Context Aware 3D Fully Convolutional Networks for Coronary Artery Segmentation

Yongjie Duan[1,2,3], Jianjiang Feng(✉)[1,2,3], Jiwen Lu[1,2,3], and Jie Zhou[1,2,3]

[1] Department of Automation, Tsinghua University, China
[2] State Key Lab of Intelligent Technologies and Systems, Tsinghua University, China
[3] Beijing National Research Center for Information Science and Technology, China

Abstract. Cardiovascular disease caused by coronary artery disease (CAD) is one of the most common causes of death worldwide. Coronary artery segmentation has attracted increasing attention since it is useful for better visualization and diagnosis. Conventional lumen segmentation methods basically describe vessels by a rough tubular model, thus presenting inferiority on abnormal vascular structures and failing to distinguish exact coronary arteries from vessel-like structures. In this paper, we propose a context aware 3D fully convolutional network (FCN) for vessel enhancement and segmentation in coronary computed tomography angiography (CTA) volumes. Combining the superior capacity of CNN in extracting discriminative features and satisfactory suppression of vessel-like structures by spatial prior knowledge embedded, the proposed approach significantly outperforms conventional Hessian vesselness based approach on a dataset of 50 coronary CTA volumes.

Keywords: Spatial prior knowledge, deep learning, lumen segmentation, coronary computed tomography angiography

1 Introduction

Over the past decades, coronary artery disease (CAD), which is caused by vessel calcification or atherosclerosis, is among the most common causes of human deaths in the world [1]. Coronary computed tomography angiography (CTA) has been widely utilized to diagnose CAD, such as plaque evaluation and stenosis assessment, thanks to the advantage of noninvasive nature and high resolution image acquisition. Most plaque evaluation and stenosis assessment, as well as estimation of fractional flow reserve (FFR), which is used to measure the influence of stenosis impeding oxygen delivery to the heart muscle, rely on a semi-automatic or automatic precise coronary artery segmentation. An inaccurate artery tree extraction will result in improper stenosis assessment while manual correction of wrong artery trees is very time-consuming.

Accurate vascular segmentation in medical image is a widely researched topic, yet remaining a challenging task because of the presence of calcifications, image

✉Corresponding author (jfeng@tsinghua.edu.cn)

artifacts, insufficient contrast, and large anatomical variations among patients. A complete vessel tree is usually considered as a combination of multi-scale and multi-orientation tubular structures, in view of the appearance the vessels present in CTA. This property is directly applied for vessel segmentation [2], centerline extracting, and lumen diameter estimation. However, such approach and similar ones fail to detect abnormal structures, namely bifurcations and lesions (e.g., calcifications, atherosclerosis, aneurysms, stents, and stenoses), due to the inferiority of using a rough tubular model [3]. To improve the segmentation performance, machine learning was applied to capture more powerful and discriminative features [4, 5]. Nonetheless, since the segmentation is regarded as a voxel-wise classification problem, these approaches may generate a lot of leaks (false positives) or holes (false negatives) in the final extraction result. Therefore, fully convolutional network (FCN), U-Net and its 3D extension [6,7] have earned increasing attention in medical image segmentation and represented better performance because of the superiority of integrating high-level constraints within the upsampling step. However, to the best of our knowledge, these networks have not yet considered corresponding holistic anatomical information, such as spatial distribution of coronary artery tree. In other words, the priori anatomical knowledge is rarely incorporated to guide the segmentation procedure.

Inspired by the conventional vessel extraction algorithms and deep learning framework, we propose an improved segmentation approach that adopts a 3D fully convolutional network and spatial prior knowledge of coronary artery tree, to predict the probability map of coronary arteries from the whole coronary CTA images. We call it context aware because unlike original 3D U-Net, our network utilizes the location information of input patches. Briefly, the main contributions of our approach are summarized as: (1) a 3D UNet-like network is customized to achieve coronary artery tree segmentation in the whole coronary CTA, and the architecture of our network is tailored to identify small objects in 3D patches by introducing a shortcut from low-level to high-level feature maps; (2) the spatial prior knowledge of coronary artery tree, estimated from training images, is incorporated to guide the segmentation within each local patch, thus reducing the complexity of model learned and increasing the performance at the same time. We evaluate our algorithm on 50 coronary CTA volumes by a five-fold cross validation. The experimental results demonstrate that our framework is robust for coronary artery segmentation and outperforms conventional Hessian vesselness based approach.

