bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ShuTu: Open-Source Software for Efficient and Accurate Reconstruction of Dendritic Morphology

Dezhe Z. Jin[1*], Ting Zhao[2], David Hunt[2], Rachel Pearcy[2], Ching-Lung Hsu[2], Nelson Spruston[2*]

1 Department of Physics and Center for Neural Engineering, the Pennsylvania State University, University Park, Pennsylvania, U.S.A
2 Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia, U.S.A

∗ Corresponding authors: Dezhe Z. Jin, dzj2@psu.edu, Nelson Spruston, sprustonn@janelia.hhmi.org

Abbreviated title: **S**huTu

Number of pages: 70
Number of figures: 16
Number of words in Abstract: 169
Number of words in Significance Statement: 43
Number of words for Introduction: 611
Number of words for Discussion: 1074

Conflict of Interest: The authors declare no competing financial interests.

Acknowledgements: Research for D.Z.J was supported by the Visiting Scientist Program at the Janelia Research Campus, Howard Hughes Medical Institute; and by the Huck Institute of Life Sciences at the Pennsylvania State University.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Neurons perform computations by integrating inputs from thousands of synapses – mostly in the dendritic tree – to drive action potential firing in the axon. One fruitful approach to understanding this process is to record from neurons using patch-clamp electrodes, fill the recorded neuron with a substance that allows subsequent staining, reconstruct the three-dimensional architecture of the dendrites, and use the resulting functional and structural data to develop computer models of dendritic integration. Accurately producing quantitative reconstructions of dendrites is typically a tedious process taking many hours of manual inspection and measurement. Here we present ShuTu, a new software package that facilitates accurate and efficient reconstruction of dendrites imaged using bright-field microscopy. The program operates in two steps: (1) automated identification of dendritic process, and (2) manual correction of errors in the automated reconstruction. This approach allows neurons with complex dendritic morphologies to be reconstructed rapidly and efficiently, thus facilitating the use of computer models to study dendritic structure-function relationships and the computations performed by single neurons.

Significance Statement

We developed a software package – ShuTu – that integrates automated reconstruction of stained neurons with manual error correction. This package facilitates rapid reconstruction of the three-dimensional geometry of neuronal dendritic trees, often needed for computational simulations of the functional properties of these structures.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

The geometry of dendritic arbors directly influences synaptic integration and the resultant firing patterns of neurons (Mainen and Sejnowski, 1996; Henze et al., 1996; Stuart and Spruston, 1998; Krichmar et al., 2002). Dendritic morphologies vary widely across and within regions of the brain (Parekh and Ascoli, 2013), so consideration of morphology is an important aspect of understanding the mechanisms by which different neurons carry out their unique functions. Intracellular recording of neurons is a common technique for studying dendritic integration of input signals (Hamill et al., 1981; Stuart and Spruston, 1998). To fully understand the implications of these experiments, numerical simulations of the recorded neurons are often needed (Jaeger, 2001; Krichmar et al., 2002; Gidon and Segev, 2012; Menon et al., 2013). Informative simulations require accurate reconstructions of the geometry of the recorded neurons, including branching structures and diameters of the branches.

The traditional method of reconstructing neuron morphology requires intensive human labor (Zandt et al., 2017). A slide containing a neuron filled with biocytin is mounted on a motorized stage and imaged using a video camera mounted to a bright-field microscope. The neuron image is displayed on a computer screen, and the reconstruction is done manually. The user clicks the mouse along the images of dendritic branches on the screen. While clicking, the user adjusts the cursor size to match the diameters, and turns the z-position knob on the microscope to keep the branches in focus. Each click records the x, y, and z positions and the radius r at a single point, and connects the point to the previously clicked point. Bifurcations are marked and followed up sequentially. The morphology is recorded in a series of these clicked points.

Manual reconstruction in this way is computationally straightforward. Since it requires no image storage or processing, the computational demand is minimal. However, there are several drawbacks, especially when the accuracy of reconstruction is crucial. Repetitive clicking while measuring the radii and turning the focus knob makes manual reconstruction labor-intensive and time-consuming. The problem is exacerbated at high magnification. To see fine processes of neurons, it is desirable to image neurons with an objective at 100X magnification and a large

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

numerical aperture (Jaeger, 2001; Brown et al., 2011). In our experience, however, it can take 10 - 15 hours or more of continuous work to reconstruct the dendritic tree of a pyramidal neuron in this way. Over this period of time, instability of the sample in the microscope can lead to problems. Furthermore, the accuracy of the reconstruction can suffer from fatigue-induced mistakes. Another problem with manual reconstruction is that the accuracy is hard to check independently because it is difficult to precisely align the previous reconstruction with the neuron image after remounting the slide.

Automatic reconstruction of neuron morphology using computer algorithms promises to reduce manual labor and increase productivity. There have been intensive efforts towards this goal, including open-source projects such as the Digital Reconstruction of Axonal and Dendritic Morphology Challenge (DIADEM)(Liu, 2011; Svoboda, 2011; Gillette et al., 2011; Gillette et al., 2011) and the BigNeuron project (Peng et al., 2015). Commercial software is also moving in this direction. In our experience, however, available software suffered from a variety of problems, including limited automation and tedious approaches for error correction. Thus, we sought to develop an open-source software platform that would overcome these limitations. In this paper, we describe our open-source software package, ShuTu (Chinese for "dendrite") – a system for reconstructing biocytin-filled neurons efficiently and accurately. To avoid the impression of marketing our software, we make no attempt to compare it to other open-source or commercial software; instead, we encourage others to try it and judge for themselves.

Results

We demonstrate the use of ShuTu by going through the steps involved in reconstructing a single CA3 pyramidal neuron from a mouse hippocampal slice. We then present reconstruction results for other cell types as well. All neurons were stained following patch-clamp recordings in brain slices prepared from 17-30 day-old male mice (C57Bl/6), using biocytin-containing intracellular solution. Following recording and staining, neuron reconstruction proceeded according to the following steps: (1) image acquisition; (2) image processing; (3) automated reconstruction; (4)

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

manual editing and error correction. Additional details regarding slice recordings and computer systems requirements are provided in the Materials and Methods section. Operational commands for ShuTu are provided in Appendix 1. Technical details regarding the algorithms used in ShuTu are provided in Appendix 2.

Image acquisition

ShuTu uses tiles of tiff stacks covering the entire neuron (Fig. 1). Nearby tiles should overlap by ∼20%, in order to facilitate accurate stitching of tiles into a single image. We imaged hippocampal neurons using a Zeiss AxioImager microscope with AxioCam and ZEN blue software. Once the boundary in the field of view (xy) and the range of the depths (z) that contain the neuron were set, the images at each tile position and depth were acquired automatically, and the positions of the these images were stored in an xml file (image metadata). Other microscope/software combinations can be used, as long as tiff stacks and their relative positions are provided to ShuTu (see Materials and Methods for details). It is also possible to use ShuTu to reconstruct neurons imaged using two-photon, confocal, or wide-field fluorescence microscopy (see Discussion). However, we have restricted our use of the software to neurons stained using biocytin and a dark reaction product.

The number of images required to capture the full three-dimensional (3D) morphology of a neuron depends on its size and the magnification of the microscope objective. The CA3 pyramidal neuron reconstructed here was relatively large and we imaged it using 100X objective (NA 1.4) (Fig. 1). A total of 51 tiles was required, with 224 images per tile (0.5 μm increments through the depth of the slice), thus yielding a total of 11,424 images. This neuron is contained in a volume of $\sim 400 \times 600 \times 100$ μm^3 and has a total dendritic length of ~ 8815 μm. The full imaging process for the CA3 pyramidal neuron took approximately 2 hours on the Zeiss microscope system we used. Faster imaging times (and fewer tiles) can be accomplished using lower magnification objectives, but in our experience 100X provides more accurate estimates of diameters for small-caliber dendrites. During imaging, care was taken to ensure that the

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

99 microscope settings were optimized to obtain images of all dendrites, including those with the
100 smallest diameter. This resulted in significant background noise, which was removed automati-
101 cally in a final step of the reconstruction process (see below). We made no attempt to image or
102 reconstruct axons, as these were of finer caliber than dendrites and for many neurons they were
103 difficult to discern beyond a short distance from their origin near the soma.

104 Image processing

105 Because the ZEN blue microscope software provides individual images files for each tile in each
106 image plane, ShuTu first converts the image files into tiff stacks using the image metadata
107 file (xml) and parsing the file names for depth information. Each tile was imaged successively
108 through the depth of the slice, so no alignment of the images is required to form a stack. As each
109 stack consists of 224 images, about five minutes of CPU time was required for each stack (see
110 Materials and Methods for the system used). The CA3 pyramidal neuron reconstructed here
111 consists of 51 tiles, for a total of just over four hours. With multiple CPU cores and sufficient
112 memory, ShuTu can automatically distribute the task across multiple cores in parallel, resulting
113 in approximately linear reduction in the real time required to construct the stacks.

114 After the tiff stacks are created, the tiles need to be stitched to find precise relative positions
115 between the tiles. ShuTu also accomplishes this task in a parallel manner, requiring a similar
116 amount of computational time as construction of the stacks. These two image processing steps
117 are performed in series, but they can be executed sequentially without user intervention. In the
118 case of our example CA3 pyramidal neuron, both of these steps were performed in just a few
119 hours by using multiple CPU cores.

120 Automated reconstruction

121 After image processing, ShuTu produces a draft reconstruction of the neuron using an automatic
122 reconstruction algorithm (Materials and Methods). We devised the algorithm to specifically deal
123 with several challenges posed by the bright-field images of biocytin-filled neurons (Fig. 2). One

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

is background noise (Fig. 2a). While patching a neuron, biocytin can spill out and create large blobs in the image stacks. Dirt or dust can be picked up, resulting in structures that look like neurites, especially when color information is not used. Second, during the process of fixing the tissue, thin dendrites can become beaded, with very faint signals between the beads (Fig. 2b). Third, close crossings of adjacent branches requires special attention to resolve (Fig. 2c). Fourth, shadows of out-of-focus branches can be as strong as signals from thin dendrites in focus (Fig. 2d), making it hard to trace some dendrites without being fooled by the shadows. These challenges make it difficult to create a perfect reconstruction from automated algorithms. Our algorithm is designed to address many of these issues, but some manual correction is ultimately required. In the following, we outline the steps involved in the algorithm, using the tile shown in Fig. 1b as an example. Technical details of the algorithm are presented in Appendix 2, which should be useful for adjusting the parameters for specific situations encountered by users.

Conversion to gray scale and 2D projection

The color images are converted into grayscale images. A minimum intensity projection of the tiff stack is then created, which has the same dimension as a single 2D plane in the stack. The intensity at each pixel is chosen to be that of the darkest pixel among all pixels in the stack having the same xy position. This minimum intensity projection reveals all neurites in the tiff stack (Fig. 3a), along with noise from the sources mentioned above. To remove smooth variations due to uneven lighting, the 2D projection is blurred by Gaussian smoothing (Fig. 3b) and subtracted from the original 2D projection (Fig. 3c). Additionally, this process makes faint branches nearly as visible as well-stained ones (Fig. 3d); the inverse peaks corresponding to the branches in the intensity profile have more even heights after the background removal (purple curve) than before (green curve).

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Binary mask

The 2D projection is used to create a mask, which is a binary image with the white pixels indicating the neurites and dark pixels the background (Fig. 3e-i). An accurate mask is crucial for our reconstruction algorithm. Considering the intensity as heights, the neurites in the original 2D projection can be viewed as meandering valleys of dark pixels. To create the mask, we evaluate the possibility that each pixel in the 2D projection belongs to a valley. This is accomplished by comparing the local patch of image centered at the pixel with valley detectors of varying orientations (Fig. 3e) (Frangi et al., 1998). A valley detector is a 2D image consisting of an oriented dark band flanked by two bright bands. The response of the detector is the sum of the products of the corresponding pixels in the detector and the local patch (Fig. 3f). The response has a maximum (λ_1) at one orientation, and a minimum (λ_2) at the orthogonal orientation (Fig. 3f). If the local patch is nearly uniform in intensity, the response is close to zero at all orientations, and λ_1 is small (Fig. 3f, blue curve, which describes the responses at the blue pixel in Fig. 3c). In contrast, if the local patch contains a valley, the maximum response (λ_1) is large and the minimum response (λ_2) is small (Fig. 3f, red curve, at the red pixel in Fig. 3c). If the patch contains a crater corresponding to a blob, λ_1 can be large, but so can λ_2, because there is no previleged orientation. These insights are used to select pixels in valleys but not in blobs or in the background through thresholding λ_1 while also factoring in the difference between λ_1 and λ_2, creating the binary mask (Fig. 3h). The mask is further smoothed to eliminate noisy speckles and rough edges in the boundaries, creating the smoothed mask (Fig. 3i).

SWC points

The mask is used to place SWC points along the neurites. The SWC points are placed along the centerlines of the binary mask (Fig. 4a). The radii of the SWC points are computed as the shortest distance to the nearest boundaries (Fig. 4b). To determine the depths of the SWC points in the original tiff stack, we dissect the centerlines into segments between end points and/or crossing points. These segments are called 'xy-paths' (e.g., Fig. 4a, red arrow). Cutting

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

through the tiff stack while following an xy-path, we create a 'z-image' for that segment (Fig. 4c). This z-image contains all pixels in the tiff stack whose xy positions lie in the xy-path. The branch whose 2D projection falls on the xy-path manifests as a dark valley in the z-image spanning from the left edge to the right edge (Fig. 4c). ShuTu finds the line through the dark valley (red dotted line in Fig. 4c), from which the depths of the neurites (and the SWC points) are determined. The distance between successive SWC points is set to roughly the sum of their radii. The distance is made shorter when the radii changes rapidly along the centerlines to reflect large changes in short distances in the dendritic morphology.

Invalid SWC points are automatically removed (see below regarding validity of SWC points), and adjacent SWC points along one xy-path are connected. If the removal creates a large distance between two consecutive SWC points, they are not connected. Biologically, sharp turns in neurites are rare. Therefore, to safeguard against possible errors, we do not connect SWC points if doing so creates sharp angles in consecutive lines of connections. To avoid connecting branches far away in depth, SWC points are not connected if the difference in z is too large. These decisions depend on parameters set by the user (Appendix 2).

Validity of SWC points

In some cases, the xy-paths from the centerlines of the binary mask are incorrect. For example, nearby branches can be merged in the mask. Checking the validity of the SWC points is thus crucial for eliminating mistakes. To check the validity of an SWC point, we use the image at the plane of the SWC point, and create intensity profiles in eight directions centered at the SWC point (Fig. 5a). For each profile, we look for a significant inverse peak after smoothing the profile (Fig. 5b-c). The significance is checked against fluctuations in the intensity. To do so, we select all points in the smoothed profile with values above the median, which contains mostly the parts of the profile in the background, and compute the standard deviation σ of the differences between the smoothed and the original profile for the selected points (Fig. 5b, green and black curves, respectively). A threshold, set to the median minus σ (Fig. 5b, dotted

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

gray line), is used to judge whether the smoothed profile has two flanks. Another threshold, set to the median minus 15σ, is used to judge whether the inverse peak is deep enough (Fig. 5b, gray line). If both criteria are met, the profile is judged to have a significant inverse peak. The width of the inverse peak is the distance between the steepest descending point and the steepest ascending point of the peak, identified by the derivatives of the smooth profile (Fig. 5e-f).

If none of the profiles have a significant inverse peak, the SWC point is invalid. Otherwise, we chose the profile with the minimum width among the valid ones. In some cases, an SWC point can be at the edge of thick dendrite or soma (see below). To eliminate them, we check wether the intensity with the half radius of the SWC point is low enough (Fig. 5d). Specifically, we check that the intensity values of the smoothed profile (Fig. 5d, violet curve) orthogonal to the chosen profile (Fig. 5d, green curve) within the half radius (Fig. 5d, dotted vertical lines) is smaller than a threshold. This threshold is set to the maximum of the chosen profile within the range plus σ. If not dark enough, the SWC point is invalid.

If the SWC point passes the validity test, we set its radius to the half width of the inverse peak in the chosen profile. Its xy position is adjusted to that of the inverse peak. To ensure that this adjusting process converges, we adjust each SWC point three times iteratively. If the final xy position shifts from the original position more than twice of the original radius, we mark the SWC point invalid since it is most likely created erroneously. Finally, if the final radius of the SWC is smaller than 0.2 μm or larger than 10 μm, the SWC is most likely due to noise and is marked invalid.

