerates that overlapped with the ground truth annotations. In addition to the RoboEM validation strategy, we also made use of subcellular type predictions to decide whether a RoboEM tracing should connect two agglomerates. Note that type predictions were also used by FFN to avoid merge errors across subcellular types[18]. After agglomerates were reconnected using RoboEM tracings, the resulting agglomeration state was evaluated as before.

To test whether RoboEM-based correction also allows for reduced split error rates when compared at the same merge error rate as the segmentation and/or agglomeration for the two multiSEM datasets, the following strategy was used to only apply subsets of RoboEM tracings thereby limiting merge error rates: prediction uncertainty using Monte Carlo dropout was quantified for every RoboEM step, then the maximum uncertainty over steps was taken and forward and validation tracings were combined with a minimum of the two uncertainty scores. The rationale was to first score every direction of the validated tracings according to the position with maximum uncertainty and then, since the forward and backward tracings yielded the same flight path, the minimum of those two uncertainty scores was taken to quantify the uncertainty of the tracing as a whole. Application of validated RoboEM tracings up to percentiles of 0.2, 0.4, 0.6 and 0.8 of tracing uncertainties then yielded partial RoboEM corrections. These were evaluated for split and merge error rates and those with minimal split error rates at same merge error rate as segmentations/agglomerations were added as intermediate points within the split merge error plane of Fig. 2b. Specifically, for the mouse cortex multiSEM dataset, application of the top 80% of validated RoboEM tracings on the agglomeration at 85% agglomeration threshold yielded the same merge error rate as the agglomeration at 90% agglomeration threshold. Similarly for the human cortex multiSEM dataset, the application of the top 40% of validated RoboEM tracings on the FFN c3 agglomeration yielded the same merge error rate as the FFN c2 agglomeration[18].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data necessary to reproduce reported results for the mouse cortex SBEM dataset[1] are available: https://wklink.org/9276 (raw data), https://L4dense2019.brain.mpg.de (code and data of the original publication); trained RoboEM weights, and evaluation data are available in the supplement subject to provisions as stated in the code availability section (below). Raw data for the mouse cortex multiSEM dataset are available: Si1L3 https://wklink.org/2458 (spine head attachment test set), Si150L4 https://wklink.org/7122 (axon test set); RoboEM training data for the mouse cortex multiSEM dataset are available: https://wklink.org/8172. Raw data and training data for the human cortex multiSEM dataset[18] were obtained from and are available as detailed in ref. 18 at https://h01-release.storage.googleapis.com/landing.html. Trained RoboEM weights for mouse and human multiSEM data are available on reasonable request to allow usage monitoring according to licensing criteria (all source code and binary files are publicly available under the limited right to use for the exclusive purpose of undertaking academic or not-for-profit research, as further detailed in the code availability section below). In addition, data needed to run the example code is part of the zipped code package provided in the supplementary material subject to provisions as stated in the code availability section (below). Source data are provided with this paper.

Code availability

All source code and binary files are publicly available in the supplement under the following provisions: any copyright or patent right is owned by and proprietary material of the Max Planck-Gesellschaft zur Förderung der Wissenschaften e.V. (MPG). MPG makes no representations or warranties of any kind concerning the source code and binary files, neither express nor implied and the absence of any legal or actual defects, whether discoverable or not. The source code and binary files available in the supplement are subject to a non-exclusive, revocable, non-transferable, and limited right to use for the exclusive purpose of undertaking academic or not-for-profit research. Use of the Code or binary files or any part thereof for commercial or clinical purposes and alterations are strictly prohibited in the absence of a Commercial License Agreement from MPG. Download and/or use of the source code and binary files under the aforementioned research license comes with the acceptance of the license agreement provided with the download.

