NeuroGPS-Tree: automatic reconstruction of large-scale neuronal populations with dense neurites

Tingwei Quan[1–4], Hang Zhou[1–3], Jing Li[1–3], Shiwei Li[1–3], Anan Li[1–3], Yuxin Li[1–3], Xiaohua Lv[1–3], Qingming Luo[1–3], Hui Gong[1–3] & Shaoqun Zeng[1–3]

The reconstruction of neuronal populations, a key step in understanding neural circuits, remains a challenge in the presence of densely packed neurites. Here we achieved automatic reconstruction of neuronal populations by partially mimicking human strategies to separate individual neurons. For populations not resolvable by other methods, we obtained recall and precision rates of approximately 80%. We also demonstrate the reconstruction of 960 neurons within 3 h.

Mapping neuronal circuits, one of the central tasks in brain studies, requires identification of individual neurons, their synaptic connections and their cell type from imaging data sets of neuronal populations[1,2]. These analyses depend largely on reconstruction (i.e., digitization) of the morphologies of individual neurons or neuronal trees in a population[3,4]. Advances in labeling and imaging techniques have enabled the generation of high-quality population imaging data[5–8], which has resulted in extensive and detailed studies of neuronal networks and has inspired the reconstruction of large neuronal populations. Manual tracing of several neurons has been used to resolve dendritic morphology, but it is time consuming and impractical for population reconstruction.

Software tools for automatic reconstruction have been described[9–14], most of which can trace individual branches of neuronal trees successfully. These techniques can be extended to sparsely labeled neurons but are not useful for dense populations, in which closely spaced neurites may come into contact, connect or overlap. If falsely inferred links are not identified, individual neuronal trees may be combined in the reconstruction (**Supplementary Fig. 1**). Even some of the best available methods[9–11] have been unable to identify individual neuronal trees in reconstruction tests using such populations (**Supplementary Fig. 2**).

When we examined the principles underlying these tools, we noticed that the use of only local signal intensity[9–14] or specific link patterns[15,16] to evaluate suspected links might not resolve problematic links. When an experienced human completes manual segmentation, he or she uses global cues in addition to local signals to progressively identify spurious links. This strategy is essential for separating individual neurons from interwoven neuronal arbors with high accuracy. Inspired by this, we developed a tool for neuronal reconstruction that integrates such a global segmentation strategy into the framework of automatic reconstruction. This work builds on our previous software[17], NeuronGlobalPosition-System (NeuroGPS), and is called NeuroGPS-Tree. The principles behind NeuroGPS-Tree are outlined in **Supplementary Figure 3**.

Using our neurite-tracing algorithm (Online Methods and **Supplementary Note 1**), we reconstruct a neuronal network from all neurites traced. We typically find many spurious links in the reconstructed network. We then decompose this network into individual neuronal trees (**Fig. 1a**) by identifying and breaking the spurious links. Superficially, this decomposition seems straightforward, but it also has a solid theoretical basis, as it is essentially a graph partitioning problem (**Supplementary Note 2**).

In contrast to published methods[9–14] that use local information from around the contact site, our strategy is to use information at scales ranging from the contact site to the entire network (**Fig. 1b**,**c**) for identification, along with corresponding statistical information extracted from neuronal morphology (**Fig. 1d**,**e**). The procedure consists of three parts (**Fig. 1c**). First, at the neurite scale, we formulate a decision rule to identify the direction of a neurite (Online Methods). The decision rule is based on statistical morphological information and results in highly accurate identification of fiber direction (**Supplementary Fig. 4**). Depending on the identified fiber directions, we define and delete three different types of spurious connections (Online Methods). Second, at the whole-network scale, we search for 'bridges' that connect different neuronal trees. We then delete the link at the end where the fluorescence signal strength is weakest. We repeat this procedure until there are no paths linking any two trees. Finally, we revise each extracted neuronal tree on the basis of tree structure information (**Fig. 1d**) and statistical morphological information. More details regarding the identification of spurious links can be found in the Online Methods.

To obtain statistical information relating to neuronal morphology, we analyzed approximately 300 neurites from three manually traced neurons. (Increasing the neurite number beyond 300 did not change the statistical conclusion.) A neuronal tree can be described as a tree-like structure (**Fig. 1d**) in which the starting

[1]Britton Chance Center for Biomedical Photonics, Wuhan National Laboratory for Optoelectronics–Huazhong University of Science and Technology, Wuhan, China. [2]Ministry of Education Key Laboratory for Biomedical Photonics, Huazhong University of Science and Technology, Wuhan, China. [3]Department of Biomedical Engineering, School of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, China. [4]School of Mathematics and Statistics, Hubei University of Education, Wuhan, China. Correspondence should be addressed to S.Z. (sqzeng@mail.hust.edu.cn) or H.G. (huigong@mail.hust.edu.cn).