Mark pixels occupied

As the SWC points are created, we mark pixels in the tiff stack in the vicinity of the SWC points as occupied (Fig. 6). Before creating a new SWC point, we check whether its center point is marked as occupied. If so, no SWC point is created. This avoids creating redundant SWC points for the same piece of dendritic branch.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Thick dendrites and soma

The widths of dendrites can vary by as much as five times from the thin terminal dendrites to the thick apical dendrite near the soma. The thick dendrites and the soma can be missing from the binary mask, which is created with the valley detector tuned for detecting thin dendrites. Instead, only edges of the thick dendrite and soma are captured in the mask, leading to invalid SWC points that are eliminated. To solve this issue, we specifically detect the presence of thick dendrites and soma. The thick dendrites and soma are typically well-stained and show up as darkest parts in the 2D projection. We use this fact to decide whether there are thick dendrites and soma that are not been covered by existing SWC points. If the lowest intensities in the pixels covered by the existing SWC points are brighter than the lowest intensities in the 2D projection, we decide that the binary mask missed the soma or thick dendrite. We create a binary mask on a 2D projection, excluding pixels around the existing SWC points. New SWC points are added based on this mask.

Extending SWC points in 3D

Gaps in the SWC structure can create broken representations of a continuous dendritic branch. They occur due to errors in the 2D projection representing the 3D branch structures, either because of weak signals or because of occlusions produced by crossing or nearby branches. To bridge these gaps, we extend the SWC points in 3D (the tiff stack) from the end points in the SWC structure.

To minimize the interference from noise, we first delete isolated SWC points that are not connected to any other SWC points. We then mark pixels nearby the existing SWC points occupied (red circles, Fig. 7a) to ensure that the extension does not create duplicated SWC points. From an end SWC point (yellow circle, Fig. 7a), we search for the next candidate SWC point. The search is done in a ring area centered at and in the plane of the end point. The search is also restrict to the pixels from which the lines to the end point (black lines, Fig. 7a) form angles smaller than $\pi/3$ to the line from the end point to its connected SWC point (yellow line,

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. 7a). The arc is divided into 64 points, and a profile is built using the shortest intensity-weighted distances from these points to the end points (black line, Fig. 7b). The profile is smoothed (green line, Fig. 7b). and the xy position of its minimum is set as the xy position of the candidate SWC point. To find z of the candidate SWC point, we build the profile of intensity in z, smooth it, and find the position of its minimum (Fig. 7c).

We test the validity of the candidate SWC point, during which the xy position and radius are adjusted. If accepted, the extension process continues from the new SWC point as the end point. If the candidate point is marked occupied, the extension stops. After the extension stops, the end SWC point is connected to the nearest SWC point if the difference in z is smaller than two times the sum of their radii.

Connecting end points

After extending SWC points in 3D, a continuous branch can still be represented with broken segments of SWC points, especially if the underlying signal is broken or there are closely crossing branches (e.g., Fig. 2b,c). We connect these segments with heuristic rules based on the distances between the end points, in order to recover the branch continuity (Appendix 2). After connecting the end points, the SWC structure for the tiff stack is complete.

The results for our example tiff stack are shown in Fig. 8, in which the SWC points are overlaid with the underlying image, and in Fig. 9, in which the SWC structure is shown in four different view angles in 3D to reveal more details. For this particular tiff stack, the automated reconstruction is complete and accurate. In other cases, however, errors remain, which need to be corrected manually (see below).

Subdivision in z

The 2D projection can be complicated when there are many branches in one tiff stack, which often leads to missed branches due to occlusions. One way of mitigating this problem is to divide the tiff stack in z into several slabs with equal heights in z. SWC points are created separately

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

for each slab as described above, and then combined for the entire stack. The extension from the end points is done with the entire stack. When branches extend across the boundaries between subdivisions of tiff stacks, they are automatically connected by extension from the end points, as described above.

ShuTu allows the user to decide how many subdivisions are necessary based on the complexity of the morphology and the thickness of the tiff stacks. The user should keep in mind that a large number of subdivision slows down the automated tracing. In our example neuron, we divided all tiff stacks into eight slabs.

Combining SWCs

The SWCs of individual tiles of tiff stacks are combined to form the SWC of the entire neuron. The positions of SWCs are shifted based on the relative coordinates obtained in the stitching process. The SWC points of individual stacks are read in sequentially. To void duplicated SWC points in the overlapping regions of adjacent stacks, pixels near the SWC points that are already read in are marked occupied. If the position of SWC points are at the marked pixels, they are deleted. After reading in the SWC points of all stacks, we extend the end points and connect them if they are nearby. Isolated short branches (< 5 SWC points) and small protrusions (< 3 SWC points) from main branches are deleted to reduce noise in the SWC structure. The results SWC structure for the example neuron is shown in Fig. 10a-d.

Manual editing and error correction

The SWC structure created by the automatic algorithm requires editing, such as removing noise, tracing thin or faint dendrites, connecting ends, and correcting mistakes in the radii and positions of the SWC points and in the connections between them. We have designed ShuTu to make these operations easy for the user. In this section we highlight a number of editing techniques.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Inspecting the reconstruction

The SWC structure can be examined in three ways: Tile Manager, Stack View, and 3D View (Fig. 11). In Tile Manager, the SWC structure is overlaid with 2D projection of the entire neuron (Fig. 11a). In this view, it is easy to identify missing, discontinuous, or incorrectly connected branches.

Double clicking on one tile in Tile Manager loads the tiff stack into Stack View (Fig. 11b), in which the SWC structure is overlaid with the image. The radii, depths and connectivity of the SWC points can be examined in detail by scrolling up and down through the z dimension of the tiff stack.

From Stack View, a 2D projection can be created by clicking on Make Projection button (Fig. 12). There is an option to subdivide the stack into multiple slabs in z, in which case separate 2D projections are created. Subdivision is useful when the branching patterns are complicated. Mistakes in the reconstruction can be easily spotted in Projection View, including missed branches, broken points, incorrect connections, and inclusion of noise (Fig. 12). Incorrect positions and diameters for the SWC points are easy to identify as well.

In 3D View, the SWC structure can be rotated and shifted in order to reveal incorrect connections, especially large jumps in z, which can be obscure in other views.

Editing can be done in Stack View, Projection View, and 3D View. In all cases, after any editing, the SWC structure is updated in all views. A selected point can be deleted or moved and its radius can be modified. A selected point in Projection View or 3D View can also be located in Stack View for further examination and modification using the stiff stack.

Adding SWC points

In Stack View, SWC points can be added in three ways. The first method is smart extension. The user selects an SWC point on a branch that needs extension, finds a target point on the branch and locates the focus plane in z, and then clicks on the target. SWC points will be added along the branch from the selected SWC point to the target point (Fig. 13a). The

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

path is computed with the shortest distance algorithm, and the radii and positions of the SWC points are automatically calculated using the automated algorithm described above. The second method is manual extension. It is the same as the smart extension, except that the only point added is at the target point and its radius needs to be adjusted manually. The third method is mask-to-SWC (Fig. 13b). In Projection View, a mask along a branch is drawn by selecting the start and end points. The path is automatically computed with the shortest-distance algorithm. The mask can also be drawn manually. After the mask is completed, it is converted to SWC points along the branch. The positions and radii of the SWC points are computed automatically.

These three ways of adding SWC points are complimentary. When the branch to be reconstructed is long, the mask-to-SWC method is efficient. However, it requires that the underlying signals is strong enough, otherwise the computation of the path and the depths can be inaccurate. When the branch to be covered is short, the smart extension method is efficient, although it also requires relatively strong signal. Manual extension always works.

ShuTu users can reconstruct the entire neuron with one of these three methods. The extension methods can be used after creating a single seed SWC point. However, the process is tedious because the focus plane must be located in every click. The mask-to-SWC method traces branches in 2D projections, and is therefore more efficient.

Modifying connections

The end points in the SWC structure are highlighted with blue or yellow colors. In some cases, it is necessary to connect nearby points that have been incorrectly identified as end points. This can be done by selecting two end points and connecting them. If the distance between the two points are more than the sum of their radii, SWC points can also be added automatically while bridging the gap. A selected end point can also be automatically connected to its nearest neighbor.

Incorrect connections can be broken after selecting two connected SWC points. The branching points are highlighted with green; these points need to be examined carefully for incorrect

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

connections, especially when branches cross.

All SWC points connected to a selected point can be highlighted (Fig. 14). This is useful for finding broken connections in the SWC structure. At the end of the reconstruction, all SWC points that belong to the neuron should be connected. At this point, all noise points can be deleted in a single step.

Reconstruction efficiency

To quantify the efficiency of reconstructing neurons through the automatic algorithm and manual editing, we counted the number of editing operations (NEO) required for achieving the final reconstructions starting from the one generated by the automatic algorithm. The results for the example neuron are shown in Fig. 15. The SWC points that are added in the editing phase are shown in red, and those from the automatic reconstruction are shown in blue (Fig. 15a,b). The added SWC points are about 9% of the total SWC points in the structure. The NEO is 850. Among the editing operations, extensions are dominant. Correcting connection mistakes are sizable as well. The manual time spent in repairing the automatic reconstruction was around 1.5 hours.

The efficiency of reconstruction depends on the image quality and the complexity of the neuron morphology. For neurons with sparse processes, the automated reconstruction captures the most of neuronal structure, and manual editing is not intensive. For the example shown in Fig. 16a, which is a mouse CA3 pyramidal neuron, simpler than the one shown in previous figures, the NEO is 137, and the time spent in editing was approximately 30 minutes. In contrast, when the processes are dense, the automated reconstruction contains many misses and mistakes, and manual editing takes more efforts. An example is shown in Fig. 16b, which is a rat CA1 pyramidal neuron; the NEO is 999, and the time spent in editing was approximately 2.5 hours. The complexity of the processes requires more time for examining the appropriate dendritic structures. Another example of a complex neuron is shown in Fig. 16c, which is a mouse Purkinje cell imaged with confocal microscope. The increased complexity decreases the

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

quality of automated reconstruction, and the NOE is 1373, leading to approximately 2.6 hours of editing.

Discussion

We have demonstrated how ShuTu can be used to reconstruct neuron morphology by converting microscope images to SWC files. Our goal is to provide a practical system that can be readily implemented and used in labs who need accurate dendritic reconstructions of neurons that have been studied and stained following recordings with patch-clamp electrodes. As an open-source software package, it can be continuously improved by the community. We have also provided raw images [SITE ADDRESS], which should be useful for testing and improving the software.

A major aim of our software package is to minimize human labor in reconstructing neurons. We introduced editing functions in ShuTu to improve the efficiency of editing, and implemented a method for counting the number of editing operations (NEO) as the measure of the success of automatic reconstruction. The algorithm for reconstruction and editing functions were developed with the goal of optimizing the automated reconstruction and reducing NEO. For example, our automatic reconstruction algorithm can be aggressive in finding neurites, including faint ones, despite the fact that this may lead to inclusion of more noise in the reconstruction. During the editing phase, this noise can be easily eliminated once the SWC points belonging to the neuron are all connected (Fig. 14). In contrast, if the aim is to analyze morphological metrics such as Sholl analysis (Sholl, 1953) based on the automatically reconstructed neurons without manual editing, such noise could be problematic.

There are many parameters in our reconstruction algorithm. The user should experiment with these parameters as the optimal settings may depend on properties of the images, which are likely to vary depending on staining and imaging procedures. Among the most important parameters are the distances between pixels and between the successive planes, which are determined by the image acquisition process. Also important is the number of subdivisions of one tiff stack. Since our algorithm relies on 2D projections, subdivision reduces overlap of neurites

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

from different depths, and improves the reconstruction quality. Checking the validity of each SWC point is critical, so the users should pay close attention to adjusting these parameters. In Appendix 2, we have pointed out other important parameters while describing the technical details of the automatic reconstruction algorithm.

There are a number of open-source software packages for reconstructing neurons, most notably Vaa3D (Peng et al., 2014) and neuTube (Feng et al., 2015). Vaa3D has extensive capabilities for processing images from various kind of sources. In contrast, we focused on optimizing our software for the particular application of neurons stained with a dark reaction product following patch-clamp recording. Although ShuTu may work for neurons stained in other ways, we have made no attempt to optimize it for use with multiple staining and imaging procedures. In addition, ShuTu was developed with a philosophy that perfect automatic reconstruction is difficult, if not impossible. Therefore, we emphasized the importance of manual annotation and error correction. In keeping with this philosophy, ShuTu includes user-friendly software to facilitate these processes.

Another open-source software package for neuron reconstruction – neuTube – also has a strong 3D capability for manipulating SWC structure. As ShuTu is based on neuTube (T. Zhao is a contributor to both), many of the features of ShuTu are adaptations of neuTube. However, ShuTu includes several important extensions. neuTube was designed to deal with single tiff stack. As such, it was not designed to deal with reconstructing entire neuron, unless the neuron is contained in a single tiff stack.

ShuTu is a complete solution that includes the capability to deal with multiple tiff stacks, including modules for processing and stitching the images. In the interactive mode, the neuron structure is represented in multiples ways that are all linked (Fig. 11), thus improving the ease and accuracy of editing the SWC structure.

Commercial solutions for neuron reconstruction also exits (e.g. Neurolucida 360, MBF Bioscience; Imaris FilamentTracer, Bitplane). Detailed comparison of ShuTu to these other software packages is difficult, as it requires mastery of all of them to be fair. We encourage authors and

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

users of other software packages to test ShuTu on their dataset or test their favorites on the images used in this work, and provide feedback.

Staining neurons with biocytin is common in patch-clamp experiments. However, methods for reconstructing neurons based on biocytin are limited. When dealing with bright-field images like the biocytin data, a common strategy is to apply some preprocessing method first (Türetken et al., 2011; Narayanaswamy et al., 2011; Zhou et al., 2015), making the images friendly for automatic reconstruction. Preprocessing, however, is often computationally intensive and does not guarantee good performance. ShuTu is specifically tailored to deal with inherent problems with images from biocytin filled neurons.

ShuTu is not restricted to biocytin-filled neurons. In principle, it can also handle images from confocal and fluorescent microscopy, simply by inverting the images. However, we made no attempt to develop this application of ShuTu. Care is likely to be necessary in ensuring that microscopy and image acquisition properties are optimized to maximize the utility of ShuTu for this application. For now, we have chosen to leave this enhancement to others and focus our efforts on one common method of staining and imaging neurons.

Improving image quality will inevitably improve the efficiency and accuracy of neuron reconstruction. Users need to make sure that high quality images are taken by following proper microscopy practices and protocols. Tissue fixation and clearing processes can influence the accuracy of the reconstructed neurons. Tissue shrinkage often occurs during the fixation process. To be accurate these factors need to be quantified for specific experimental settings and the dimensions of the reconstructed neurons need to be adjusted to account for shrinkage and distortion.

ShuTu has some limitations. It is not designed to trace axons, which are often too thin and faint following patch-calmp recording to trace automatically. Spines are not marked. In addition, ShuTu cannot reliably handle multiple neurons stained simultaneously. It is possible that editing operations currently requiring human judgements, such as when dendritic branches closely cross each other, could be automated in the future using machine learning approaches

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(Turaga et al., 2010).

In conclusion, we have shown that ShuTu provides a practical solution for efficient and accurate reconstructions of neuron morphology. The open-source nature of the software will allow the research community to improve the tool further, and increased efficiency in neuronal reconstruction should facilitate more studies incorporating quantitative metrics of dendritic morphology and computer simulations of dendritic function.

Materials and Methods

Whole-cell recording and neuron staining

All experiments were performed according to protocols approved by the Institutional Animal Care and Use Committee of the Janelia Research Campus. Acute brain slices were prepared from mice (17-30 days old). After animals were deeply anesthetized with isoflurane, they were decapitated and the brain rapidly removed into chilled cutting solution consisting of (in mM) 215 sucrose, 2.5 KCl, 20 glucose, 26 $NaHCO_3$, 1.6 NaH_2PO_4, 1 $CaCl_2$, 4 $MgCl_2$, and 4 $MgSO_4$. Hippocampi were dissected out and cut into 400 μm thick transverse sections on a Leica VT 1200s vibrating microslicer (Leica, Ltd., Germany). The cutting solution was slowly exchanged with artificial cerebrospinal fluid (ACSF) containing (in mM) 124 NaCl, 2.5 KCl, 10 glucose, 26 $NaHCO_3$, 1.0 NaH_2PO_4, 2.0 $CaCl_2$, and 1.0 $MgCl_2$. Both cutting and ACSF solutions were saturated with 95% O_2 and 5% CO_2 (pH 7.4). The slices were incubated at room temperature for at least 1 hour before recording, and then were transferred as needed to a submersion-type recording chamber perfused with ACSF at 2 ml/min.