References

60. Briggman, K. L., Helmstaedter, M. & Denk, W. Wiring specificity in the direction-selectivity circuit of the retina. *Nature* **471**, 183–188 (2011).
61. Staffler, B. et al. SynEM, automated synapse detection for connectomics. *eLife* **6**, e26414 (2017).
62. Hua, Y., Laserstein, P. & Helmstaedter, M. Large-volume en-bloc staining for electron microscopy-based connectomics. *Nat. Commun.* **6**, 7923 (2015).
63. Ronneberger, O., Fischer, P. & Brox, T. U-Net: convolutional networks for biomedical image segmentation. In *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention* Vol. 9351 (eds Navab, N. et al.) 234–241 (Springer, 2015).
64. Silversmith, W., Bae, J. A., Li, P. H. & Wilson, A. M. Kimimaro: skeletonize densely labeled 3D image segmentations. *Zenodo* https://doi.org/10.5281/zenodo.5539913 (2021).
65. LeCun, Y. et al. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1**, 541–551 (1989).
66. Glorot, X., Bordes, A. & Bengio, Y. Deep sparse rectifier neural networks. In *Proc. 14th International Conference on Artificial Intelligence and Statistics (AISTATS)* Vol. 15, 315–323 (2011).
67. Clevert, D.-A., Unterthiner, T. & Hochreiter, S. Fast and accurate deep network learning by exponential linear units (ELUs). Preprint at https://doi.org/10.48550/arXiv.1511.07289 (2016).
68. Abadi, M. et al. TensorFlow: a system for large-scale machine learning. In *Proc. 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16)* 265–283 (2016).
69. Tieleman, T. & Hinton, G. Overview of mini-batch gradient descent. *Neural Networks for Machine Learning, Lecture 6a* https://www.cs.toronto.edu/~hinton/coursera/lecture6/lec6.pdf (2015).

70. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986).
71. Sutskever, I., Martens, J., Dahl, G. E. & Hinton, G. E. On the importance of initialization and momentum in deep learning. In *Proc. International Conference on Machine Learning* Vol. 28, 1139–1147 (MLResearchPress, 2013).
72. Kingma, D. P. & Ba, J. L. Adam: a method for stochastic optimization. Preprint at https://doi.org/10.48550/arXiv.1412.6980 (2015).
73. He, K., Zhang, X., Ren, S. & Sun, J. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In *Proc. IEEE International Conference on Computer Vision* 1026–1034 (IEEE, 2015).
74. Glorot, X. & Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In *Proc. 13th International Conference on Artificial Intelligence and Statistics (AISTATS)* 249–256 (2010).
75. Gal, Y. & Ghahramani, Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. Preprint at https://doi.org/10.48550/arXiv.1506.02142 (2016).
76. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
77. Piegl, L. & Tiller, W. *The NURBS Book* (Springer, 1996).

Acknowledgements

We thank B. Staffler and M. Berning for discussions; T. Colnaghi and K. Reuter for discussions and code contributions related to compute performance; C. Guggenberger for management of compute infrastructure (Max Planck Compute and Data Facility, Garching, Germany) and V. Pinkau, N. Rzepka, J. Striebel and G. Wiese (scalable minds GmbH, Potsdam, Germany) for contributions to segmentation, agglomeration and subcellular type predictions on the mouse cortex multiSEM dataset. This research was funded by the Max Planck Society.

Author contributions

M.H. conceived, initiated and supervised the study. M. Schmidt conceptualized, developed and implemented all aspects of RoboEM and performed analyses. A.M. contributed to RoboEM development and adapted it for axon reconstruction in large-scale datasets. M. Sievers provided mouse ATUM-multiSEM data. M. Schmidt and M.H. wrote the paper with contributions by all authors.

Funding

Open access funding provided by Max Planck Society.

Competing interests

A patent application (EP22188663.3-1210) and a PCT application (PCT/EP2023/071656) have been filed by the Max Planck Society at the European Patent Office naming M. Schmidt, A.M. and M.H. as inventors. A.M., M. Schmidt, M. Sievers and M.H. are beneficiaries of a licensing agreement between the Max Planck Society and scalable minds, Inc., relating to automated image processing of 3D-EM data.

