2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Coronary Artery Extraction from CT Coronary Angiography with Augmentation on Partially Labelled Data

Ziqing Wan, Weimin Huang^, Su Huang, Zhongkang Lu, Liang Zhong, Zhiping Lin

Abstract— **Coronary artery disease (CAD) is an important cause of morbidity and mortality. CT coronary angiography is considered as first-line of investigation in patients suspected of having CAD. Coronary artery centerline extraction is a challenging prerequisite for coronary artery stenosis evaluation. These challenges include the small and complex structure, variation of plaques and imaging noise. Deep learning methods often require adequate annotated data to build a good model. This work aims to adopt a dataset that has partial annotation to augment the data to train a Coronary Neural Network (CorNN) to track the coronary artery centerline. We combined a small training dataset with densely labelled centerline and radius, augmented with a larger dataset with only the centerline sparsely labelled to train networks to track centerlines from 3D computed tomography coronary angiography. The vessel orientation estimation is patch based, with or without additional radius prediction. The patch data are carefully positioned and sampled, which are tagged with the orientations computed based on the centerlines. Our experiment results show that, with the augmentation of the new data, although partially annotated, nearly 10% or more improvement has been achieved for the coronary artery extraction by the proposed approach.**

I. Introduction

According to the World Health Organization, 17.5 million lives are taken each year by CVDs (Cardiovascular diseases), an estimated 31% of all deaths worldwide [1]. Coronary artery disease (CAD), the most common type of CVD, is the narrowing or blockage of coronary arteries. The cause of CAD is the plaque deposited in the walls of coronary arteries. The buildup of plaque reduces the supply of blood or in some cases block blood flow to heart muscles, which leads to heart attack and even heart failure [2].

For patients suspected of having CAD, computed tomography coronary angiography (CTCA) can be used to examine and evaluate the severity of coronary stenosis. However manual coronary artery tree extraction is not only time-consuming but also skill demanding. Therefore, researchers developed (semi-) automatic approaches that can minimize the user interaction. In [3], Tek et al proposed a graph-based algorithm that used a multi-scale medialness filters and minimum-path algorithm to compute the tree of coronary artery centerlines. The algorithm could lead to short false positive section and leak to veins. Other conventional methods proposed include semi-automatic or fully automatic methods as summarized in [4], by region growing, fitted shape and model, tube detection, connectedness tree algorithm, and minimum path approaches. In [5], Yang et al proposed a method based on improved vesselness filter. An interesting idea is to use rays casted to a sphere to describe the geometric features.

Wolterink et al [6] proposed a deep learning-based approach to centerline extraction. The method trained a convolutional neural network (CNN) to predict the direction and radius of coronary artery centerlines. With ostia and candidate centerline seed point detection, fully automatic coronary artery centerlines can be extracted. No hand-crafted filter or appearance feature is needed. However, this method requires the training datasets with good labels of centerlines and the radius, not only for the radius estimation but also for guidance to process the training data to obtain the ground truth of the orientations. We adopted the networks and used data augmentation for the task by tracking the centerline only without the radius estimation. As a result, we improve the tracking accuracy by leveraging on a larger dataset without the tedious radius labeling on the centerlines by a simple CorNN.

II. Related Works

A. *Minimal Cost Path (MCP)*

Many past works are based on Minimal Cost Path (MCP) method e.g. the work in [7], which is based on two manually selected points along the coronary artery, and a minimal cost on a path between two predefined points on each centerline. However, MCP is easily affected by other short-cut paths between points.

B. *3D segmentation approach*

The type of methods obtains the tree of coronary arteries using lumen segmentation first, and extract the centerlines following the segmentation of vessel [8]. This method needs to have a segmentation model [9]. It is time-consuming because the lumen segmentation is done in the full 3D CT if no special preprocessing is applied to limit the searching space.

III. Method

A. *CNN for Direction and Radius Prediction*

The original CorNN predicts the direction and radius of an input vessel patch [6]. The output consists of 500 channels corresponding to all the possible 500 directions and 1 channel for radius. The determination of direction becomes a problem of classification instead of regression, while the determination of radius is a regression problem. The architecture of this network is shown in Table I.