Whole-cell recordings were obtained by visualized patch technique under IR-DIC optics. The recording pipette resistance ranged between 4 and 6 $M\Omega$. Series resistance (6 - 15$M\Omega$) and input resistance were monitored throughout each voltage-clamp recording. Recordings with >10% change in series resistance were excluded. All experiments were performed in the current-clamp configuration. The intracellular pipette solution consisted of (in mM) 135 K-gluconate,

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

5 KCl, 1 $CaCl_2$, 0.1 EGTA-Na, 10 HEPES, 10 glucose, 5 MgATP, and 0.4 Na3GTP, 0.1% biocytin, pH 7.2 280-290 mOsm. Resting potential ranged from -69 to -58 mV. Maximal recording time after dissection was 6 hr. Recording temperature was set to 32.0 $\pm$ 0.1 C$^\circ$ using a TC-344A single-channel temperature controller (Warner Instruments, Inc, Hamden, CT, USA). All experiments were executed with a Dagan BVC-700 amplifier, digitized (3 - 5 kHz) using an ITC-16 analog-to-digital converter (Instrutech) and analyzed using custom-made software for IgorPro (Wavemetrics Inc., Lake Oswego, OR, USA). All chemicals were purchased from Sigma-Aldrich (St. Louis, MO, USA). Neurons were filled with biocytin and fixed (12-24 hours) with paraformaldehyde (4%) after recording, then washed in 1X PBS solution. Biocytin staining was carried out with vector PK4000 and SK4100 kits (Vector Laboratories, Burlingame, CA, USA).

System requirements and installation

The software requires installation of `Python` and Open MPI. The software package has been tested on a desktop computer with Intel Core i7-4770 CPU@3.40GHz CPU and 16 GB memory, running Ubuntu 14.04 LTS. `Python` was version 2.7.6. These are typical settings for current high-end desktop computers. Multiple processors are desirable since the algorithms are designed to utilize multiple processors to speed up computation. However, the memory usage must be monitored to make sure that the demand on memory does not exceed 100%. The number of processors used is specified in `ShuTu.py` (variable `nProc`).

ShuTu can be downloaded from

`https://www.janelia.org/shutu,`

or

`http://personal.psu.edu/dzj2/ShuTu/.`

An installation script is provided for Ubuntu and Mac OSX systems. In the directory of ShuTu, one can run

`sudo ./build.sh,`

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

which checks and installs necessary software including `python` and necessary modules, as well as Open MPI. The C programs are also compiled.

The source code for ShuTu is available at https://github.com/tingzhao/ShuTu.

Image acquisition and processing

The software works with tiles of tiff stacks covering the entire neurons. Nearby tiles overlap, typically by 20%, to help fine tune the relative positions of the tiles ("stitching"). The names of the tiff stacks use the convention of a common string (`filenameCommon` followed by a number and `.tif`. With the x, y positions of the tiles specified, one can use the function `stitchShiftTiles` to stitch the tiles. The results are stored in file `filenameCommon.json`.

Modern microscopes often allow automatic generations of overlapping tiles of tiff stacks. In our case, we imaged hippocampal neurons with Zeiss Axio Imager with AxioCam and Zen blue software. Once the boundary in the field of view (XY) and the range of the depths (Z) that contain the neuron are set, the images at each tile position and depths are automatically taken, and the positions of the image are stored in an `xml` file. The filenames of these images contain information about the tile number and depth. Using them, we assemble all images at different depths for each tile into one tiff stack (script `createTiffStacksZeiss.py`). The command is

```
python createTiffStacksZeiss.py dataDirectory filenameCommon
```

Here `dataDirectory` is the path to the directory in which the `xml` file resides. The images of the planes are stored in a subdirectory. `filenameCommon` is the common part of the names given to the created tiff stacks.

The user can generate the overlapping tiff stacks in other ways. The files should be named in the format of `filenameCommon1.tif`, `filenameCommon2.tif`, etc.

The tiff stacks are preprocessed using the command

```
python  processImages.py dataDirectory
```

If the images are dark-field, the command should be

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

```
python processImages.py dataDirectory 1
```

In this case, the images are inverted into bright-field images. The original images are renamed by adding `.org,tif` to the end of the original file names, and are moved to a directory `OriginalImages`.

Stitching the images is done with function `stitchShiftTiles`, which takes lists of the tile numbers, the x, y positions of the top left corners of the tiles and the size of the images `nx, ny`. The x, y positions can be in arbitrary unit and need not be precise; all that matters is that they convey which tiles are neighbors and roughly how much they overlap. Functions are also provided in cases the tiles are on gird and offset percentage is known (function `gridStitchImages`); or the offset percentage and images sequences are specified in a text file (function `tileSequencesStitch`). In our case, all these information can be read from the `xml` file generated by the Zen Blue software during automatic image acquisation (function `xmlStitchImages`). Stitching is done with the script `stitchTilesZeissXML.py` using the command

```
python stitchTilesZeissXML.py dataDirectory
```

After stitching is done, one can proceed to reconstruct the neuron semi-automatically using the software `ShuTu` (see Appendix 1). Another choice is to run the automatic reconstruction algorithm to create a draft reconstruction and edit it using `ShuTu` (see below).

In stitching, the precise offsets of nearby tiles are computed by maximizing phase correlation (Zitova and Flusser, 2003). Using the maximum spanning tree algorithm (Graham and Hell, 1985), a tree graph connecting all tiles and maximizing the sum of phase correlations along the connected nearby tiles is computed and used to set the relative coordinates of all tiles.

Automated reconstruction

The code for automated reconstruction is written in C (`ShuTuAutoTrace.c`). The C code is parallelized with MPI protocol, and runs with the command

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

`mpirun -n np ./ShuTuAutoTrace dataDirectory ShuTu.Parameters.dat,`

where `np` is the number of processors used and `ShuTu.Parameters.dat` is a text file that contains the parameters. This creates a SWC file `fnamecommon.auto.swc` in `dataDirectory`, which can be loaded in `ShuTu` for manual editing (`File`→`expand current`→`swc`).

Appendix 1: Editing commands for ShuTu

Loading a project

A reconstruction project can be opened by clicking on `Open Project` icon or `File` → `Open Project`. In the directory of the neuron, there should be a file `filenameCommon.tiles.json`, which is created after stitching the tiff stacks. Clicking on it opens `Tile View`, in which the 2D projections of the tiff stacks are shown. The 2D projection of the neuron should be visible. If there is a previous reconstruction of the neuron, which is stored a file `filenameCommon.swc`, it will be automatically loaded and overlaid onto the 2D projection. The SWC file generated by the automated algorithm, `filenameCommon.auto.swc`, can be loaded by selecting `File` → `Expand current` → `SWC`.

Double clicking on any tile in the `Tile View` loads the corresponding tiff stack in `Stack View`. The loaded SWC points are overlaid onto the tiff stack. To go up and down in the z-dimension, use the right and left arrow keys.

Clicking on `Make Projection` button creates 2D projection of the tiff stack. The user can specify the number of subdivisions used in the projection. All of the projections of the subdivisions are contained in the `Projection View`, which can be browsed with the left and right arrow keys.

The SWC structure is also displayed in `3D View`. It can be rotated with the arrow keys, and shifted with the arrow keys while pressing the `Shift` key,

In all views, zoom is controlled with `+` and `-` keys. After zooming in, different parts of the images can be navigated by pressing-dragging the mouse.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The functions of the arrow keys can be also performed with mouse wheel or track pad when available.

Editing SWC points

The SWC structure can be edited in `Stack View`, `Projection View`, and `3D View`. All editing can be reversed by `Ctrl`-z (or `Command`-z). Colors of SWC points indicate their topological roles in the structure: yellow and blue indicate the end points of branches; green at the branching points; and red the interior points. Lines between SWC points indicate their connectivity.

In `Stack View`, an SWC point is plotted with a circle at its xyz position in the tiff stack. The radius of the circle the same as that of the SWC point. As the focus plane shifts away from the z of the SWC point, the circle shrinks with its color fading. This helps the user to visually locate the z of the SWC points and inspect whether the positions and radii of the SWC points match the underlying signals of the neurites in the tiff stack.

Extension is the most used editing function. In `Stack View`, it can be done in two ways. The first is manual extension. Click an SWC point to extend, and the cursor becomes a circle connected to the SWC point. Focus on the target neurite using the arrow keys, and match the radius of circle with that of the neurite using `e` and `q`. `Ctrl`-clicking on the target points creates a new SWC point connected to the starting SWC point. (In Mac, use `Command` instead of `Ctrl`.) The second is smart extension. It is the same as manual extension, except that the user clicks without pressing `Ctrl`. This method allows clicking far from the starting SWC points; the algorithm fills in additional connected SWC points along the neurites with the radii and depths automatically calculated. Smart extension works well when the underlying signal is strong.

To change the properties of a particular SWC point, select it by clicking on it and pressing `Esc` to come out of the extension mode. The radius can be changed with `e` and `q`. It can be moved with `w,s,a,d` for up, down, left, right. Pressing `x` deletes it.

To connect two SWC points, click on the first point and `Shift`-click on the second point, then press `c`. Pressing `Shift`-c after selecting two points automatically fills additional SWC

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

points, similarly as in the smart extension. To disconnect two SWC points, select them then press `b`.

In `Projection View`, 2D projections of the subdivisions of the tiff stack are overlaid with the SWC points. In this view it is easier to spot missed branches and incorrect connections. There is also a mask-to-SWC method for tracing branches. To draw a mask along a branch, press `r`. The cursor becomes a red dot. Roughly match the radius of the dot with that of neurite with `e` and `q`. Click on the start point, then `Shift`-click on the target. A red mask will be drawn along the branch. Clicking on `Mask` → `SWC` button converts the mask into SWC points, which can be examined in detail in the `Stack View`. The mask can also be drawn manually by press-dragging the mouse along the branch. To get of out the mask drawing mode, press `Esc`.

Clicking on an SWC point selects it. Pressing `z` locates the selected point in the `Stack View`, and its z position and other properties can be further examined with the tiff stack.

The user can directly modify the connections in the `Projection View`. The operations are the same as in the `Stack View`.

In `3D View`, the user can examine and modify the connections between SWC points. Connecting or breaking connections between two SWC points is the same as in the `Stack View` and the `Projection View`. Selecting an SWC point and pressing `z` locates it in the `Stack View` for further examination and extension. This operation also loads a new tiff stack if the selected point is not in the current tiff stack.

A useful way of locating broken points in the SWC structure is the operation that selects all connected SWC points to the selected SWC point. It is done by pressing `s-3`, or right-clicking the mouse and selecting `Select` → `All connected nodes`.

After correctly connecting all SWC points belonging to the neuron, the user can delete all noise points simply by selecting all SWC points in the neuron, right-clicking the mouse, and performing `delete unselected`.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Annotating, saving, and scaling the SWC structure

After the reconstruction is done, the user needs to annotate the SWC points as soma, axon, apical dendrite, basal dendrite. This is best done in the `3D View`. In the panel `control and settings`, change `Color Mode` to `Branch Type` to reveal the types of SWC points. To annotate the soma, the user can select one point in the soma, right-click the mouse, and select `Change Property` → `Set as root`. More SWC points belong to the soma can be selected by `Shift`-clicking. Then right-click to bring up the menu, then select `Change type` and set the value to 1. The SWC points in the soma are shown in blue.

To annotate the axon, select the one SWC point closet to the soma, and press `s-1`. This selects all SWC points down stream of the selected point. Then change type to 2. Basal dendrites and apical dendrite can be similarly annotated, and their types are 3 and 4, respectively.

In the panel `control and settingings`, setting `Geometry` to `normal` produces the volume representation of the SWC structure, with surface rendered between adjacent SWC points.

To save the reconstruction, click on the objects in the panel `Objects`, which selects the corresponding SWC points. Then in the window of the SWC structure, left-click and do `save as`. It is best to use the default filename `filenameCommon.swc`.

The dimensions of the SWC points in `filenameCommon.swc` are pixel based. To convert them into physical dimensions in μm, type in the terminal

```
python scaleSWC.py dataDirectory
```

This process uses `xyDist` and `zDist` in `ShuTu.py`, which specify in μm the xy pixel distance and z distance between successive planes. The results are saved in `filenameCommon.scaled.swc`.

Right after finishing the reconstruction and with `ShuTu` closed, the number of various editing operations can be analyzed using the command

```
python analyzeNEO.py
```

A plot similar to Fig. 15c will be generated. The script `analyzeNEO.py` parses the log file generated by `ShuTu`. The log file can contain several neuron reconstruction sessions, but the

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

script only parses the most recent one. When estimating the total time of manual editing, idle times of the user are excluded if they are detected in the log file.

There are many more editing functions in ShuTu. The user can refer to `Help` for more instructions.

Appendix 2: Technical details of automated reconstruction

Here we provide technical details of the automated reconstruction algorithm presented in the main text. These details should help the users to adjust parameters for their specific needs, and facilitate further development of the algorithm. The parameters in each step are summarized in series of tables. The algorithm is explained with the same example used in the main text.

Coordinate system

A tiff stack consists of successive 2D images (referred to as planes) taken at increasing depths at regular intervals. We denote a pixel in a tiff stack with coordinates (x, y, z). Here x, y are the pixel positions in the planes, and z is the depth. We take the convention that in a plane, the x axis points vertically downwards and the y axis horizontally to the right (Fig. 1).

The distance between neighboring pixels in x and y is denoted as d_{xy}. The distance between successive planes is denoted as d_z. In the example, $d_{xy} = 0.065$ μm, and $d_z = 0.5$ μm (Table 1).

Prepocessing

Our algorithm requires that the images are grayscale with bright background. Other image types must be converted into bright-field grayscale images, and this is done in preprocessing. In particular, color images are converted into grayscale according to

$$I_g(x, y, z) = 0.21 I_r(x, y, z) + 0.72 I_g(x, y, z) + 0.07 I_b(x, y, z),$$

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

where I_g is the intensity of the grayscale and I_r, I_g, I_b are those of the red, green, and blue channels. Dark-field images are inverted by subtracting the grayscale intensity at each pixel from the maximum intensity of the tiff stack.

To reduce pixel noise, each plane is smoothed with 2D Gaussian filter with $\sigma = 1$ pixel. The intensity is linearly scaled so that the range is from 0 to 1 for the tiff stack.

2D projection

We identify neurites in a tiff stack from its minimum-intensity 2D projection. The intensity $I(x, y)$ of the 2D projection is taken as the minimum intensity among all pixels with the same z. Projections of dendritic branches form dark paths in $I(x, y)$ (Fig. 3a). Shadows of branches in out-of-focus planes (Fig. 2d) do not create separate dark paths in the 2D projection; instead, their projections flank those of the branches, forming smooth decay of intensity away from the center lines of the branches. The problem of confusing the shadows of the branches as neurites in the out-of-focus planes, as shown in Fig. 2d, does not exist in the 2D projection.

To eliminate smooth variations of the background due to uneven lighting, we subtract from $I(x, y)$ a background, which is obtained by blurring $I(x, y)$ with a Gaussian filter with standard deviation $\sigma_b = 2$ μm (Fig. 3b). We then normalize the range of $I(x, y)$ to $(0, 1)$ (Fig. 3c). Smaller σ_b enhances weak signals relative to strong signals (Fig. 3d). This is because the background with smaller σ_b tracks the signal strength more closely, and when subtracted, takes away more from the strong signals. But σ_b should be large enough to ensure that the subtracted background is smooth and does not weaken the signals.

Binary mask

From the 2D projection we create a binary image $b(x, y)$ to indicate pixels that belong to neurites. Specifically, $b(x, y) = 1$ for pixels in the neurites (foreground pixels) and $b(x, y) = 0$ for those in the background (background pixels). We call the area defined by the foreground pixels as binary mask.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The first step in creating the mask is convolving $I(x, y)$ with valley detectors with varying orientations, and finding the maximum and minimum responses to the detectors (Fig. 3e). A valley detector $f(x, y)$ is a patch of 2D image (or filter) consisting of an oriented dark band flanked by two bright bands. Mathematically the filter is expressed as

$$f(x,y) = \frac{1}{2\pi\sigma^2}\frac{\partial^2}{\partial\tau^2}e^{-(x^2+y^2)/2\sigma^2},$$

which is a directional second derivative of a Gaussian with standard deviation σ. Here

$$\frac{\partial}{\partial\tau} = \hat{\tau}\cdot\nabla = \tau_x\frac{\partial}{\partial x} + \tau_y\frac{\partial}{\partial y},$$

where $\hat{\tau} = \tau_x\hat{x} + \tau_y\hat{y}$ is a unit vector perpendicular to the orientation of the dark band.

