
Proceedings Paper:
Wang, C, Hayashi, Y, Oda, M et al. (4 more authors) (2019) Tubular Structure Segmentation Using Spatial Fully Connected Network with Radial Distance Loss for 3D Medical Images. In: Lecture Notes in Computer Science. MICCAI 2019: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, 13-17 Oct 2019, Shenzhen, China. Springer Verlag , pp. 348-356. ISBN 9783030322250

https://doi.org/10.1007/978-3-030-32226-7_39

Tubular Structure Segmentation Using Spatial Fully Connected Network With Radial Distance Loss for 3D Medical Images

MICCAI 2019 Submission #1939

No Institute Given

Abstract. This paper presents a new spatial fully connected tubular network for 3D tubular-structure segmentation. Automatic and complete segmentation of intricate tubular structures remains an unsolved challenge in the medical image analysis. Airways and vasculature pose high demands on medical image analysis as they are elongated fine structures with calibers ranging from several tens of voxels to voxel-level resolution, branching in deeply multi-scale fashion, and with complex topological and spatial relationships. Most machine/deep learning approaches are based on intensity features and ignore spatial consistency across the network that are otherwise distinct in tubular structures. In this work, we introduce 3D slice-by-slice convolutional layers in a U-Net architecture to capture the spatial information of elongated structures. Furthermore, we present a novel loss function, coined *radial distance loss*, specifically designed for tubular structures. The commonly used methods of cross-entropy loss and generalized Dice loss are sensitive to volumetric variation. However, in tiny tubular structure segmentation, topological errors are as important as volumetric errors. The proposed radial distance loss places higher weight to the centerline, and this weight decreases along the the radial direction. Radial distance loss can help networks focus more attention on tiny structures than on thicker tubular structures. We perform experiments on bronchus segmentation on 3D CT images. The experimental results show that compared to the baseline U-Net, our proposed network achieved improvement about 24% and 30% in Dice index and centerline over ratio.

Keywords: tubular structure segmentation, spatial FCN, radial distance loss, blood vessel, bronchus

1 Introduction

The extraction and quantification of bronchi and blood vessels remains a fundamental problem with relevance to many computer aided diagnosis (CAD) systems. Automatic and complete segmentation of intricate tubular structures remains a major challenge in medical image-processing field. Airways and vasculature pose high demands on image analysis as they are elongated fine structures with calibers ranging from several tens of voxels to voxel-level resolution, branching in deeply multi-scale fashion, and with complex topological and spatial relationships. Over the past few years, deep learning methods have become the dominant approach in many data analysis fields and have achieved

remarkable advances in the medical image analysis [1]. Recently, airways and vasculature image analysis have also been approached using deep convolutional neural networks (ConvNets) instead of the traditional filter banks and machine learning methods.

Many studies have been made using 2D/2.5D image patches to train their networks [2–4]. Oda et al. [2] used multiple fully connected networks (FCNs) to tackle the problem of segmenting tiny abdominal arteries. Three 2D U-Net networks were employed to respectively segment 2D patches on three anatomical planes, and three 2D segmentation results were merged as a final volumetric segmentation. Similarly, Yun et al. [3] used 2D patches on three anatomical planes to train three ConvNets as classifiers to classify whether a voxel belongs to the bronchus. To explore the inherent relationship between such orthogonal 2D patches, Tetteh et al. [4] presented 2D orthogonal cross-hair filters to make use of 3D context information but with a reduced computational burden.

The above approaches made use of local contexts to address the tubular-structure segmentation. Although, they achieved good performance, the problem remains of taking full advantages of large-range information in the field of view (FOV), which can provide global queues unavailable from small local patches. Meng et al. [5] attempted to use an original 3D U-Net combined with a traditional bronchus tracking method to improve bronchus segmentation accuracy. Huang et al. [6] presented a liver-vessel segmentation method using a 3D U-Net with variant Dice loss function. Both works used a large 3D sub-volume size of around $100 \times 100 \times 100$ voxels.

