TreeNet: Multi-Loss Deep Learning Network to Predict Branch Direction for Extracting 3D Anatomical Trees

Mengliu Zhao and Ghassan Hamarneh
{mengliuz, hamarneh}@sfu.ca

School of Computing Science, Simon Fraser University, Canada

Abstract. Calculation of blood vessel or airway direction is important for the task of tree tracking in 3D medical images. However, most existing works treat branch direction estimation as only a by-product of vesselness or tubularness computation. In this work, we propose a deep learning framework for predicting tracking directions of anatomical tree structures. We modify the deep V-Net architecture with extra layers and leverage a novel multi-loss function that encodes direction as well as cross sectional plane information. We evaluate our method on both 3D synthetic and 3D clinical pulmonary CT datasets. On the synthetic dataset, we outperform state of the art methods by at least 10% in direction estimation accuracy. For the clinical dataset, we outperform competing methods by $1-4\%$ in mean direction accuracy and $4-10\%$ in corresponding standard deviation.

1 Introduction

Tree extraction is a crucial task in 3D medical image analysis, and accurately extracted circulatory and respiratory trees can be further utilized in surgery planning, registration and tree space analysis [1,2,3]. However, the automation and accuracy of tree extraction still remains an open problem due to the natural complexity and variability of the topologies of anatomical tree structures [4], the various imaging reconstruction artifacts [5], varying image intensities along branches, the similarity between tubular structure lumen and background tissue lumen, and the changing geometry due to different pathologies [6,7,8].

One major category of tree extraction algorithms is based on iterative tracking, which usually starts from a given seed point near the root of the tree, predicts the direction of the branch to track along it, detects bifurcations to spawn children trackers, and progresses down the tree hierarchy to smaller branches until some stopping criteria are met [9]. Although there have been several works that tackle the important bifurcation detection step of the tracker [10,11], very few are specifically designed to determine the direction of the branch at a given point. Most works simply treat the problem of direction estimation as a by-product of vesselness or tubularness calculation [12,13,14].

Most existing methods on vesselness (with explicit/implicit direction prediction) rely on certain assumptions made on the geometry of tubular structures.

Most Hessian based vessel enhancement filters, e.g., Frangi et al. [15] and Jerman et al. [13,14], assumed the vessels were elongated structures. Cetin et al. [10] measured intensity distribution within an oriented cylinder-sphere combined model and constructed a corresponding tensor representation to optimize for vessel directions, however, their success relied on a good match between the cylinder and the actual vessel shape. Law et al. [12] used a gradient based tensor to model oriented flux flow and the vessel direction was also approximated by the eigenvector – intrinsically their assumption of vessel shapes were still straight tubes. However, in clinical datasets, especially those exhibiting various pathologies or abnormalities such as narrowing (e.g., in COPD [6]), aneurysms [7], and high tortuosity (which might indicate diseases like arterial hypertension and strokes [8]), the aforementioned shape assumptions might no longer hold true, which results in incorrect direction estimates.

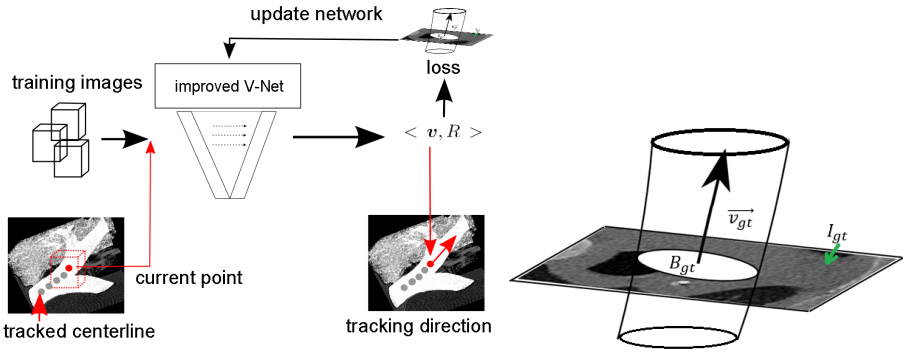
In contrast to the above deterministic methods, stochastic and learning based tracking methods provide more flexibility by adjusting the prediction retrospectively during the tracking process, or by using prior information learnt from the training data [16]. Lee et al. [17] proposed to use particle filtering to track vessel contours slice by slice, with the per-slice contour obtained by the Chan-Vese model. Lesage et al. [18] also proposed to use particle filtering method, but in contrast, utilized geometric flow, image features, as well as radius and direction prior distributions to perform Bayesian inference. In these tracking processes, vessel directions were found implicitly by subtracting neighboring points along the detected centerlines. On the other hand, a significant number of machine learning based methods ignore directional information completely by focusing on pixel-wise classification [19,9,20].

