Neuroinformatics manuscript No.
(will be inserted by the editor)

Rivulet: 3D Neuron Morphology Tracing with Iterative Back-Tracking

Siqi Liu* · Donghao Zhang · Sidong Liu · Dagan Feng · Hanchuan Peng · Weidong Cai*

Received: date / Accepted: date

Abstract The digital reconstruction of single neurons from 3D confocal microscopic images is an important tool for understanding the neuron morphology and function. However the accurate automatic neuron reconstruction remains a challenging task due to the varying image quality and the complexity in the neuronal arborisation. Targeting the common challenges of neuron tracing, we propose a novel automatic 3D neuron reconstruction algorithm, named Rivulet, which is based on the multi-stencils fast-marching and iterative back-tracking. The proposed Rivulet algorithm is capable of tracing discontinuous areas without being interrupted by densely distributed noises. By evaluating the proposed pipeline with the data provided by the Diadem challenge and the recent BigNeuron project, Rivulet is shown to be robust to challenging microscopic imagestacks. We discussed the algorithm design in technical details regarding the relationships between the proposed algorithm and the other state-of-the-art neuron tracing algorithms.

Keywords 3D Neuron Reconstruction · Neuron Morphology

Availability: This method has been implemented both as a plugin for Vaa3D (`https://github.com/Vaa3D/vaa3d_tools/tree/master/bigneuron_ported/siqi/rivuletv3d`) and a standalone Matlab GUI toolbox (`https://github.com/lsqshr/Rivulet-Neuron-Tracing-Toolbox`).

Siqi Liu*, Donghao Zhang, Sidong Liu, Dagan Feng, Weidong Cai*
School of Information Technologies, University of Sydney, NSW, Australia
E-mail: siqi.liu@sydney.edu.au, tom.cai@sydney.edu.au

Hanchuan Peng
Allen Institute for Brain Science, Seattle, WA, USA.
E-mail: hanchuan.peng@gmail.com

1 Introduction

Neuron morphology is a core neuroscience interest. 3D microscopic images are used to visualise the neuronal architectures. Neuron tracing is a primary way to digitalise the tree-like branching of axons and dendrites as a sequence of intersected cylinders from optical microscopies. Within the scope of computational neuroscience, the reconstructed morphological models are acquired for purposes such as neuronal identity, anatomically and biophysically realistic simulations, morphometric and stereological analysis and determining potential connectivity (Parekh and Ascoli, 2013). A great proportion of the digitalised neurons so far were acquired by manual tracing which is a highly labour intensive procedure.

Most of the existing tracing algorithms require a certain level of user intervention. The fully automatic and precise neuron tracing remains a challenging task mainly due to the poor quality of neuron images caused by the fundamental limits in confocal microscopy. The dentritic structures often have highly varying contrast due to the uneven distribution of fluorescent markers within the neuron cells, resulting in discontinuity and broken shapes of neuronal fibres. The image noises come from different sources which generally do not follow a Gaussian distribution, mainly because the excitation power of the laser scanning device is often limited to protect the cellular structures. Different levels of anisotropic distortion is also caused by the Point Spread Function (PSF) imposed by the optics of the microscope (Santamaría-Pang et al, 2015; Pawley, 2006). Thus, it is non-trivial to approach the imaging limits simply with conventional de-noising or deconvolution algorithms in automated pipelines due to the varying types and levels of the noises and distortions. The main challenges affect most of the existing neuron tracing algorithms can be summarised as (1) The irrelevant structures and noisy points which cause over-tracing non-existent arbors from the background; (2) Gaps in continuous arbors which cause under-tracing arbors of interest; (3) Wrongly wired topology between different branches and (4) Non-smooth surface of the arbors violating the geometric assumptions.

The recent state-of-the-art algorithms are often pipelines combining preprocessing, branch tracing, and post-processing methods. A number of semi-automatic/automatic 3D tracing algorithms and softwares have been proposed to enable large-scale data collection in recent years (Zhao et al, 2011; Mukherjee et al, 2015; Santamaría-Pang et al, 2015; Chen et al, 2015; Mukherjee and Stepanyants, 2012; Basu and Racoceanu, 2014; Türetken et al, 2011; Xiao and Peng, 2013; Wang et al, 2011; Yang et al, 2013; Peng et al, 2011; Ming et al, 2013). Many of the algorithms were supported by the hackathon events such as the Diadem challenge (Brown et al, 2011) and the recent BigNeuron project (Peng et al, 2015b).

Hessian based image restoration methods are widely used as a way of preserving the curvilinear structures and eliminating the noise points (Zhou et al, 2014; Frangi et al, 1998). The neuronal structures are then segmented from the background voxels with adaptive or manual threshold. Some recent voxel-wise

learning methods based on Hessian measurements (Santamaría-Pang et al, 2015) or multi-scale wavelet representation (Chen et al, 2015) would further increase the segmentation results, though there would be a trade-off of running-time especially for non-parametric classifiers such as support vector machine (SVM). Although preprocessing methods can be helpful to enhance the image quality to a certain level, the above mentioned difficulties for tracing algorithms still remain.

Neuronal arbours in light microscopic images are often not with perfect 3D tubular shapes and the termini do not form ideal hemispheres. Algorithms rely on over-complicated geometrical assumptions would have difficulties dealing images with noise affected arbours (Mukherjee et al, 2015; Santamaría-Pang et al, 2015; Wang et al, 2011). Methods rely on the precision of seed detections tend to have missing arbours and unconnected branches. The tracing methods based on the original fast-marching algorithm (Santamaría-Pang et al, 2015; Mukherjee and Stepanyants, 2012; Basu and Racoceanu, 2014) or minimum spanning tree (Türetken et al, 2011; Yuan et al, 2009; González et al, 2010) tend to produce over-traced branches and wrong topology. The combination of fast-marching and gradient descent was shown to be effective for jumping the gaps seen in poorly segmented foreground by iteratively re-initialising the start point for tracing based on previous traced branches (Mukherjee and Stepanyants, 2012). However the reinitialization method might be also risky to jump between spatially closed branches and noise points. Recent fast-marching based methods, such as APP2 (Xiao and Peng, 2013) depending on a grey-scale weighted distance transform (GWDT) and post-processing criteria designed with prior knowledge of neuronal morphology, effectively reduced the disadvantages in previous fast-marching based algorithms.

To tackle the challenges mentioned above, we present the Rivulet algorithm for automatic and precise 3D neuron tracing, which showed promising results over the state-of-the-art algorithms in a pilot study (Zhang et al, 2016). Rivulet uses Hessian-based measurements to enhance the neuron segmentation and performs multi-stencils fast marching (MSFM) on a speed image obtained from GWDT. A gradient descent approach based on RK4 (Jameson et al, 1981) is used to trace fibres from the resulted time crossing map with sub-voxel precision. The iterative tracing of Rivulet was originally inspired by an arbour skelontonisation method which was proposed for medical images with better resolution and less noises (Van Uitert and Bitter, 2007). Comparing to the original sub-voxel skelontonisation method, Rivulet is more robust to noises and gaps in poorly segmented foreground map, and also has a lower time complexity which is important since single neurons tend to have more complex arborisations than the tissues of medical interests, e.g. vessels and intestines. Rivulet iteratively traces a branch from the location with the farthest geodesic distance in the remaining foreground with RK4 gradient descent. This provides a higher chance to find the long branches in early iterations that are less risky to be affected by noises. The gradient descent stops when a set of stopping criteria are met or the soma location is reached. The whole tracing process terminates when a large proportion of the segmented foreground has been

(a) Original Image (b) Thresholding (c) Classification Result

Fig. 1 The example of the foreground images produced by thresholding and anisotropic classification. (a) is the original image contaminated with noises; (b) is the foreground binary map produced by the conventional thresholding; (c) is the foreground binary map produced by the a classifier trained with 6 Hessian-based features. It is noticeable that the classifier is helpful for removing most of the background noise points and preserving the tubular structures.

discovered by the traced branches. The risk of over-tracing is controlled by a newly proposed confidence score which measures the proportion of foreground voxels stepped by a traced branch. In the experiments, Rivulet is shown to be robust to both synthetic and real challenging images that posed different challenges for the compared algorithms. Both a Vaa3D (Long et al, 2012; Peng et al, 2014, 2010) neuron tracing plugin and a standalone Matlab GUI toolbox have been released with the proposed algorithm implemented.