Compared to 2D local patch-based methods, 3D FCNs for tubular-structure segmentation has some limitations that need to be addressed. Foremost is the severe imbalance in the sizes of the large background area when compared to the small foreground (tubular structure). Second, it's unclear whether FCNs can learn useful features from the abundant contexts of 3D data that while capturing coarse-scale long-range interrelationships are also accurate to detect tubular details and topological correctness at finer scales. The motivation of this work is to improve the accuracy of tubular-structure segmentation in respect to these two concerns.

The main contributions of this work can be summarized as follows. **1)** We introduce 3D recurrent convolutional layer in FCN architecture for tubular-structure segmentation. **2)** We propose a novel radial distance loss for 3D tubular-structure segmentation that helps the networks to recover tiny tubular structures. **3)** Both 3D recurrent convolution layer and radial distance loss are generic and flexible, so they can be easily incorporated in other state-of-the-art networks and used in other applications.

2 Method

2.1 Overview

Our main contribution is twofold from the viewpoint of tackling both network architecture and its loss function. The motivation of this work is to design a new end-to-end FCN for 3D-tubular structure segmentation. Compared with traditional layer-by-layer convolutional layers, the proposed slice-by-slice convolutional layer permits messages to efficiently pass slices. Our proposed loss function places more weight on the centerline

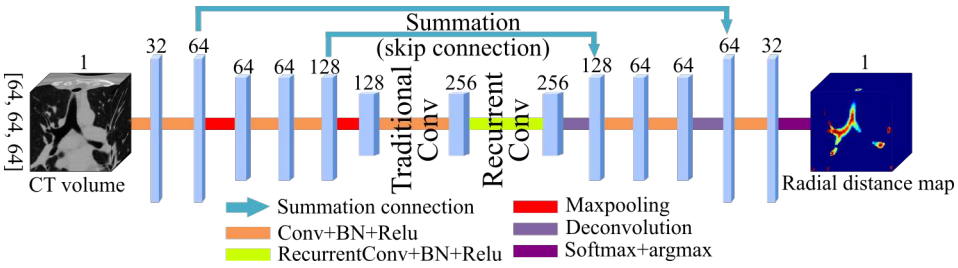


Fig. 1. Overview architecture

of a tubular structure than on the outer border which can help the network pay more attention to tiny structures. A simple illustration of our architecture is shown in Fig. 1.

2.2 3D spatial FCN

Spatial-CNN was first proposed to address the traffic lane detection problem [7]. The main contribution of spatial-CNN is the introduction of a slice-by-slice convolutional layer. Different from the traditional layer-by-layer CNN, slice-by-slice convolutions perform like recurrent neural networks. The in-layer recurrent convolutions provide more efficient message-passing between neurons in the same layer. This can help networks reinforce structures with strong spatial constraints [7].

In this work, we incorporated Recurrent Convolutional Layers (RCLs) in a 3D FCN architecture for 3D medical imaging applications. Detailed structure of RCLs is illustrated in Fig. 2. Convolutions are performed in both forward and backward directions along the width, height, and depth dimensions. As shown in Fig. 2, RCL_w, RCL_h, and RCL_d denote RCLs with convolutions in front-to-back direction along the three dimensions. $\mathrm{RCL}_w^{'}$, $\mathrm{RCL}_h^{'}$, and $\mathrm{RCL}_d^{'}$ denote RCLs with convolutions in the reverse direction. Forward computation of RCL_w can be defined as

$$\boldsymbol{z}_{c,i,j,k} = \begin{cases} \boldsymbol{x}_{c,i,j,k}, & \text{if } i=0 \\ \boldsymbol{x}_{c,i,j,k} + f(\boldsymbol{z}_{c,i-1,j,k} * K), & \text{if } 0<i<W, \end{cases} \tag{1}$$

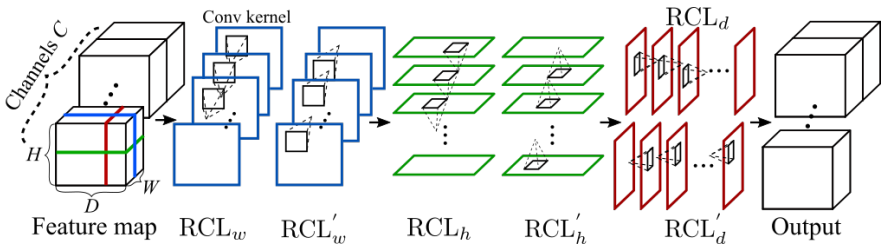


Fig. 2. RCL architecture: Feature map represents the output of the last convolutional layer with a shape of $N \times C \times W \times H \times D$. Slice-by-slice recurrent convolutions are performed in each RCL.