RECEIVED 21 DECEMBER 2014; ACCEPTED 22 OCTOBER 2015; PUBLISHED ONLINE 23 NOVEMBER 2015; DOI:10.1038/NMETH.3662

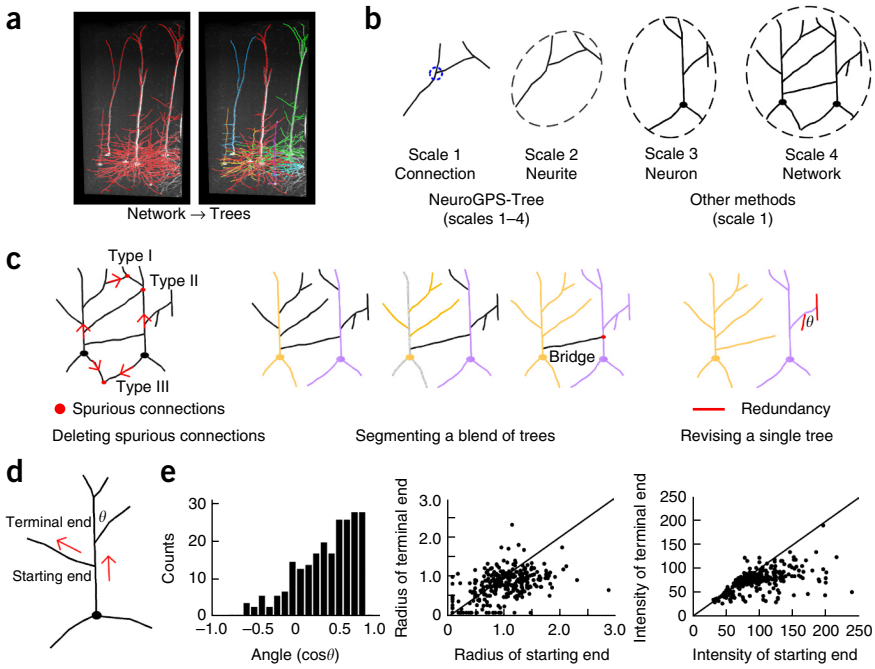

Figure 1 | Integrating information at different scales to progressively detect spurious links in neuronal population reconstructions. (**a**) Deconstruction of a reconstructed network into six neuronal trees. (**b**) NeuroGPS-Tree uses signal information at different scales to identify spurious links. (**c**) The procedure for identifying spurious links. In the left-hand panel, spurious links (red dots) are defined and deleted after confirmation of the neurites' directions (red arrows). The center panel illustrates the segmentation of the blended trees. First, neurites that connect directly to the soma are assigned to their own soma. Second, the assigned neurites (gray) are deleted, and the neurites that have no links to any other soma (purple) are assigned to the soma directly connected to the deleted neurites. Third, a bridge from the unassigned neurite is identified (black), and one end of the bridge (red dot) is deleted on the basis of comparison of the signal strengths of the two ends of the bridge. In the right-hand panel, a neuronal tree is revised after the redundant neurites (red) have been identified. (**d**) A neuronal tree in which the directions of all neurites can be determined (red arrows). (**e**) The statistical morphological information extracted from the experimental data set (three neurons with approximately 300 neurites) includes the angles between two neurites that come into contact, as well as the size and intensity of the ends of the neurites. The minimum radius was set to 0.1 μm. For values plotted below the diagonal lines, the *x*-component is larger than the *y*-component.

end of a neurite connects to another neurite and the terminal end terminates without a link; this allowed us to uncover differences between the signals from the two ends of a given neurite. We concluded that in most cases, the size and signal intensity of the starting end were greater than those of the terminal end (72% versus 28% for size and 80% versus 20% for intensity; **Fig. 1e**). We also computed angles between two directly linked neurites and found that the angle was likely to be less than 90° (83% versus 17%; **Fig. 1e**). We referred to the above information as statistical information describing neuronal morphology.