Additional information

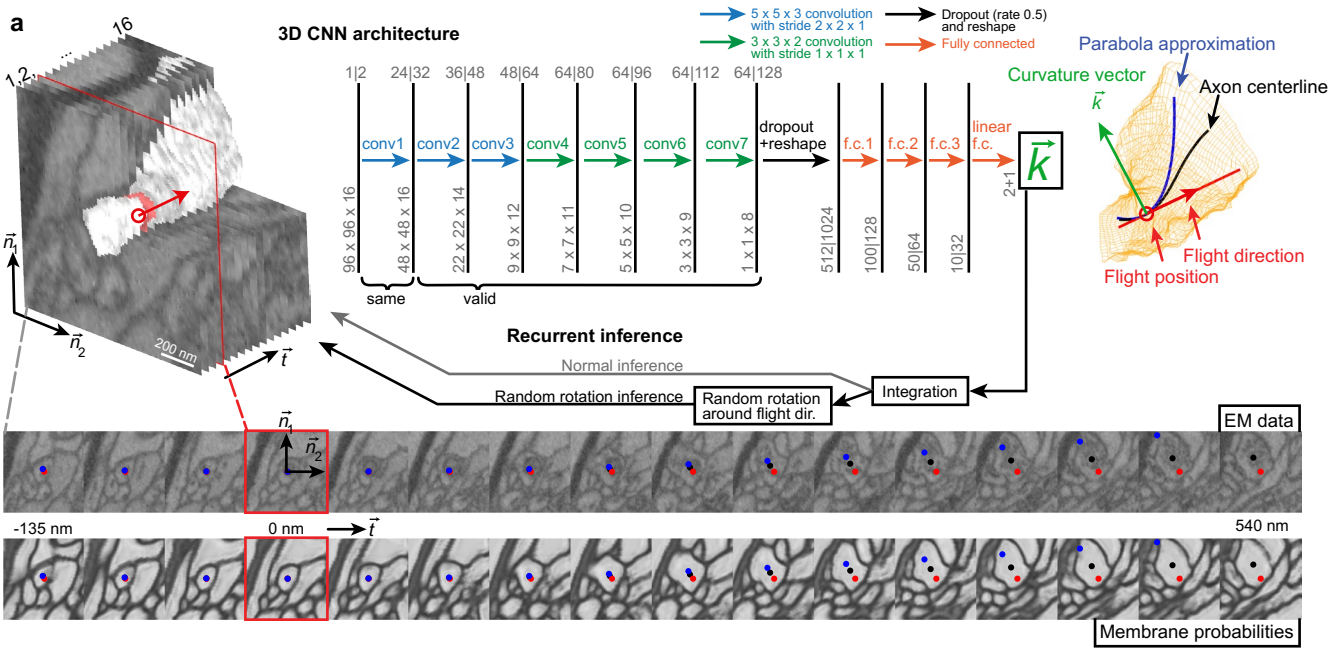
Extended data is available for this paper at https://doi.org/10.1038/s41592-024-02226-5.

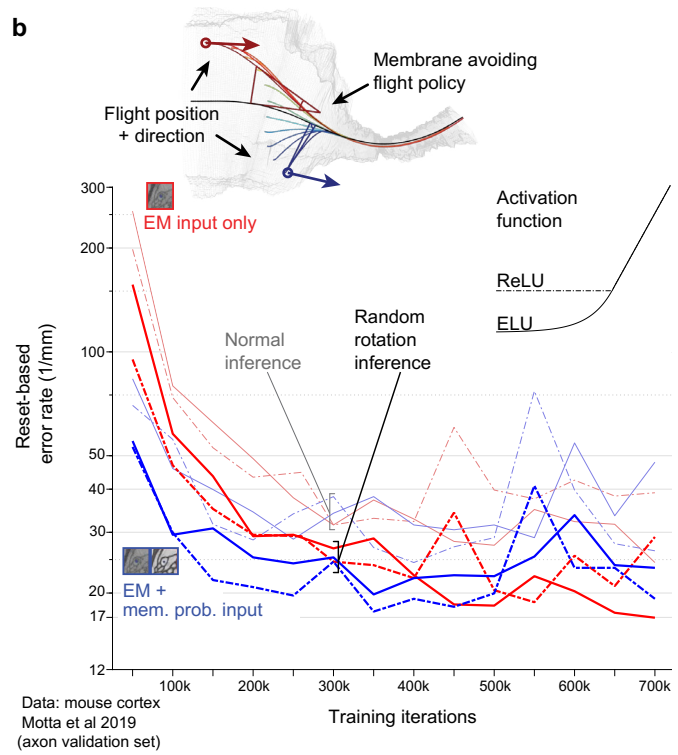
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41592-024-02226-5.