Convolving $I(x, y)$ with the filter creates the response $R(x, y)$:

$$\begin{aligned} R(x,y) &= \int dx'dy' I(x+x', y+y') f(x', y') \\ &= I_{xx}\tau_x^2 + 2I_{xy}\tau_x\tau_y + I_{yy}\tau_y^2, \end{aligned} \tag{1}$$

where

$$I_{xx} = \frac{1}{2\pi\sigma^4}\int dx'dy' I(x+x', y+y')\left(\frac{x'^2}{\sigma^2} - 1\right)e^{-(x'^2+y'^2)/2\sigma^2},$$

$$I_{xy} = \frac{1}{2\pi\sigma^4}\int dx'dy' I(x+x', y+y')\frac{x'y'}{\sigma^2}e^{-(x'^2+y'^2)/2\sigma^2},$$

$$I_{yy} = \frac{1}{2\pi\sigma^4}\int dx'dy' I(x+x', y+y')\left(\frac{y'^2}{\sigma^2} - 1\right)e^{-(x'^2+y'^2)/2\sigma^2}.$$

We obtain the maximum or minimum response at (x, y) using the Lagrange multiplier method:

$$R' = I_{xx}\tau_x^2 + 2I_{xy}\tau_x\tau_y + I_{yy}\tau_y^2 - \lambda(\tau_x^2 + \tau_y^2 - 1).$$

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

At the extrema we have

$$0 = \frac{\partial R'}{\partial \tau_x} = 2(I_{xx} - \lambda)\tau_x + 2I_{xy}\tau_y,$$

$$0 = \frac{\partial R'}{\partial \tau_y} = 2I_{xy}\tau_x + 2(I_{yy} - \lambda)\tau_y.$$

These are linear equations, which can be expressed in matrix form as

$$\begin{pmatrix} I_{xx} - \lambda & I_{xy} \\ I_{xy} & I_{yy} - \lambda \end{pmatrix} \begin{pmatrix} \tau_x \\ \tau_y \end{pmatrix} = 0. \tag{2}$$

To have none-zero solutions for τ_x and τ_y, we must have

$$\begin{vmatrix} I_{xx} - \lambda & I_{xy} \\ I_{xy} & I_{yy} - \lambda \end{vmatrix} = 0,$$

where λ is the eigenvalue of the Hessian matrix. There are two solutions:

$$\lambda_1(x, y) = \frac{1}{2}\left(I_{xx} + I_{yy} + \sqrt{(I_{xx} - I_{yy})^2 + 4I_{xy}^2}\right),$$

$$\lambda_2(x, y) = \frac{1}{2}\left(I_{xx} + I_{yy} - \sqrt{(I_{xx} - I_{yy})^2 + 4I_{xy}^2}\right).$$

Here we chose $\lambda_1(x, y) > \lambda_2(x, y)$. Solving τ_x and τ_y from Eq. (2) and plugging in to Eq. (1), we find the response at the extrema:

$$R(x, y) = \tau_x^2 I_{xx} + 2I_{xy}\tau_x\tau_y + I_{yy}\tau_x^2 = \lambda(\tau_x^2 + \tau_y^2) = \lambda.$$

Hence the maximum $R_m(x, y)$ of the responses $R(x, y)$ to valley detectors at varying orientations is given by

$$R_m(x, y) = \lambda_1(x, y).$$

To see how we can create the mask from $\lambda_1(x, y)$ and $\lambda_2(x, y)$, we exam three simple examples

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of synthetic 2D images containing some aspects of 2D projections of the real images containing neurites.

The first example is a Gaussian valley in y direction:

$$I(x,y) = I_0 - \frac{I_1}{\sqrt{2\pi}\sigma_s} e^{-x^2/2\sigma_s^2}.$$

Here σ_s is the scale of the widths of the valley; I_0 is the baseline intensity; and I_1 is the amplitude. This is an idealized model of the 2D projection of a dendritic segment with half-width σ_s. An ideal mask for this Gaussian valley is a rectangular strip spanning the y direction, centered long y-axis and with half-width σ_s.

$\lambda_1(x,y)$ and $\lambda_2(x,y)$ are easily calculated. We find that

$$I_{xx} = \frac{I_1}{\sqrt{2\pi}(\sigma^2+\sigma_s^2)^{3/2}} \left(1 - \frac{x^2}{\sigma^2+\sigma_s^2}\right) e^{-x^2/2(\sigma^2+\sigma_s^2)},$$

and

$$I_{xy} = I_{yy} = 0.$$

Therefore,

$$\lambda_1 = \begin{cases} I_{xx}, \text{ if } I_{xx} \geq 0, \\ 0, \text{ if } I_{xx} < 0. \end{cases}$$

$$\lambda_2 = \begin{cases} 0, \text{ if } I_{xx} \geq 0, \\ I_{xx}, \text{ if } I_{xx} < 0. \end{cases}$$

We can obtain a mask close to the ideal mask by thresholding $\lambda_1(x,y)$. If we set the foreground pixels as those with $\lambda_1(x,y) > 0$, the boundary of the mask is given by

$$x_b = \pm\sqrt{\sigma^2+\sigma_s^2}.$$

The half-width of the mask is $\sqrt{\sigma^2+\sigma_s^2}$, and it is larger than σ_s. Taking $\sigma \to 0$ leads to the

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ideal mask. For finite σ, it is possible to set a higher threshold for $\lambda_1(x, y)$ and obtain the ideal mask; but this requires a threshold that depends on the width of the valley.

From this example we see that we can obtain a mask that closely follow dendritic branches by thresholding the maximum responses to the valley detectors, $\lambda_1(x, y)$. The threshold should be larger than 0. Larger σ for the detectors tends to broaden the mask; therefore it is desirable to have small σ to obtain masks that closely cover the dendritic branches.

The second example is a Gaussian blob:

$$I(x, y) = I_0 - \frac{I_1}{2\pi\sigma_s^2} e^{-(x^2+y^2)/2\sigma_s^2}.$$

This is an idealized model for the 2D projections of spills created during the staining process (Fig. 2a). Such spills are noise that should be eliminated; therefore the ideal mask for a Gaussian blob should be empty.

We find that

$$I_{xx} = \frac{I_1}{2\pi(\sigma^2+\sigma_s^2)} \left(1 - \frac{x^2}{\sigma^2+\sigma_s^2}\right) e^{-(x^2+y^2)/2(\sigma^2+\sigma_s^2)},$$

$$I_{xy} = -\frac{I_1 xy}{2\pi(\sigma^2+\sigma_s^2)^2} e^{-(x^2+y^2)/2(\sigma^2+\sigma_s^2)},$$

$$I_{yy} = \frac{I_1}{2\pi(\sigma^2+\sigma_s^2)} \left(1 - \frac{y^2}{\sigma^2+\sigma_s^2}\right) e^{-(x^2+y^2)/2(\sigma^2+\sigma_s^2)}.$$

Therefore,

$$\lambda_1(x, y) = \frac{I_1}{\pi(\sigma^2+\sigma_s^2)} e^{-(x^2+y^2)/2(\sigma^2+\sigma_s^2)},$$

$$\lambda_2(x, y) = \frac{I_1}{\pi(\sigma^2+\sigma_s^2)} \left(1 - \frac{x^2+y^2}{\sigma^2+\sigma_s^2}\right) e^{-(x^2+y^2)/2(\sigma^2+\sigma_s^2)}.$$

We see that thresholding the maximum responses λ_1 creates a circular mask, which is far from the desired empty mask. To suppress creating foreground pixels for the Gaussian blob, additional criteria for the mask are needed. We notice that near the center of Gaussian blob, λ_1 and λ_2 are approximately equal. This motivates another criterion for the mask: in addition to λ_1 being

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

greater than a threshold, the foreground pixels must satisfy the condition $\lambda_1 > \alpha_\lambda|\lambda_2|$, where $\alpha_\lambda > 1$ is a factor. This criterion should suppress foreground pixels for the Gaussian blob, except around a ring near the radius $\sqrt{\sigma^2 + \sigma_s^2}$, where λ_2 is close to zero. This is not the ideal mask for Gaussian blob, but it is close. Note that this additional criterion does not affect the mask for the Gaussian valley in the first example, hence does not interfere with detection of dendritic branches.

The third example is a random image, which has a mean intensity I_0 and no correlations between the pixels:

$$< (I(x,y) - I_0)(I(x',y') - I_0) >= \sigma_I^2 \delta(x - x', y - y').$$

Here σ_I^2 is the variance of the pixel intensity. This is an idealized model for random pixel noise in the real images. The ideal mask should be empty.

It is easy to see that

$$< I_{xx} >=< I_{xy} >=< I_{yy} >= 0.$$

Additionally,

$$< I_{xx}^2 >=< I_{yy}^2 >= \frac{3\sigma_I^2}{16\pi\sigma^6},$$

$$< I_{xy}^2 >=< I_{xx} I_{yy} >= \frac{\sigma_I^2}{16\sigma^6},$$

$$< I_{xx} I_{xy} >=< I_{yy} I_{xy} >= 0.$$

From these we find the mean of the responses

$$< R(x,y) >= 0,$$

and the variance

$$\sigma_R^2 =< R(x,y)^2 >= \frac{3\sigma_I^2}{16\pi\sigma^6},$$

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

where we have used $\tau_x^2+\tau_y^2=1$. σ_R represents the range of the responses expected from random fluctuations of the intensity. To avoid creating foreground pixels for random fluctuations, we should set the threshold for λ_1 larger than σ_R. But a threshold that is too large diminishes the mask for dendritic branches. Therefore the threshold for λ_1 must be chosen to preserve the signals while suppressing random noise. Inevitably, some foreground pixels to noise are unavoidable, which creates random speckles in the mask (Fig. 3h).

A guideline for selecting the length scale σ in the valley detectors can be devised by combining the insights from the Gaussian valley and the random image. The peak of the maximum responses from the Gaussian valley, which occurs at $x=0$, is given by

$$\lambda_{1,\max}=\frac{I_1}{\sqrt{2\pi}(\sigma^2+\sigma_s^2)^{3/2}}.$$

Comparing this to the variance of the responses to the random image, we can define the signal-to-noise ratio as

$$\rho_s=\frac{\lambda_{1,\max}}{\sigma_R}=\frac{4I_1}{\sqrt{6}\sigma_I}\left(\frac{\sigma^2}{\sigma^2+\sigma_s^2}\right)^{3/2}.$$

This ratio is an increasing function of σ. Therefore, a large σ is useful for suppressing noise. However, a large σ overestimates the width of the valley (given by $\sqrt{\sigma^2+\sigma_s^2}$), leading to widening of the foreground pixels. For real images, such widening can create a mask that merges nearby branches, leading to an inaccurate representation of the neuronal structure. Hence the choice of σ is a compromise between enhancing the signal-to-noise ratio while avoiding the merger of nearby branches in the mask. When the intensity fluctuation is small, we can select a small σ, leading to an accurate mask. If the fluctuation is large, we need to choose a large σ and live with the imperfect mask.

Based on the insights gained from the examples discussed above, we formulate the following procedure for creating the mask $b(x,y)$ from $\lambda_1(x,y)$ and $\lambda_2(x,y)$. Select σ of the valley detector such that the neurites are clearly visible in $\lambda_1(x,y)$ (Fig. 3d). Set $b(x,y)=0$ with $\lambda_1<\alpha_\lambda|\lambda_2|$ to suppress circular blobs in the 2D projection. Select a threshold θ_λ above the noise level, and

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

set $b(x,y) = 1$ if $\lambda_1(x,y) > \theta_\lambda$ and $b(x,y) = 0$ otherwise (Fig. 3e). Since pixels belonging to the neurites tyically have higher λ_1 compared to those with random fluctuations, we set the threshold θ_λ such that the fraction of pixels selected to the mask is f_λ. For our example neuron, the parameter values are: $\sigma = 0.1$ μm, $\alpha_\lambda = 10$, and $f_\lambda = 0.1$ (Table 2).

The binary mask generated as above is noisy, and the boundaries for neurites are rugged (Fig. 3h). To clean up noise and smooth the boundaries, we use the sparse-field level-set method outlined in (Lankton, 2009), which is a technical report based on (Whitaker, 1998). The details of level-set smoothing is as follows.

For the 2D projection $I(x,y)$ after background subtraction and normalization, we compute the gradient

$$g_r(x,y) = \sqrt{I_x^2 + I_y^2}.$$

We rescale the gradient so that the range is from 0 to 1. An edge indicator is defined as

$$g(x,y) = \frac{1}{1+g_r^\beta},$$

where β is an exponential, typically smaller than 1, for compressing the gradient values. This function is minimal at edges of branches, where the gradients are larger. We seek a contour $\mathcal{C}$ such that the energy function

$$\mathcal{E} = \mu L[\mathcal{C}] + \oint_{\mathcal{C}} dl g(l)$$

is minimized, where $L[\mathcal{C}]$ is the total length of the contour and μ is weight parameter that controls the smoothness of the contour. The curve that minimize this energy function will be smooth and sit along the maximum gradient boundaries between the branches and the background.

The contour can be expressed as the zero-crossing points of a level set function $\phi(x,y)$. Inside $\mathcal{C}$, we have $\phi > 0$, and outside $\phi < 0$. Note that

$$L[\mathcal{C}] = \oint_{\mathcal{C}} dl.$$

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The unit vectors normal to the contours in ϕ are given by

$$\hat{n} = -\frac{\nabla\phi}{|\nabla\phi|}.$$

Hence

$$L[\mathcal{C}] = \oint_{\mathcal{C}} dl\hat{n}\cdot\hat{n} = -\oint_{\mathcal{C}} dl\hat{n}\cdot\frac{\nabla\phi}{|\nabla\phi|} = -\int_{\mathcal{C}} dxdy\nabla\cdot\frac{\nabla\phi}{|\nabla\phi|}.$$

The last step uses the divergence theorem. Note that

$$-\int_{\mathcal{C}} dxdy\nabla\cdot\frac{\nabla\phi}{|\nabla\phi|} = -\int dxdyH(\phi)\nabla\cdot\frac{\nabla\phi}{|\nabla\phi|}.$$

Here $H(\phi)$ is the step function; it is 1 if $\phi > 0$ and 0 if $\phi < 0$. Integration by part gives

$$-\int dxdyH(\phi)\nabla\cdot\frac{\nabla\phi}{|\nabla\phi|} = \int dxdy\frac{\nabla\phi}{|\nabla\phi|}\cdot\nabla H(\phi) = \int dxdy\frac{\nabla\phi}{|\nabla\phi|}\cdot\nabla\phi\delta(\phi) = \int dxdy\delta(\phi)|\nabla\phi|.$$

The surface term is zero because H is zero at the boundary. Here $\delta(\phi)$ is the Dirac δ-function.

Therefore, we have

$$L[\mathcal{C}] = \int dxdy\delta(\phi)|\nabla\phi|.$$

Similarly, we can derive

$$\oint_{\mathcal{C}} dlg(l) = \oint_{\mathcal{C}} dl\hat{n}\cdot(g\hat{n}) = \int dxdy\delta(\phi)g(x,y)|\nabla\phi|.$$

Hence, we have

$$\mathcal{E} = \int dxdy\left(\mu + g(x,y)\right)\delta(\phi(x,y))|\nabla\phi(x,y)|.$$

We use the variational method to find the ϕ that minimizes $\mathcal{E}$. Noting that

$$|\nabla(\phi+\delta\phi)| = \sqrt{|\nabla\phi|^2 + 2\nabla\phi\cdot\nabla\delta\phi} = |\nabla\phi| + \frac{\nabla\phi\cdot\nabla\delta\phi}{|\nabla\phi|},$$

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

we find

$$\delta|\nabla\phi| = \frac{\nabla\phi \cdot \nabla\delta\phi}{|\nabla\phi|}.$$

Applying integration by part, we have

$$\delta\mathcal{E} = \int dxdy \left[-\delta(\phi)\nabla \cdot \left((\mu + g)\frac{\nabla\phi}{|\nabla\phi|}\right)\right]\delta\phi.$$

Setting $\mathcal{E} = 0$, we find

$$-\delta(\phi)\nabla \cdot \left((\mu + g)\frac{\nabla\phi}{|\nabla\phi|}\right) = 0.$$

The surface term in the integration vanishes if we impose the Neumann boundary condition

$$\frac{\partial\phi}{\partial n} = 0.$$

At equilibrium and on $\mathcal{C}$ we have

$$-\nabla \cdot \left((\mu + g)\frac{\nabla\phi}{|\nabla\phi|}\right) = 0.$$

At other points, ϕ can be arbitrary. Minimization of

$$\delta\mathcal{E} = \int dxdyf[\phi]\delta\phi$$

can be done by solving the equation

$$\frac{\partial\phi}{\partial t} = -f[\phi].$$

This equation implies

$$\delta\phi = -dtf[\phi]$$

at each time step. Therefore

$$\delta\mathcal{E} = -\int dxdyf[\phi]^2dt < 0,$$

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

leading to decreasing $\mathcal{E}$. In our case, we need to evolve

$$\frac{\partial \phi}{\partial t} = \nabla \cdot \left((\mu + g) \frac{\nabla \phi}{|\nabla \phi|} \right) = F$$

on the boundary. For ϕ at points other than the boundary, we need to change ϕ such that it remains a smooth function around the boundary and the second derivatives can be computed. We used the sparse-field implementation for solving this differential equation, which iteratively updates the sets of points near the boundary (Lankton, 2009). After $N_{levelset}$ number of iterations, we obtain a new binary mask by setting $b(x, y) = 1$ if $\phi(x, y) > 0$, and $b(x, y) = 0$ otherwise.