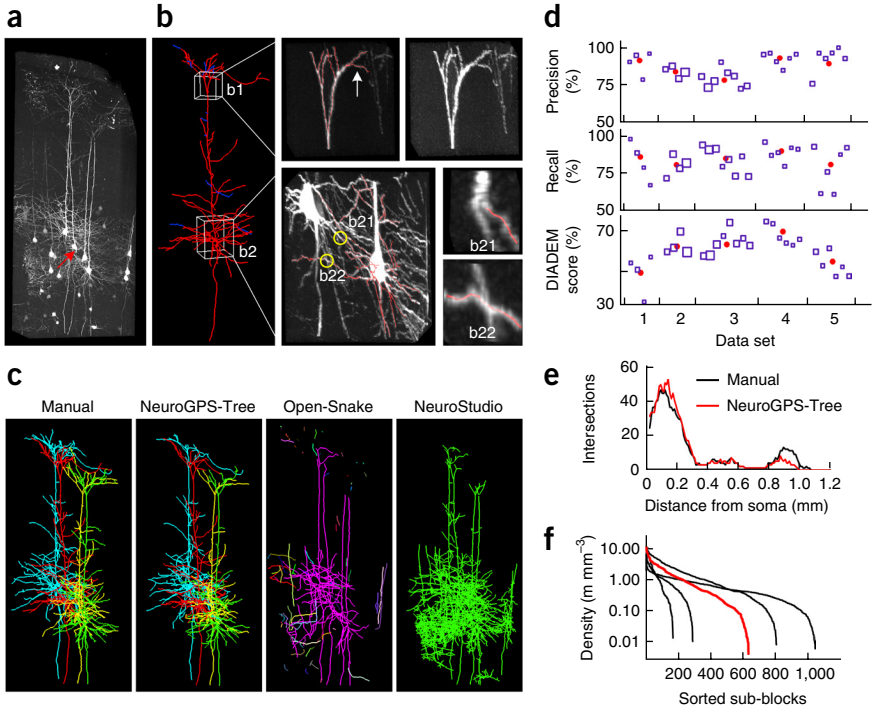

Figure 2 | Reconstruction performance of NeuroGPS-Tree. (**a**) Image stack of a neuronal population from neocortex. The red arrow indicates a neuron. (**b**) Left, a neuronal tree from the population reconstructed by three expert annotators. Red represents agreement among the annotators, and blue denotes disagreement. The center and rightmost panels show enlarged views of the reconstructed results and the original signals. The top row of zoomed images, corresponding to the boxed area labeled "b1" in the leftmost panel, shows that by reducing the range of voxel color, we were able to obtain high contrast between foreground and background. The bottom row, corresponding to the boxed area labeled "b2," shows manually resolved details of the dense neurites, with examples marked by yellow circles (b21 and b22). (**c**) Comparison of manual, NeuroGPS-Tree, Open-Snake and NeuronStudio reconstructions of a neuronal population. Individual neuronal trees from the population are delineated in different colors. (**d**) Evaluating the reconstruction by NeuroGPS-Tree on five data sets (**Supplementary Figs. 5–7**). Each open square represents one neuronal tree reconstruction, and each red dot represents the averaged value for each group. The size of the square has a positive linear relationship to the total length of the neuronal processes of the manually reconstructed neuronal tree. The reconstruction of data set 2 is shown in **c**. (**e**) Sholl intersections calculated from the reconstruction of a neuronal tree (red arrow in **a**) plotted against the distance from the soma. (**f**) Neurite density distribution, measured as the total length of process in a sub-block; the red curve corresponds to data set 2.