2 Rivulet Tracing

2.1 Preprocessing

Nonlinear Anisotropic filter For images corrupted by strong noises, it is non-trivial to segment neuronal structure only based on an intensity threshold. We apply a nonlinear anisotropic filter $e^{-(|\nabla u|)^2} f(\lambda_1, \lambda_2, \lambda_3)$ to filter out the image noises, where $\lambda_1, \lambda_2, \lambda_3$ are the eigenvalues of the Hessian matrix $H(u)$ at position u. $f(x)$ is a vesselness filter defined with the eigenvalues as

$$f(x) = \begin{cases} \sum_{i=1}^{3} a_i \lambda_i k_i, & \text{if } \lambda_1 \approx 0, \lambda_1 \gg \lambda_2, \lambda_1 \gg \lambda_3 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $i = 1, 2, 3$ and a_i are predefined as $(a_1 = 0.5, a_2 = 0.5, a_3 = 25)$ and $k_i = e^{-\lambda_i^2 / \sum_i \lambda_i^2}$ (Yang et al, 2013). All voxels with values greater than 0 are marked as the foreground in the filtered image. Though the Rivulet tracing introduced in Section 2.2 is robust to most of the small noise points left in the segmentation, the filtering is mainly helpful for reducing the running time of Rivulet.

Neuron Segmentation For images with moderate noises, a manually chosen threshold would reasonably segment the neuron. For images with low signal to noise ratio (SNR), we apply a parametric classifier on 6 Hessian based measurements extracted from the preprocessed images to extract the curvilinear structures. The voxel classification is performed with Quadratic discriminant analysis (QDA) which the classifier is formed as

$$y = a_0 + a_1^T x + x^T a_2 x \tag{2}$$

where x is an input vector of length 6 for each voxel and a_i are the decision surface coefficients for different input orders. The features include the three Hessian eigenvalues ($\lambda_1 \leq \lambda_2 \leq \lambda_3$) extracted at the scale with the maximum anisotropic response, the Frangi vesselness score (Frangi et al, 1998), an modified Krissian vesselness score (Krissian et al, 2000), and the Fractional Anisotropy (FA) (Alexander et al, 2007) which was originally used as a diffusion indicator in diffusion tensor images (DTI). The vesselness score can be represented as

$$\mathcal{V}^{Frangi} = \begin{cases} 0 & \text{if } \lambda_2 > 0 \text{ or } \lambda_3 > 0, \\ (1 - exp(-\frac{\mathcal{R}_\mathcal{A}^2}{2\alpha^2}))exp(-\frac{\mathcal{R}_\mathcal{B}^2}{2\beta^2})(1 - exp(-\frac{\mathcal{S}^2}{2c^2})) \end{cases} \tag{3}$$

where $\mathcal{R}_\mathcal{A} = |\lambda_2|/|\lambda_3|$; $\mathcal{R}_\mathcal{B} = |\lambda_1|/\sqrt{|\lambda_2\lambda_3|}$; S is the Frobenius matrix norm of the Hessian $\mathcal{S} = \|\mathcal{H}\|_\mathcal{F} = \sqrt{\sum_{j \leq D} \lambda_j^2}$; α, β and c are thresholds controlling the sensitivity of $\mathcal{R}_\mathcal{A}$, $\mathcal{R}_\mathcal{B}$ and $\mathcal{S}$. The modified Krissian score used in this study is defined as

$$\mathcal{V}^{Krissian} = \begin{cases} 0 & \text{if } \lambda_1 + \lambda_2 + \lambda_3 \geq 0 \\ -\frac{\lambda_2}{\lambda_3}(\lambda_2 + \lambda_3) \end{cases} \tag{4}$$

The FA diffusion measure can be shown as

$$FA = \sqrt{\frac{1}{2}} \frac{\sqrt{(\lambda_3 - \lambda_2)^2 + (\lambda_2 - \lambda_1)^2 + (\lambda_3 - \lambda_1)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \tag{5}$$

The input vector x is standardised with zero-mean and rescaled within the range from 0 to 1.

In our experiment we also tried non-parametric classifiers such as support vector machine (SVM) as the classifier which slightly outperforms the QDA in segmentation accuracy. However the computational time for kernel computation scales up with the number of the training samples as well as the number of the voxels to be segmented. A parametric classification model can be helpful to constrain the running time. The segmented image is further processed with the linear level-set algorithm (Sethian, 1999) to eliminate the independent noise point. For the training set, the neurons were firstly manually traced as ground truths. Then the foreground voxels are sampled within the estimated radius of the fibres. The background training voxels are only sampled within a certain distance away from the fibre boundary (in our experiment within a distance of

3 voxels), since only such background voxels may affect the proposed tracing algorithm. A visual comparison between the thresholding and the classification results on a Olfactory Projection (OP) Fibre image provided by the Diadem challenge is shown in Fig. 1. The image shown in Fig. 1 was not included in the training set containing 7 other Olfactory Fibre Images. It is noticeable that the classifier is helpful for removing most of the background noise points and preserving the tubular structures. We evaluated the effectiveness of the classification based segmentation method with the leave-one-out (LOO) evaluation using 8 OP images. Both SVM and the quadratic classification could achieve average accuracies greater than 99.9% when considering all the voxels. With only the voxels within a distance of 3 voxels away from the thresholded region area (intensity of 10), the average accuracies of SVM and quadratic classification were close, respectively 93.5% and 93.3%. However the average time taken by SVM per image was 53.6 seconds per 3D imagestack; while the quadratic classification only took 0.03 seconds per imagestack on a Intel Core i7-6700 CPU.

2.2 Rivulet Tracing

Overview of Rivulet Tracing Fast-marching algorithm (Sethian, 1999) has been used in neuron tracing by growing the discovered region progressively from the soma location, because it is helpful to jump the gaps between discontinuous neuron segments (Xiao and Peng, 2013; Van Uitert and Bitter, 2007; Basu and Racoceanu, 2014). Since the tubular shapes of the neurons can be broken due to the poor image quality, we preprocess the segmented image with GWDT used in APP2 (Xiao and Peng, 2013) to obtain a distance transform map with bright voxels near the centrelines. Thus, the tracking procedure can be performed independently of the exact shapes of the neuron fibres. An enhanced fast-marching algorithm, multi-stencils fast-marching (Hassouna and Farag, 2007), is used to obtain a more accurate estimation of the geodesic distances with sub-voxel precision. The sub-voxel precision can be helpful to generate smooth neuronal curves when the image resolution is limited. Many of the previous algorithms progressively discover the neuron branches by growing the discovered region from the soma location to the outer region of the image. Since it can be unclear when the tracing procedure should stop if the tracing starts from the soma location to the outer unknown region, Rivulet traces each branch by back-tracking from the outer most region to the soma location. The tracing procedure mainly stops when it reaches the soma location or merges into a previously discovered branch. Each back-tracking procedure starts from the locations with the longest geodesic distances in the remaining undiscovered regions which are generally termini of neurons. The need of seed detection is thus eliminated. We propose a confidence score for each traced neuron segment, which indicates the proportion of the traced nodes generated

on the foreground voxels. To filter out the branches that may contain serious tracing errors caused by the noise points or gaps in the neuron structure, the branches are merged into the trunk only if they have high confidence scores. Rather than repeatedly computing the results of fast-marching (Van Uitert and Bitter, 2007), we only perform the multi-stencils fast-marching once and reuse the results by excluding the voxels covered by the traced branch from the choices of the start points for the following iterations. Thus, each voxel in the image is only traced once at most. This also enables measuring the proportion of the segmented foreground that has been explored by the discovered branches. When a high coverage rate is enforced, Rivulet would be capable of automatically discovering most of the major branches represented by the segmented foreground voxels.

Multistencils Fast Marching We apply the grey-scale weighted distance transform (GWDT) originally used in APP2 (Xiao and Peng, 2013) on the segmented foreground to obtain a weighted distance map $D(u)$. In $D(u)$, the foreground voxels far away from the boundaries of the segmented foreground map are brighter than the voxels close to the foreground boundaries. The background voxels in the distance map were assigned 10^{-10} instead of 0, allowing jumping between different foreground boundaries in the fast-marching method.

The fast-marching (FM) method has been used in neuron tracing algorithms and was proven to be robust to reconstruct the geometric information of curvilinear structures (Xiao and Peng, 2013; Van Uitert and Bitter, 2007; Basu and Racoceanu, 2014). FM tracks moving interfaces by solving the Eikonal equation (Adalsteinsson and Sethian, 1995; Tsitsiklis, 1995)

$$|\Delta T|F = 1, T(\Gamma_0) = 0 \tag{6}$$

where the arrival time T of the initial position of the front of the boundary Γ_0 is set to 0; the speed image $F = (D(u)/D_{max})^4$ in our study where D_{max} is the maximum value $D(u)$.