In practice, we observe that it is sufficient to smooth the initial mask by minimizing the length of the boundary alone. Hence we set the edge indicator

$$g(x, y) = 0.$$

This is because the boundaries of the initial mask are already near the neurite boundaries.

The parameter μ controls the smoothness of the boundary. Smoothing deletes small noisy speckles. Larger μ creates smoother boundaries but can also cause small neurites to disappear. We set $\mu = 0.1$. Also important is the number of iterations $N_{levelset}$. It should be large enough to reduce noise and smooth the boundaries, but small enough not to loose structures due to over-smoothing. In our example we set $N_{levelset} = 500$.

As the final step, we remove connected pixels with total area smaller than A_s. This removes noise and cleans up the mask (Fig. 3f). In the example we set $A_s = 1$ μm^2.

Parameters for creating the mask are listed in Table 2.

Creating SWC points from the mask

Using the mask, we create the SWC points that describe the dendritic structure. The (x, y) positions of the SWC points are placed along the centerlines of the mask. The radii r are set as

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the shortest distances to the boundaries of the branches from the centerlines. The z positions are computed using the centerlines and the tiff stack.

The centerlines of the mask are obtained by skeletonization (Zhang and Suen, 1984). The skeleton is computed by iterative thinning of the mask based on the pixel values in the 8 neighboring points (Zhang and Suen, 1984) (Fig. 4a). The distance from a pixel in the centerline to the nearest boundary is computed using the Euclidian distance transformation (Danielsson, 1980) (Fig. 4b).

The depths z of the points on the centerlines of the mask are found in two steps. First, the centerlines are dissected into xy-paths (Fig. 4a). The xy-paths with length smaller than $l_{sm} = 0.5$ μm are considered as noise and excluded. Second, a z-image is created by following the xy-path and cutting through the tiff stack (Fig. 4c). A dark valley in the z-image spanning from the left edge to the right edge indicates the branch whose 2D projection falls on the xy-path (Fig. 4c). A line through the valley can be found by evaluating all paths from the left edge to the right edge (red dotted line in Fig. 4c). Specifically, a left-right path in the z-image starts from a point at the left edge. The next point is selected from the three nearby points to the right (the change in z is -1, 0, or 1). This process iterates until the right edge is reached. For each left-right path, we compute the weighted distance, which is the sum of the distance between consecutive pixels multiplied by a weight $e^{\alpha_d I}$, where $\alpha_d = 20$ is a parameter and I is the intensity of the right point. The weight penalizes bright pixels, and encourages the left-right path to go through dark pixels. The left-right path with minimal weighted distance, or the shortest path, is selected. The path follows the dark valley spanning from the left edge to the right edge, as shown in Fig. 4c (dotted red line). The z values of this shortest path gives the depth for each point in the xy-path. A large α_d ensures that the shortest path follows dark pixels. But a value too large can lead to distortions of the path due to dark spots from other branches or noise.

Linked SWC points are placed on the xy-path. The distance between successive SWC points

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

is set to

$$\frac{r_1 + r_2}{1 + \alpha_p |r_1 - r_2|},$$

where r_1, r_2 are the radii of the two SWC points, and $\alpha_p = 0.1$ is a factor for adjusting the distance based on the difference of the radii. When the radii are almost the same, the distance is roughly the sum of the radii, and the SWC points touch each other. If the radii are quite different, however, the SWC points are placed closer in order to better reflect the rapid changes in the diameters along the branch. To avoid overlapping SWC points at the crossing points in the skeleton, the first and the last SWC points are placed away from the respective end points of the xy-path by the radii of the SWC points.

We check the validity of SWC points, remove any invalid ones, and connect adjacent SWC points along the xy-path. Removal of invalid SWC points may have created a large distance between two consecutive SWC points; in this case, we do not connect them. The criterion is

$$d_{12} > \alpha_{xy}(r_1 + r_2),$$

where d_{12} is the Euclidian distance in xy between the two SWC points; r_1, r_2 are the radii; and $\alpha_{xy} = 2.0$ is a factor. Sharp turns in xy-path often result from errors in the skeleton due to crossing branches. To avoid this problem, we do not connect the two SWC points if doing so creates a lage angle (greater than $\theta_{thr} = \pi/3$) between consecutive lines connecting the SWC points. We also do not connection the SWC points if the z difference between them is too large, as this often leads to errors in connecting branches far way in z. The criterion is

$$d_z |z_1 - z_2| > \alpha_{zj} d_{xy}(r_1 + r_2),$$

where $d_z = 0.5\ \mu m$ is the distance between successive planes in the stiff stack; $d_{xy} = 0.065\ \mu m$ is the pixel distance in xy; and $\alpha_{zj} = 2.0$ is a factor for adjusting the threshold.

The parameters for creating SWC points are listed in Table 3.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Checking the validity of an SWC point

The SWC points created from the mask can be incorrect. For example, if some branches are parallel to each other and are very close, they can be merged in the mask, leading to incorrect SWC points. Therefore it is important to check the validity of the SWC points and reject incorrect ones. Since the mask can be imperfect, we check the validity not with the mask but with the original tiff stack. The main idea is that a valid SWC point should sit on the centerline of a valley in the plane at the depth of the SWC point.

Consider an SWC point at (x_p, y_p, z_p) with radius r_p. The validity of the point is tested with the intensity $I(x, y)$ of pixels in the plane at $z = z_p$ (Fig. 5a). Ideally, the SWC point should be at a local center of a valley in $I(x, y)$ of a dendritic branch. To test this, we create a profile of intensity along a line through (x_p, y_p) and at angle θ relative to the x-axis (Fig. 5a), and test the existence of an inverse peak. The profile is a one-dimensional curve $I_\theta(d)$ (black line, Fig. 5b), where d is the coordinate of a pixel point on the line, with (x_p, y_p) set as the origin. $I_\theta(d)$ is the intensity value at the pixel point. The range of $|d|$ is limited to $d_{\max} = \min(r_m, 4r_p)$. Here $r_m = 3\ \mu m$ is the lower bound for the range. We obtain a smoothed profile $I_{s,\theta}(d)$ by convolving $I_\theta(d)$ with a Gaussian filter with $\sigma_s = 0.3\ \mu m$ (green line, Fig. 5b), and detect the inverse peak in $I_{s,\theta}$. We take θ to be multiples of $\pi/8$. Hence there are 8 profiles (Fig. 5a,c).

The existence of an inverse peak in the profile I_θ is evaluated relative to the fluctuation level in I_θ. To quantify the fluctuation level, we compute the standard deviation σ_p of the difference $I_\theta - I_{s,\theta}$ for all points with $I_{s,\theta} > I_m$, where I_m is defined as the $100(1 - \theta_{th})$ percentile of $I_{s,\theta}$. Here the parameter θ_{th} is set to 0.5. This threshold is set such that the points in a potential inverse peak, which should have low values of $I_{s,\theta}$ and large differences between I_θ and $I_{s,\theta}$ due to rapid changes in the profile, do not contribute to and distort the evaluation of the fluctuation level. We evaluate the existence of an inverse peak in the smoothed profile $I_{s,\theta}$ using two criteria. The first criterion ensures that the inverse peak is deep enough. Specifically, we require that the local minimum of $I_{s,\theta}$ near (x_p, y_p) is smaller than a threshold $I_{th} = I_m - \alpha_{th}\sigma_p$ (gray line, Fig. 5b), where $\alpha_{th} = 15$ is a factor. If not, we decide that profile does not contain a valid

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

peak. The second criterion checks that the two flanks of the smoothed profile rise above $I_m - \sigma_p$ (dotted gray line, Fig. 5b). If not, the inverse peak is invalid.

For a valid peak, we determine the width of the peak starting from the minimum of $I_{s,\theta}$. We take a derivative of $I_{s,\theta}$ and smooth it to get dI_s (Fig. 5e). Starting from the minimum point, we trace the negative part of of the derivative and record the maximum dI_{s-} of the absolute value of $dI_{s,\theta}$ reached during the tracing. The tracing stops once $dI_{s-} > \alpha_{deriv}\sigma_p$ and $|dI_s| < dI_{s-} - \alpha_{deriv}\sigma_p$. Here $\alpha_{deriv} = 0.1$ is a factor. This way of stopping ensures that the tracing picks out the first significant peak in the derivative dI_s and does not stop because of small bumps in $dI_{s,\theta}$. Similarly, we trace the positive part of dI_s and record the maximum dI_{s+} of dI_s. The stop criterion is $dI_{s+} > \alpha_{deriv}\sigma_p$ and $dI_s < dI_{s+} - \alpha_{deriv}\sigma_p$. The width w of the peak is measured as the distances between the positions of dI_{s+} and dI_{s-} (black vertical lines, Fig. 5e). The derivatives of all eight smoothed profiles are shown in Fig. 5f.

If none of the profiles have valid inverse peaks, the SWC point is invalid. Otherwise, we chose the profile with the minimum width among the valid ones, and assign its peak position as the position (x_m, y_m) of the SWC point, and set $r_m = \alpha_r w/2$, where the factor $\alpha_r = 0.8$. This factor is introduced to match the radii of SWC points with those of the branches according to subjective judgement. If the distance between the original position (x_p, y_p) and (x_m, y_m) exceeds $\alpha_{shift} r_m$, where $\alpha_{shift} = 2$, then the SWC point has shifted too much, and it is flagged as invalid. If r_m is smaller than $r_{min} = 0.2$ μm or larger than $r_{max} = 10$ μm, the radius of the SWC point is too small or too large, and the point is invalid.

Finally, we check whether the pixels with a radius $0.5r_m$ from the center are dark enough. Specifically, we check whether the maximum value of the smoothed profile in the orthogonal direction (red line,Fig. 5d) to the chosen profile (cyan line,Fig. 5d) within $0.5r_m$ are smaller than a threshold (green horizontal line, Fig. 5d), which is set as the maximum smoothed intensity of the chosen profile at r_m plus σ_p. If not, the SWC point in invalid. This check ensures that the SWC points created along edges of thick dendrites or the soma are eliminated.

If the SWC point passes all the tests described above, it is judged as valid, and (x_m, y_m, z)

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and r_m are set as the new position and radius. The position and radius of the SWC are thus corrected after the validity check. To ensure that the corrections converge, the checking procedure is iterated three times with the positions and radius updated. The SWC point is accepted if it passes the tests all three times. The angle θ of the profile denotes the orientation perpendicular to the branch (red line, Fig. 5a).

The parameters used in checking the validity of an SWC point are listed in Table 4.

Mark pixels occupied

To avoid creating duplicated SWC points, we mark pixels in the tiff stack in the vicinity of the existing SWC points as occupied. Before creating a new SWC point, we check whether the pixel at its center is marked as occupied; if so, the SWC point is not created.

The marked pixels around two connected SWC points (x_1, y_1, z_1, r_1) and (x_2, y_2, z_2, r_2) are in a volume formed by two half cylinders with radii $\alpha_{occ} r_1$ and $\alpha_{occ} r_2$, respectively, and a trapezoidal prism that fits with the half cylinders (Fig. 6). Here $\alpha_{occ} = 1$ is a parameter for adjusting the extent of exclusion in xy. The z extent of the marked volume is large enough to contain the two SWC points: the distances from the top and bottom planes of the volume to the nearest SWC points are set to the maximum of $\alpha_{occ} r_1$, $\alpha_{occ} r_2$, or $z_{occ} = 2$ μm. Increasing the volume of the marked pixels prevents creations of spurious SWC points. However, if the volume is too large, correct SWC points can be eliminated, especially when branches are close to each other. These opposing constraints should guide the choice of the appropriate size for the volume.

The parameters for marking pixels occupied are in Table 5.

Thick dendrites and the soma

Thick dendrites and the soma can be missing from the mask created with the valley detectors. There are two main reasons: (1) their dimensions are much larger than the length scale of the valley detectors; (2) the intensities of the pixels inside them are uniform. Consequently, only the pixels at their boundaries show up in the mask. This leads to the error of no SWC points

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

for these structures. To correct this problem, we check the existence of thick dendrite and the soma in each tiff stack. The check is based on the observation that these structures are typically well stained and the pixels in them are very dark.

Specifically, we create and compare two 2D projections of the tiff stack. In the first one, we only project the pixels that are marked occupied because they are in the vicinity of the existing SWC points. We enlarge the marked volume by increasing α_{occ} and z_{occ} to three times of the original values. This is to ensure that all pixels associated with the dendrites already covered the SWC points, including the shadows the dendrites in the out-of-focus planes, and completely marked. From this 2D projection we determine a threshold, which is set to the intensity of the darkest 5% of pixels. This threshold indicates the darkness of the pixels covered by the existing SWC points.

In the second 2D projection, we project only the pixels that are not marked occupied. We then count the number of pixels that are darker than the threshold determined from the first 2D projection. If this number is larger than the number of pixels in an area $0.5L^2_{\text{soma}}$, where $L_{\text{soma}} = 3\ \mu m$ is the length scale of the soma, we decide that the tiff stack contains thick dendrite and/or the soma since there are significant number of dark pixels that are uncovered by the SWC points.

To place SWC points on thick dendrites and the soma, we create a 2D projection of the tiff stack with all pixels, smooth it, and create a mask by selecting top $\theta_{soma} = 0.05$ fraction of the darkest pixels. From the mask we create SWC points as described before. SWC points are created only if their positions are not marked occupied by the existing SWC points, using the original values of α_{occ} and z_{occ}.

The parameters for creating SWC points for thick dendrites and soma are listed in Table 6.

Extending SWC points in 3D

The SWC structures created from the masks are often incomplete, mostly due to the limitations of the masks in separating nearby branches in 2D projections. These branches could be well

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

separated in the tiff stack although their 2D projections are not; therefore it is useful to extend the SWC structure using the tiff stack.

To minimize the interference from noise, we delete isolated SWC points that are not connected to any other SWC points. To ensure that the extension does not create duplicated SWC points, we mark pixels nearby the existing SWC points occupied (red circles, Fig. 7a). From an end SWC point (x, y, z, r) (yellow circle, Fig. 7a), we search the plane at z for the candidate for the next point. To reduce noisy fluctuations, the pixel intensity in the plane is smoothed with a Gaussian filter with $\sigma = 2$ pixels. The search is done in a ring area centered at (x, y), with inner and outer radii set to $\alpha_{smin}r$ and $\alpha_{smax}r$ (blue arcs, Fig. 7a), where $\alpha_{smin} = 2.5$ and $\alpha_{smax} = 6$ are the factors determining the range of the ring area. We also restrict the search to a range of angle ($\theta_{\text{thr}} = \pi/3$), where the angle is between the line from a pixel in the ring area to the end point (black lines, Fig. 7a) and the line from the end point to its connected SWC point (yellow line, Fig. 7a). The weighted distance is obtained by summing the geometric distance between consecutive pixel points multiplied by a factor $e^{\alpha_d I}$, where $\alpha_d = 20$ is the factor and I is the pixel intensity.

We search the potential point for extension along an arc of radius r_s centered at the end point and spanning the allowed angle range. We start with $r_s = \alpha_{smin}r$. We divide the arc into 64 points, and build a profile of the shortest distances at these points (black line, Fig. 7b). The profile is smoothed with a Gaussian filter with $\sigma = 4$ (green line, Fig. 7b). We find the minimum of the smoothed profile, and take the corresponding pixel point (x_c, y_c, z, r) as the candidate SWC point. We then adjust z of the point by building the profile of pixel intensity from $z - z_e$ to $z + z_e$, where z_e is the larger of $z_b = 2.5$ μm or r (black line, Fig. 7c). The profile is Gaussian smoothed with $\sigma = 1$ (green line, Fig. 7c). We find the local minimum in the smoothed profile near z as the depth of the candidate SWC point (x_c, y_c, z_c, r).

We test the validity of the candidate SWC point, which also adjusts x_c, y_c and r_c. This process of adjusting z_c and testing validity is done three times to ensure the convergence of x_c, y_c, z_c, r_c. If the validity is confirmed all three times, we check whether x_c, y_c, z_c is marked

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

occupied. If not, a new SWC point (x_c, y_c, z_c, r_c) is created and connected to the end point. The extension process then continues from the new SWC point as the end point. If the candidate point is marked occupied, the extension stops; we find the existing SWC point nearest to the candidate point, and if the difference in z is not too large as described previously, we connect the nearest point to the end point.

If the candidate point fails the validity test, we increase the radius r_s of the arc by 2 pixels or $(\alpha_{smax}r - \alpha_{smin}r)/10$, whichever is larger, and repeat the search process. This increase of r_s stops when r_s reaches $\alpha_{smax}r$.

The parameters used in extending SWC points are listed in Table 7.

Connecting end points

After extending SWC points in 3D, a continuous branch can still be represented with broken segments of SWC points, especially if the underlying signal is weak or there are closely crossing branches (Fig. 2b,c). We connect these segments with heuristic rules to recover the branch continuity. To do so, we compute the Euclidean distance between all pairs of end points that are not connected and the differences in z are within the allowed range as described before. If the distance of the pair in xy is smaller than $1.5(r_1 + r_2)$, where r_1, r_2 are the radii, they are judged to be close to each other and are connected.

