Coronary Artery Segmentation by Deep Learning Neural Networks on Computed Tomographic Coronary Angiographic Images

Weimin Huang, Lu Huang, Zhiping Lin, Su Huang, Yanling Chi, Jiayin Zhou, Junmei Zhang, Ru-San Tan, Liang Zhong

***Abstract*—Coronary artery lumen delineation, to localize and grade stenosis, is an important but tedious and challenging task for coronary heart disease evaluation. Deep learning has recently been successful applied to many applications, including medical imaging. However for small imaged objects such as coronary arteries and their segmentation, it remains a challenge. This paper investigates coronary artery lumen segmentation using 3D U-net convolutional neural networks, and tests its utility with multiple datasets on two settings. We adapted the computed tomography coronary angiography (CTCA) volumes into small patches for the networks and tuned the kernels, layers and the batch size for machine learning. Our experiment involves additional efforts to select and test various data transform, so as to reduce the problem of overfitting. Compared with traditional normalization of data, we showed that subject-specific normalization of dataset was superior to patch based normalization. The results also showed that the proposed deep learning approach outperformed other methods, evaluated by the Dice coefficients.**

I. Introduction

Around 6% of adult population is affected by coronary heart disease (CHD), which can often be life-threatening when total occlusions of the artery vessels occur. Computed tomography coronary angiography (CTCA) is increasingly being used for non-invasive diagnosis of CHD. Extraction of coronary artery lumen from CTCA is a critical step for analysis of CTCA images. Manual segmentation performed on each CTCA slices is very tedious and time-consuming. As only 2D slices of medical images are shown to user, annotation of such data with segmentation labels would be difficult without expert knowledge and contextual information. Furthermore, human operators may fatigue easily from doing prolonged annotation. Low productivity, inter-operator variability, and reliance on specialized expertise constitute serious limitations. We believe that recent advances of machine learning especially deep convolutional neural networks (CNNs), have potential to replace or reduce manual segmentation of coronary artery lumen.

In machine learning, deep CNNs have been widely used in image analysis that has showed the capability for to perform complex tasks like object detection, classification, and semantic segmentation. The success is due to the ability of CNNs to learn hierarchical representations of raw input data, without relying on hand-crafted features [1]. In medical imaging, deep CNNs are also being applied for various domains with promising results reported [2-12].

Moeskops et al [7] reported multi-task deep CNNs for multi-organ segmentation. Three 2D-patches at the orthogonal views from 3D volumes are input to multi-layer CNNs. Two fully connected networks (FCNs) then perform pixel-level segmentation. Santini et al [8] applied multi-layer CNNs followed with ReLu layer after each CNN layer for coronary artery calcium segmentation and scoring on 2D slices. Kjerland reported the result of coronary artery segmentation [9] based on 3D DeepMedic architecture [10]. Other CNNs for medical image segmentation include 2D/3D U-Net for cell and brain tumor segmentation [3,4], 3D V-Net [5] and ConvNet (Residue U-Net) [2,12] for prostate, myocardial and blood pool magnetic resonance imaging (MRI) segmentation.

In the literature, coronary artery segmentation has been developed using different image processing approaches — graph-cut, level-set, Gaussian distribution of lumen with level-set, and active contour are some of the methods [13,11], that usually require critical pre and post-processing steps.

This paper presents a deep learning based approach to the coronary artery lumen segmentation of CTCA volume datasets. The basic 3D U-Net architecture is adopted on two settings of coronary data prepared for lumen segmentation, with or without the centerline of vessels. The deep learning neural networks are tested on two datasets respectively. One has 18 CTCA, with partial artery centerline and segmentation annotated. The other one has 34 CTCA data, with the major vessel and branches labelled, without centerline annotation.

Results obtained in our experiments were quantitatively compared with other methods. Without any post-processing, Dice measurement showed that the 3D deep neural networks generated very promising results on both centerline based lumen segmentation as well as segmentation in original CTCA images. All experiments in the paper involving human subjects are approved by the Institutional Review Board.