B. *CNN for Direction Prediction only*

For weakly annotated data, we have only the centerlines without the radius. Thus, the network can predict only the directions of the centerline given a patch around a centerline

*Research is supported partially by FS-2020-016 and MOH-000358 by NMRC.

Weimin Huang, Su Huang, Zhongkang Lu are with the Institute for Infocomm Research, A*Star, Singapore. Ziqing Wan and Zhiping Lin are with the School of EEE, Nanyang Technological University, Singapore. Liang Zhong is with National Heart Centre Singapore and Duke-NUS Medical School. (^Corresponding author: wmhuang@i2r.a-star.edu.sg).

point. The output of last layer becomes a 500-dimensional vector corresponding to all the possible 500 directions. There is no output for radius. The architecture of this network is adjusted in Table II.

TABLE I. CNN FOR DIRECTION AND RADIUS PREDICTION

Layer	Input	Output	Kernel	Dilation
1	1x19x19x19	32x17x17x17	3x3x3	1
2	32x17x17x17	32x15x15x15	3x3x3	1
3	32x15x15x15	32x11x11x11	3x3x3	2
4	32x11x11x11	32x3x3x3	3x3x3	4
5	32x3x3x3	64x1x1x1	3x3x3	1
6	64x1x1x1	64x1x1x1	1x1x1	1
7	64x1x1x1	501x1x1x1	1x1x1	1

TABLE II. CNN FOR DIRECTION PREDICTION ONLY

Layer	Input	Output	Kernel	Dilation
1	1x19x19x19	32x17x17x17	3x3x3	1
2	32x17x17x17	32x15x15x15	3x3x3	1
3	32x15x15x15	32x11x11x11	3x3x3	2
4	32x11x11x11	32x3x3x3	3x3x3	4
5	32x3x3x3	64x1x1x1	3x3x3	1
6	64x1x1x1	64x1x1x1	1x1x1	1
7	64x1x1x1	500x1x1x1	1x1x1	1

C. Loss Functions

Cross-entropy loss [10] is used to optimize the model by minimizing the difference between two probability distributions. Given y, and $\hat{y}$ where y is the ground truth and $\hat{y}$ is the predicted value. The cross-entropy function is defined by

$$L(y, \hat{y}) = -(ylog(\hat{y}) + (1-y)log(1-\hat{y}))$$

D. Improvement Using Partially Labeled Data

The size of fully labeled dataset (with both centerline and radius at each centerline point) CTCA08 [11, 6] is small. With only CTCA08 for training, the performance of the trained model is not good enough for an automatic approach. One way is to enlarge the size of the training data e.g., including another CTCA13 dataset [4]. However, CTCA13 dataset comes with only centerline without radius information.

To address this problem, we can manually set the value of the radius used for CTCA13 dataset as 3mm, which is a slightly larger than the actual vessel radius of most coronary arteries. The radius R of a sphere at each centerline point x determines the associated direction to its previous point and the direction to its next point located on the sphere. The two directions will be set as the ground truth direction for the training patch. The reasons are as follows to set a radius slightly larger than the actual radius, so it does not affect much on the computing of the proper direction d.

If the radius is much larger than the actual one, it will result in the inability to find the correct orientation and points given a patch as shown in Figure 1(a), especially for the skewed arteries.

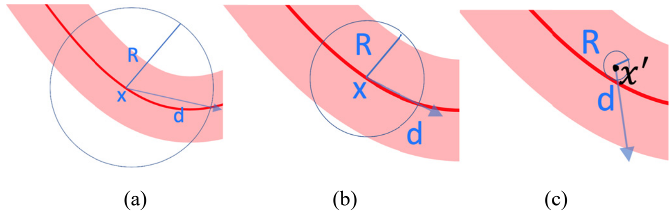

Figure 1. (a) Direction d from point x to next point with R >> actual radius. (b) Direction d from point x to next point with R > actual radius. (c) Direction d from point x' to next point with R << actual radius.The orientation is totally wrong (with augmentation).

For setting a radius slightly larger than the actual one, the correct directions and the next neighboring centerline point can still be found along the direction, shown in Figure 1(b).