2 Methods

In this paper, we aim to extract complete coronary artery tree from coronary CTA volume. A statistical model is obtained first to estimate the spatial distribution of coronary arteries, which contributes to the reduction of false positives and false negatives. Then our proposed segmentation network utilizes this prior knowledge as an additional input channel and produce voxel-wise probability map, which is thresholded to obtain the final segmentation result. No morpho-

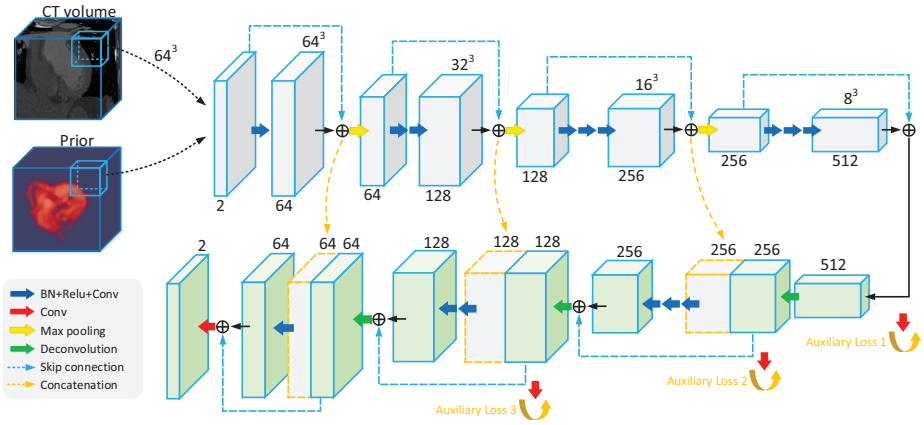
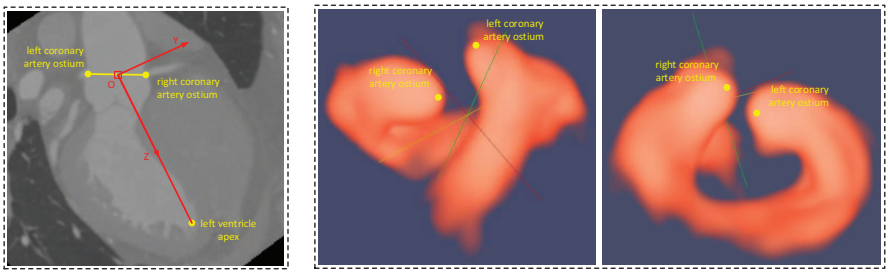


Fig. 1. Network architecture of our proposed framework.

logical post-processing is used to clean up small disconnected components. The schematic illustration of our framework is shown in Fig. 1.

2.1 Spatial Distribution Prior

Normally, the coronary arteries start from the coronary ostia, then bifurcating into two main branches along the heart surface. Obviously, this anatomical knowledge can be introduced as an auxiliary constraint to help to reject most wrong extractions scattered out of the heart or inside. In order to acquire the spatial distribution model, all the annotated labels in training dataset, are aligned to the same coordinate system defined by three anatomical landmarks, namely, the ostium of left coronary artery, the ostium of right coronary artery, and the left ventricle apex. These landmarks are extracted using a boosting-based detection algorithm [8]. As shown in Fig. 2, the origin is defined as the middle point between the ostia of left and right coronary artery. The z axis is defined as the direction pointing from the origin to the left ventricle apex. The y axis is defined as the vector perpendicular to the z axis and lies inside the plane determined by three landmarks. Then the x axis is defined as the cross product of previous two axes. After the alignment, a spatial probability map of coronary arteries is obtained by some non-parametric kernel density estimation, such as Parzen window. Therefore, given a test volume, the corresponding coordinate system attached is estimated following detecting the three cardiac landmarks. Then the statistical priori distribution, transformed based on the coordinate system defined previously, is used as an additional input channel for training our network. A typical improvement of using this prior knowledge is shown in section 3. Moreover, the convergence becomes faster due to the constraints introduced by priori spatial distribution.



(a) (b)

Fig. 2. (a) Normalized coordinate system defined by three anatomical landmarks: the ostium of left and right coronary artery and the left ventricle apex. (b) Generated spatial distribution of coronary arteries shown in two different viewpoints.

2.2 Network Architecture

Illustrated as Fig. 1, our proposed context aware network is customized from the original 3D U-Net [6]. Similar to its standard version, our main network is composed of 3D convolution, max pooling, up sampling (deconvolution), and shortcut connections from layers in down-sampling path to the ones in up-sampling path with equal resolution. However, considering the growth of model depth and number of trainable parameters in 3D kernels, training the network mentioned can be extremely difficult owing to the vanishing of forward and backward propagating signals. Therefore, residual skip connection is added between the input and output of each layer block to preserve flowing signals in our network. Each layer is composed of batch normalization, ReLU activation, and convolution in order, according to its superiority demonstrated in previous study [9]. Moreover, max pooling rather than convolution with stride 2 is utilized to reduce the resolution of feature maps, thus leading signals flowing directly from feature maps with high-resolution to the ones with low-resolution. In addition, the kernel size adopted in our network is set to 3^3 to keep the number of trainable parameters at a low level while preventing performance declining significantly.