The fast development of deep learning methods provides vast opportunities in exploring the structures in 3D images from coarse to fine scales [21], however, limited work has been done on analyzing 3D vasculature images, and none of them estimated tree branch directions. Mirikharaji et al. [22] proposed to use an artificial neural network trained on 2D patches to learn the probability map of airway bifurcation locations; instead of tracking new branches, they connected the bifurcations by minimal paths to form the whole tree. Fu et al. [23] proposed to combine a convolutional neural network architecture with a conditional random field model to achieve a smooth binary segmentation for retinal vessels, but their method was only performed on 2D retinal images and no vessel direction was estimated. Chen et al. [24] proposed to use a convolutional autoencoder for voxel-wise cerebral arteries segmentation while completely ignoring directional information.

We claim the following contributions are made in this paper: i) The proposed method, which extends V-Net [21], is the first tree branch direction prediction deep learning method; ii) The proposed multi-loss function is novel and specially designed for tracking 3D tubular structures; iii) The proposed model is trained and tested in a branch-specific way, which takes advantage of the "anatomical tree statistics" [25,16] and fully utilizes statistical and geometrical information.

2 Methodology

Architecture: Our proposed deep learning architecture is an improvement of V-Net [21]. The choice of V-Net is two-fold: i) its encoding-decoding process propagates contextual information into higher resolution layers – in our case, the context information is the tubularity of the neighboring points; and, ii) our multi-loss function (introduced below) relies on cross sectional plane information, so the prediction process implicitly involves plane segmentation and reconstruction. We rescale all input volumes to $64 * 64 * 64$ voxels with histogram equalization. We use batch normalization and add three extra fully connected layers (FCs, with output channels 1024, 64 and 4) at the end of the V-Net and output a 4-element vector $< \boldsymbol{v}, R >$ where $\boldsymbol{v}$ predicts the direction of the vessel in the center of the input cube, and a radius R that serves as intermediate input in training the loss layer. The overall methodology is illustrated in Fig. 1a.
At testing time: a region of interest (ROI, or 3D patch, which we use in the context interchangeably) is generated and input into the network (as shown in red dotted box in Fig. 1a), and the output is the predicted vector of the corresponding branch (shown as $\overrightarrow{v_{gt}}$ in Fig. 1b).



(a) (b)

Fig. 1: (a): The proposed architecture and tracking process. (b): Illustration of B_{gt}, $\overrightarrow{v}_{gt}$ and I_{gt} in Equation 1.

