NETracer: A Topology-Aware Iterative Tracing Approach for Tubular Structure Extraction

Supplementary Material

1. Datasets Details

We evaluate our model on both 2D and 3D datasets. Here is the details of the datasets.

Massachusetts Roads Dataset: The Massachusetts Roads Dataset is a widely used public dataset of overhead imagery for road network extraction in computer vision research. It consists of 1116 high-resolution 8-bit RGB aerial images, each measuring 1500×1500 pixels, and accompanying road segmentation annotations. To facilitate training and testing, we select 200 images as the training set and 13 images as the testing set.

DRIVE: The DRIVE dataset is widely used public datasets for retinal vessel image segmentation. It contains 40 retinal images with segmentation annotations, captured using 8-bit RGB color at a resolution of 565×584 pixels. Of these images, 33 show no signs of diabetic retinopathy, while 7 exhibit mild early diabetic retinopathy. The dataset is divided into a training set and a test set, each with 20 images.

CHASE_DB1: The CHASE_DB1 dataset includes 28 8-bit RGB retina images, each with a resolution of 999×960 pixels, obtained from the left and right eyes of 14 school children. Two independent experts manually annotated each image for segmentation, with the first annotation usually being considered the ground truth. The dataset is also divided into a training set and a test set, with the 20 images used for training and the 8 used for testing.

fMOST-VTA: This dataset was extracted from the Ventral Tegmental Area (VTA) of a mouse brain imaged by fMOST with the size of $9216 \times 14336 \times 24576$ voxels and has a spatial resolution of $0.325 \times 0.325 \times 1\mu m^3$. We preprocessed it into the 16-bit gray image stacks with the size of $512 \times 512 \times 512$ voxels. With the aid of neuTu GUI software [1], a semi-automatic neuron reconstruction software, 28 image stacks were manually annotated by experts to serve as the ground truth. Among them, 20 were used for training and 8 for testing.

DIEDAM-CCF: The Cerebellar Climbing Fibers (CCF) dataset is one of six datasets used in the DIADEM competition. It consists of three 8-bit RGB image stacks of single cerebellar climbing fibers from Long-Evans rats, named CF_1, CF_2, and CF_3, each with a corresponding gold standard reconstruction, and all neuronal structures in the dataset exhibit a tree-like distribution. CF_2 and CF_3 are used for training the model, and CF_1 is used for testing.

Algorithm 1 Position Error Metric

Input: Gold graph: $\mathbf{G}_g$, Test graph $\mathbf{G}_t$, $\varepsilon \in \mathbb{R}^+$
Output: PE score: PE

Initialize $N \leftarrow 0$, $S_{\text{PE}} \leftarrow 0$
for $\mathbf{n}_t$ in $\mathbf{G}_t$ **do**
  // Identify the closest nodes on the gold graph and calculate the Euclidean distance.
  $d_t \leftarrow \min_{\mathbf{n}_g \in \mathbf{G}_g} ||\mathbf{n}_t - \mathbf{n}_g||$
  // Check whether met the matching criterion.
  if $d_t < \varepsilon$ **then**
    $S_{\text{PE}} \leftarrow S_{\text{PE}} + d_t$
    $N \leftarrow N + 1$
  end if
end for
PE $\leftarrow S_{\text{PE}}/N$
return PE

2. Iterative Tracing Metric

Position Error Metric: The Position Error (PE) measures the Euclidean distances between the positions of matching nodes in the test and ground truth branches. For each node $\mathbf{n}_t$ in the test graph, we identify the matching nodes $\mathbf{n}_g^*$ on the ground truth graph. The criterion for matching is when their Euclidean distance is less than ε (set to 2). The average of these Euclidean distances between matched nodes is calculated as the Position Error. The detail of the Position Error is laid out in Algorithm 1.

Jump Error Metric: The Jump Error (JE) aims to evaluate the correctness of the topology in the tracing process. For each node $\mathbf{n}_t$ in the test graph and its corresponding parent $\rho(\mathbf{n}_t)$, we identify the matching nodes $\mathbf{n}_g^*$ and $\mathbf{n}_{g_p}^*$ on the ground truth graph. The criterion for matching is when their Euclidean distance is less than ε (set to 2). If the graph distance between these two matched nodes is greater than their straight-line distance, a jump error is considered to have occurred. The detail of the Jump Error is laid out in Algorithm 2.