Apparently, class imbalance, namely the background in cardiac CT volume vastly outnumbering the coronary arteries to be extracted, results in serious misclassifications. To deal with the problem and mitigate easy example dominant, the focal loss, proposed in [10], is modified as follows in our framework.

$$\mathcal{L}_{\alpha Focal}(\mathcal{X};\mathcal{W}) = -\frac{1}{|\mathcal{X}|}\sum_{y}\sum_{x}\alpha(1-p_t)^{\gamma}\log(p_t) \tag{1}$$

$$p_t = yp + (1-y)(1-p),\ \alpha = 1 - \frac{|\mathcal{X}_y|}{|\mathcal{X}|}$$

Given an input patch $\mathcal{X}$, $|\mathcal{X}|$ is the size of patch $\mathcal{X}$, and $|\mathcal{X}_y|$ is the size of class y within patch $\mathcal{X}$ similarly. $\mathcal{W}$ is the parameters to be trained within the network proposed. y and p denote the ground truth label and corresponding

probability prediction of sample x after softmax, respectively. Besides, we use a trade-off parameter w to balance the importance of positive/negative examples, and γ, which is set as 2 in this paper, is introduced to mitigate easy example dominant and focus on hard examples consequently. In addition, we also employ a two-stage training strategy to train such an imbalance problem. We extract patches whose centers have a 0.8 probability of being on foreground initially, and decrease it to 0.5 after certain epochs.

In deep network, early layers are always under-tuned because of gradient vanishing problem. The residual structure mentioned above alleviates this problem by adding skip paths. On the other hand, enhancing the gradient flow for shallow layers with deep supervision is also demonstrated to be effective in segmentation task. Similar to [11], we incorporate three side-paths auxiliary loss to shorten the backpropagation path of gradient flow signals. So as to obtain the final formulation of loss function in our proposed network.

$$\mathcal{L}(\mathcal{X};\mathcal{W}) = \mathcal{L}_{\alpha Focal}(\mathcal{X};\mathcal{W}) + \sum_{s=1,2,3} \beta_s \mathcal{L}^s_{\alpha Focal}(\mathcal{X};w^s) \tag{2}$$

where β_s is the weight of different side-paths w^s and set as 0.3, 0.6, 0.9 from coarse to fine.

3 Experiments and Results

Our proposed approach is evaluated on a total of 50 coronary CTA volumes, which are collected from different scanners, by a five-fold cross validation. The images are randomly divided into 5 groups, and one of them is selected for testing then the rest for training in turn. In term of ground truth, some public datasets, such as Rotterdam Coronary Artery Algorithm Evaluation Framework, only provide annotated contours on cross-sections of some main branches. It is not feasible to use these annotations for training or testing our networks. In this paper, therefore, the complete coronary artery trees in 50 CTA volumes are annotated by radiologists from cooperative hospital. And the spatial prior calculated uses only training cases in each of the five folds of cross validation. Considering the variations of image spacing, we resampled these images to the same voxel size of 0.4^3 mm^3. Moreover, data augmentation is applied on 30 percent training samples, including ± 25 degrees rotation around z axis and ± 10 pixels translation along orthogonal axes.

3.1 Implementation Details

We implemented our 3D network using a NVIDIA GeForce 1080Ti in *Tensorflow*. Considering the limitation of available GPU memory and superiority of training on mini-batch, cropped $64 \times 64 \times 64$ sub-volumes are randomly selected as the input to train our network with a batch size of 6. We update the weights of network using an Adam optimizer with initial learning rate of 0.001 and momentum of 0.5. To avoid overfitting, a L2 regularization is used in our network

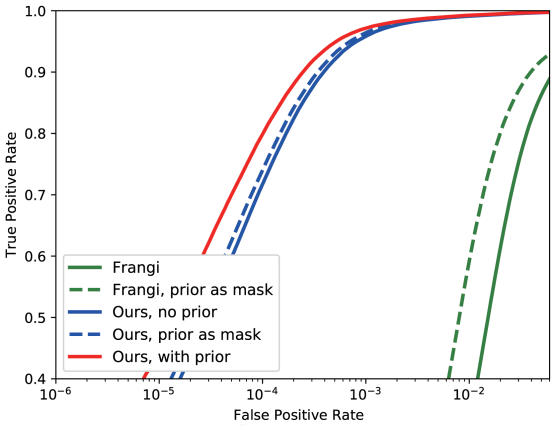


Fig. 3. The quantitative performance comparison of our proposed algorithm and multi-scale Hessian vesselness based method [2]. The curves of Frangi filtering are consistent with its performance reported in [4].