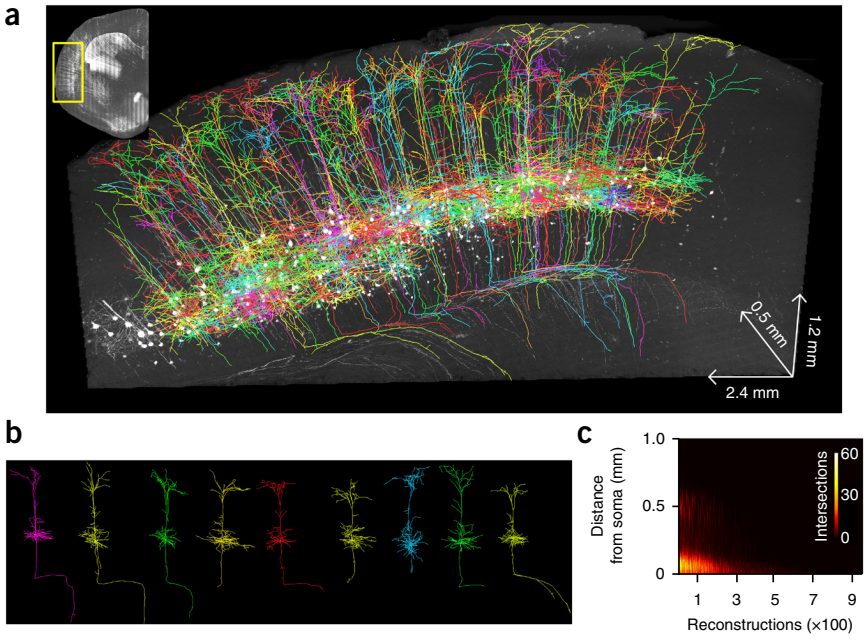

Figure 3 | Reconstruction of a neuronal population from dense and large-scale data using NeuroGPS-Tree. (**a**) A neuronal population including approximately 960 neuronal trees was reconstructed in 3 h; here 126 typical neuronal trees are displayed in different colors. The image stack was selected from the cortical region outlined by the yellow rectangle (top left). (**b**) Single neurons selected from the reconstruction in **a** shown in detail. (**c**) Sholl analysis of the 960 reconstructed neurons as an example of high-throughput quantitative analysis.

We used manual reconstructions of neurons as ground truths to evaluate our automatic algorithm using a previously published method[18] (**Fig. 2a**,**b** and **Supplementary Video 1**). Three skilled annotators each manually reconstructed several fluorescently labeled cortical neurons. The ratio of agreement to disagreement among annotators was greater than 50:3, indicating >94% accuracy for manual reconstruction. This high accuracy can be attributed to the following factors: (1) annotators can reduce the range of voxel color to make a process clearer, and (2) human analysis can identify spurious links between closely neighboring processes.

We then evaluated the performance of NeuroGPS-Tree using data sets from different brain areas, including neocortex, hippocampus and thalamic nucleus. We compared the reconstruction performance of manual reconstruction, NeuroGPS-Tree and two reference algorithms (Open-Snake[9] and NeuronStudio[10]). For better visualization, we compared portions of reconstructed neuronal trees with many neurites (**Fig. 2c** and **Supplementary Figs. 5–7**). On the basis of these comparisons, NeuroGPS-Tree reconstructions were similar to manual reconstructions, whereas other software tools could not reconstruct neuronal populations from these data sets. NeuroGPS-Tree was able to reconstruct all neuronal trees of a neuronal population in the analyzed domain. Additional reconstructions by NeuroGPS-Tree are provided in **Supplementary Figure 8**.

We also quantified differences between the manual and NeuroGPS-Tree reconstructions. The results showed that both average recall and precision were approximately 80%, and the DIADEM score[19] was approximately 0.6 (**Fig. 2d**). The major loss might have resulted from the dense interwoven fibers near the soma and the low-contrast far-end dendrites (**Fig. 2e**). Note that these evaluations of NeuroGPS-Tree for multiple neurons (**Fig. 2d**) are on par with those of Open-Snake, NeuronStudio and Vaa3D for a single neuron (**Supplementary Fig. 9**). For data sets containing multiple neurons (**Supplementary Fig. 8**), the precision rates for Open-Snake and NeuronStudio were approximately 30%, and the recall rates were approximately 50% (**Supplementary Fig. 10**). In addition, we determined why the DIADEM score sometimes falls considerably below the recall and precision rates (**Supplementary Fig. 11** and **Supplementary Discussion**).

We also counted Sholl intersections for the neuronal trees derived by manual reconstruction and by NeuroGPS-Tree and calculated the correlation coefficient. We found small differences between Sholl measurements for the automated and manual reconstructions (**Fig. 2e**). For the reconstructions in **Figure 2c**, the correlation coefficients ranged from 0.96 to 0.98. In addition, we estimated the neurite density of these five data sets (**Fig. 2f**); the highest process density ranged from 10 to 12 m mm^{-3}.

We used NeuroGPS-Tree to reconstruct a neuronal population from a dense, large-scale data set. Our data set from a cortical region contained 2,400 × 1,200 × 500 voxels (1.36 GB in size), in which many neuronal processes were densely distributed (**Supplementary Video 2**). Without any manual intervention, a neuronal population containing approximately 960 neuronal trees could be reconstructed within 3 h (Intel Xeon CPU 3.46 GHz computing platform). For better visualization, we displayed 126 typical neuronal trees from the reconstructed neuronal population (**Fig. 3a** and **Supplementary Video 3**). These reconstructions provided detailed neuronal structures (**Fig. 3b**) and enabled further morphological quantitative measurements such as Sholl analysis (**Fig. 3c**). In summary, our results suggest that NeuroGPS-Tree is capable of high-throughput characterization of the morphological features of a neuronal population. We also demonstrated that NeuroGPS-Tree can be successfully applied to data sets from various cortical regions (**Supplementary Fig. 12**).