3. Ablation Study

Seed Generation: The seed generation module in our model is pluggable and can be adapted to different methods, such as manually selected nodes or automatic seed generator. In this work, we employed the simplest thresholding method on the predicted centerline probability map to generate the seed nodes. Any pixel or voxel on the predicted

Algorithm 2 Jump Error Metric

Input: Gold graph: $\mathbf{G}_g$, Test graph $\mathbf{G}_t$, $\varepsilon \in \mathbb{R}^+$
Output: JE score: N

Initialize $N \leftarrow 0$
for $\mathbf{n}_t$ in $\mathbf{G}_t$ **do**
  $d_t \leftarrow \min_{\mathbf{n}_g \in \mathbf{G}_g} ||\mathbf{n}_t - \mathbf{n}_g||^2$
  if $d_t < \varepsilon$ **then**
    $\mathbf{n}_g^* \leftarrow \arg\min_{\mathbf{n}_g \in \mathbf{G}_g} ||\mathbf{n}_t - \mathbf{n}_g||$
    $\mathbf{n}_{g_p}^* \leftarrow \arg\min_{\mathbf{n}_g \in \mathbf{G}_g} ||\rho(\mathbf{n}_t) - \mathbf{n}_g||$
    $d_1 \leftarrow \text{GraphDistance}(\mathbf{G}_g, \mathbf{n}_g^*, \mathbf{n}_{g_p}^*)$
    $d_2 \leftarrow ||\mathbf{n}_g^* - \mathbf{n}_{g_p}^*||$
    if $d_1 > \alpha \cdot d_2$ **then**
      $N \leftarrow N + 1$
    end if
  end if
end for
return N

Thresholds	SSD-F1	L-F1	PE ↓	JE ↓	ABL ↑	Seeds
0.99	0.907	0.899	0.698	2.0	225.3	206
0.95	0.919	0.913	0.698	2.3	222.7	399
0.9	0.919	0.913	0.693	2.3	219.3	909
0.7	0.918	0.912	0.704	2.3	206.9	1361

Table 1. Different thresholds for seed generation.

Decay Rate	SSD-F1	L-F1	PE ↓	JE ↓	ABL ↑
0.75	0.920	0.913	0.695	2.4	222.1
0.50	0.919	0.913	0.693	2.3	219.3
0.25	0.919	0.912	0.708	2.3	217.9

Table 2. Different decay rates for $\mathbf{Q}$ set selection.

centerline with an intensity greater than the threshold 0.9 is selected as a seed node. We provide an analysis of the effect of different thresholds in the following Table 1. The thresholds are set as 0.99, 0.95, 0.9, and 0.7. In the range of the threshold value between 0.7 and 0.95, the tracing results are of similar quality, indicating that our method is *not dependent* on the particular threshold.

Q Sets Selection: For the selection of the $\mathbf{Q}$ sets in GDS strategy, we iteratively mask the local peaks in the predicted future node probability map. When the maximum value in the masked probability map decays to half of its original maximum (0.5), it indicates that most of the likely peak nodes have been selected. Here, we present an analysis of the impact of the decay rate in the following table, as shown in Table 2. The decay rates are set as 0.75, 0.5, and 0.25. It can be observed that variations in the decay rate have a minimal impact on the final results, indicating that this module is *not sensitive* to this parameter.

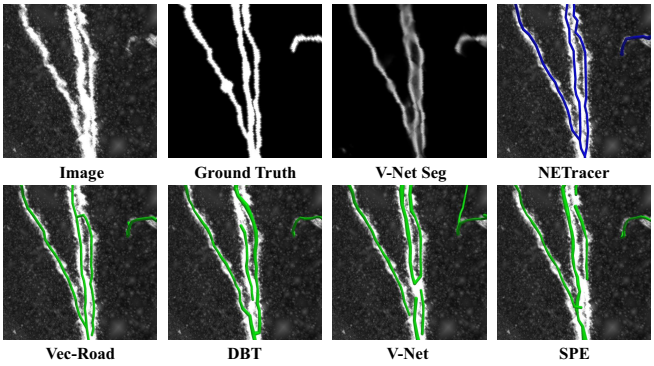

Figure 1. Comparison of various methods on neuron datasets.

4. Visual Comparison with SOTA Methods

We also give more visualization results of our method, as shown in Fig. 1 2 3. Various segmentation-based and tracing-based method are choose as the comparison methods. When confronted with low confidence in centerline segmentation outcomes, our method consistently preserves the accuracy of the topology.

5. Limitations

Despite the promising performance of our NETracer model in extracting tubular structures, several limitations need to be addressed. Firstly, our NETracer model mainly focuses on the downstream modules and the topology loss, and does not include network modules dedicated to enhancing the learning ability of the model, such as the attention and deformable modules. Secondly, as an iterative tracing method, the geodesic distance-based search strategy slightly increases the computational effort of the model, although it is effective in improving the topological accuracy of the results. This could become a problem in future applications for large-scale 3D brain neuron image reconstruction. Thus, our future work will focus on developing adaptive connectivity strategies based on graph models, allowing for more flexible and context-aware connections between nodes.