Table 1. Quantitative evaluation for coronary segmentation

Metrics	Frangi	Frangi +mask	Ours no prior	Ours +mask	Ours with prior
DSC[%]	2.5±0.9	9.5±3.6	74.7±5.6	75.8±4.7	**79.5±3.6**
Precision[%]	1.3±0.5	5.2±3.2	70.2±8.5	72.0±7.2	**78.5±6.0**
Recall[%]	74.1±10.9	74.0±10.9	80.9±7.6	80.9±7.6	**81.3±3.6**

with the value of $5e^{-4}$. Finally, the whole network is trained from scratch, and the weights are initialized from a normal distribution $\mathcal{N}(0, 0.01)$. Within the test stage, we adopted sliding windows with an overlapping ratio of 0.4 to generate input patches from the whole volume, and consequently overlap-tiling strategy is utilized to reconstruct the whole predicted probability map.

3.2 Results

Three metrics are used to evaluate the segmentation performance of our proposed approach, including dice similarity coefficient (DSC), precision, and recall. Quantitative comparison of our proposed network with/without spatial priori distribution (**Ours with/no prior**) and Hessian vesselness based method (**Frangi**) [2] [1], is shown in Fig. 3 and Table 1.

To explore the effectiveness of spatial prior knowledge usage as well as the way of incorporating prior knowledge, our generated priori spatial distribution is thresholded as a binary mask to reduce negative voxels (**Frangi+mask**, **Ours+mask**), thereby achieving a fair comparison with our proposed context aware method. It is observed that our proposed network gets an obvious improvement, compared with Hessian vesselness based method. Concentrating on the usage of spatial prior knowledge, we notice that utilizing the priori spatial

[1] Public implementation by Kroon, D.J. https://ww2.mathworks.cn/matlabcentral/fileexchange/24409-hessian-based-frangi-vesselness-filter.

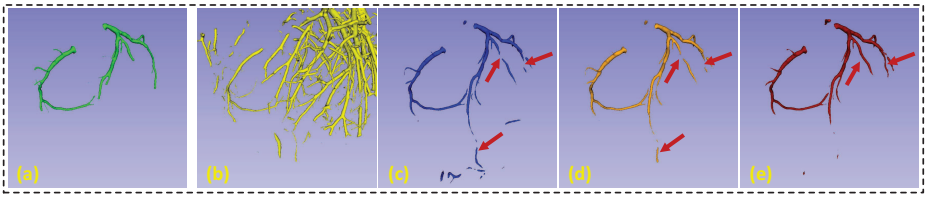


Fig. 4. Qualitative comparison of different algorithms on one example. No morphological post-processing is used to clean up small disconnected components. (a) annotated label, (b) filtered result of Frangi [2] using mask, (c) result of our network without prior knowledge, (d) filtered result of (c) using mask, (e) result of proposed method.

distribution by just using it as a binary mask is inferior to incorporating it as an additional input channel.

In Fig. 4, the probability maps predicted by different approaches are thresholded to visually compare the performance. Similarly, we could observe that satisfactory performance is achieved by our proposed approach, in which the false extractions are significantly less than the ones in Hessian vesselness based method with binary priori mask. Besides, some disconnected segments caused by artifacts or insufficient contrast could also be reconnected due to the prior knowledge incorporated. It is reasonable to assume that the priori spatial distribution, estimated from finite samples, may not be consistent with the coronary arteries distribution within a new image exactly. Therefore, compared with using the prior knowledge as a binary mask, taking it as an additional input channel achieves better performance, especially on those segments referred by red arrows in Fig. 4.

For a better view of results, contours of arteries segmented by different approaches on some 2D planes, including common, calcified, and bifurcated, are shown in Fig. 5. Compared to Hessian vesselness based method, our proposed method is highly consistent with manual annotations, especially on those abnormal vascular structures. Note that the DSC and Precision in Table 1 are not that high. It is because the radiologists are somewhat conservative, thereby omitting many thin vessels in annotations, while the algorithms extract the complete coronary tree (including thin vessels).

Average segmentation time of our approach for the complete coronary Artery tree is about 159 seconds per volume on a computer with a NVIDIA GeForce 1080Ti GPU. The Hessian vesselness method consumes about 226 seconds per volume on a computer with 2.1GHz CPU.