We use the multi-stencils fast marching method (MSFM) to obtain a more accurate solution to Eq. (6) in 3D Cartesian domain (Hassouna and Farag, 2007) by computing the solution at each grid point along several stencils that cover its entire neighbour points. Let U_1, U_2, U_3 be the directional derivatives along three unit vectors r_1, r_2 and r_3 in the grid system. α, β and γ are the rotating angles between stencils S and the unit vectors r_1, r_2 and r_3. T_1, T_2, T_3 are three adjacent neighbours reached by a certain orientation of the rotated stencils. Then,

$$U^T(RR^T)^{-1}U = \frac{1}{F^2(x)} \tag{7}$$

and

$$RR^T = \begin{pmatrix} 1 & cos\alpha & cos\gamma \\ cos\alpha & 1 & cos\beta \\ cos\gamma & cos\beta & 1 \end{pmatrix} \tag{8}$$

Here $RR^T = (RR^T)^{-1} = I$ when substituting $\alpha = \beta = \gamma = \frac{\pi}{2}$ into Eq. (8). If $T(x)$ is greater than the values of the three adjacent neighbours T_1, T_2, T_3 that participated in the solution, $T(x)$ is derived from the approximation of the directional derivatives as

$$\sum_{v=1}^{3} g_v(h)(a_v T^2(x) + b_v T(x) + c_v) = \frac{1}{F^2(x)} \tag{9}$$

where coefficients a_v, b_v and c_v are given as $[a_v\ b_v\ c_v] = [1\ -2T_v\ T_v^2]$; $g_v(h)$ is the orientation schemes of the stencils that cover the entire neighbour points defined in (Hassouna and Farag, 2007). Otherwise $T(x)$ is

$$min(T_v + \frac{\|x - x_v\|}{F(x)}),\ v = 1, 2, 3 \tag{10}$$

The point with D_{max} is chosen as the source point p_s for MSFM and is considered as the coordinate where soma locates. Practically it would not affect the tracing results even p_s is not positioned exactly at the soma. Also the choice of p_s can be replaced with the soma centre detected by other soma detection algorithms.

Gradient Back-Tracking The tracing starts from the furthest geodesic distance point $p_f^{(1)}$ with $max(T(x))$ in the segmented foreground (Van Uitert and Bitter, 2007). $p_f^{(i)}$, the start point of the i-th iteration, is considered as the globally optimum starting point since the curve $C(p_f^{(i)}, p_s)$ may be the longest branch of the target neuron remain undiscovered. From $p_f^{(i)}$ it tracks back to p_s by designating points along the gradient descent of $\Delta T(x)$ with the classical 4th order Runge-Kutta (RK4) method as

$$\begin{aligned} p_{n+1} &= p_n + \tfrac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4) \\ k_1 &= f(p_n) \\ k_2 &= f(p_n + \tfrac{h}{2}k_1) \\ k_3 &= f(p_n + \tfrac{h}{2}k_2) \\ k_4 &= f(p_n + hk_3) \end{aligned} \tag{11}$$

where p_n is the traced point at step n; $f(.)$ is the normalised 3D interpolation of $\Delta T(x)$; h is the step size and is practically set as 1. The back-tracking stops when one of the following stopping criteria is met: (1) more than G continuous points are traced without stepping on a foreground voxel; (2) the Euclidean distance $D(p_n, p_s)$ is less than the voxel size; (3) p_n is out of the image boundary; (4) the tracing has not moved away from the current position in 15 steps and (5) $T(p_n) \leq 0$. The gradient back-tracking procedure is illustrated in Fig. 2.

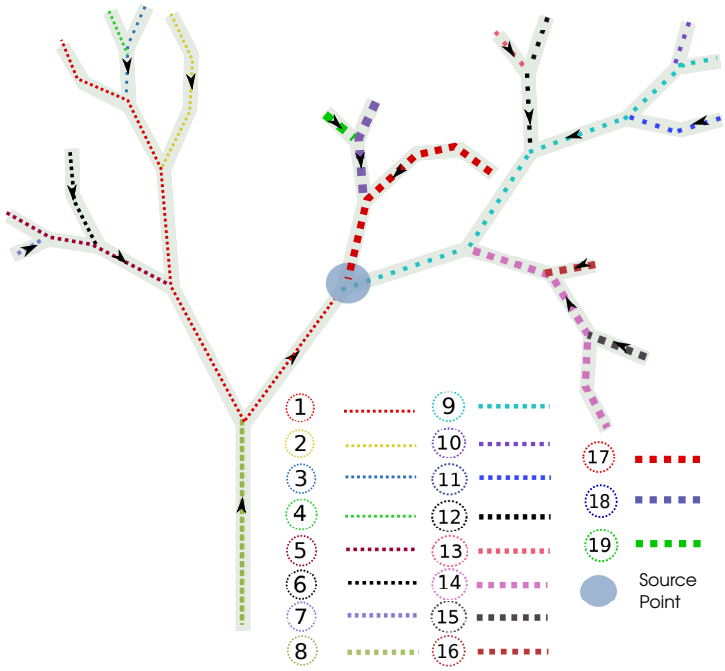
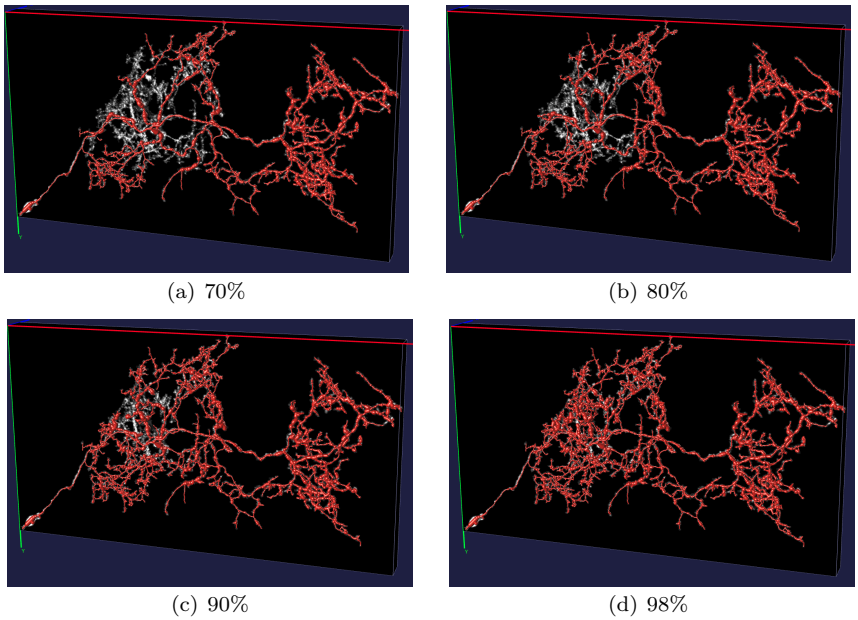


Fig. 2 The illustration of gradient back-tracking procedure. The branches with different colours represent the fibres reconstructed in different iterations. The number indicates the order in which the branch is expected to be discovered.

Branch Erasing When a branch is traced, the radius r_i at each p_i is estimated based on the foreground image with the sphere growing method (Peng et al, 2011). We do not use the radius obtained from the GWDT since GWDT can be sensitive to noises sometimes. Then for each p_n, spherical region u_n with size $\frac{4}{3}\pi r_i^3$ is defined. $T(U) = -1$ where $U = \{u_1 \cup u_2 \cup u_3 \cdots \cup u_n\}$. Due to the tracing stopping criterion (5) declared in last section, the back-tracking of new branches stops at the region covered by existing branches. After the branch is erased from $T(x)$, the traced branch is added to the trunk if it meets the criteria defined later and another back-tracking process starts over from the point with the furthest geodesic distance on the erased $T(x)$.

Voxel-Based Confidence Score Inspired by the confidence score developed in SmartTracing (Chen et al, 2015), we propose a simple voxel-based confidence score for Rivulet to select the branches to be added to the neuronal tree. Only the branches with high confidence score will be kept in the tracing result. For each branch $\{p_1, p_2, p_3, \ldots, p_n\}$ with length l, we define a percentile $C = \sum_{i=0}^{l} B_i/l$ to measure the overall confidence of the tracing process, where B_i is the voxel value $\{0, 1\}$ of the segmented image in which p_i stays. C represents the proportion of the endpoint decisions made based on the foreground. When the back-tracking starts from a far away noise point rather than the



(a) 70% (b) 80%

(c) 90% (d) 98%

Fig. 3 The reconstruction at stages of different proportions (%) of foreground image covered by the traced branches.

neuron body, C is expected to be low. Also a branch traced by filling many big gaps may be more risky to be added to the trunk rather than keeping it unconnected.