References

[1] Ting Zhao, Jun Xie, Fernando Amat, Nathan Clack, Parvez Ahammad, Hanchuan Peng, Fuhui Long, and Eugene Myers. Automated reconstruction of neuronal morphology based on local geometrical and global structural models. *Neuroinformatics*, 9:247–261, 2011. 1

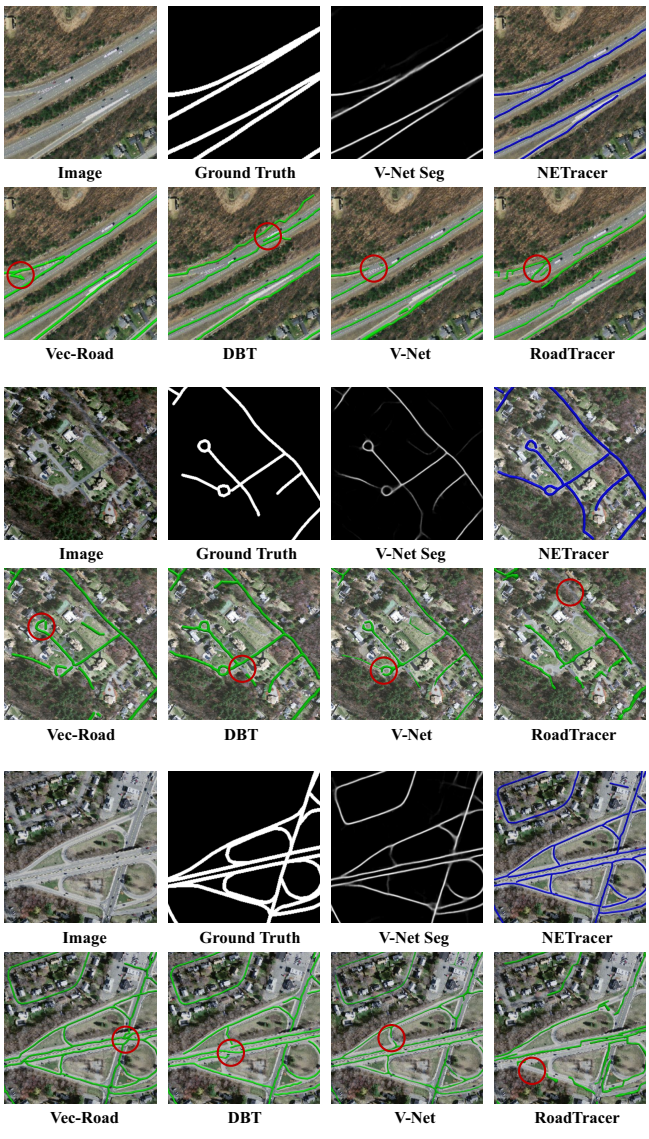

Figure 2. Comparison of various methods on road datasets.

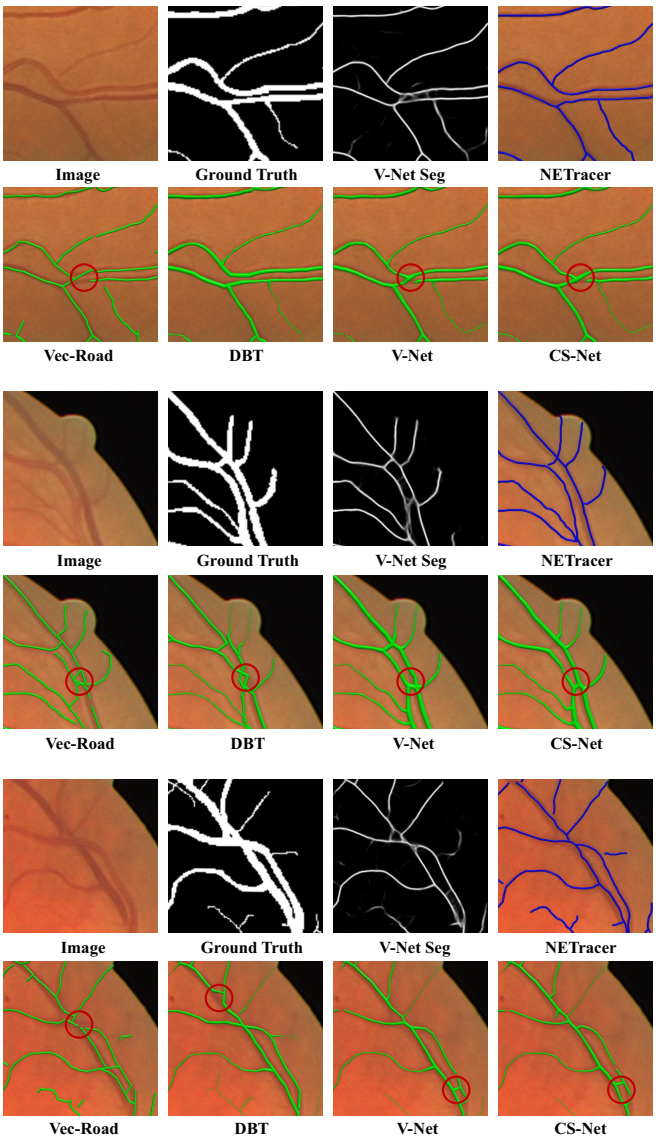

Figure 3. Comparison of various methods on vessel datasets.