NeuroGPS-Tree is suitable for processing image stacks acquired by different image modalities, such as DIADEM data sets[20] (**Supplementary Figs. 13–16**), and by commercial microscopy with proper parameter selection (**Supplementary Software**). NeuroGPS-Tree is also compatible with other suitable tracing methods (**Supplementary Fig. 17**). Nevertheless, a high-quality image data set is a prerequisite for a better reconstruction (**Supplementary Fig. 18**).

NeuroGPS-Tree reconstructs neuronal populations by partially mimicking the strategy used by experienced annotators and progressively approaches an accurate reconstruction by repetitively using statistical information about neuronal morphology at multiple scales

(**Supplementary Discussion**). These features make NeuroGPS-Tree an effective tool for analyzing data sets in which the neurite density is too complex for previously established methods. NeuroGPS-Tree is also suitable for the analysis of large-scale data sets, and it may be useful for mapping neuronal circuits.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank the members of the Britton Chance Center for Biomedical Photonics for advice and help in experiments. We also thank S.L. Hill and Y. Wang for suggesting Sholl analysis, as well as H. Peng and G.A. Ascoli for help with software usability and paper quality. This work is supported by the National Basic Research Program of China (grant 2011CB910401), the National Natural Science Foundation of China (grants 81327802 and 91432116), the Science Fund for Creative Research Group of China (grant 61421064), the National Key Scientific Instrument & Equipment Development Program of China (grant 2012YQ030260) and the the Director Fund of the Wuhan National Laboratory for Optoelectronics.

AUTHOR CONTRIBUTIONS

S.Z. and H.G. conceived of the project. S.Z. and T.Q. designed the model and wrote the manuscript. T.Q. developed the algorithm. H.Z. wrote the software. J.L., H.Z. and S.L. performed image analysis and processing. A.L. and Y.L. constructed the computing platform for image preprocessing. S.Z., H.G., X.L. and Q.L. produced data. All authors revised the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Lichtman, J.W. & Denk, W. *Science* **334**, 618–623 (2011).
2. Helmstaedter, M. & Mitra, P.P. *Curr. Opin. Neurobiol.* **22**, 162–169 (2012).
3. Meijering, E. *Cytometry A* **77**, 693–704 (2010).
4. Donohue, D.E. & Ascoli, G.A. *Brain Res. Rev.* **67**, 94–102 (2011).
5. Li, A. *et al. Science* **330**, 1404–1408 (2010).
6. Ragan, T. *et al. Nat. Methods* **9**, 255–258 (2012).
7. Silvestri, L., Bria, A., Sacconi, L., Iannello, G. & Pavone, F. *Opt. Express* **20**, 20582–20598 (2012).
8. Osten, P. & Margrie, T.W. *Nat. Methods* **10**, 515–523 (2013).
9. Wang, Y., Narayanaswamy, A., Tsai, C.-L. & Roysam, B. *Neuroinformatics* **9**, 193–217 (2011).
10. Wearne, S.L. *et al. Neuroscience* **136**, 661–680 (2005).
11. Peng, H., Ruan, Z., Long, F., Simpson, J.H. & Myers, E.W. *Nat. Biotechnol.* **28**, 348–353 (2010).
12. Türetken, E., Gonzalez, G., Blum, C. & Fua, P. *Neuroinformatics* **9**, 279–302 (2011).
13. Zhao, T. *et al. Neuroinformatics* **9**, 247–261 (2011).
14. Chothani, P., Mehta, V. & Stepanyants, A. *Neuroinformatics* **9**, 263–278 (2011).
15. Lee, Y.-H., Lin, Y.-N., Chuang, C.-C. & Lo, C.-C. *Neuroinformatics* **12**, 487–507 (2014).
16. Gala, R., Chapeton, J., Jitesh, J., Bhavsar, C. & Stepanyants, A. *Front. Neuroanat.* **8**, 37 (2014).
17. Quan, T. *et al. Sci. Rep.* **3**, 1414 (2013).
18. Helmstaedter, M., Briggman, K.L. & Denk, W. *Nat. Neurosci.* **14**, 1081–1088 (2011).
19. Gillette, T.A., Brown, K. & Ascoli, G.A. *Neuroinformatics* **9**, 233–245 (2011).
20. Brown, K.M. *et al. Neuroinformatics* **9**, 143–157 (2011).