Branch Merging A new branch is dumped when (1) it has less than 8 nodes; (2) the confidence C is less than 50%; (3) the tracing was stopped because G steps were stepped on the background. The point closest to p_s is considered as the root node. After the first branch is added, the endpoints of the newly discovered branches p_n are connected to the trunk either the Euclidean distance $D(p_n, p_{min}) < R \times (r_n + 3)$ or $D(p_n, p_{min}) < R \times (r_{min} + 3)$, where p_{min} is the previously added node with the minimum euclidean distance from p_n and R (default 1.5) is a wiring threshold which can be chosen according to the image quality. If the connection criteria are not met, the branch stays unconnected, since branch connection with low confidence may result in even worse topology error in neuron tracing. Rivulet tracing stops only when a high proportion of the foreground $T(x)$ has been erased by -1 (default 98%). The coverage of the foreground area ensures that Rivulet is not likely to under-trace the neurons which makes Rivulet powerful to reveal very densely distributed arbors. The tracing with different coverage proportion is shown in Fig. 3. At the same time, it is noticeable that the tracing from noises is controlled by the confidence score C and the gap threshold G in back-tracking.

Time Complexity Comparing to a previous skelontonisation algorithm which originally proposed back-tracking for medical images with tubular structures (Van Uitert and Bitter, 2007), Rivulet is much faster which can be explained in the time complexity. In (Van Uitert and Bitter, 2007), the augmented fast marching is performed in each back-tracking process to find the point with the furthest geodesic distance on $T(x)$. This led to a complexity of $O(kn \log n)$ where k is the number of branches and n is the number of foreground voxels. The $O(n \log n)$ term comes with the fast-marching in each back-tracking iteration. Since the complexity scales linearly with the number of branches, it results in impractically long running time for neurons which are likely to have hundreds of branches. Because of the branch erasing and the branch merging components of Rivulet, fast-marching is only performed once before all the back-tracking iterations and the gradient of $T(x)$ is reused as well. Hence, Rivulet has a time complexity of $O(n \log n)$ which is the same as other fast-marching based tracing methods.

3 Software

The presented Rivulet algorithm has been implemented as a Vaa3D neuron tracing plug-in and a standalone Matlab GUI Toolbox named 'Rivulet'. The Vaa3D plugin was written in C++ thus faster and less memory consuming than the Matlab Toolbox. The segmentation and filtering can also be easily conducted with other Vaa3D plug-ins. The Rivulet Matlab Toolbox is capable of visualising segmented images and SWC files. It allows users to examine the preprocessed results at several stages, such as thresholding, classification and filtering, easing the parameter choosing for different datasets. It also supports I/O with Matlab workspace for flexibility and is compatible with multiple image formats, such as Vaa3D-Raw, TIF, NIFTI and MAT extensions.

4 Experimental Results

4.1 Materials

The data used in this study were acquired from the DIADEM challenge (Brown et al, 2011)[1] and the BigNeuron Project (Peng et al, 2015a,b)[2]. We compared the proposed algorithm with 5 other state-of-the-art algorithms, including Neuron Studio (Wearne et al, 2005), Snake (Wang et al, 2011), NeuTube (Zhao et al, 2011; Feng et al, 2015), MOST (Ming et al, 2013) and APP2 (Xiao and Peng, 2013). We used the Vaa3D implementations for all the compared algorithms, including Rivulet, for fair comparison. We tuned the parameters of each algorithm with exhaust search and validated the results with visual validation when there were parameters available in their corresponding Vaa3D

[1] http://diademchallenge.org/
[2] http://alleninstitute.org/bigneuron/about/

plugins. We firstly investigated Rivulet with visual inspections on synthetic images to evaluate its robustness against close fibres, gaps and noise points. 8 tracing results were obtained on the widely used Olfactory Projection (OP) Fibres dataset from the DIADEM challenge. For the Diadem datasets, we presented the results of Rivulet visually and the quantitative analysis across other widely used 5 methods. We also chose 3 challenging cases provided by the BigNeuron Project to compare Rivulet against other state-of-the-art neuron tracing methods. For the BigNeuron datasets, we presented the visual inspections for all the compared methods and the gold standard ground truth reconstructions provided by the BigNeuron neuron annotation workshops. The spatial distance (SD), substantial spatial distance (SSD) and the percentage of substantial nodes (SSD%) were computed for all quantitative analysis (Peng et al, 2011). SD is the average reciprocal minimal spatial distance of the nodes between a pair of reconstructions; SSD is the average spatial distance between nodes with spatial distances greater than 2 voxels, which the discrepancy is considered visible; SSD% is the percentage of SSD nodes in a pair of reconstructions.

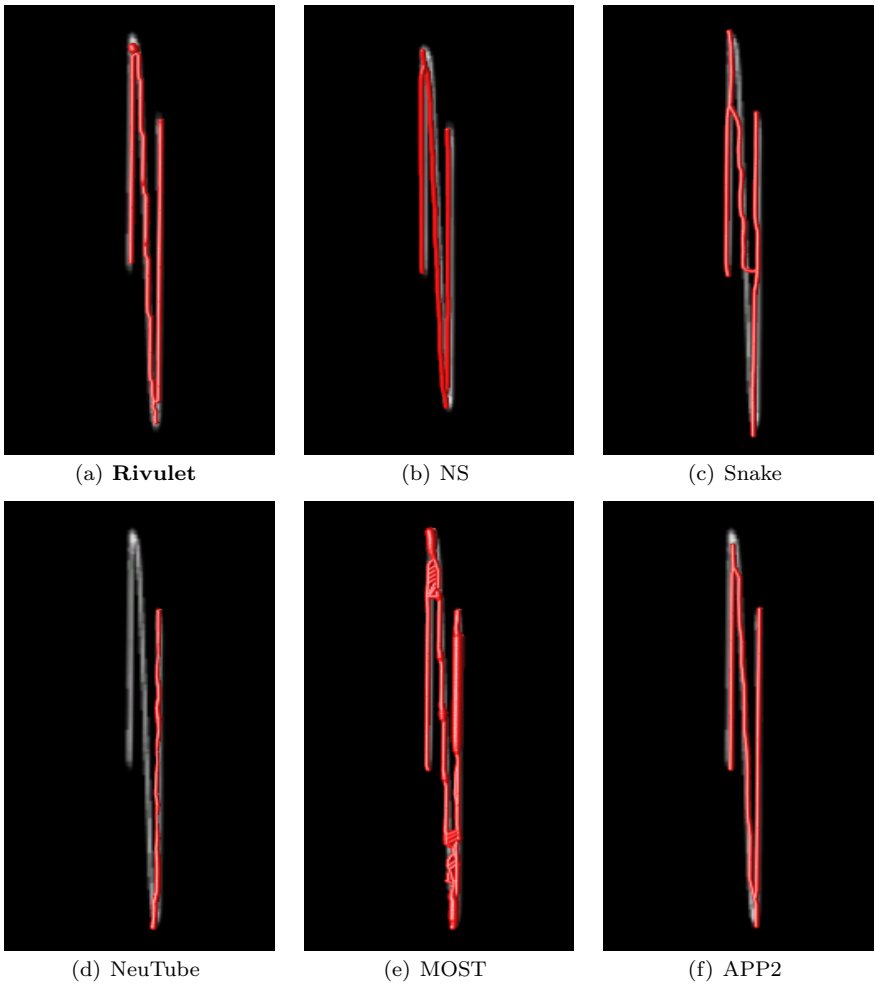
All the reconstructions were visualised with Vaa3D 3.060. The quantitative analysis was performed with the Vaa3D Neuron Toolbox 2.0.

4.2 Results on Synthetic Tubular Structures

For each synthetic image, a 2D grey scaled slice was manually made and replicated in the z axis to simulate the tubular radius. A Gaussian filter was applied to the synthetic volumes to smooth the corners to produce a tubular structure. All the following three synthetic images were 3D volumes with tubular structures. To evaluate the effectiveness of the proposed tracing component of Rivulet, all the segmentations used below were only performed with intensity thresholding.

Close Fibres In images with dense fibres, miswiring was often seen between two closed branches in previous methods. Also it is common that fast-marching based methods are likely to suffer from jumping wrong gaps between closed fibres (Mukherjee and Stepanyants, 2012). We performed the Rivulet tracing and other compared algorithms on a slim 'Z' shaped tube as shown in Fig. 4. The Rivulet tracing completely traced the 'Z' shaped tube and did not jump between the close branches near the two sharp corners.

Discontinuity and Noises A tube with small gaps was synthesised to simulate the discontinuous neuron segments which is shown in Fig. 5. We added salt and pepper noises, which can not be eliminated by thresholding, with density of 2% (the second row of Fig. 5) and 5% (the third row of Fig. 5) to the image to simulate the affection of noise points. Rivulet was able to fill the gaps with or without the noise points. Though a few extra redundant small branches were wrongly traced by Rivulet when the noises were dense (5%), Rivulet was able to preserve the overall shape of the tube without being much affected.