Augmentation with Partially Labeled Data

Most of the time, machine learning method can achieve much better result with increased training samples. For centerline tracking [6], it augments the samples by shift and rotation around the reference centerline points.

If the radius is smaller than the actual one, after augmentation, the current point is shifted to x' which is not on the centerline. The previous point and the next point of x' may be on the upper or lower side of x'as shown in Figure 1(c), which outputs a wrong orientation for the training patch.

- Shift: To get augmented samples that do not locate at the centerlines exactly, it shifts the center point randomly using a 3D normal distribution with $\mu = 0.0$ and $\sigma = 0.25R'$. *R'* is the preset vessel radius.

For the partially labeled data when the augmented center point is not on the centerline, the value of radius has a greater impact on the computation of the vessel prediction directions. *R'*=3 mm is too large for data augmentation for small vessels. Thus, for data augmentation using partially labeled data, *R'*=1 mm is used.

- Rotation: After shifting center points and obtaining the 3D patch centered at the new point, it rotates the patch randomly to generate more patch samples.

IV. EXPERIMENTS

A. Data Preprocessing

Two datasets were used in the experiments, i.e., CTCA08 dataset [11, 6] and CTCA13 dataset [4].

There are 8 data available in CTCA08 dataset (dataset00 to dataset07) labeled. Each of them contains one CTCA image and labels of a few vessels as reference centerlines with the centerline coordinates and radius of the vessel at each point. Due to no ground truth available in the rest set in CTCA08, we use leave-one-out for training and testing. In each round of training, we also keep one subject data as validation set. In CTCA13 dataset, we use 18 data of them for our study due to the training time limitation. Each of them contains one 3D CTCA scan and 4 to 5 vessels with labeled centerline coordinates, without radius information. All CT images are resampled with isotropic resolution 0.5mm.

The following steps are repeated for all the vessels of the 8 datasets from CTCA08 dataset.

1. For each current center point P_c, another two center points P_k and P_h related to P_c are searched as follows and recorded as the previous point and the next point of P_c. Firstly, P_c is taken as the starting point and we look backward along the centerline. When the distance between P_c and the found point is slightly larger than the given radius R, the found point is retained as the previous point P_k. Then P_c is still taken as the starting point, but we look forward along the centerline. Another point is found and retained as the next point P_h.
$$0 < (P_c\text{-}P_k) - R < 0.01$$
$$0 < (P_h\text{-}P_c) - R < 0.01$$
2. For each P_c located at the centerline, a 3D image patch of size 19 × 19 × 19 centered at P_c is extracted. Its two referenced directions need to be calculated as follows. A sphere with the given R centered at P_c is used to store all the 500 possible directions by distributing 500 points evenly on its surface [6]. For each point of the 500 points, there is a vector from P_c to it, with orientation $D_i, i = 1, \ldots, 500$. The two nearest vectors in D_i to the vector P_k-P_c and P_h-P_c are set as the vessel orientation at the patch with the probabilities of these two directions as 0.5. All the other 498 directions have set with probability zero.

For CT images in CTCA13, the radius information is missing. We assign a reasonable value R=3mm as the radius to get P_k and P_h for vessel orientation estimation. And R'=1mm for augmented data selection.

The centerline tracking in all experiments is fully automatic, by using the centerline seed point detection and ostia detection proposed in [6] to initiate the tracking.

B. Training, Validation and Testing

Data batches are generated from CTCA08 dataset and CTCA13 dataset. Cross-validation is used to estimate how accurately the predictive model will perform in practice.

V. Results and Discussions

A. Predicted Centerlines visualization

The reference centerlines and the respective predicted centerlines of one example of CTCA08 dataset are plotted in Figure 2. We verified that it actually predicts more correct arteries than the reference, shown in Figure 2 (right image). In CTCA08, we will use 6 subjects in training, 1 in validation and 1 in testing. All CTCA13 data are used in training.

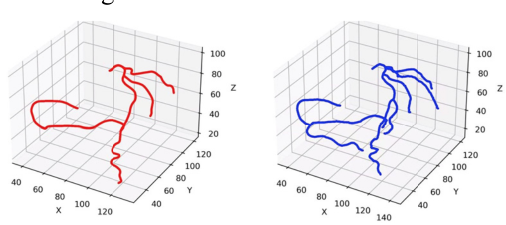

Figure 2. The reference centerlines (left) and the predicted results (right)

B. Evaluation Measures

Centerline extraction performance is evaluated using the following three metrics [11].