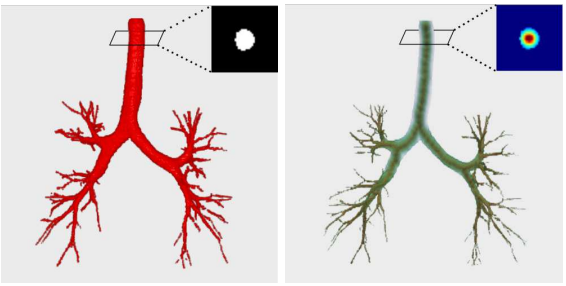


Fig. 3. Volume rendering of binary ground truth and corresponding radial distance map. Radial distance map shown on the right is rendered with pseudo color mapping 0 to blue, and 1 to red.

where $\boldsymbol{\mathcal{X}}$ denote an input 4D feature tensor of size $C \times W \times H \times D$, and $\boldsymbol{\mathcal{Z}}$ denotes the output of RCL_w. Let c,i,j,k denote the index of channel, width, height and depth dimensions. K is a 3D convolution kernel of size $w \times w \times w$. $f(\cdot)$ denotes a nonlinear activation function. RELU is used in this work. Similarly, forwarding for RCL_h and RCL_d can be easily derived based on Eq. 1.

In this work, we chose U-Net architecture [8, 9] with three multi-scale levels as our backbone FCN. As shown in Fig. 1, we placed the RCLs directly after the deepest feature map, i.e. the compressed low-dimensional representation. In previous spatial-CNN [7], the authors placed the RCLs after the output of CNN, under the assumption that the top hidden layer with rich semantic information is an ideal place to apply RCLs. However, in our FCN architecture, applying RCLs to the deepest representation provides better performance than using top hidden layer.

2.3 Radial distance loss

The use of Dice loss for medical segmentation tasks was first proposed by Milletari et al. [10]. They showed that Dice loss outperformed other losses, especially in severe data-imbalance situations [10]. More recently, Hausdorff distance loss and contour loss were proposed for localization and segmentation tasks [11, 12]. Dice loss measures volumetric variation, and Hausdorff distance loss and contour loss measure the distance between boundaries. However, no losses were specifically designed for tubular structures. In this work, we propose a novel radial distance loss (RD loss) for tubular structures. The motivation of our proposed loss is to capture the geometric topology loss believed to be more important than volume loss in the tubular-structure segmentation.

The proposal of RD loss is inspired by the centerline overlap (CO) metric used for evaluating blood vessel segmentation [13]. Wang et al. used the CO metric to give a more accurate description of tiny blood vessel segmentation than conventional Dice similarity coefficient (DSC). To take advantage of the CO metric while keeping a volumetric measurement, our proposed RD loss is defined as

$$L = -\frac{1}{2}\sum_{k=0}^{1} \mathcal{W}_k \left(\frac{2\sum_i^N p_{i,k} d_{i,k}}{\sum_i^N p_{i,k}^2 + \sum_i^N d_{i,k}^2} \right), \tag{2}$$

where $p_i \in \mathbf{P}$ and $d_i \in \mathbf{D}$ denote the i-th voxel predicted binary result and the radial distance map. Voxel index $i \in [1,N]$. $k \in [0,1]$ denote class. Class weights $\mathcal{W}$ is defined as reciprocal volume ratio of each class. Notice that Eq. 2 is similar to traditional Dice loss [10] except that we use radial distance map $\mathbf{D}$ instead of binary ground truth $\mathbf{G}$. $\mathbf{D}$ is defined as:

$$\mathbf{D} = -\frac{1}{\max(\mathbf{F})}\mathbf{F} + 1, \tag{3}$$

where $\mathbf{F} = \{f_i, i \in N\}$ denotes a distance map created by a Euclidean distance transformation from the centerline of ground truth. $\mathbf{F}$ is defined as:

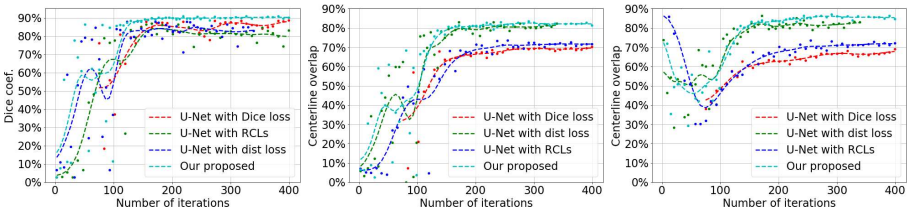
$$f_i = \begin{cases} \min\{d(f_i, s_j); s_j = 1\}, & \text{if } s_j = 0, \\ 0, & \text{if } s_j = 1 \text{ or } g_i = 0. \end{cases} \tag{4}$$

Here, $s_j \in \mathbf{S}$ denotes the j-th voxel of centerline data $\mathbf{S}$ extracted from $\mathbf{G}$. Finally, we obtained radial distance map $\mathbf{D}$ by normalizing F from 0 to 1 using a simple monomial form. An example of a radial distance map is shown in Fig. 3. RD loss can be weighted combination of centerline loss and Dice loss.

3 Experiments and Results

To evaluate our proposed method, we performed bronchus segmentation experiments on 3D clinical CT scans acquired with a standard dose. The sizes of CT slices were 512×512 pixels with a resolution of $0.63-0.97$ mm. The number of CT slices ranged from 238 to 851 with varying thickness of $0.63-1.00$ mm. Monte Carlo cross-validation (MCCV) was conducted three times. All 38 CT images were randomly divided into training and validation subsets containing 35 and 3 cases. The model with the best validation accuracy is tested on three unseen datasets acquired in a different hospital than those used for training.

Since we chose U-Net as our backbone FCN, we performed comparison experiments using U-Net with Dice loss as a baseline. U-Net with only RCL structures and U-Net with only radial distance loss were also validated to prove the effectiveness of each proposal. We also compared our proposed method with two other methods, one



(a) DSC' **(b)** CO' **(c)** CO''

Fig. 4. Validation results. Dot markers represent actual validation scores. For a clearer visualization, we plot smoothed results in dashed line using Gaussian filter.

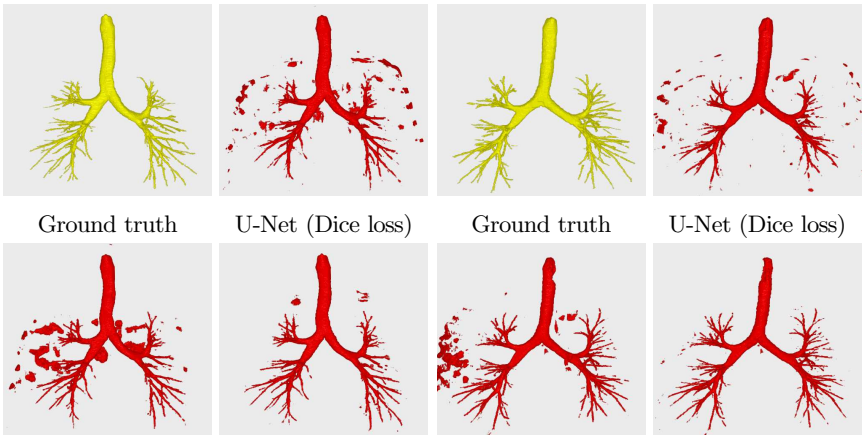


Fig. 5. Volume rendering of segmentation results of two validation cases. No post-processing was performed. Trained models at 300^{th} iteration were used for prediction.