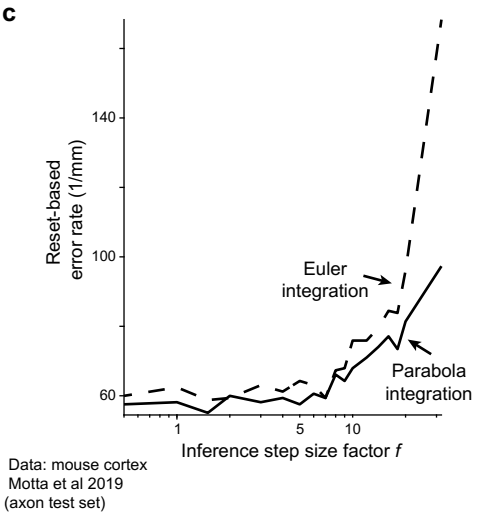
Correspondence and requests for materials should be addressed to Martin Schmidt or Moritz Helmstaedter.

Peer review information *Nature Methods* thanks Albert Cardona and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Nina Vogt, in collaboration with the *Nature Methods* team.

https://doi.org/10.1038/s41592-024-02226-5

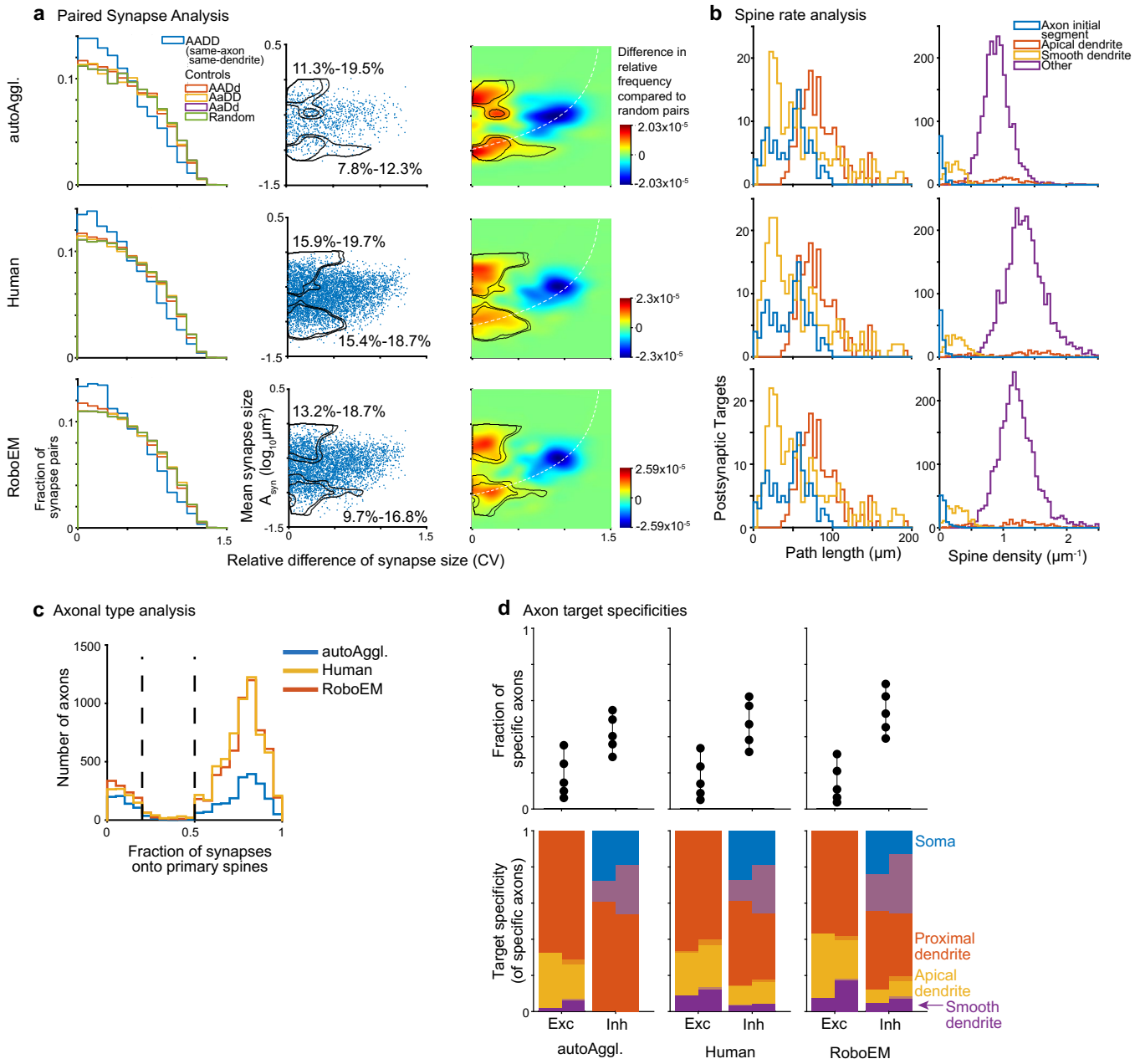






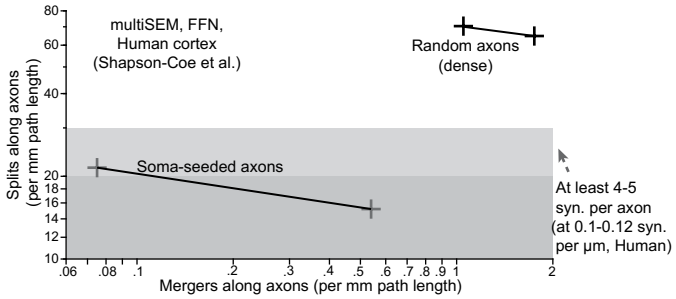
Extended Data Fig. 1 | Automated flight reconstruction. (a) 3D Convolutional neural network architecture variants processing projections of neurite-aligned subvolumes of EM input data and making predictions about neurite continuation in terms of a curvature vector; EM data from Motta et al.[1]. This curvature vector is representative of a parabola approximation to the neurite centerline and can be integrated during recurrent inference to yield the next position and orientation and thereby the next subvolume to be processed. Performing random rotations around the flight direction during inference can be considered zero-cost test time augmentation that allows to decorrelate subsequent inputs with respect to the orientation. The sequence of visited points during recurrent inference represents a skeleton reconstruction of the neurite. **(b)** Training iterations using a membrane avoiding flight policy for off-center and off-direction inputs and comparing ReLU and ELU activation functions[66,67], as well as EM only input versus EM and membrane probability maps as input. EM data was used for predicting neurite continuation. Random rotation inference mode as depicted in (a) was used. **(c)** Reset-based evaluation of RoboEM for different step size factors f on the test set axons from the mouse cortext dataset[1]. (f = 1.5).

https://doi.org/10.1038/s41592-024-02226-5



Extended Data Fig. 2 | Connectomic analyses compared across reconstruction states. Analyses based on EM data and reconstructions from Motta et al.[1] **(a)** Paired same-axon same-dendrite analysis yielding similar fractions of paired synapses that are consistent with Hebbian learning. **(b)** Spine rate analysis yielding underestimates of apical dendrite spine densities for the reconstruction state before RoboEM or manual-annotation-based spine attachment. **(c)** Connectomic definition of excitatory and inhibitory axons with at least 10 synapses based on fraction of primary spine innervation, compared for the automated state (AutoAggl.) before RoboEM, the manual-annotation-based split and merge resolution of axons, and RoboEM application. **(d)** Axonal target specificities of excitatory and inhibitory axons.

https://doi.org/10.1038/s41592-024-02226-5



Extended Data Fig. 3 | Reconstruction difficulty in proximal (soma-seeded) vs. dense (randomly-seeded) axons. Split and merge error rates for random and soma-seeded axons in the two published FFN-segmentations on multiSEM data from human cortex[18]. Note that soma-seeded axons are substantially less error-prone and therefore not representative for dense axon reconstruction performance (proximal axons are usually wider and less convoluted, thus easier to reconstruct).