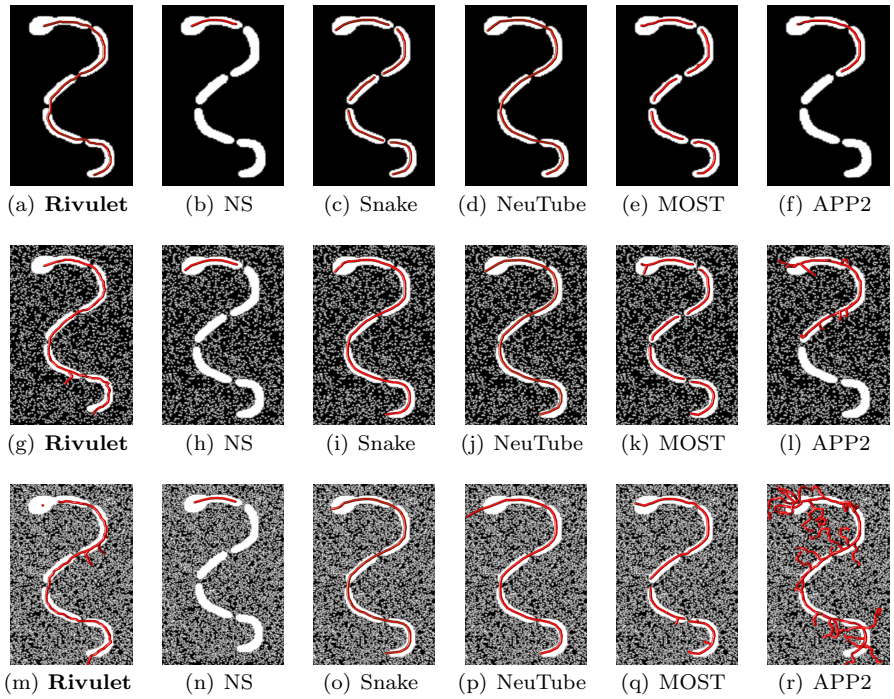


(a) **Rivulet** (b) NS (c) Snake

(d) NeuTube (e) MOST (f) APP2

Fig. 4 The reconstructions on a synthetic 3D 'Z' shaped tube with closed parallel fibres and sharp corners.

Tree Structure with Broken Tubular Shapes We synthesised a tree structure with dense branches to simulate the real neurons with densely distributed arbours. We deleted different proportions of voxels from the image, 40% (the second row of Fig. 6) and 70% (the third row of Fig. 6), to simulate broken shapes of the neuron arbours with non-smooth surfaces and small holes. Since Rivulet does not infer well-shaped tubular shapes of neuron arbours, it was able to preserve the overall morphology of the tree structure even when the tubular structure was relatively broken. It also did not over-trace the tree structure by generating non-existing branches.



(a) **Rivulet** (b) NS (c) Snake (d) NeuTube (e) MOST (f) APP2

(g) **Rivulet** (h) NS (i) Snake (j) NeuTube (k) MOST (l) APP2

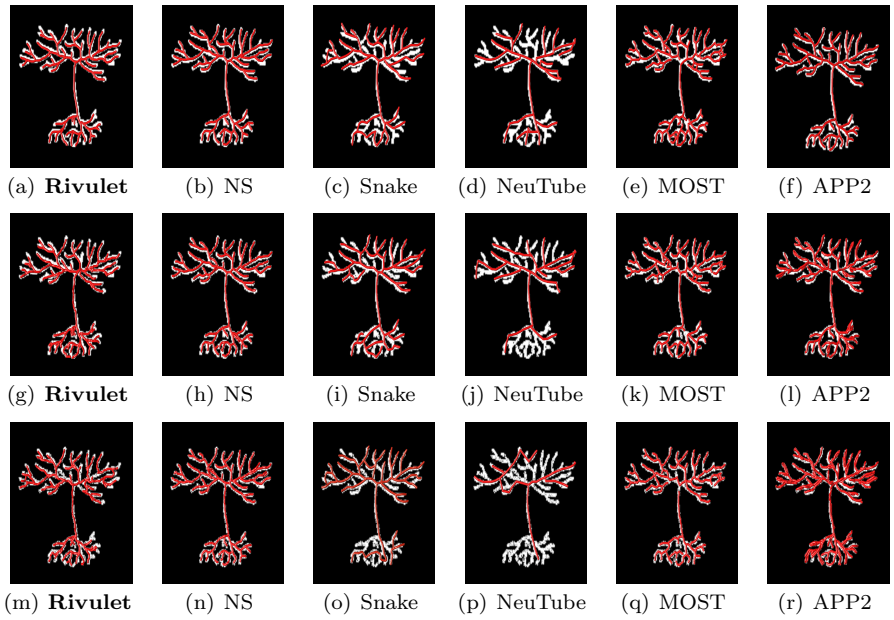
(m) **Rivulet** (n) NS (o) Snake (p) NeuTube (q) MOST (r) APP2

Fig. 5 The reconstructions of a synthetic broken tube with progressively added salt and pepper noises. The first row is the original tubular image without noises added; The second and third row are images contaminated with salt and pepper noises of density 2% and %5 respectively.

4.3 Tracing Olfactory Projection Fibres from Diadem Challenge

The dataset of Olfactory Project (OP) Fibres is one of the 6 open-access datasets provided by DIADEM challenge (This dataset is publicly available at `http://diademchallenge.org/olfactory_projection_fibers_readme.html`). The OP dataset was widely used in previous studies to evaluate the neuron tracing results. This dataset contains 9 axons of drosophila olfactory bulb neurons acquired with 2-channel confocal microscopy. In Fig. 7, we present 8 reconstructions of OP dataset which are shown together with the manually reconstructed ground truth. Rivulet was able to successfully output results almost identical to the ground truth. The OP2 image was intentionally excluded from the evaluation because it contains many irrelevant structures.

The quantitative evaluation of the 8 OP images is shown in Fig. 8. Rivulet achieved low SD and SSD in most cases. The SSD% score was slightly higher since the sub-voxel tracing produced more nodes than the other compared methods.

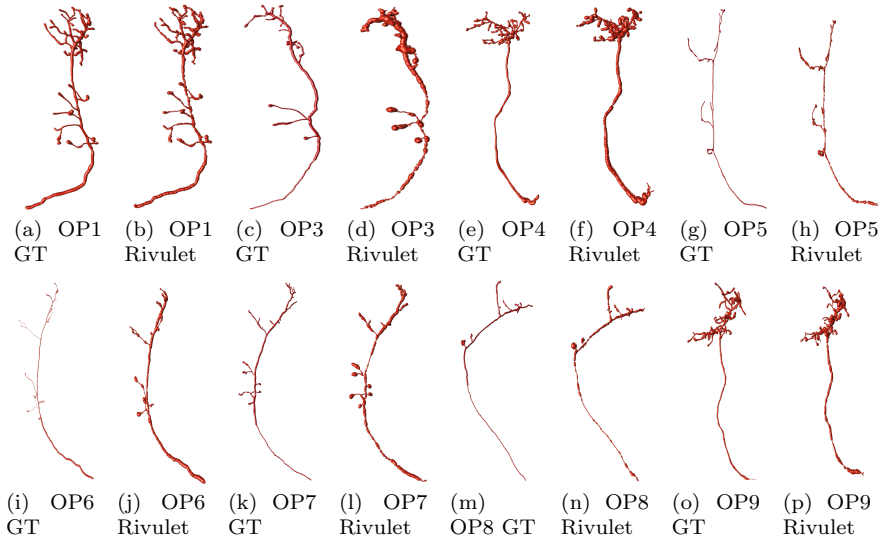


(a) **Rivulet** (b) NS (c) Snake (d) NeuTube (e) MOST (f) APP2

(g) **Rivulet** (h) NS (i) Snake (j) NeuTube (k) MOST (l) APP2

(m) **Rivulet** (n) NS (o) Snake (p) NeuTube (q) MOST (r) APP2

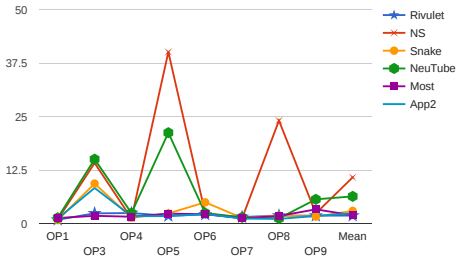
Fig. 6 The comparison between the state-of-the-art tracing algorithms on a synthesised tree image with densely distributed branches. The first row is the reconstructions based on the original image; the second and the third row are the reconstructions based on the image with 40% and 70% voxels deleted.