II. Related works

A. Coronary artery lumen segmentation

Lumen segmentation has been studied by researchers using region growing, graph-cut, level-set and other methods [14,11,13]. Due to inherent CTCA image noise and the inter-subject variations, it is challenging to segment automatically. As summarized in [13], a level-set method

*The work is supported partially by A*STAR BEP grant 152 148 0030, and NMRC grant NMRC/BnB/0017/2015.

Weimin Huang, Su Huang, Yanling Chi, Jiayin Zhou are with the Institute for Infocomm Research, Singapore 138632, e-mail: {wmhuang, huangs, chiyl, jzhou}@i2r.a-star.edu.sg.

Lu Huang, Zhiping Lin are with the School of EEE, Nanyang Technological Univ., Singapore, e-mail:{lhuang016, ezplin}@ntu.edu.sg.

Junmei Zhang, Liang Zhong and Ru San Tan are with National Heart Centre, Singapore e-mail: {zhang.junmei, zhong.liang}@nhcs.com.sg; tan.ru.san@singhealth.com.sg.

978-1-5386-3646-6/18/$31.00 ©2018 IEEE

achieved the best, with certain pre- and post-processing applied to the 2D data. The Dice is up to 0.73 for centerline based lumen segmentation [15], which uses learned tissue probability map on lumen and followed by level-set method for the segmentation. In [11], kernel regression followed by graph-cut method achieved 0.65-0.68 in Dice on lumen segmentation.

A recent effort is using deep neural networks applied for coronary artery segmentation [7], which uses 25 CNNs and 2 FCNs on 3 orthogonal views of the 2D images. It showed on multiple organ segmentation tasks, including coronary artery. While in [9], DeepMedic structure, which is also direct CNNs without decoding, is used for 3D artery segmentation.

B. U-Net: Convolutional Neural Networks

Deep CNNs are a type of deep neural network which are comprised of one or more convolutional layers, followed by one or more fully connected layers as in a standard multilayer neural network. Deep CNNs typically consist of the following layers, convolution, pooling, rectified linear unit (ReLU), fully connected layer with a loss function defined on it. Essentially, U-Net [3] and V-net [13] like structures are similar. Both have contracting paths to encode the data and expansion (up-sampling/deconvolution) paths to decode the data. This encoder-decoder architecture outputs a dense end-to-end segmentation map.

Input image is fed into the network at the beginning. The data are propagated through the network along the layers, and a segmentation output will be obtained at the other end. Most of the data operations are 3x3 convolutions followed by a non-linear ReLU and 2x2 max pooling operation with stride 2 for down sampling. The max pooling operation reduces the size of feature map and as a result, it lowers the spatial resolution by a factor of 2. By following the sequences, we increase the number of feature map and reduce size of the data. The U-net also has an expansion path to decode and restore the high resolution segmentation map. It is a sequence of up sampling layers combined with CNNs. By concatenation with high-resolution features from the contracting path, it decodes the fine details of the objects. Note that the 2x2 up sampling or deconvolution will be followed by a ReLU operation. The loss is defined as negative Dice coefficient: the lower, the better.

The source code applied was an extension to the experiment of Multimodal Brain Tumor Image Segmentation Benchmark (BRATS) [4]. Using Keras, a high-level neural networks API, running on top of Tensorflow as a backend, it allowed us to train U-Nets with original 2D image data as well as reconstructed 3D data.

III. CORONARY ARTERY SEGMENTATION BY DEEP CNNS

Coronary artery centerline, obtained by automatic method and/or manual correction, allowed us to train the CNNs for specific processed CTCA data. The 3D stacks of vessel cross sections can be tailored for the 3D U-Net. The image resolution is around 0.3 to 0.6mm per pixel. A coronary vessel can be as small as 2 pixels in diameter. Thus we could not have a very deep architecture [6]. To encompass the aortic root where the coronaries originate, a suitable cross section window is fixed at 64x64. The inter-slice thickness (pitch) of the 2D images is around 1mm, so we selected 16 cross section images to stack into a 64x64x16 volume for training and testing.