ONLINE METHODS

Tracing of neuronal processes using the constrained principal curve. In this section we describe the procedure for automatically tracing neuronal processes in a generalized manner. A more detailed description of the tracing procedure is provided in **Supplementary Note 1**. We developed an automatic method based on the constrained principal curve to trace neuronal processes in three-dimensional image stacks. Briefly, automatically tracing a neuronal process requires the use of an algorithm to extract the neuronal-process centerline formed by a series of traced points. In the constrained-principal-curve method, in addition to the local signal of the current point in the neuronal process, the previously traced points and the detected direction of the current point determine the direction of the next point. In contrast, the principal-curve method[21] uses only the local signal for this purpose. This modification provides more opportunities to correctly determine the next direction from the current position (**Supplementary Fig. 19a**,**b**). When the direction of one part of the neuronal process changes sharply, the constraint of the current direction may lead to an incomplete trace. The rayburst sampling method[22] is a supplemental method that handles this, allowing one to obtain complete traces (**Supplementary Fig. 19c**,**d**). We used a typical method for detecting collisions to build the links between neuronal processes (**Supplementary Note 1**). Moreover, the procedure was designed to identify the crossings and bifurcations of two neuronal processes. Because of these features, our tracing method can generate an effective reconstruction when two neuronal processes cross, when two processes are positioned in parallel and when the direction of one portion of a neuronal process changes sharply (**Supplementary Fig. 20**).

Reconstruction of a neuronal population. Population reconstruction is essential to separate individual trees from the reconstructed network and can be regarded as a special graph partitioning problem (**Supplementary Note 2** and **Supplementary Fig. 21**). The reconstructed network consists of traced neuronal processes, of which any two can connect directly or indirectly. For a sparse reconstruction, the reconstructed network usually corresponds to a neuronal tree. A network that is reconstructed from high-density data sets usually contains a few to a few hundred neuronal trees. The sections below describe the decomposition of a reconstructed network into neuronal trees, including (1) deletion of spurious link patterns in the reconstructed network, (2) extraction of individual neuronal trees of a neuronal population and (3) revision of the extracted neuronal trees.

Deleting spurious link patterns in the reconstructed network. This step consists of three parts: (1) extracting statistical information describing neuronal morphology, (2) identifying the direction of a neuronal process and (3) searching for and deleting the spurious links in the reconstructed network. Here spurious link patterns are determined by the information provided in the first and second parts, and by the asserted identity of the spurious link candidates.

To extract the statistics of neuronal morphology, one needs to know the direction of the neuronal processes in a neuronal tree. This direction is defined as the direction from the starting end of a neuronal process to the distal end. Here the starting end is the end connected to the soma or to another neuronal process. According to this definition, the direction of the neuronal process can be obtained easily from the topological structure of a neuronal tree. In detail, the soma is assigned as the root of the tree, the neuronal processes connected to the soma are assigned as the first-grade nodes of the tree, and the neuronal processes connected to first-node neuronal processes are considered second-grade nodes. This procedure is repeated until all neuronal processes have been assigned and the tree structure can be obtained. From this tree structure, one can identify the starting end, which connects to higher-grade neuronal processes, and thus can also identify the direction of each neuronal process. Once the directions of all neuronal processes in a single tree have been fully identified, one can extract statistical features of neuronal morphology from the neuronal trees. From the analysis of three typical neuronal trees, we obtained the following statistical information: in most cases, the signal intensity and the radius of the starting end of a neuronal process are stronger and larger than those at the other end. In general, the angle between two connected neuronal processes is less than 90°.

Using the above statistical information for a single tree, we designed a decision rule and a procedure for identifying the direction of neuronal processes in a reconstructed network that contains two or more neuronal trees. For each neuronal process in the reconstructed network, the directional values of its two ends are defined as

$$F_i = r_i + \frac{v_i}{\min(v_1, v_2)} + c_i, (i = 1,2)$$

where r_i and v_i are the radius and the signal intensity, respectively, of the end of neuronal process i. If the end of i has a link, c_i is set to 0.7; otherwise, it is set to 0. Note that the value of c_i is determined experimentally, which ensures that the direction of a neuronal process can be identified with high precision (**Supplementary Fig. 4**). We selected five initial traced points at the end of i, and we used five spheres centered on these selected points to fit the neuronal process. The values of r_i and v_i were set to the average of the fitting radius of the spheres and the average signal intensity of the selected points, respectively. The direction of a neuronal process was determined on the basis of our condition that the starting end of a neuronal process is the end with a directional value 0.5 greater than that of the other end.

Using the above directional identification in the reconstructed network, we searched for and deleted three types of link patterns when the two ends of a neuronal process had these links. In type I, the link occurs between the terminal end of one neuronal process and the middle of another. In type II, the link occurs when the two ends of a neuronal process without an identified direction each connect to a neuronal process with an identified direction. In this case, we also assume that a neuronal process that does not have an identified direction has two directions that both start at the link, thus indicating that two links have full directional information and their angles can be computed if the following conditions are satisfied: (a) one angle is less than 90° and the other is more than 90°, and (b) the difference between the cosines of the two angles is more than 0.1 and the spurious link corresponds to the link with the smaller cosine value. In type III, the link occurs between the terminal ends of two neuronal processes.