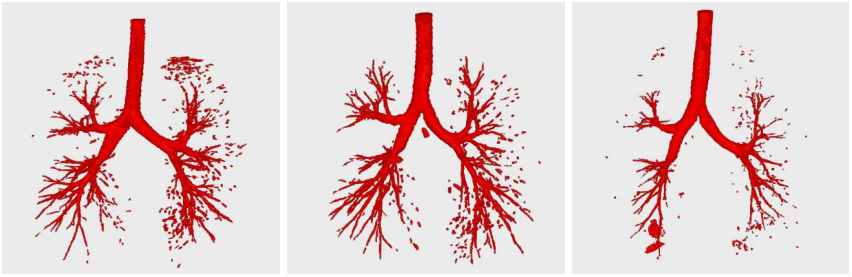
is a variant U-Net architecture, V-Net [10] (Dice loss and RD loss), and the other is 3D voxelwise residual networks, VoxResNet [14].

In training phase, for each epoch, 4 sub-volumes with a size of $64 \times 64 \times 64$ voxels were randomly cropped from each CT image for all 35 training cases. No data augmentation was performed in our experiments. Random cropping was performed in this work. The initial learning rate was set to 0.01, and it decayed by 0.2 every 150 epochs. The optimization was realized via stochastic gradient descent (SGD).

DSC and *CO* were used for quantitative validation. To validate the general segmentation ability of each method, DSC' was measured on the segmentation results

Table 1. Quantitative comparison results. All measurement was shown with mean ± standard values.

Method	Dataset	DSC' (%)	DSC''(%)	Se (%)	CO' (%)	CO'' (%)
(1) In-house dataset						
Yun et al. [3]	Train: 59 Test:8	89.9±8.9	-	-	-	-
Meng et al. [5]	Train: 30 Test:20	86.6	-	79.6	-	-
(2) Our bronchus dataset						
VoxResNet (Dice loss) [14]	MCCV Test:3	79.6±3.7	90.0±3.4	72.3±5.0	39.2±2.7	31.0±2.1
V-Net (Dice loss) [10]		65.4±9.9	91.0±2.0	69.0±2.0	28.3±3.9	19.8±1.1
V-Net (RD loss)		83.3±2.0	88.4±0.7	76.3±4.6	53.8±1.0	66.6±4.9
U-Net (Dice loss) [9]		64.0±19.5	92.4±1.6	82.9±5.7	47.2±18.1	54.3±9.0
Our proposed		88.7±1.2	94.5±0.8	86.5±1.0	76.6±6.0	80.6±5.6



(a) Test case 1 **(b)** Test case 2 **(c)** Test case 3

Fig. 6. Segmentation results of three test data. For a clearer visualization, we removed outliers under volume size of 10 voxels.

with no post-processing, and DSC'' was only measured on thick branches (before 2^{nd} generation of dichotomous branching). We computed CO scores on the results with two post-processing strategies. One is measured on the largest connected component extracted from the segmentation results. The other one is measured on tiny bronchi after the 2^{nd} generation of dichotomous branching and masked by a dilated ground truth with 5 voxels to remove false positives. These two measures are denoted CO' and CO''. CO' measures the ability to segment tubular structures, while CO'' measures the ability of tiny structures. Figure 4 shows validation results. Volume rendering of two validation cases are shown in Fig. 5. Quantitative comparison results are shown in Table 1. Other than DSC and CO, we also compute sensitivity (Se). To demonstrate robustness of method, segmentation results of three test data are shown in Fig. 6.

4 Discussion and Conclusions

In this work, we proposed 3D recurrent convolutional layer and radial distance loss, and demonstrated the implementation of these proposals in a widely used U-Net architecture. As experimental results showed, our proposed approaches achieved significant improvement over our baseline architecture, and obtained competitive results with state-of-the-art methods. Our proposed extensions, viz. 3D RCL and radial distance loss, are generic and flexible component that they can be easily incorporated in other deep learning architectures. We demonstrated an application on V-Net architecture using our RD loss, remarkable improvement was obtained comparing with the one use Dice loss.