Based on 3D U-Net [3], with the encoder and decoder, a network with 4 CNNs on encoding and 4 CNNs on decoding was used for coronary segmentation. In encoding, 3 MaxPooling layers were used, correspondingly 3 UpSampling layers were used in the decoding path to restore the resolution. Table I shows the main data flow of the architecture.

Layer (Type)	Input Size	Output Size	Kernel Size	Filter Number
Conv3D_1	1x64x64x16	32x64x64x16	3x3x3	32
Conv3D_1	32x64x64x16	64x64x64x16	3x3x3	64
MaxPooing1	64x64x64x16	64x32x32x8	2x2x2	-
Conv3D_2	64x32x32x8	64x32x32x8	3x3x3	64
Conv3D_2	64x32x32x8	128x32x32x8	3x3x3	128
MaxPooling2	128x32x32x8	128x16x16x4	2x2x2	-
Conv3D_3	128x16x16x4	128x16x16x4	3x3x3	128
Conv3D_3	128x16x16x4	256x16x16x4	3x3x3	256
MaxPooling3	256x16x16x4	256x8x8x2	2x2x2	-
Conv3D_4	256x8x8x2	256x8x8x2	3x3x3	256
Conv3D_4	256x8x8x2	512x8x8x2	3x3x3	512
UpSampling1	512x8x8x2	512x16x16x4	2x2x2	-
Concatenate1	512x16x16x4	768x16x16x4	-	+256
Conv3D_5	768x16x16x4	256x16x16x4	3x3x3	256
Conv3D_5	256x16x16x4	256x16x16x4	3x3x3	256
UpSampling2	256x16x16x4	256x32x32x8	2x2x2	-
Concatenate2	256x32x32x8	384x32x32x8	-	+128
Conv3D_6	384x32x32x8	128x32x32x8	3x3x3	128
Conv3D_6	128x32x32x8	128x32x32x8	3x3x3	128
UpSampling3	128x32x32x8	128x64x64x16	2x2x2	-
Concatenate3	128x64x64x16	192x64x64x16	-	+64
Conv3D_7	192x64x64x16	64x64x64x16	3x3x3	64
Conv3D_7	64x64x64x16	64x64x64x16	3x3x3	64
Conv3D_8	64x64x64x16	1x64x64x16	1x1x1	1
Activation	1x64x64x16	1x64x64x16	-	-

Table I 3D U-Net structure for lumen segmentation with centerline known for the CTCA

After convolution, non-linear ReLU and max pooling operation were applied. Our kernel size was 3×3×3, and convolutional layers were separated by 2×2×2 max pooling with stride 2 for down sampling. The input data shape was in size of (1, 64, 64, 16) where 1 represented the number of input channel. After the first UpSampling layer, it concatenated with the output obtained before the MaxPooling layer to produce same size of input. All convolutional layers were followed by one activation layer using sigmoid function.

The loss function was based on the dice similarity coefficient. Assume a volume of N voxels, the predicted binary $p_i \in P$ and the ground truth $g_i \in G$, the dice coefficient between the two binary volumes is written as:

$$\text{Dice} = \frac{2*(P \cap G)}{|P|+|G|},$$

Given p_i or $g_i \in \{0,1\}$, it can be rewritten in following format for easy differentiation:

$$\text{Dice} = \frac{2\sum_i^N p_i g_i}{\sum_i^N {p_i}^2 + \sum_i^N {g_i}^2}.$$

In the case where no pre-extracted centerline was available, we adjusted the networks to process the cube patch

of 32×32×32 cropped directly from the CTCA with the corresponding annotation data for the model training and testing. The main architecture is similar to Table I. Details are omitted here due to space limitation.