After connecting the nearby pairs, we consider more distant ones. If the two end points are within $\alpha_{xy}(r_1 + r_2)$ in xy, where $\alpha_{xy} = 2$ is a factor; and if the angles between the two lines, linking the end points to their respective connected SWC points, is smaller than $\theta_{thr} = \pi/3$; then the two points are preserved in the candidate pool for potential connections (see the section for create SWC points from xy-paths for these criteria). We then iteratively connect the pairs of end points in the pool, connecting the closest available pairs first. Once connected, the end points are excluded from further connections. This pairwise connection stops if the pairs are all considered or if further connections create loops in the SWC structure.

The results for our example tiff stack are shown in Fig. 8.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Subdividing tiff stack in z

The 2D projection can be complicated when there are many branches in one stack. This often leads to occlusions in the 2D projection and missed branches in the reconstruction. One way of mitigating this problem is to divide the tiff stack in z into n_{div} slabs with equal heights in z. We create SWC points separately for each slab. When creating the 2D projection for a slab, we include extra volume in z by extending the height in both directions by $z_{ext} = 5$ μm. This is useful for getting good projections of branches that are near the dividing planes between the slabs. The z-image also includes the extended volume. Any SWC points whose depths are beyond the slab boundary are deleted. The extension from the end points are done with the entire tiff stack, which helps to connect SWC points that belong to the same branch but are cut by the subdivision. In our example neuron, we set $n_{div} = 8$.

The parameters of subdivision are listed in Table 8.

Combining SWCs for the entire neuron

The image of an entire neuron consists of multiple tiff stacks stitched together (Fig. 10a). The coordinates of the stacks relative to the first stack are determined during the stitching. For each stack we obtain the SWC structure, and shift the positions of the SWC points by the relative coordinates of the stack. The SWC points of individual stacks are read-in sequentially. To avoid duplicated SWC points in the overlapping regions of adjacent stacks, pixels near the SWC points that are already created are marked occupied by setting the parameters z_{occ} 5 times of the usual value and r_{occ} 2 times. If the position of SWC points are at the marked pixels, they are deleted. For individual stacks, the step of connecting the end points is omitted. Instead, after reading in the SWC points of all stacks, we extend the SWC points from the end points, and then connect the new end points. To eliminate noise, we delete very short leaf branches in the SWC structure (those that have fewer than $n_{dmin} = 3$ SWC points). In addition, we eliminate isolated branches that have fewer than $n_{imin} = 5$ SWC points. Increasing this number reduces noise in the reconstruction, but can also delete some of the correct reconstruction. The

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

reconstructed SWC structure for the example neuron is shown in Fig. 10b-d.

The parameters are listed in Table 9.

References

Brown KM, Barrionuevo G, Canty AJ, De Paola V, Hirsch JA, Jefferis GSXE, Lu J, Snippe M, Sugihara I, Ascoli GA (2011) The DIADEM data sets: representative light microscopy images of neuronal morphology to advance automation of digital reconstructions. *Neuroinformatics* 9:143–57.

Danielsson PE (1980) Euclidean distance mapping. *Computer Graphics and image processing* 14:227–248.

Feng L, Zhao T, Kim J (2015) neutube 1.0: a new design for efficient neuron reconstruction software based on the swc format. *eneuro* 2:ENEURO–0049.

Frangi AF, Niessen WJ, Vincken KL, Viergever MA (1998) Multiscale vessel enhancement filtering In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 130–137. Springer.

Gidon A, Segev I (2012) Principles governing the operation of synaptic inhibition in dendrites. *Neuron* 75:330–341.

Gillette TA, Brown KM, Ascoli GA (2011) The DIADEM metric: comparing multiple reconstructions of the same neuron. *Neuroinformatics* 9:233–45.

Gillette TA, Brown KM, Svoboda K, Liu Y, Ascoli GA (2011) DIADEMchallenge.Org: A Compendium of Resources Fostering the Continuous Development of Automated Neuronal Reconstruction. *Neuroinformatics* 9:303–304.

Graham RL, Hell P (1985) On the history of the minimum spanning tree problem. *Annals of the History of Computing* 7:43–57.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Hamill OP, Marty A, Neher E, Sakmann B, Sigworth F (1981) Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflügers Archiv European journal of physiology* 391:85–100.

Henze DA, Cameron WE, Barrionuevo G (1996) Dendritic morphology and its effects on the amplitude and rise-time of synaptic signals in hippocampal CA3 pyramidal cells. *The Journal of comparative neurology* 369:331–44.

Jaeger D (2001) Accurate reconstruction of neuronal morphology. *Computational neuroscience: Realistic modeling for experimentalists* pp. 159–178.

Krichmar JL, Nasuto SJ, Scorcioni R, Washington SD, Ascoli Ga (2002) Effects of dendritic morphology on CA3 pyramidal cell electrophysiology: a simulation study. *Brain research* 941:11–28.

Lankton S (2009) Sparse Field Methods - Technical Report. *Georgia Institute of Technology Techniacal Report* .

Liu Y (2011) The DIADEM and beyond. *Neuroinformatics* 9:99–102.

Mainen ZF, Sejnowski TJ (1996) Influence of dendritic structure on firing pattern in model neocortical neurons. *Nature* 382:363–366.

Menon V, Musial TF, Liu A, Katz Y, Kath WL, Spruston N, Nicholson DA (2013) Balanced synaptic impact via distance-dependent synapse distribution and complementary expression of ampars and nmdars in hippocampal dendrites. *Neuron* 80:1451–1463.

Narayanaswamy A, Wang Y, Roysam B (2011) 3-D image pre-processing algorithms for improved automated tracing of neuronal arbors. *Neuroinformatics* 9:219–31.

Parekh R, Ascoli Ga (2013) Neuronal morphology goes digital: a research hub for cellular and system neuroscience. *Neuron* 77:1017–38.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Peng H, Bria A, Zhou Z, Iannello G, Long F (2014) Extensible visualization and analysis for multidimensional images using Vaa3D. *Nature protocols* 9:193–208.

Peng H, Hawrylycz M, Roskams J, Hill S, Spruston N, Meijering E, Ascoli GA (2015) Bigneuron: large-scale 3d neuron reconstruction from optical microscopy images. *Neuron* 87:252–256.

Sholl DA (1953) Dendritic organization in the neurons of the visual and motor cortices of the cat. *Journal of anatomy* 87:387.

Stuart G, Spruston N (1998) Determinants of voltage attenuation in neocortical pyramidal neuron dendrites. *The Journal of neuroscience* 18:3501–3510.

Svoboda K (2011) The past, present, and future of single neuron reconstruction. *Neuroinformatics* 9:97–8.

Turaga SC, Murray JF, Jain V, Roth F, Helmstaedter M, Briggman K, Denk W, Seung HS (2010) Convolutional networks can learn to generate affinity graphs for image segmentation. *Neural computation* 22:511–538.

Türetken E, González G, Blum C, Fua P (2011) Automated reconstruction of dendritic and axonal trees by global optimization with geometric priors. *Neuroinformatics* 9:279–302.

Whitaker RT (1998) A Level-Set Approach to 3D Reconstruction from Range Data. *International Journal of Computer Vision* 29:203–231.

Zandt BJ, Losnegård A, Hodneland E, Veruki ML, Lundervold A, Hartveit E (2017) Semi-automatic 3d morphological reconstruction of neurons with densely branching morphology: Application to retinal aii amacrine cells imaged with multi-photon excitation microscopy. *Journal of neuroscience methods* 279:101–118.

Zhang TY, Suen CY (1984) A fast parallel algorithm for thinning digital patterns. *Communications of the ACM* 27:239–242.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1068 Zhou Z, Sorensen S, Zeng H, Hawrylycz M, Peng H (2015) Adaptive image enhancement for
1069 tracing 3d morphologies of neurons and brain vasculatures. *Neuroinformatics* 13:153–166.

1070 Zitova B, Flusser J (2003) Image registration methods: a survey. *Image and vision comput-*
1071 *ing* 21:977–1000.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure Legends

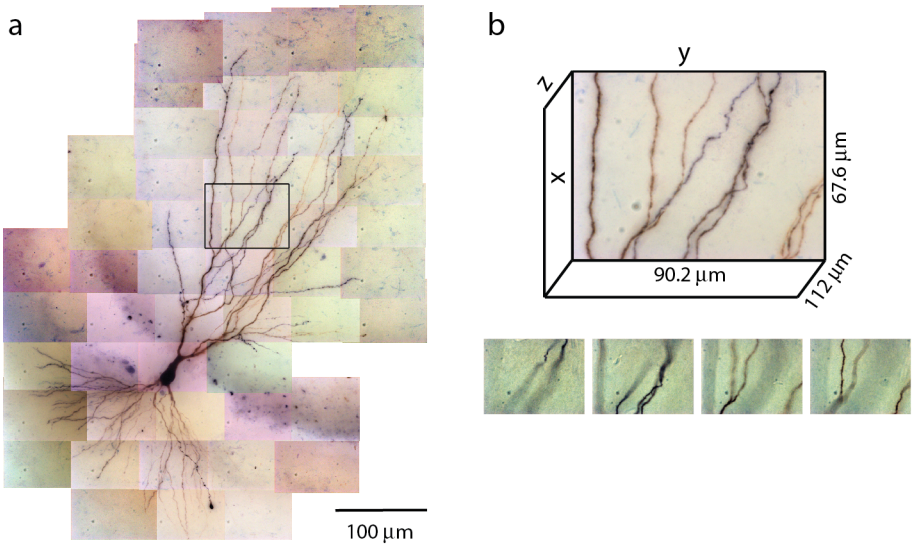

Figure 1: Tiles of tiff stacks covering the entire neuron. a. A mouse CA3 neuron imaged at 100X, NA 1.4. There are 51 tiles covering the entire neuron. 2D projection is shown. b. Dimensions of one tile. Each tiff stack consists of 224 planes of images. The distance between successive image planes is 0.5 μm. Four planes at different depths in the tiff stack indicated by the black rectangle in (a) are shown below.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

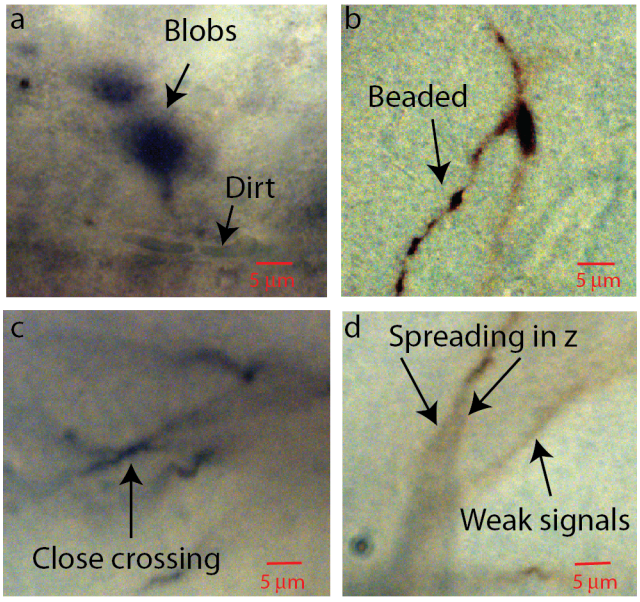

Figure 2: Aspects of bright-field images of biocytin-filled neurons that makes automatic reconstruction challenging. Parts of images in single planes of the tiff stacks are shown. a. Biocytin can be spilled and create spurious signals. Dirt or dust can also add noise. b. Thin branches can be broken into "beads". c. Close crossing between adjacent neuron branches. d. A branch can cast bifurcating shadows in z with darkness level comparable to weak signals from nearby faint branches.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

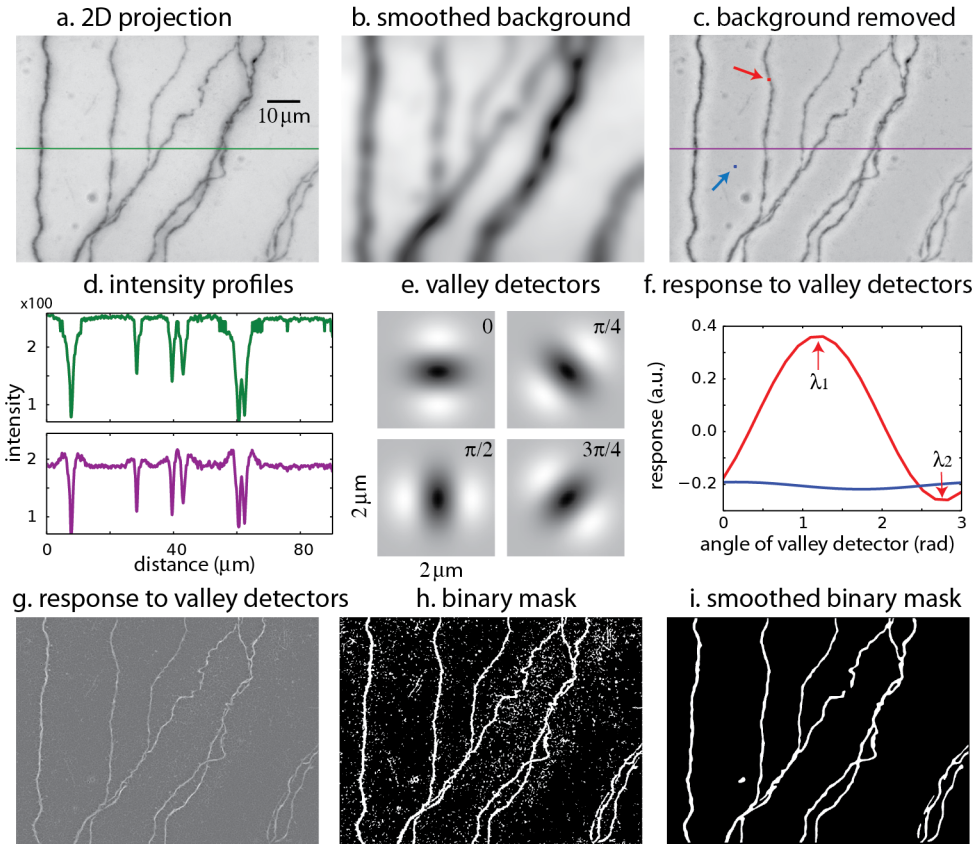

Figure 3: The process of creating a binary mask from 2D projection. a. 2D projection from the image stack. The intensity profile across the green line is shown in (d). b. Smoothed background obtained from Gaussian smoothing of the 2D projection. c. 2D projection after removing the smoothed background. The intensity across the purple line is shown in (d). The red and blue arrows indicate the points to be tested with valley detectors in (f). d. Intensity across the midline in the original 2D projection (green line in (a)) and after removal of the background (purple line in (c)). e. Images of valley detectors at four orientations. f. Responses to valley detectors at varying orientations for two points shown in c (red point, red line; blue point, blue line). λ_1 and λ_2 are the maximum and minimum responses, respectively. g. The maximum responses to the valley detectors (λ_1) for all pixels. h. The binary mask obtained from thresholding λ_1. i. The smoothed binary mask.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

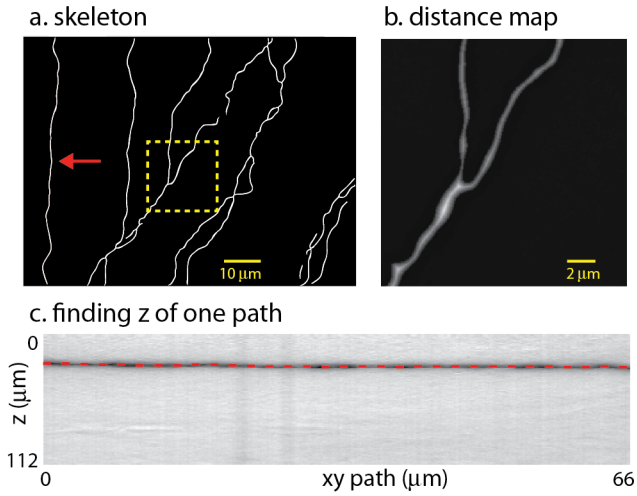

Figure 4: Creating SWC structure from the mask. a. The skeleton obtained by thinning the mask. b. Distance map computed from the mask. The square region highlighted in (a) is shown. The brightness of each pixel is proportional to the distance to the nearest boundary in the mask. c. Finding the depth of the path. The image is constructed by cutting through the stack in z-dimension following the xy path indicated by the arrow in (a). The dark band is the neurite along the path. The dotted line is the depth (z) computed using the left-right shortest path algorithm.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