Extract individual neuronal trees of a neuronal population. Using the soma localization method[17], one can count the number of somas in the reconstructed network. A single neuron contains only one soma; thus, the reconstructed network should be decomposed into neuronal trees that have the same number of somas. We perform the decomposition procedure as follows. First, in a reconstructed network, we assign the neuronal processes that directly connect to the soma to their own somas. Second, for a single soma, we delete the assigned neuronal processes and the corresponding links in the reconstructed network. In this case, we search for the unassigned neuronal processes that have no links with other somas, and we assign the resultant neuronal processes to the given soma. These two steps produce a reconstructed network that consists of somas as well as assigned and unassigned neuronal processes. Third, we use these somas and neuronal processes to build a connecting graph. Each point of the graph represents a soma and its assigned neuronal processes, called a soma point, or the unassigned neuronal process. A link in the graph corresponds to the link between a soma point and unassigned neuronal processes or a link between unassigned neuronal processes. Fourth, we search for a bridge that connects different neuronal trees, and we delete the spurious link between the ends of the bridge. In the connecting graph, we compute the shortest paths between any two soma points using the Dijkstra algorithm[23]. From these paths, we determine the frequency of a bridge neuronal process for which the two ends have links, we compute the directional value of the bridge that occurs most frequently, and we delete the link at the end of the bridge that has the smaller directional value. From the remaining links, the graph is reconstructed and the link is deleted again. This procedure is repeated until there is no path between any two soma points. Finally, in the connected graph, the soma point and its connected points can be converted to a neuronal tree.

Properties 1 and 2 (**Supplementary Note 2**) indicate that a candidate spurious link is the link between one end of a neurite with two links and another neurite. Here, a spurious link is a link that makes two soma nodes connected by a path. Thus, the procedure aims to identify neuronal processes for which both ends have links and then delete one of the two links. This procedure is effective at extracting neuronal trees from the reconstructed network and has been verified using a large-scale test data set (**Fig. 3** and **Supplementary Fig. 12**).

Revise neuronal trees. The extracted neuronal tree may contain neuronal processes from other neuronal soma that are not contained in the analyzed data set. Thus, the neuronal tree must be revised. A neuronal tree is a tree structure with different levels of nodes; thus, the direction of a neuronal process can be identified, and the angle between two connected neuronal processes can be computed. Using this information, we searched for and revised two cases in the tree structure (**Supplementary Fig. 22**). If the end of a neuronal process in a node connected with the middle part of a process in a lower node, we deleted the lower node and all of its child nodes (i.e., we deleted some neuronal processes). The weighted mean of the values (i.e., the inner products of the direction of a neuronal process in the node and of its connected processes in the lower node) was less than –0.3. The weight coefficient was proportional to the length of the neuronal process in the lower node. All lower nodes and their child nodes were deleted. We thereby obtained a neuronal tree that corresponded to the revised tree structure.

Sample preparation. The experiments were performed in accordance with the guidelines of the Experimental Animal Ethics Committee at Huazhong University of Science and Technology. We used the Thy1-YFP-H transgenic mouse line (adult P21 male mice) for our analysis. The procedures for sample preparation followed a previously published protocol; a detailed description can be found in ref. 24.

Data acquisition. The whole brain was imaged using a two-photon fluorescence micro-optical sectioning tomography system[25] (2p-fMOST) with chemical reactivation[24]. The experimental spatial resolution of 2p-fMOST is approximately 0.45 μm in the lateral dimension and approximately 1.8 μm in the axial dimension. The original voxel size of $0.5 \times 0.5 \times 1\ \mu m^3$ was merged to $1 \times 1 \times 1\ \mu m^3$ for most of the analysis, except for the analysis in **Figure 2a**, for which it was merged to $1 \times 1 \times 2\ \mu m^3$. These parameters are suitable for the reconstruction of a neuronal population.