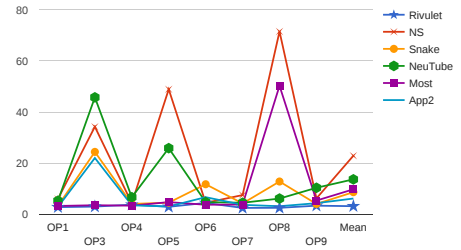
(a) OP1 GT (b) OP1 Rivulet (c) OP3 GT (d) OP3 Rivulet (e) OP4 GT (f) OP4 Rivulet (g) OP5 GT (h) OP5 Rivulet

(i) OP6 GT (j) OP6 Rivulet (k) OP7 GT (l) OP7 Rivulet (m) OP8 GT (n) OP8 Rivulet (o) OP9 GT (p) OP9 Rivulet

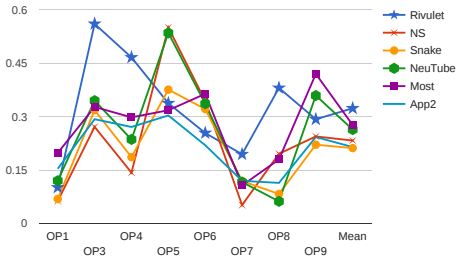
Fig. 7 The reconstructions of the DIADEM-OP dataset presented with the manually traced ground truth (GT). The OP2 image was intentionally excluded from the evaluation because it contained many irrelevant structures.



(a) Spatial distance (SD)



(b) Substantial Spatial distance (SSD)



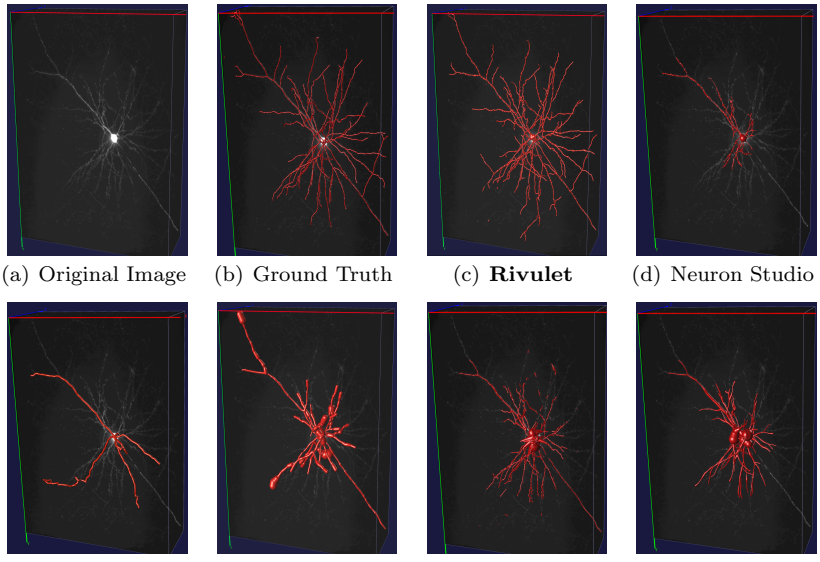
(c) The percentage of SSD nodes (SSD%)

Fig. 8 The quantitative analysis on 8 images from the OP dataset.

4.4 Tracing the Images from BigNeuron Project

To compare Rivulet with other tracing methods on modelling animals, we used a subset of the data provided by the BigNeuron Project including neurons of fruit fly and mouse. We selected 3 very challenging images to compare Rivulet with other state-of-the-art algorithms. The first image shown in Fig.9 is an

apical neuron of mouse contributed by Tufts University. This image was corrupted by strong and dense background noises with high intensities, thus many noise points remained in the segmented foreground. The other algorithms tend to under-trace the neuron since the neuron fibres were highly blended with the background. The algorithms that march from the soma location to the outer boundaries are likely to generate non-existing cones caused by the noise points. Rivulet reconstructed most of the fibres successfully mainly because of its back-tracking procedure and the capability of filling gaps.



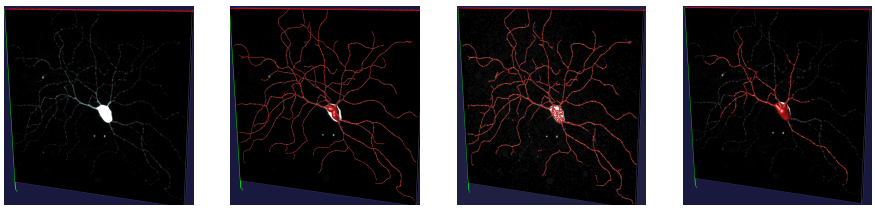
(a) Original Image (b) Ground Truth (c) **Rivulet** (d) Neuron Studio

(e) Snake (f) NeuTube (g) MOST (h) APP2

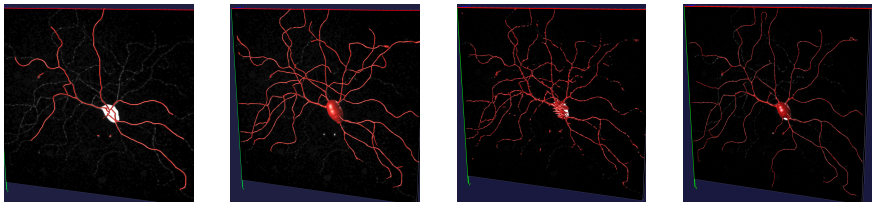
Fig. 9 The comparison between the state-of-the-art tracing algorithms on a noisy image of mouse apical neuron. The background noises are highly indistinguishable to the neuron related signals.

The neurons shown in Fig. 10 and Fig. 11 are two mouse retinal ganglion cell images contributed by University of Washington. Due to the extremely thin fibres, there are many small gaps in the foreground which can not be trivially fixed by preprocessing techniques. The compared methods normally output under-reconstructed results since the discontinuity caused early stops of tracing. The methods using the segmented foreground as boundary walls are more likely to generate short discontinuous segments in such images. Because of the coverage proportion embedded in Rivulet, it did not under-trace the broken fibres. Even when the tracing procedure stops wrongly in the mid-way, the rest of the same fibre would be traced in another back-tracking iteration. It is observable that Rivulet successfully reconstructed the most fibres among the compared algorithms. The traced fibres were consistent with the image voxels and the gold standard ground truth.

The quantitative analyses are shown in Fig. 12. Rivulet outperformed other compared methods regarding to the SD and SSD in all the 3 images. It is noticeable that though Rivulet generated more nodes than the other methods, it achieved comparable low SSD% scores.

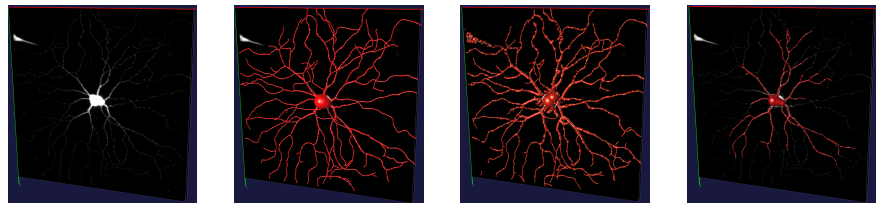


(a) Original Image (b) Ground Truth (c) **Rivulet** (d) NeuronStudio

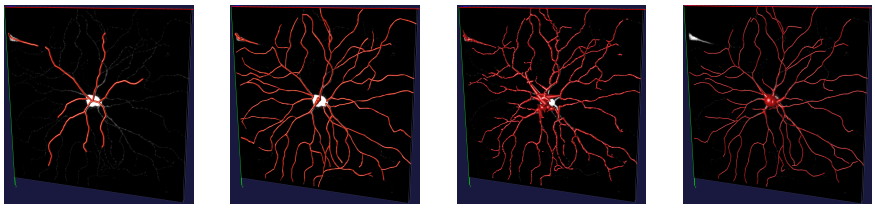


(e) Snake (f) NeuTube (g) MOST (h) APP2

Fig. 10 The comparison between the state-of-the-art tracing algorithms on a mouse retinal ganglion cell image. The fibres were corrupted with many small gaps that are likely to cause under-reconstruction.



(a) Original Image (b) Ground Truth (c) **Rivulet** (d) NeuronStudio



(e) Snake (f) NeuTube (g) MOST (h) APP2

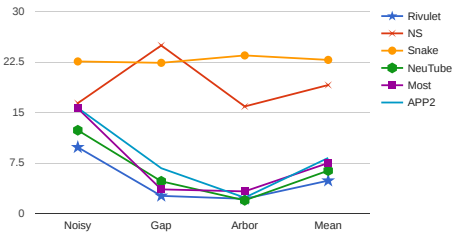
Fig. 11 The comparison between the state-of-the-art algorithms on a mouse retinal ganglion cell image with complex arborisation. Due to the dark foreground were blended with the background noises, some fibres are easily missed by the tracing algorithms.