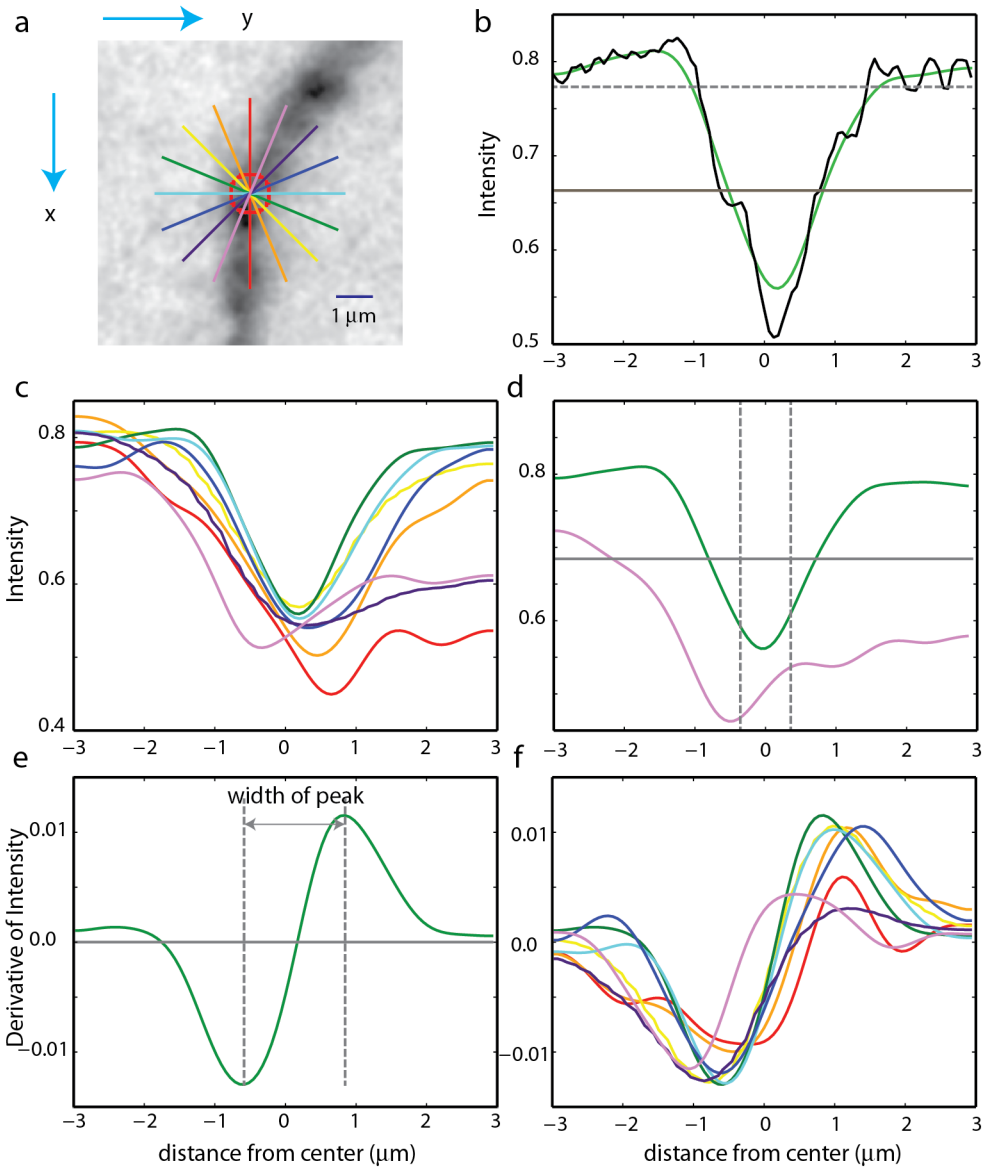

Figure 5: Checking the validity of an SWC point. a. A patch of image around an SWC point to be examined. The image is taken from the z-plane of the SWC point. Profiles of the intensities along eight directions are taken (straight lines; colors indicate angles). The green line is the profile chosen to adjust the SWC point. The red circle indicates the radius of the SWC point. b. The profile (black, raw; green, smoothed) along the green line in (a). The dotted gray line is the baseline, and the solid gray line is the threshold. An inverse peak is judged valid if the flanks of the smoothed profile go above the baseline, and the minimum value goes below the threshold. c. Smoothed profiles at all eight directions. d. The chosen profile (green) and the profile at the orthogonal direction (violet). The vertical lines are at the half radius points. Note that the center of the profiles are slightly shifted compared to those in (c). For the SWC to be valid, the minimum intensity of the profile at the orthogonal direction must be below the threshold (gray line) within the vertical lines. e. Smoothed derivative of the smoothed profile in (b). The vertical lines indicate the local maxima of the derivatives. The distance between the vertical lines is the width of the peak. f. Smoothed derivatives of the profiles for all eight directions. The profile with the minimum width is chosen.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

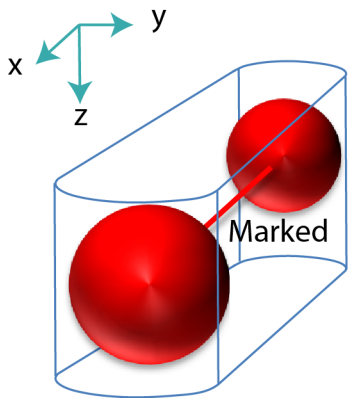

Figure 6: Mark pixels in tiff stack occupied. The pixels around two connected SWC points (red spheres), formed by two half cylinders and a trapezoidal prism, are marked as occupied.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

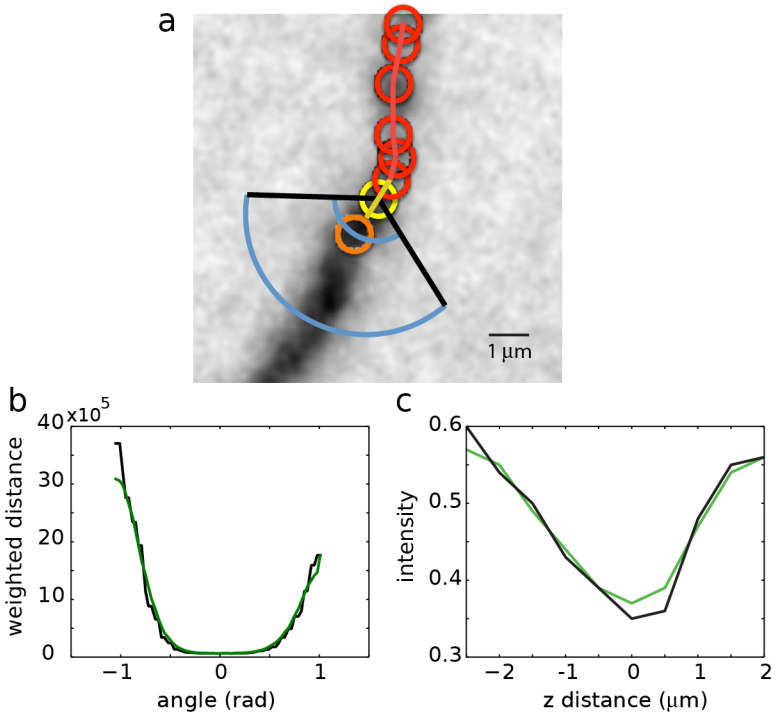

Figure 7: Extending SWC structure in 3D. a. The candidate point for extension is searched in the plane of an end SWC point (yellow circle). Red circles are the SWC points that are connected to the end point. The search is restricted to the pixels in the area enclosed by the two blue arcs and two black lines. The intensity-weighted shortest distances from the end point to pixels in the search area are computed. b. The profile of the shortest distances along the inner arc (black line in in (a)). The green line is the smoothed version. The angle is measured relative to the line connecting the end point to its connected SWC point (yellow line in (a)). The angle at the minimum value of the smoothed profile is selected as the position for the candidate point (orange circle in (a)). c. The depth of the candidate point is adjusted using the intensity profile in z at the xy position of the candidate point (black line). The point of minimum intensity in the smoothed profile (green line) is set as the depth of the candidate point. The z distance in the graph is relative to the z position of the end point.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

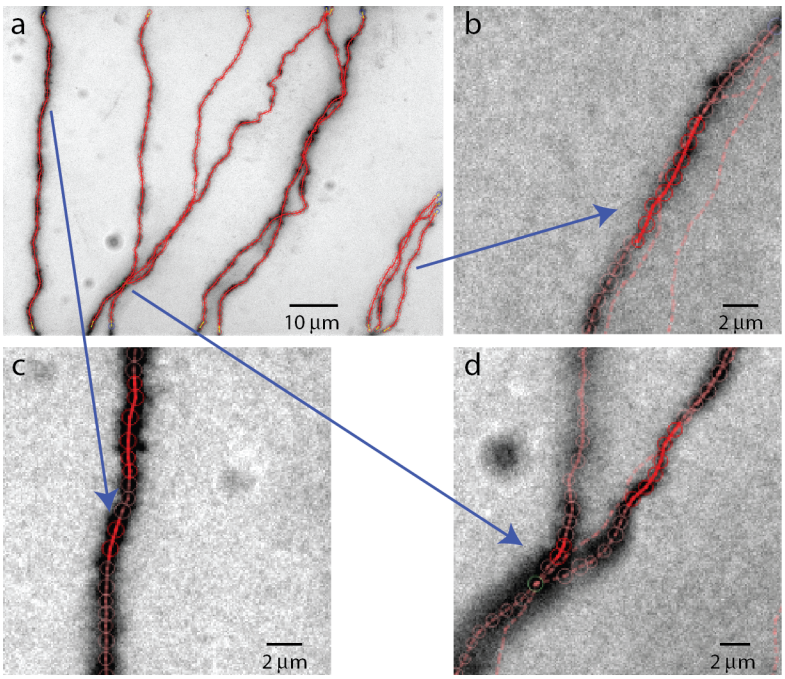

Figure 8: The SWC structure overlaid on the image. a. The SWC structure projected on to the image of 2D projection. Red circles are SWC points. Connections between them are indicated with red lines. b-d. SWC structure overlaid at the specific planes in the tiff stack, zoomed in to show more details. Arrows indicate the corresponding regions in the 2D projection.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

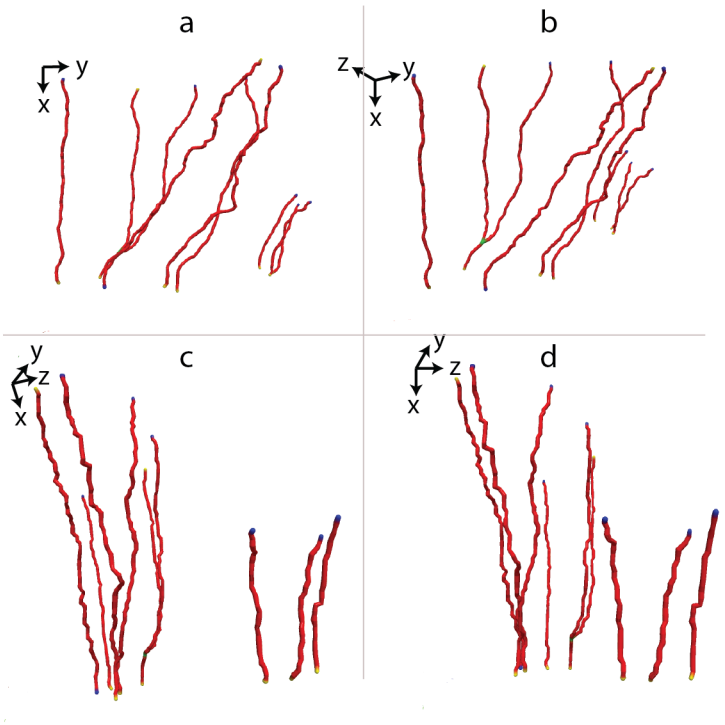

Figure 9: The SWC structure viewed from four different angles to reveal the 3D structure. The viewing angles are indicated with the directions of xyz coordinates. The view angle in (a) corresponds to the 2D projection view.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

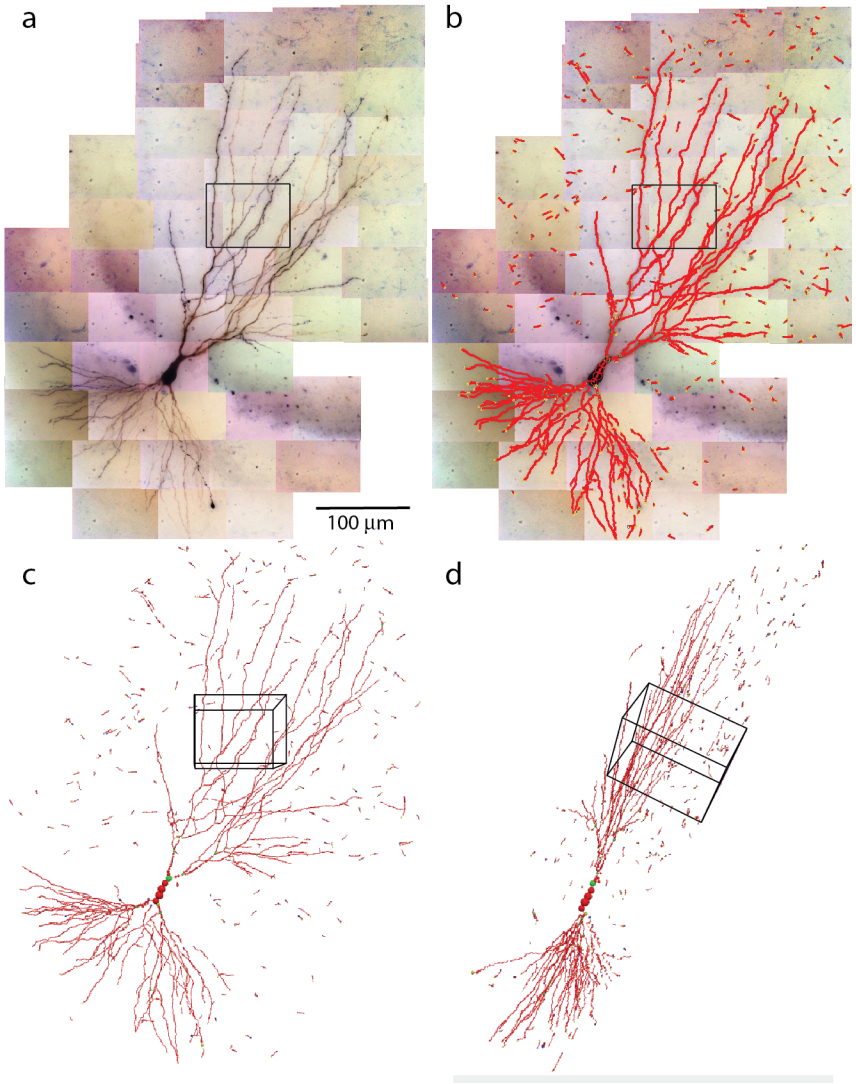

Figure 10: Combing SWCs from the stacks for the entire neuron. a. 2D projection of the entire neuron. Individual tiff stacks are stitched together to obtain their relative coordinates. The stack in Fig.9 is highlighted with a black rectangle. b. The SWC of the entire neuron is obtained by combining the SWCs of individual tiff stacks. The SWC points are overlaid onto the 2D projection. c. 3D view of the SWC structure. The 3D box corresponds to the highlighted stack in (a). d. The 3D view from a different angle.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

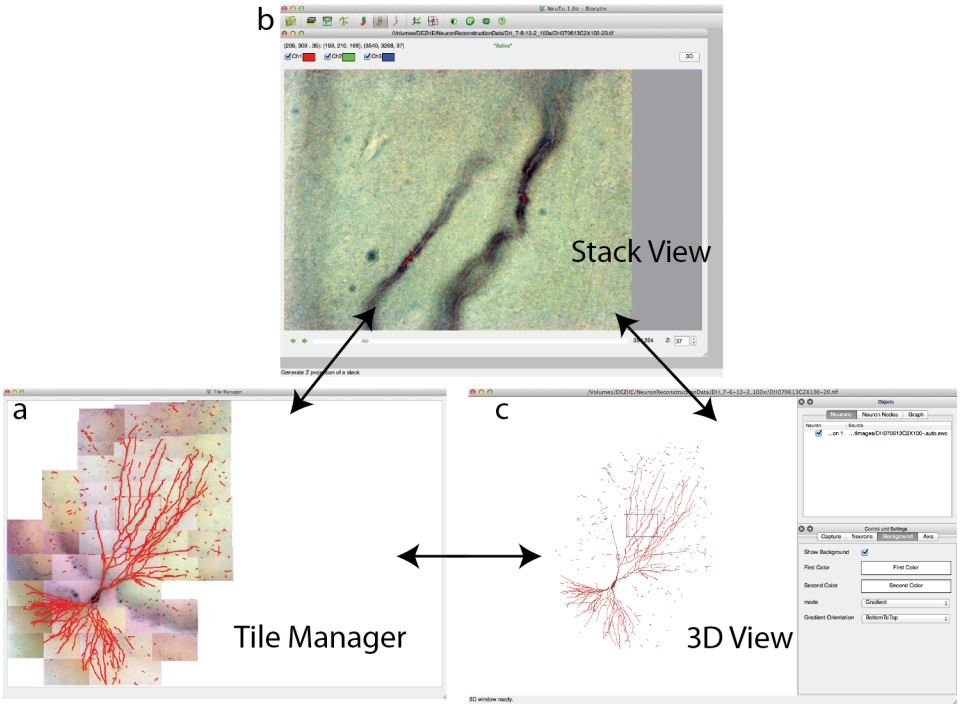

Figure 11: Three views for examining the SWC structure. a. Tile manager. 2D projection of the stitched stacks is superimposed with the 2D projection of the SWC structure. b. Stack view. One stack is loaded, with the SWC points in the stack overlaid onto the image. The 2D projection view of the stack can be created within this window. c. SWC view. The 3D structure can be viewed from different angles and edited.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