Evaluation of reconstruction and visualization. We regarded manual reconstruction as the gold standard. Manual reconstruction was performed using the strategy described in the redundant skeleton consensus procedure[18] (RESCOP) with minor modifications to suit our data. It included the following steps: (1) overlaying of multiple independent manual reconstructions obtained using commercial Amira software, (2) identification of the locations of agreement and disagreement from an overlay of the reconstructions and (3) re-inspection of the locations of disagreement and use of a voting method to reach an agreement among annotators as to whether the reconstruction of the re-inspected locations was accepted. The voting method obeyed the simple-majority rule, which differs from the statistical model used with the RESCOP. The main reason for this is that the tested experimental data set was relatively simple compared with the data set collected by volume electron microscopy, and if the disagreement locations contained neurite signal, high contrast between the foreground and background could be generated by reducing the range of voxel color. Note that for more complicated data sets, the statistical model provided in the RESCOP is very valuable.

After obtaining manual reconstructions, we evaluated the automated reconstruction on the basis of the recall and precision rates as well as the DIADEM score. Both automated and manual reconstructions of a single neuronal tree are represented by skeletons consisting of a series of points. For any given point on the automatically reconstructed skeleton, we searched for the point on the manually reconstructed skeleton that was nearest to that point. If the distance between the given point and the found point was less than the predetermined value (8 μm), the given point was regarded as a true positive point on the automatically reconstructed skeleton. Similarly, true positive points on the manually reconstructed skeleton can be obtained. The recall rate is defined as the ratio of the number of true positive points to the number of points on the manually reconstructed skeleton. The precision rate is defined as the ratio of the number of true positive points to the number of points on the automatically reconstructed skeleton. The calculation of the DIADEM score is described in ref. 19. Please note that these indexes are used to evaluate the

doi:10.1038/nmeth.3662

reconstruction of one neuronal tree, rather than a population. The recall rate, precision rate and DIADEM score of the neuronal population reconstruction were calculated using the weighted average of the recall rate, the precision rate and the DIADEM score of the reconstructions of neuronal trees with many neurites, in which the weight was proportional to the total length of the neuronal processes of an individual neuron identified in the manual reconstruction.

This parameter was set to 8 μm in our application on the basis of the following considerations. First, the dendrite near the soma is very thick, and the corresponding nodes identified by an algorithm may deviate from those that are manually identified. A small value of this parameter will cause this type of node to be regarded as a false node. Second, the total neurite length of each neuron is very large, so the precision and recall rates are quite stable across a wide range of the distance parameter.

Empirical results also support the above facts. Using two neurons in **Figure 2a** as an example, we demonstrated that the recall and precision rates are functions of the distance parameter (**Supplementary Fig. 23**). We observed that when the parameter was less than 6 μm, the recall and precision rates varied sharply, and when the parameter ranged from 6 μm to 10 μm, the fluctuation was approximately 2%. We therefore set it to 8 μm, which might not be optimal but is reasonable.

NeuroGPS-Tree can be used to preliminarily visualize and examine the reconstructed results. Third-party, professional visualization software is recommended for displaying the results.

Sholl analysis. In Sholl analysis[26], a series of sphere surfaces in 10-μm radial increments were centered at the neuronal soma, and the reconstructed neurites intersecting each sphere surface were counted. Using this operation, we were able to plot the Sholl intersections versus the distance from the soma.

Calculating neurite density. Each data set was divided equally into hundreds of sub-blocks. The size of the sub-blocks was 50 × 50 × 50 μm^3. The total length of the manually detected neuronal processes in each sub-block was computed and sorted. "Neurite density" refers to the ratio of the calculated total length and volume size of the sub-block.

Testing the efficiency of NeuroGPS-Tree. For the data set shown in **Figure 3a** (960 reconstructed neuronal trees, 1.36 GB), testing on an Intel Xeon CPU 3.46 GHz computing platform required approximately 3 h of processing for neuronal population reconstruction, including soma localization and soma shape reconstruction; thus, hundreds of neuronal trees can be obtained within an hour using NeuroGPS-Tree. For a data set with a size of 0.65 GB (**Supplementary Fig. 12**), processing lasted approximately 1 h. In addition, the amount of memory used by NeuroGPS-Tree was approximately ten times the size of the tested data set.

Code availability. NeuroGPS-Tree is open source and available as **Supplementary Software**. Additional test data sets and a user guide can be downloaded from https://sourceforge.net/projects/neurogps-tree or https://vbn.org.cn/gosoftware.

21. Bas, E. & Erdogmus, D. *Neuroinformatics* **9**, 181–191 (2011).
22. Ming, X. *et al. PLoS One* **8**, e84557 (2013).
23. Dijkstra, E.W. *Numerische Mathematik* **1**, 269–271 (1959).
24. Xiong, H. *et al. Nat. Commun.* **5**, 3992 (2014).
25. Zheng, T. *et al. Opt. Express* **21**, 9839–9850 (2013).
26. Sholl, D.A. *J. Anat.* **87**, 387–406 (1953).