4 Conclusion

In this paper, we propose a context aware 3D fully convolutional network for extracting coronary artery tree in the whole cardiac 3D CTA volumes. The proposed approach integrates the strength of deep networks in extracting effective features and spatial prior knowledge constraint in guiding segmentation

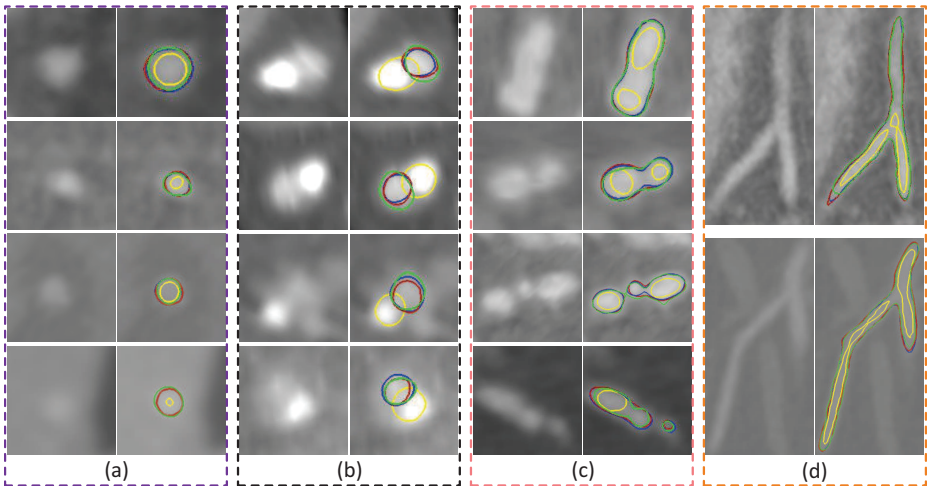


Fig. 5. Comparison of segmented lumen on 2D planes, including annotations (green), Hessian vesselness (yellow), our segmentation with (red) and without (blue) prior. (a) common, (b) calcifications, (c)(d) bifurcations from different projections.

procedure, thus reducing vast majority negative voxels. Our evaluation on 50 CTA volumes shows obvious suppression of vessel-like structures and accurate segmentation of coronary arteries. For further improvement, the spatial prior distribution could be estimated by a more accurate heart alignment algorithm. Besides, Zheng et al. have proved that it is effective to use extracted heart surface as a constraint of coronary artery segmentation [4]. We intend to introduce the heart surface as another anatomical prior knowledge to achieve a further improvement.

Acknowledgement. This work is supported by the National Natural Science Foundation of China under Grant 61622207.

References

1. Roger, V.L., Go, A.S., Lloyd-Jones, D.M., Benjamin, E.J., Berry, J.D., Borden, W.B., Bravata, D.M., Dai, S., Ford, E.S., Fox, C.S., et al.: Heart disease and stroke statistics-2012 update: a report from the american heart association. Circulation **125**(1) (2012) e2–e220
2. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale vessel enhancement filtering. In: MICCAI. (1998) 130–137
3. Lesage, D., Angelini, E.D., Bloch, I., Funka-Lea, G.: A review of 3D vessel lumen segmentation techniques: models, features and extraction schemes. Medical Image Analysis **13**(6) (2009) 819–845
4. Zheng, Y., Loziczonek, M., Georgescu, B., Zhou, S.K., Vega-Higuera, F., Comaniciu, D.: Machine learning based vesselness measurement for coronary artery seg-

mentation in cardiac CT volumes. In: Proc. SPIE 7962, Medical Imaging 2011: Image Processing. 79621K
5. Chen, F., Li, Y., Tian, T., Cao, F., Liang, J.: Automatic coronary artery lumen segmentation in computed tomography angiography using paired multi-scale 3D cnn. In: Proc. SPIE 10578, Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging. 105782R
6. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: MICCAI. (2016) 424–432
7. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: 3D Vision (3DV), 2016 Fourth International Conference on, IEEE (2016) 565–571
8. Zhou, S.K.: Discriminative anatomy detection: classification vs regression. Pattern Recognition Letters **43** (2014) 25–38
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR. (2016) 770–778
10. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollar, P.: Focal loss for dense object detection. In: ICCV. (2017) 2980–2988
11. Yang, X., Bian, C., Yu, L., Ni, D., Heng, P.A.: Hybrid loss guided convolutional networks for whole heart parsing. In: International Workshop on Statistical Atlases and Computational Models of the Heart. (2017) 215–223