Loss Function: We define the following multi-loss function:

$$L(\boldsymbol{v}^i_{dt}, R^i_{dt}) = \omega_1 L_{dir} + \omega_2 L_{mask} + \omega_3 L_{image} + \omega_4 L_{radius} \quad (1)$$
$$L_{dir}(\boldsymbol{v}^i_{dt}, R^i_{dt}) = -|\boldsymbol{v}^i_{gt} \cdot \boldsymbol{v}^i_{dt}|^2, \quad L_{mask}(\boldsymbol{v}^i_{dt}, R^i_{dt}) = ||B^i_{gt} - B^i_{dt}||^2_{L_2}$$
$$L_{image}(\boldsymbol{v}^i_{dt}, R^i_{dt}) = ||I^i_{gt} - I^i_{dt}||^2_{L_2}, \quad L_{radius}(\boldsymbol{v}^i{}_{dt}, R^i_{dt}) = |R^i_{gt} - R^i_{dt}|^2$$

where i is the training index, gt refers to ground truth value, dt refers to model prediction. I^i_{gt} and I^i_{dt} are corresponding gt and predicted cross sectional planes (going through the center voxel). B^i_{gt} and B^i_{dt} are the ground truth and predicted (using radius R^i_{dt} and circular expression) branch masks on the cross sectional

planes, R^i_{gt} and R^i_{dt} are the ground truth and predicted radii, as illustrated in Fig. 1b. The four terms L_{dir}, L_{image}, L_{mask}, L_{radius} capture the errors in, respectively, direction estimation, cross sectional image plane reconstruction, branch internal area estimation and radius estimation. We normalize L_{image} and L_{mask} by the patch cube size and set the weights empirically to $\omega_1 = \omega_2 = \omega_3 = \omega_4 = 1$. The total loss is optimized over v^i_{dt} and R^i_{dt}. Since accurate direction prediction leads to accurate cross section plane prediction, using multiple loss terms should theoretically increase the direction prediction accuracy.

Tree Tracking: We follow the tracking procedure in [9], which starts from a given seed point in a branch and tracks along vessel/airway detected by the proposed architecture. Additional tracking details are given in Section 3.

3 Evaluation

Synthetic Dataset: We use three different types of synthetic dataset to mimic pathologies such as narrowing and aneurysms, and high tortuosity [6,7,8]: i) Occlusion, ii) Torus and iii) Leakage. Examples are shown in Fig. 2. We augment the data by rotating the volumes along each axis randomly between $[0, 60°]$, using two radius values, translating along each dimension separately by three values ([-1, 0, 1], so 9 cases in total) and adding Gaussian noise with standard deviation from 0.005 to 0.1 (20 cases). This brings the total number of image volumes per each category to 360. We then run a 3-fold cross validation ensuring that augmentations of any volume are not split across the train and test sets.

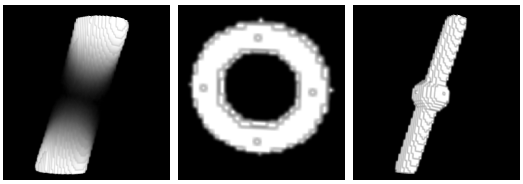


Fig. 2: Synthetic examples (noise free).

Clinical Dataset: The clinical dataset is from the Extraction of Airways from CT (EXACT) 2009 challenge[1]. Sixteen training volumes with binary segmentations were provided by the organizers. We extracted two categories of data: 1) ROIs, each is a cuboid containing one of the following 7 anatomical structures: trachea, right main bronchi (RMB), left main bronchi (LMB), right superior lobar bronchus (RSLB), right intermediate bronchus (RIB), left superior lobar bronchus (LSLB) and left inferior lobar bronchus (LILB); 2) patches, each is a cube randomly sampled from the branch centerlines, with radii twice the mean radii of the branch, intensities normalized to $[0, 1]$, and augmented by adding Gaussian noise with standard deviation $[0.01, 0.04]$ with step size= 0.01. red We perform a 4-fold cross validation on the patients for training and testing.