IV. Experiments

A. Data Preprocessing

The first data set was from 34 subjects with segmentation labelling but without centerline. We used 30 for training of the deep neural network model and 4 for testing. As the original data is 512×512 per slice, with 200 to 500 slices per CTCA, they were too big for the model training but too few for deep learning. We cropped small 3D patch cubes at 32×32×32 from the original CTCA, with stride 16 for the training. Discarding those with fewer than 160 voxels as vessel in a cube, we obtained 12364 cubes for the U-Net training.

The second set of CTCA images from [13] comprises 18 CTCA volume data from different subjects. The ground truth masks available were not on full vessel trees but for a few vessel segments in each subject. With the centerline, we extracted the stack of images (cross sections) along the centerline to form 3D data. In this way, there were in total 78 image sections and masks from 18 patients' data that could be used for training and testing. To use deep learning training, the data were further cropped into smaller sections along the centerline and data augmentation was performed by cropping data images along x, y and z-axes plus shifting.

After augmentation, all training data patch were cropped in the same size of (64,64,16), by stride of 1 along the center line. We also rescaled the intensity of data to between minimum of 0 and maximum of 2000. For those pixels where intensity was below 800, we changed them to 799, because most useful information on artery resides at intensity values from 800 to 2000. Before training, we normalized the data patches by subtracting the mean value and dividing the standard deviation. 8547 patches in 3D were obtained from 15 subjects' CTCA volumes for the model training. CTCA of another 3 subjects were used for testing. The resolution at the cross section is interpolated using the minimum resolution of the original CTCA data, and the distance between the cross section was around 1mm.

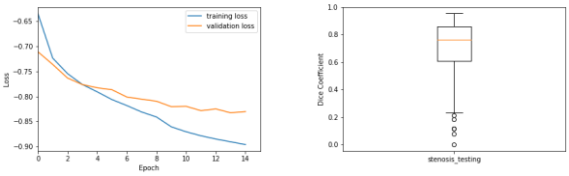
B. Training and validation

In our experiment, all the training and testing were conducted on a workstation, with Intel Xeon E5-2650 @ 2.2G, 192G RAM, and an NVIDIA Geforce GTX 1080 GPU, 8G RAM. It took approximately 6 days and 2 days to complete training respectively. All input data were shuffled and split into training data and validation data at rate 0.8, with 80% data used for training and 20% for validation. In total, we adopted 15 epochs to train our model. Epoch refers to the number of times the algorithm dealt with the entire data set: an epoch was completed each time the algorithm went through all samples in the training set.

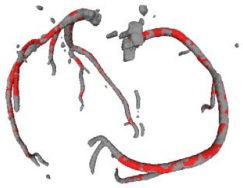
V. Results and Discussion

A. Results obtained from original CT images

On dataset 1 without center line, based on the training data, the training loss was -0.896 and validation loss was -0.829. Tested on another four testing subjects' CTCA data, we obtained the mean of Dice coefficient of 0.7146. Fig.1. shows the results of training and testing: (a) Loss graph of training and validation; (b) Boxplot of dice coefficient on testing cube data, on which the median dice is plotted, which is around 0.75.



(a) Training / validation loss (b) Dice box plot on test data

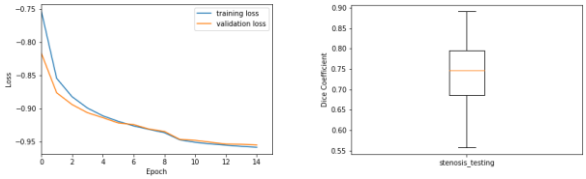


(c) Segmentation by 3D U-Net (gray mesh) on one test data, and Manual annotation (red mesh).