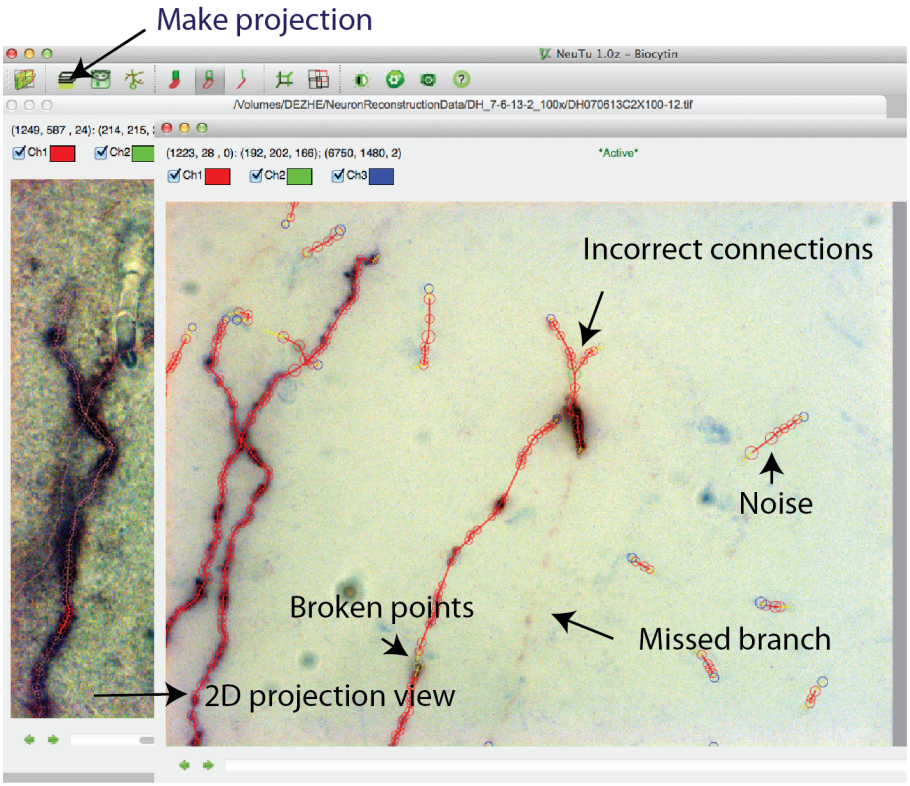

Figure 12: In the Stack View, clicking on the Make Projection button creates the 2D projection of the tiff stack and the SWC structure. It is easy to spot mistakes in this view. SWC points can be removed and their properties changed. The connections between SWC points can be modified. Selecting one SWC point and pressing `z` locates the points in the Stack View for further examination and modification.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

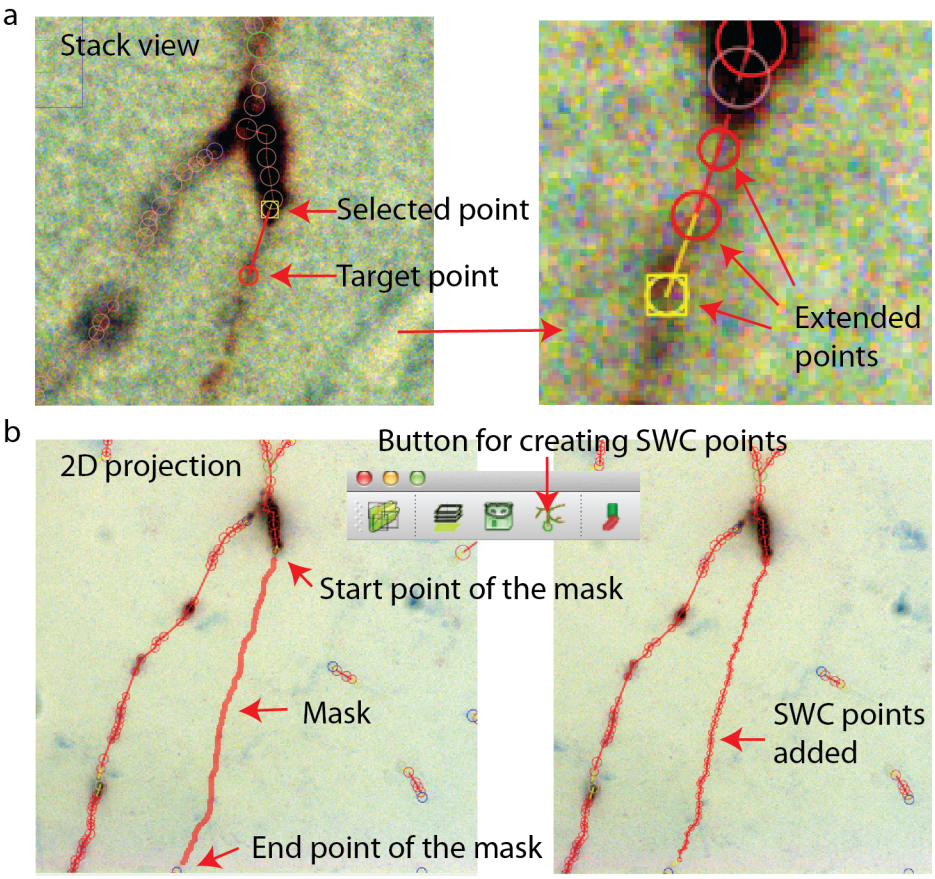

Figure 13: Creating SWC points. a. In Stack View, an SWC point is selected. Find the target point by finding the focus plane of the branch. Clicking on the target point creates SWC points connecting the target point to the selected point along the branch. b. In Projection View, pressing `r` starts mask creation. Click on the starting point and `Shift-click` on the end point along a branch creates a mask. Clicking on the `Mask` → `SWC` button creates SWC points along the mask.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

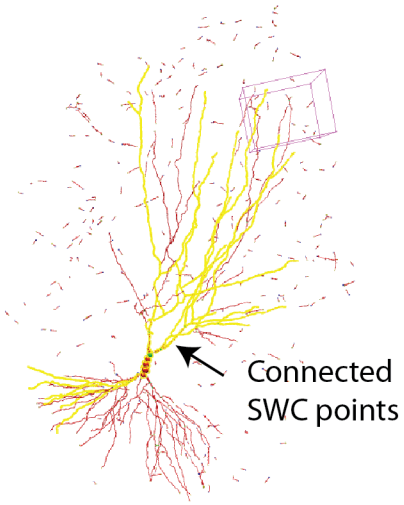

Figure 14: In 3D View, selecting one SWC point and pressing `s-4` selects all SWC points connected to the selected. This operation is useful for detecting broken connections.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

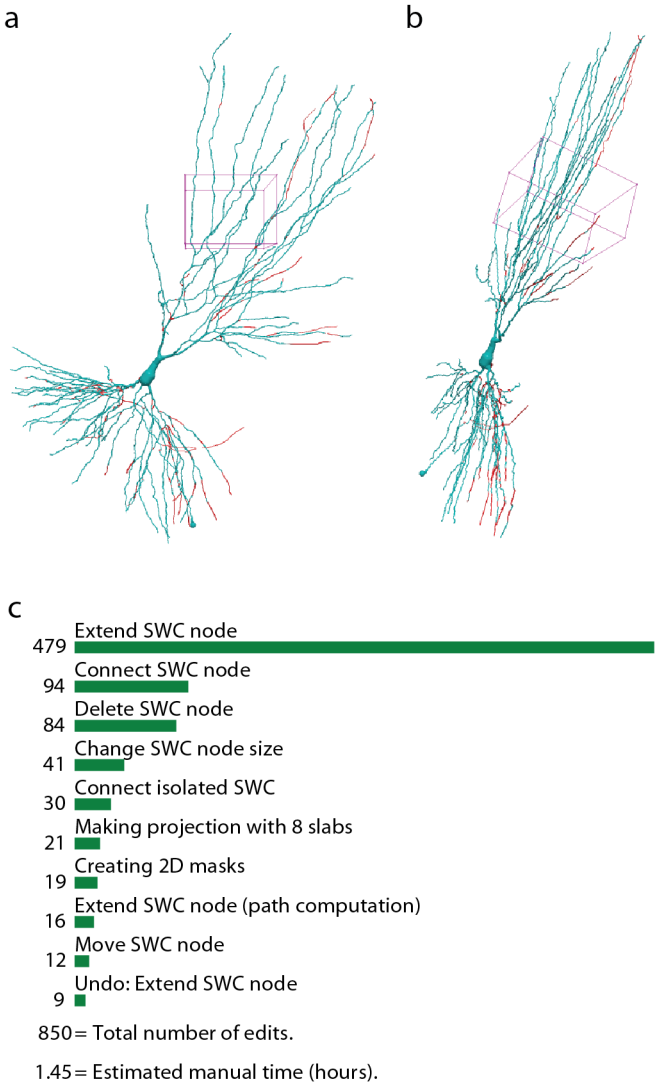

Figure 15: Reconstructed neuron after editing. a,b. Two different views of the reconstructed neurons. The SWC points from the automatic reconstruction are in blue, and those added in the editing process are in red. c. Top ten operations done in the editing process and the total number of edits.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

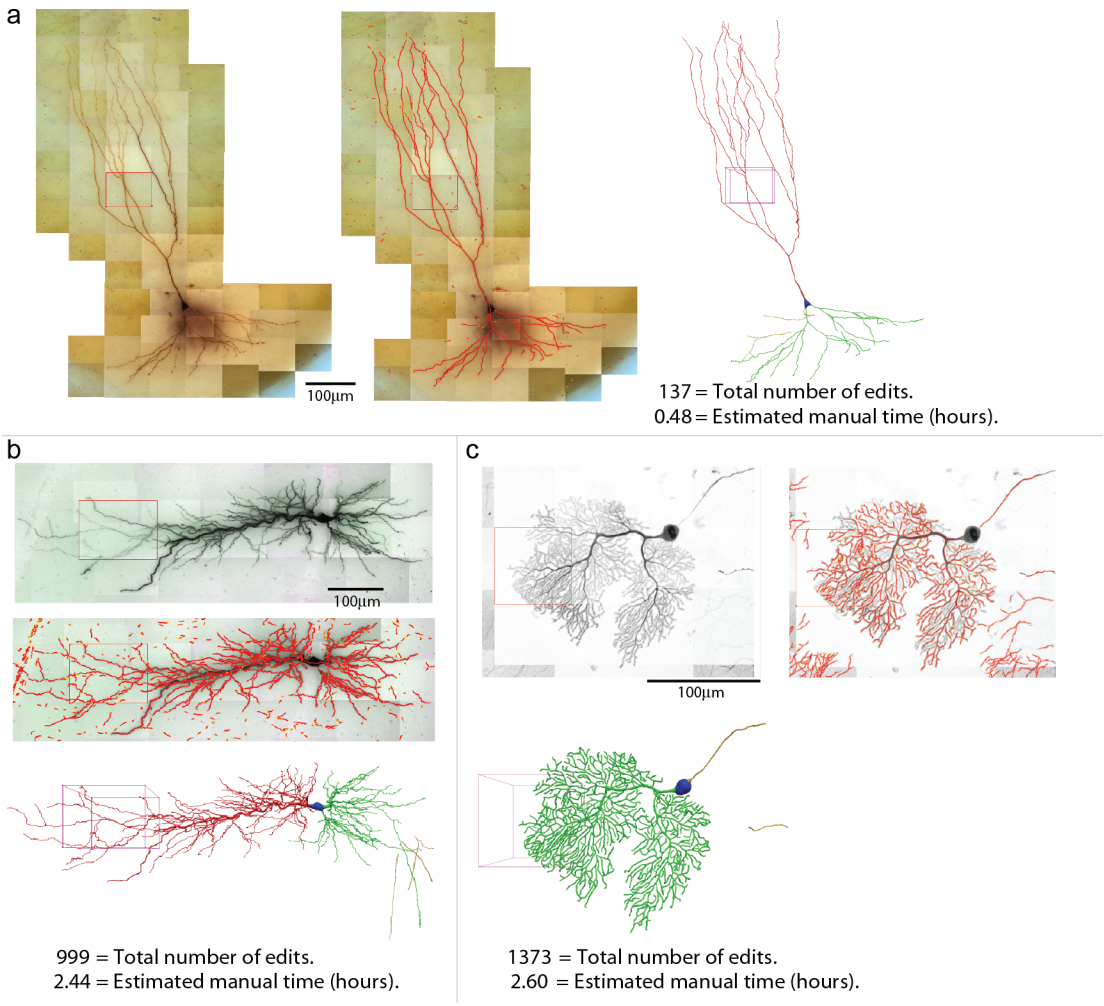

Figure 16: Three examples of reconstructions. Shown for each neuron are the 2D projection of the images, automated reconstruction on top of the 2D projection, and the final reconstruction (blue, soma; red, apical dendrite; green, basal dendrite; gold, axon). a. A pyramidal neuron in the mouse CA3 region imaged at 100X (biocytin). b. A pyramidal neuron in the rat CA1 region imaged at 63X (biocytin). c. A mouse purkinje cell imaged at 63X with confocal microscope.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Tables

Parameter	Value	Meaning
d_{xy}	0.065 μm	pixel distance in xy
d_z	0.5 μm	z-distance between successive planes
σ_b	2 μm	length scale for smooth background

Table 1: Parameters for tiff stacks and 2D projections.

Parameter	Value	Meaning
σ	0.1 μm	length scale of valley detector
α_λ	10	factor for eliminating circular blobs
f_λ	0.1	fraction for determining the threshold
μ	0.1	parameter for level set smoothing
$N_{levelset}$	500	number of level set iterations
A_s	1.0 μm^2	threshold for small areas removed

Table 2: Parameters for creating the binary mask.

Parameter	Value	Meaning
l_{sm}	0.5 μm	smallest length of xy-path used
α_d	20	factor for weighting distance between pixels
α_p	0.1	factor for decreasing distance between consecutive SWC points
α_{xy}	2	factor for disconnecting two consecutive SWC points
θ_{thr}	$\pi/3$	maximum angle allowed between consecutive lines
α_{zj}	2	factor for z jump threshold

Table 3: Parameters for creating SWC points from 2D mask.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Parameter	Value	Meaning
r_m	3 μm	lower bound for the range of the profiles
σ_s	0.3 μm	length scale for smoothing profiles
θ_{th}	0.5	fraction for determining the threshold for peak
α_{th}	15	factor for determining the threshold for peak
α_{deriv}	0.1	factor for determining the stop criterion for derivatives
α_{shift}	2	factor for allowing shifts in the position
r_{min}	0.2 μm	lower bound for the radius
r_{max}	10 μm	upper bound for the radius
α_r	0.8	factor for adjusting the radius

Table 4: Parameters for checking validity of an SWC point.

Parameter	Value	Meaning
α_{occ}	1	factor for adjusting radius of the marked volume
z_{occ}	2 μm	lower bound for extend of the volume in z

Table 5: Parameters for marking pixels near the SWC points occupied.

Parameter	Value	Meaning
L_{soma}	3 μm	length scale of the soma
θ_{soma}	0.05	fraction for the threshold of creating mask

Table 6: Parameters for creating SWC points for thick dendrites and the soma.

Parameter	Value	Meaning
α_{smin}	2.5	factor for the minimum radius of the search area
α_{smax}	6	factor for the maximum radius of the search area
z_s	2.5 μm	the minimum of the range in z for adjusting the depth

Table 7: Parameters for extending SWC points in 3D.

Parameter	Value	Meaning
n_{div}	8	number of subdivisions
z_{ext}	5 μm	extension of sub-slabs when creating SWC points

Table 8: Parameters for subdividing a tiff stack.

bioRxiv preprint doi: https://doi.org/10.1101/226548; this version posted November 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Parameter	Value	Meaning
n_{dmin}	3	minimum number of SWC points allowed in leafs
n_{imin}	5	minimum number of SWC points allowed in isolated branches

Table 9: Parameters for reducing noise in SWC structure.