[1] http://image.diku.dk/exact/index.php

Competing Methods: We compare with 4 state-of-the-art algorithms: i) OOF [12]; ii) Tensor [10]; iii) Jerman [13,14]; 4) Particle filtering [17]. Since particle filtering doesn't directly predict the vessel direction, we use a primitive tracking method to first trace the branch centerline, then estimate the directions. For multiscale methods, the scale ranges are set according to mean branch scales learnt from the dataset, and all other parameters are set according to the original paper (for airways, i.e., dark-on-bright, some parameters are inverted accordingly). Note that although i) and ii) are not direction prediction methods per se, they leverage direction estimates to filter branches, which makes the comparison fair.
Tracking Details: Both the proposed method and the competing ones use the same initial seeds, which are selected from the ground truth branch centerlines. By calculating the mean radii $\bar{R}$ of the corresponding branch, the ROI radii are set automatically as $2\bar{R}$.
Evaluation Metrics: Two metrics are used to evaluate the results. For the tracking method, we use the asymmetric distance function proposed in [9] to compare the ground truth centerline to the extracted centerline:

$$D(C_1, C_2) = \{ \min_{s_2 \in C_2} dist(s_1, s_2) | \forall s_1 \in C_1 \} \tag{2}$$

where $dist(s_1, s_2)$ is 3D Euclidean distance between voxels s_1 and s_2, C_1 the ground truth centerline and C_2 the detected centerline (by either the proposed method or particle filtering). $D(C_1, C_2)$ returns a set of distance values for all the points on C_1, so a smaller mean value and standard deviation would indicate a better result. For other competing methods, since they return a per-voxel direction estimate, we use the following symmetric accuracy metric:

$$accu(\boldsymbol{v}_1, \boldsymbol{v}_2) = \boldsymbol{v}_1 \cdot \boldsymbol{v}_2 \tag{3}$$

where $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are the branch direction vectors to be compared.
Experiments: The evaluation result on the synthetic dataset is shown in Table 1. We can see a marked improvement in the proposed method over the competing ones by at least 10% in mean direction accuracy. For the Occlusion category, all competing methods performed poorly, since all these methods assume that the foreground is always luminous. For the Torus category, we can see the Tensor method [10] performing the worst, as it modeled the blood vessel as cylindrical tubes, which were very different from the torus shapes in the given images. On the contrary, the Tensor method performed much better than other competing methods on the Leakage category, as a long cylinder might overcome the small leakage (but not good enough to overcome the occlusion) and found the correct direction. It is worth noting that, although the Jerman filter could achieve good enhancement results at tortuous and bulging branches [13,14], the filter was not designed to deal with the direction estimation task.

We observe that by removing only one of the loss terms (other than L_{dir}) actually performs worse than using only L_{dir}. This is not surprising when we remember that the cross sectional plane and the direction together serve as the Frenet frame, so removing one term would invalidate the frame representation.

Since L_{image} contains the most information on the cross sectional plane, removing it leads to the worst performance. The improvement in prediction accuracy of the multi-loss function supports our hypothesis that all four terms contribute to the result, given their complementary nature.

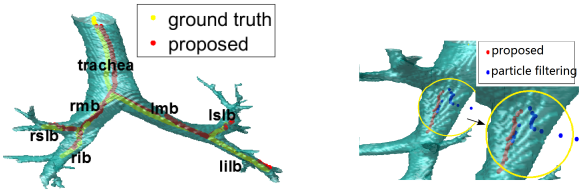
Fig. 3a shows an example where an airway centerline tree is extracted using our proposed method (red curves) and compared to the ground truth tree centerlines (yellow curves). Fig. 3b shows a qualitative comparison between the tracking result, along branch LIB, between the particle filtering and the proposed method. The mean and standard deviation of distance errors of each branch are shown in Fig. 4. The proposed method achieves a lower error mean and standard deviation on every anatomical branch.

The results in Table 2 are consistent with the synthetic data results. The proposed method outperforms all the competing methods on all branches.

We run our experiments on a Nvidia GTX GeForce 12 GB TITAN GPU, and the processing time per patch at testing stage is 0.04 second.