Figure 1 Results of training loss and testing Dice coefficient from original CTCA datasets

B. Results obtained from pre-processed data with centerline

For dataset 2, the training curve of the Dice loss and boxplot are shown in Fig. 2. Fig.2: (a) Loss graph of training and validation; (b) Boxplot of Dice coefficient of testing data. The training loss was -0.958 and validation loss was -0.954. The mean of dice coefficient was 0.7406.



(a) Training / validation loss (b) Dice boxplot on testing data

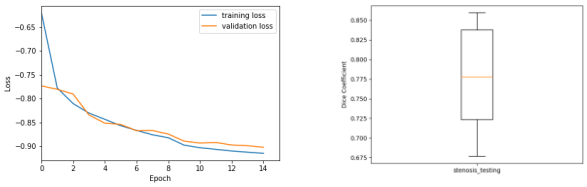
Figure 2 Results of training loss and testing dice coefficient from pre-processed data with centerline

Based on training and testing results from two different datasets, we can clearly see that segmentation obtained from preprocessed data with centerline performed better than the original data. From loss graph, we can also see the validation loss of the first dataset dropped slowly after epoch 4 and only ended up with 0.829. On the other hand, the dropping rate of validation loss from the second dataset was nearly proportional to that of training loss. The final Dice loss improved considerably compared to the first dataset.

C. Improvement

The difference between the training and the testing (Fig. 2) shows the overfitting of the model, probably due to data augmentation. Another explanation is data normalization which is on the patch basis. Due to the variable intensity at different vessel location, it might be better to normalize the data on the subject (CTCA) basis. By doing so, we then window the CTCA intensity between -200 to 1000 in Hounsfield units (HU), and normalize using 70 HU as the

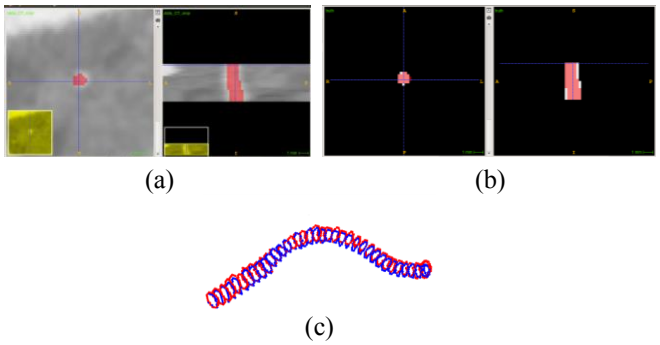
mean lumen intensity and 160 HU as the standard deviation. To prevent overfitting, we discarded the step of shifting as data augmentation. In this way, we obtained 1533 data images for training. The results are shown in Fig.3: (a) Loss graph of training and validation; (b) Boxplot of Dice coefficient of testing data. From the figure, the training loss and validation loss are -0.914 and -0.901, respectively, which seemed not as good as the result in Fig. 2. However, the testing result showed an improved mean of Dice coefficient of 0.7755, 4.7% change compared to the previous model.



(a) Training and validation loss (b) Dice boxplot of testing data

Figure 3 Results of training loss and testing dice coefficient after intensity modification and reduced data augmentation.

As the training data seemed relatively small, we proceeded to use 17 out of 18 subjects for training and left only one for testing, and changed the split rate from 0.8 to 0.9, i.e., 90% for training, 10% for validation. The Dice was 82.9% for the testing. Fig.4. shows a sample of segmentation output: (a) One section of CT image sample with segmented prediction as mask; (b) Segmented prediction overlap ground truth.



(a) (b)

(c)

Figure 4 Predicted segmentation output shown over CT and ground truth mask. (a) CT image with the prediction mask, (b) Ground truth label (white) and the prediction (red), (c) extracted contour in physical world, red contours are ground truth and blue contours are the prediction.

Table II shows the mean Dice using different methods, measured against the ground truth labels that are manually generated by experienced observers. Compared with others, where the Dice ranged from 0.59 to 0.73, the 3D deep learning approach presented in this paper showed much better results despite the training and testing data are still limited.