The image shown in Fig. 13 is a fruit fly neuron with densely distributed fibres with complex arborisation that may be barely traceable by hand. The shapes of the thin arbours do not meet the tubular structure assumed by many previous algorithms. Rivulet was able to trace such broken arbours. Since Rivulet only stops tracing when a certain coverage of the foreground is achieved, thus, it is powerful to reveal the small-scaled meaningful details from the complex fruit fly neurons shown in Fig. 14.

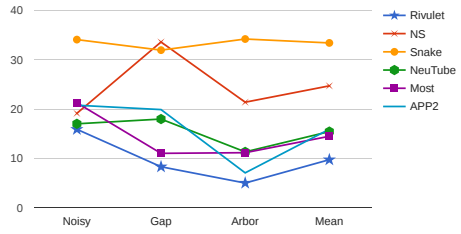
5 Conclusions and Discussions

A combination of adaptive preprocessing strategies was applied in this study which is important for accurate neuron reconstructions. For images with a moderate SNR, we only use the anisotropic filter and a foreground threshold to segment the neurons from the background. The presented voxel classification framework is only used when the foreground threshold is not sufficient to provide a reasonable segmentation. The image quality may vary significantly between datasets provided by different sites. Thus, it is non-trivial to produce ideal segmentations for neuron reconstruction. We designed the Rivulet tracing to reconstruct meaningful neuronal models from challenging images. We do not infer assumptions on the smoothness and the continuity of the neuron fibres to make Rivulet algorithm more robust to noisy images. Combining the fast-marching and gradient descent, Rivulet is able to trace discontinuous areas. Since the back-tracking iterations attempt to find the longest branches left in the image, Rivulet is not likely to miss the major branches before finishing. The coverage threshold (default 98%) avoids under-tracing although some branches are not fully traced in a single back-tracking iteration. Rivulet is also robust to over-tracing by considering the gaps between the noise points and the traced trunk. The traced noise points will be dumped after one back-tracking iteration. Since the multi-stencils fast-marching embedded in Rivulet enables the sub-voxel precision for tracing, the curves traced by Rivulet do not require post-smoothing.

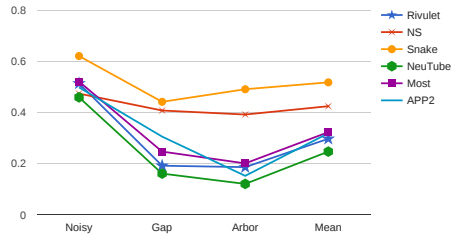
Aside from the preprocessing phase, there are only 2 parameters, the gap threshold G and the wiring threshold R, to be tuned to adjust different datasets. G is determined by the largest possible gap area to be allowed to jump. A larger gap threshold is preferred when the neuron image has large discontinuity and meanwhile has few noise points. Practically, only small gap jumping is preferred, because it has a higher chance to produce topological error when a large gap is jumped. In our experiments, G ranges from 4 to 20 depending on the levels of discontinuity of the neurons in microscopic images. R is determined by the largest possible distance between broken segments. R is expected to range from 1 to 3 practically and it is set 1.5 as default. Large R values are needed when the neuron structures are not in ideal tubular shapes or the foreground signals are weak near the bifurcations. To segment the images, an intensity threshold is needed which can be visually tuned via the Rivulet Matlab Toolbox. The intensity threshold will also be used to define an area,



(a) Spatial distance (SD)



(b) Substantial spatial distance (SSD)

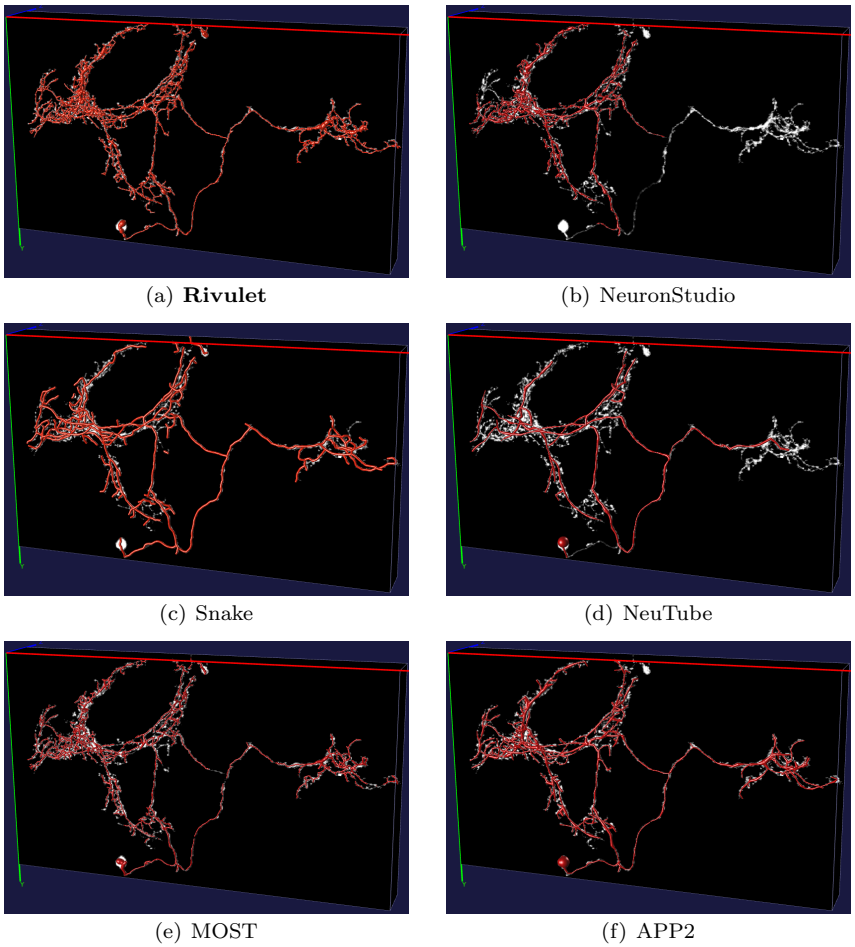


(c) The percentage of SSD nodes (SSD%)

Fig. 12 The quantitative analysis of the results between different algorithms provided by the BigNeuron project.

slightly larger than the initial segmentation, to speed up the voxel classification. To extract the Hessian eigenvalues, only the maximum σ is needed for setting. Large σ values may generate better filtering effects and require more computational resources.

Rivulet shares some similar patterns with the previous state-of-the-art algorithms regarding the algorithm design. It uses the vesselness based filtering and segmentation which are helpful for eliminating irrelevant structures before



(a) **Rivulet** (b) NeuronStudio

(c) Snake (d) NeuTube

(e) MOST (f) APP2

Fig. 13 The comparison between the state-of-the-art tracing algorithms on a fruit fly neuron with densely distributed fibres. The small and thin fibres do not meet the tubular shape assumption embedded in many previous tracing algorithms.

the tracing phase (Wang et al, 2011; Yang et al, 2013). Rivulet preserves the long branches as performed in the pruning strategies of APP2 (Xiao and Peng, 2013). Branch merging and branch erasing can be also found in Snake (Wang et al, 2011). Resembling MOST (Ming et al, 2013), Rivulet does not grow from the soma area, instead it traces the neuron from different locations in order to discover most of the major structures. The strategy of firstly processing the outer most dendrites can also be found in an early 2D neuron reconstruction framework proposed in Cesar-Jr. and Costa (1999), which analysed the outer most and smallest dendrites first. The Rivulet algorithm instead tracks the long branches in early iterations from the outer termini, since the accuracy of long bright branches is often with higher priority than the small branches

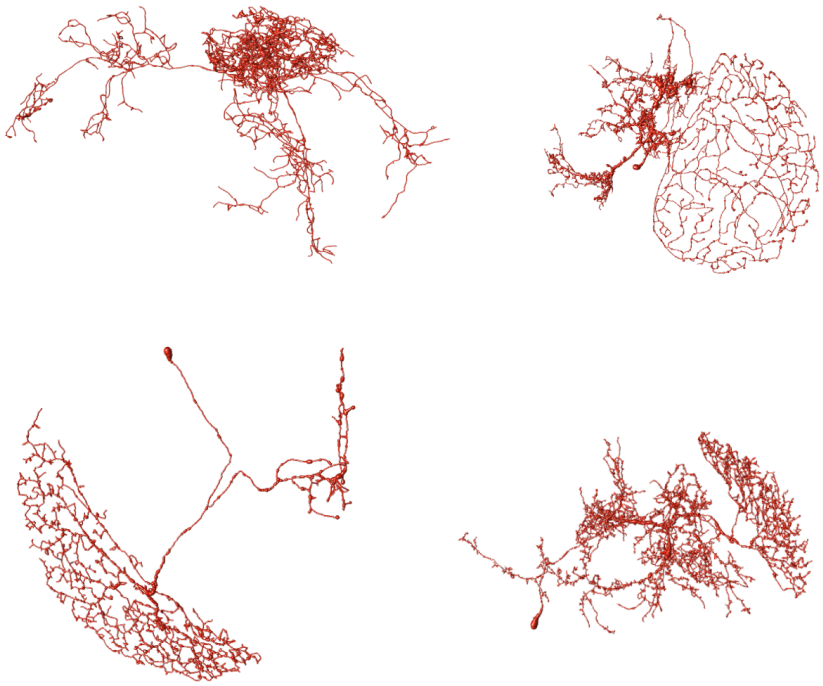


Fig. 14 Some example reconstructions of fruit fly neurons with complex arborisation obtained automatically using Rivulet.