	Year	Occlusion	Torus	Leakage
OOF [12]	2008	0.11(0.07)	0.69(0.28)	0.21(0.12)
Tensor [10]	2015	0.47(0.1)	0.48(0.037)	0.89(0.03)
Jerman [13,14]	2016	0.44(0.46)	0.62(0.48)	0.45(0.07)
Proposed with L_{dir} only		0.95(0.06)	0.96(0.09)	0.97(0.04)
Proposed w/o L_{image}		0.90(0.07)	0.95(0.12)	0.94(0.04)
Proposed w/o L_{mask}		0.93(0.07)	0.93(0.13)	0.97(0.04)
Proposed w/o L_{radius}		0.94(0.07)	0.93(0.10)	0.97(0.05)
Proposed		**0.97(0.02)**	**0.97(0.06)**	**0.99(0.04)**

Table 1: Three fold cross validation result on synthetic dataset.



(a) (b)

Fig. 3: (a): Whole tree extracted. (b): Centerlines tracked by proposed algorithm and competing particle filtering algorithm on LIB.

4 Conclusion and future work

We proposed the first deep learning architecture for estimating anatomical tree branch directions, which is a critical step for the common tracking-based

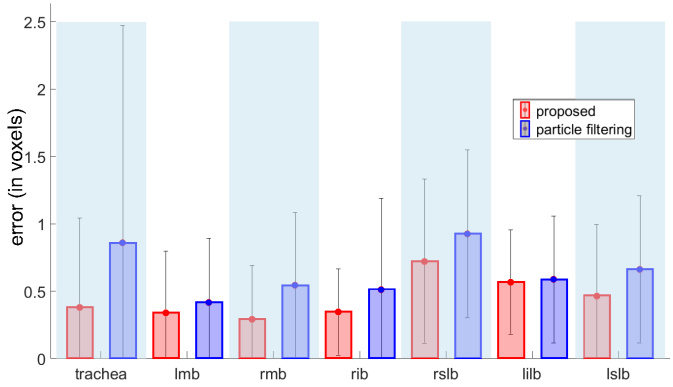


Fig. 4: Distance error bar between GT centerlines and detected centerlines.

	Level 1	Level 2		Level 3			
	Trachea	LMB	RMB	LILB	LSLB	RSLB	RIB
OOF [12]	0.19(0.18)	0.24(0.24)	0.29(0.26)	0.30(0.24)	0.42(0.31)	0.40(0.31)	0.43(0.28)
Tensor [10]	0.86(0.15)	0.60(0.20)	0.78(0.15)	0.68(0.22)	0.34(0.29)	0.33(0.26)	0.82(0.13)
Jerman [13,14]	0.91(0.17)	0.93(0.16)	0.90(0.18)	0.87(0.22)	0.87(0.17)	0.86(0.21)	0.88(0.18)
Proposed	**0.92(0.11)**	**0.95(0.07)**	**0.93(0.08)**	**0.89(0.15)**	**0.91(0.08)**	**0.90(0.11)**	**0.90(0.10)**

Table 2: Direction accuracy (mean and standard deviation) on airway branches of different levels.

tree extraction methods. Our proposed loss function is unique in that it follows the geometry of the target structure (i.e. the curvilinear tree branches) by using branch direction agreement and cross sectional image information, based on a Frenet frame of reference. In future work, we intend to apply our model on other anatomical trees, such as cerebral vasculature and coronary vessels.

Acknowledgments. We thank NVIDIA Corporation for the donation of Titan X GPUs used in this research and the Natural Sciences and Engineering Research Council of Canada (NSERC) for partial funding.