Our Method (DL, 3D, with or without centerline)	[7] DL, 3 views of 2D CTCA (no centerline)	[9] DL, DeepMedic, 3D (no centerline)	[11] Regression, graph-cut, (with centerline)	[15] Tissue probability, level-set (with centerline)
0.71-0.78	0.65	0.60	0.65-0.68	0.69-0.73

Table II Comparison of methods by mean Dice

VI. Conclusion

In this paper, we presented approaches to coronary artery lumen segmentation using 3D U-Net for both CTCA data with and without centerline. We compared the segmentation performances based on dice coefficient of testing batch from two datasets. The results showed that preprocessed data with centerline trained a better model than the original data. It also showed a better testing output with Dice of 0.8291. In experiments we showed that simple data augmentation might not benefit the testing data. From the training curve, we believe that with better quality training data, the performance of coronary segmentation can be further improved.

References

[1] Matthew D. Zeiler, R. Fegus, "Visualizing and understanding convolutional networks," ECCV 2014, pp. 818–833.

[2] L. Yu, J.-Z Cheng, Q. Dou et al. (2017) Automatic 3D Cardiovascular MR Segmentation with Densely-Connected Volumetric ConvNets. MICCAI 2017. LNCS 10434, pp.287-295.

[3] O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, LNCS, Vol.9351: 234—241. (2015)

[4] B. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, et al., "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," IEEE Transaction on Medical Imaging, 2014, pp.1993-2024.

[5] F. Milletari, N. Navab, and S.-A. Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. 2016, arXiv:1606.04797v1 [cs.CV].

[6] W. Mengqiao, Y.Jie, C.Yilei, et al., "The multimodal brain tumor image segmentation based on convolutional neural networks," IEEE international conference on Computational Intelligence and Applications (ICCIA), 2017.

[7] P. Moeskops, J.M. Wolterink, B.H.M. van der Velden, K.G.A. Gilhuijs,T. Leiner, M.A. Viergever, and I. I_sgum, "Deep learning for multi-task medical image segmentation in multiple modalities," MICCAI 2016, Part II, pp. 478-486

[8] G. Santini, D. Della Latta, N. Martini et. al., "An automatic deep learning approach for coronary artery calcium segmentation," arXiv:1710.03023v1 [cs.CV] 9 Oct 2017.

[9] Øyvind Kjerland, "Segmentation of Coronary Arteries from CT-scans of the heart using Deep Learning," Thesis for M.Sc in Computer Science, Norwegian Univ of Sci and Tech, June 2017.

[10] K. Kamnitsas, C. Ledig, V. F.J. Newcombe et. al., "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," Medical Image Analysis, 36 (2017), pp.61-78.

[11] R. Shahzad, H. Kirişli, C. Metz et. al., "Automatic segmentation, detection and quantification of coronary artery stenoses on CTA," Int J Cardiovasc Imaging (2013) 29:1847–1859.

[12] L. Yu, X. Yang, H. Chen, J. Qin, P.-A. Heng, "Volumetric ConvNets with Mixed Residual Connections for Automated Prostate Segmentation from 3D MR Images, AAAI'17, pp.66-72.

[13] H. A. Kirişli, M. Schaap, C. T. Metz, et al. "Standardized evaluation framework for evaluating coronary artery stenosis detection, stenosis quantification and lumen segmentation algorithms in computed tomography angiography," Medical Image Analysis, 17 (8), 2013, pp.859-876.

[14] P. Kitslaar, M. Frenay, E. Oost E. et. al., "Connected component and morphology based extraction of arterial centerlines of the heart (CocomoBeach)," MICCAI Workshop S4, 2008.

[15] B. Mohr, S. Masood, C. Plakas, "Accurate stenosis detection and quantification," in coronary CTA," Proc. of MICCAI Workshop on 3D Cardiovascular Imaging: a MICCAI segmentation Challenge, 2012.