Figure 5 and 6 show thick bronchi are under-segmented. This is a side effect of radial distance loss, since we decreased the weight of the most peripheral voxels. The normalization strategy in Eq. 3 can be improved to use more complex functions beyond a simple monomial. However, from Table. 1, DSC'' shows our segmentation accuracy of thick branches is still better than baseline U-Net. Segmentation results of three unseen datasets acquired in different hospital illustrated the robustness of our method. Visually, good segmentation accuracy was achieved.

In conclusion, we choose the challenging bronchus segmentation task to prove the effectiveness of our proposed method regarding general tubular structure segmenta-

tion. Experimental results showed that our proposed approaches are proven to be effective in bronchus segmentation task. However, we only evaluate the segmentation performance on bronchus segmentation. Our approaches should theoretically work on blood vessel segmentation problem. Application on blood vessel segmentation will be one of our future works. Additionally, more state-of-the-art networks incorporated with our approaches need to be investigated.

References

1. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Med. IA 42, 60–88 (2017)
2. Oda, M., Kitasaka, T., Misawa, K., Fujiwara, M., Mori, K.: Abdominal artery segmentation from CT volumes using fully convolutional network for small artery segmentation. In: International Journal of Computer Assisted Radiology and Surgery, Sup. 1. vol. 13, pp. 20–21 (2018)
3. Yun, J., Park, J., Yu, D., Yi, J., Lee, M., Park, H.J., Lee, J.G., Seo, J.B., Kim, N.: Improvement of fully automated airway segmentation on volumetric computed tomographic images using a 2.5 dimensional convolutional neural net. Med. IA 51, 13–20 (2019)
4. Tetteh, G., Efremov, V., Forkert, N.D., Schneider, M., Kirschke, J., Weber, B., Zimmer, C., Piraud, M., Menze, B.H.: Deepvesselnet: Vessel segmentation, centerline prediction, and bifurcation detection in 3-D angiographic volumes. arXiv preprint arXiv:1803.09340 (2018)
5. Meng, Q., Roth, H.R., Kitasaka, T., Oda, M., Ueno, J., Mori, K.: Tracking and segmentation of the airways in chest CT using a fully convolutional network. In: MICCAI. pp. 198–207 (2017)
6. Huang, Q., Sun, J., Ding, H., Wang, X., Wang, G.: Robust liver vessel extraction using 3D U-Net with variant dice loss function. Computers in Biology and Medicine 101, 153–162 (2018)
7. Pan, X., Shi, J., Luo, P., Wang, X., Tang, X.: Spatial as deep: Spatial CNN for traffic scene understanding. In: AAAI. pp. 7276–7283 (2018)
8. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: MICCAI. pp. 424–432 (2016)
9. Roth, H., Oda, M., Shimizu, N., Oda, H., Hayashi, Y., Kitasaka, T., Fujiwara, M., Misawa, K., Mori, K.: Towards dense volumetric pancreas segmentation in CT using 3D fully convolutional networks. In: SPIE MI 2018. vol. 10574, p. 105740B (2018)
10. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). pp. 565–571 (2016)
11. Ribera, J., Güera, D., Chen, Y., Delp, E.: Weighted hausdorff distance: A loss function for object localization. arXiv preprint arXiv:1806.07564 (2018)
12. Pennec, X., Jaıs, P., Cochet, H., Sermesant, M.: Automatically segmenting the left atrium from cardiac images using successive 3D U-Nets and a contour loss. Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges p. 221
13. Wang, C., Oda, M., Hayashi, Y., Yoshino, Y., Yamamoto, T., Frangi, A.F., Mori, K.: Tensor-based graph-cut in Riemannian metric space and its application to renal artery segmentation. In: MICCAI. pp. 353–361 (2016)
14. Chen, H., Dou, Q., Yu, L., Qin, J., Heng, P.A.: Voxresnet: Deep voxelwise residual networks for brain segmentation from 3D mr images. NeuroImage 170, 446–455 (2018)