References

1. Foruzan et al.: Analysis of CT images of liver for surgical planning. Analysis **2**(2) (2012) 23–28 1
2. Baka et al.: Oriented Gaussian mixture models for nonrigid 2D/3D coronary artery registration. TMI **33**(5) (2014) 1023–1034 1
3. Feragen et al.: Tree-space statistics and approximations for large-scale analysis of anatomical trees. In: IPMI. (2013) 74–85 1
4. Kelch et al.: Organ-wide 3D-imaging and topological analysis of the continuous microvascular network in a murine lymph node. Scientific Reports **5** (2015) 16534 1

5. Rodriguez et al.: CT reconstruction techniques for improved accuracy of lung CT airway measurement. Medical Physics **41**(11) (2014) 1
6. Wiggs et al.: A model of airway narrowing in asthma and in chronic obstructive pulmonary disease. Am Rev Respir Dis **145** (1992) 1251–1258 1, 2, 4
7. Baráth et al.: Anatomically shaped internal carotid artery aneurysm in vitro model for flow analysis to evaluate stent effect. American Journal of Neuroradiology **25**(10) (2004) 1750–1759 1, 2, 4
8. Han, H.C.: Twisted blood vessels: symptoms, etiology and biomechanical mechanisms. Journal of Vascular Research **49**(3) (2012) 185–197 1, 2, 4
9. Zhao, M., Hamarneh, G.: Bifurcation detection in 3D vascular images using novel features and random forest. In: ISBI. (2014) 421–424 1, 2, 4, 5
10. Cetin, S., Unal, G.: A higher-order tensor vessel tractography for segmentation of vascular structures. TMI **34**(10) (2015) 2172–2185 1, 2, 5, 6, 7
11. McIntosh, C., Hamarneh, G.: Vessel crawlers: 3D physically-based deformable organisms for vasculature segmentation and analysis. In: CVPR. Volume 1. (2006) 1084–1091 1
12. Law, M., Chung, A.: Three dimensional curvilinear structure detection using optimally oriented flux. In: European Conference on Computer Vision. (2008) 368–382 1, 2, 5, 6, 7
13. Jerman et al.: Enhancement of vascular structures in 3D and 2D angiographic images. TMI **35**(9) (2016) 2107–2118 1, 2, 5, 6, 7
14. Jerman et al.: Blob enhancement and visualization for improved intracranial aneurysm detection. IEEE Trans on Visualization and Computer Graphics **22**(6) (2016) 1705–1717 1, 2, 5, 6, 7
15. Frangi et al.: Multiscale vessel enhancement filtering. In: MICCAI. (1998) 130–137 2
16. Zhao et al.: Leveraging tree statistics for extracting anatomical trees from 3D medical images. In: CRV. (2017) 131–138 2
17. Lee et al.: Enhanced particle-filtering framework for vessel segmentation and tracking. Computer Methods and Programs in Biomedicine **148** (2017) 99–112 2, 5
18. Lesage et al.: Medial-based Bayesian tracking for vascular segmentation: Application to coronary arteries in 3D CT angiography. In: ISBI. (2008) 268–271 2
19. Khorshed, R., Celso, C.: Machine learning classification of complex vasculature structures from in-vivo bone marrow 3D data. In: ISBI. (2016) 1217–1220 2
20. Zhou et al.: Vascular structure segmentation and bifurcation detection. In: ISBI. (2007) 872–875 2
21. Field et al.: V-Net: A framework for a versatile network architecture to support real-time communication performance guarantees. In: Fourteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Volume 3. (1995) 1188–1196 2, 3
22. BenTaieb, A., Hamarneh, G.: Uncertainty driven multi-loss fully convolutional networks for histopathology. In: Intravascular Imaging and Computer Assisted Stenting, and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis. Springer (2017) 155–163 2
23. Fu et al.: DeepVessel: Retinal vessel segmentation via deep learning and conditional random field. In: MICCAI. (2016) 132–139 2
24. Chen et al.: 3D intracranial artery segmentation using a convolutional autoencoder. In: IEEE International Conference on Bioinformatics and Biomedicine. (2017) 2
25. Mirikharaji et al.: Globally-optimal anatomical tree extraction from 3D medical images using pictorial structures and minimal paths. In: MICCAI. (2017) 242–250 2