nature portfolio

Corresponding author(s): Helmstaedter, Moritz

Last updated by author(s): Nov 23, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used.
Data analysis	All source code and binary files are publicly available in the supplement under the following provisions: Any copyright or patent right is owned by and proprietary material of the Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V. ("MPG"). MPG makes no representations or warranties of any kind concerning the source code and binary files, neither express nor implied, and the absence of any legal or actual defects, whether discoverable or not. The source code and binary files available in the supplement are subject to a non-exclusive, revocable, non-transferable, and limited right to use for the exclusive purpose of undertaking academic or not-for-profit research. Use of the Code or binary files or any part thereof for commercial or clinical purposes and alterations are strictly prohibited in the absence of a Commercial License Agreement from MPG. Download and/or use of the source code and binary files under the aforementioned research license comes with the acceptance of the license agreement provided with the download. Any further distribution is prohibited. Publicly available software used in this study: * Segmentation software SegEM (https://github.com/mhlabCodingTeam/SegEM/, version: 1.0) from Berning et al. Neuron (2015) * Synapse detection software SynEM (https://gitlab.mpcdf.mpg.de/connectomics/SynEM) from Staffler et al. eLife (2017) * Analysis algorithms (https://gitlab.mpcdf.mpg.de/connectomics/L4dense/) from Motta et al. Science (2019) * Skeletonization software Kimimaro (https://github.com/seung-lab/kimimaro, version: 0.2.1) from Silversmith et al. (2021) * Deep learning software Tensorflow (https://www.tensorflow.org/, version: 2.3.0) from Abadi et al. OSDI'16 (2016)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data necessary to reproduce reported results for the mouse cortex SBEM dataset (Motta et al. 2019) are available: https://wklink.org/9276 (raw data), https://L4dense2019.brain.mpg.de (code and data of the original publication); trained RoboEM weights, and evaluation data are available in the supplement subject to provisions as stated in the code availability section (see above).
Raw data for the mouse cortex multiSEM dataset are available: Si11L3 https://wklink.org/2458 (spine head attachment test set), Si150L4 https://wklink.org/7122 (axon test set); RoboEM training data and trained RoboEM weights to reproduce results for the mouse cortex multiSEM dataset are available upon reasonable request. Raw data for the human cortex multiSEM dataset (Shapson-Coe et al. 2021) are available: https://h01-release.storage.googleapis.com/landing.html; RoboEM training data and trained RoboEM weights to reproduce results for the human cortex multiSEM dataset are available upon reasonable request.
In addition, data needed to run the example code is part of the zipped code package provided in the supplementary material subject to provisions as stated in the code availability section (see above).

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	n/a
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	n=3 three dimensional electron microscopy datasets were used to ensure variabilities from acquision/imaging modalities (SBEM versus multiSEM) and species (mouse versus human) have no negative impact on performance; however, the main conclusions of this study rely on the sample sizes and/or path lengths from manually proof-read axons (see below); n=90 ending queries and n=100 chiasma queries of axons were used to compare RoboEM neurite tracings with human neurite tracings; n=5 random axons for the two multiSEM datasets and n=10 random axons for the SBEM dataset were used with similar total path lengths of over 1 millimeter were used for evaluation of split and merge error rates; n=91 spine heads were used in the multiSEM dataset by Sievers et al. to evaluate attachment of spine heads using RoboEM
Data exclusions	No data were excluded.
Replication	Reconstruction algorithms as described in this study were successfully run on multiple datasets by different scientists confirming the results of this study. Source code and data required to reproduce results are made available as indicated in the code and data availability sections.
Randomization	All samples were selected randomly without input from investigators. Randomization was done using pseudo-random number generators in MATLAB or Python.
Blinding	N/A

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	Experiments were performed on laboratory animals, species mus musculus, strain C57BL/6-J from age P5 until 6 months.
Wild animals	No wild animals were used.
Reporting on sex	Experiments were approved for male mice. Sex was approved in animal husbandry and only male mice were used for experiments and sample preparation. Reason for exclusion of females was comparability with already published data, no further knowledge so far about the influence of estrus, combined with low sample numbers prepared in the studies.
Field-collected samples	No field collected samples were generated/used.
Ethics oversight	Experiments were performed as part of the approved animal research project “Nervenzellschaltkreise Grosshirnrinde Nager”. Approval was granted from the Animal Welfare Agency, Regional Council of Darmstadt/Tierschutzbehörde, Regierungspräsidium Darmstadt, file number: V54-19c20/15-F126/1002. The approval was granted from Feb. 18th 2015 until Feb 17th 2020. Ethical guidance on the study protocol was provided by the Animal Welfare Officer of the Max Planck Institute for Brain Research.

Note that full information on the approval of the study protocol must also be provided in the manuscript.