in noisy images. Also we avoided the dependence of detecting neuronal tips and bifurcation points which can be challenging in 3D images. Comparing to a medical image skelontonisation algorithm also with back-tracking embedded (Van Uitert and Bitter, 2007), the branch erasing and branch merging of Rivulet reduced the time complexity from $O(kn \log n)$ to $O(n \log n)$. Since the image quality of 3D confocal microscopic images raised challenges, such as discontinuity, strong noises and ambiguous topology, we improved the original sub-voxel back-tracking with several additional components for constructing meaningful and precise neuronal models.

Though we have shown in experiments that Rivulet successfully reconstructed neurons with different types, the tracing accuracy of Rivulet might be further enhanced by embedding newly proposed machine learning based modules such as Smart Tracing (Chen et al, 2015) and Topology Rewiring (Zhou et al, 2014). Also when the resolution of the images is limited, a 2D framework proposed by Leandro et al (2009) could be extended to 3D to further distinguish bifurcation, crossing and kissing of intersected fibres in a blurry area.

6 Information sharing statement

The source code of Rivulet Matlab Toolbox is openly available at `https://github.com/lsqshr/Rivulet-Neuron-Tracing-Toolbox`. The source code of Rivulet Vaa3D Plugin is openly distributed along with the Vaa3D code repository together with other plug-ins ported in the BigNeuron Hackathon.

References

Adalsteinsson D, Sethian JA (1995) A fast level set method for propagating interfaces. Journal of Computational Physics 118(2):269–277

Alexander AL, Lee JE, Lazar M, Field AS (2007) Diffusion tensor imaging of the brain. Neurotherapeutics 4(3):316–329

Basu S, Racoceanu D (2014) Reconstructing neuronal morphology from microscopy stacks using fast marching. In: Image Processing (ICIP), 2014 IEEE International Conference on, pp 3597–3601

Brown KM, Barrionuevo G, Canty AJ, De Paola V, Hirsch JA, Jefferis GS, 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(2-3):143–157

Cesar-Jr R, Costa L (1999) Computer-vision-based extraction of neural dendrograms. Journal of Neuroscience Methods 93(2):121 – 131

Chen H, Xiao H, Liu T, Peng H (2015) Smarttracing: self-learning-based neuron reconstruction. Brain Informatics 2(3):135–144

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(1):ENEURO–0049

Frangi AF, Niessen WJ, Vincken KL, Viergever MA (1998) Multiscale vessel enhancement filtering. In: Medical Image Computing and Computer-Assisted Interventation (MICCAI), Springer, pp 130–137

González G, Türetken E, Fleuret F, Fua P (2010) Delineating trees in noisy 2D images and 3D image-stacks. In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, IEEE, pp 2799–2806

Hassouna MS, Farag AA (2007) Multistencils fast marching methods: A highly accurate solution to the eikonal equation on cartesian domains. Pattern Analysis and Machine Intelligence, IEEE Transactions on 29(9):1563–1574

Jameson A, Schmidt W, Turkel E (1981) Numerical solutions of the euler equations by finite volume methods using runge-kutta time-stepping schemes. AIAA paper 1259:1981

Krissian K, Malandain G, Ayache N, Vaillant R, Trousset Y (2000) Model-based detection of tubular structures in 3D images. Computer vision and image understanding 80(2):130–171

Leandro J, Cesar-Jr R, Costa LF (2009) Automatic contour extraction from 2D neuron images. Journal of Neuroscience Methods 177(2):497 – 509

Long F, Zhou J, Peng H (2012) Visualization and analysis of 3D microscopic images. PLoS Comput Biol 8(6):e1002,519–e1002,519

Ming X, Li A, Wu J, Yan C, Ding W, Gong H, Zeng S, Liu Q (2013) Rapid reconstruction of 3D neuronal morphology from light microscopy images with augmented rayburst sampling. PLoS ONE 8(12):e84,557

Mukherjee A, Stepanyants A (2012) Automated reconstruction of neural trees using front re-initialization. In: SPIE Medical Imaging, International Society for Optics and Photonics, pp 83,141I–83,141I

Mukherjee S, Condron B, Acton ST (2015) Tubularity flow field - a technique for automatic neuron segmentation. IEEE Transactions on Image Processing 24(1):374–389

Parekh R, Ascoli GA (2013) Neuronal morphology goes digital: a research hub for cellular and system neuroscience. Neuron 77(6):1017–1038

Pawley JB (2006) Handbook Of Biological Confocal Microscopy, Springer US, Boston, MA, chap Fundamental Limits in Confocal Microscopy, pp 20–42

Peng H, Ruan Z, Long F, Simpson JH, Myers EW (2010) V3D enables real-time 3D visualization and quantitative analysis of large-scale biological image data sets. Nature Biotechnology 28(4):348–353

Peng H, Long F, Myers G (2011) Automatic 3D neuron tracing using all-path pruning. Bioinformatics 27(13):i239–i247

Peng H, Bria A, Zhou Z, Iannello G, Long F (2014) Extensible visualization and analysis for multidimensional images using Vaa3D. Nature protocols 9(1):193–208

Peng H, Hawrylycz M, Roskams J, Hill S, Spruston N, Meijering E, Ascoli G (2015a) BigNeuron: Large-scale 3D neuron reconstruction from optical microscopy images. Neuron 87(2):252 – 256

Peng H, Meijering E, Ascoli G (2015b) From DIADEM to BigNeuron. Neuroinformatics 13(3):259–260

Santamaría-Pang A, Hernandez-Herrera P, Papadakis M, Saggau P, Kakadiaris IA (2015) Automatic morphological reconstruction of neurons from multiphoton and confocal microscopy images using 3D tubular models. Neuroinformatics 13(3):297–320

Sethian JA (1999) Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science, vol 3. Cambridge University Press

Tsitsiklis JN (1995) Efficient algorithms for globally optimal trajectories. Automatic Control, IEEE Transactions on 40(9):1528–1538

Türetken E, González G, Blum C, Fua P (2011) Automated reconstruction of dendritic and axonal trees by global optimization with geometric priors. Neuroinformatics 9(2-3):279–302

Van Uitert R, Bitter I (2007) Subvoxel precise skeletons of volumetric data based on fast marching methods. Medical Physics 34(2):627–638

Wang Y, Narayanaswamy A, Tsai CL, Roysam B (2011) A broadly applicable 3-D neuron tracing method based on open-curve snake. Neuroinformatics 9(2-3):193–217

Wearne S, Rodriguez A, Ehlenberger D, Rocher A, Henderson S, Hof P (2005) New techniques for imaging, digitization and analysis of three-dimensional neural morphology on multiple scales. Neuroscience 136(3):661 – 680

Xiao H, Peng H (2013) APP2: automatic tracing of 3D neuron morphology based on hierarchical pruning of a gray-weighted image distance-tree. Bioinformatics 29(11):1448–1454

Yang J, Gonzalez-Bellido PT, Peng H (2013) A distance-field based automatic neuron tracing method. BMC Bioinformatics 14(1):93

Yuan X, Trachtenberg JT, Potter SM, Roysam B (2009) MDL constrained 3D grayscale skeletonization algorithm for automated extraction of dendrites and spines from fluorescence confocal images. Neuroinformatics 7(4):213–232

Zhang D, Liu S, Liu S, Feng D, Peng H, Cai W (2016) Reconstruction of 3D neuron morphology using Rivulet back-tracking. In: The IEEE International Symposium on Biomedical Imaging: From Nano to Macro, IEEE

Zhao T, Xie J, Amat F, Clack N, Ahammad P, Peng H, Long F, Myers E (2011) Automated reconstruction of neuronal morphology based on local geometrical and global structural models. Neuroinformatics 9(2-3):247–261

Zhou Z, Sorensen S, Zeng H, Hawrylycz M, Peng H (2014) Adaptive image enhancement for tracing 3D morphologies of neurons and brain vasculatures. Neuroinformatics 13(2):153–166