Overlap (OV): Overlap between a tracked coronary artery centerline and a manually annotated centerline.

Overlap until first error (OF): Before an error occurs, how much of a coronary artery centerline has been extracted accurately.

Overlap with the clinically relevant part of the vessel (OT): Overlap between a tracked coronary artery centerline and a manually annotated centerline whose vessels are both clinically relevant. A segment of a vessel is clinically relevant if its diameter is larger than or equal to 1.5 mm.

C. Results

The results for centerline extraction using CTCA08 dataset with radius information are shown in Figure 3, which includes average OV, OF, and OT over each dataset. The average OV, OF and OT of all the vessels from CTCA08 dataset are 0.8104, 0.6002, and 0.8193 respectively.

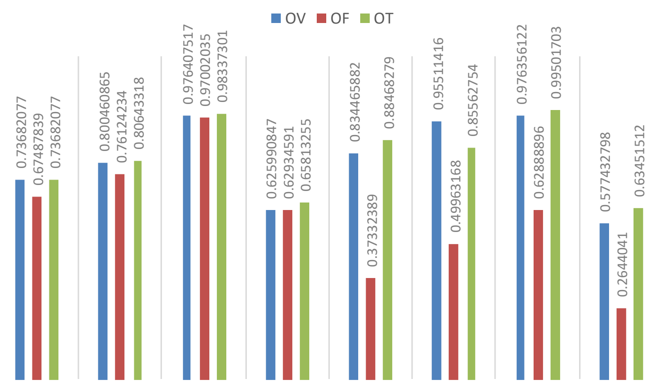

Figure 3. Average OV, OF, and OT of each dataset trained with CTCA08 dataset with radius information

Ideal OV, OF or OT is 1 for perfect tracking. Many of the above results showed good performance with the metrics more than 0.90. However, for some data, the OV, OF and OT values are much smaller than one. Main reasons for the low accuracy are as follows.

1. The CT scan has poor image quality due to motion. The CT images at arteries may also not be enhanced properly.
2. The intensity of veins and arteries in CT images can be close and, in some subjects, the veins can be very near to arteries, which is leading to erroneous tracking.
3. For some datasets, there are many soft and calcified plaques on the arteries.
4. The training datasets are limited.

The results for centerline extraction trained using CTCA08 Dataset with raidus information and augmented with CTCA13 Dataset without radius inforamtion are shown in Figure 4. The average OV, OF and OT of all the vessels from CTCA08 dataset are 0.9108, 0.7611, and 0.9052 respectively. Compared to the results using only CTCA08 dataset, the average OV, OF and OT have been improved by 10%, 16%, and 8.6%.

To compare the accuracy of centerline extraction with and without radius information, the results for centerline extraction using CTCA08 Dataset and CTCA13 Dataset both without radius information are shown in Figure 5. The average OV, OF and OT of all the vessels from CTCA08 dataset are 0.8100, 0.6025, and 0.8494 respectively. The total mean results for OV and OF are close to the trained model using CTCA08, while we see 3% increase in OT.

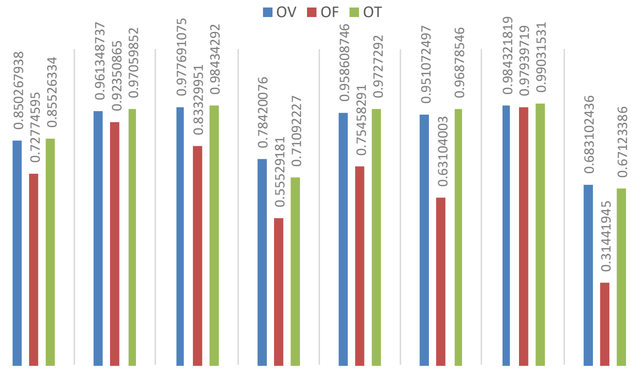

Figure 4. Average OV, OF, and OT of each dataset trained with CTCA08 dataset with radius information and CTCA13 dataset without radius information

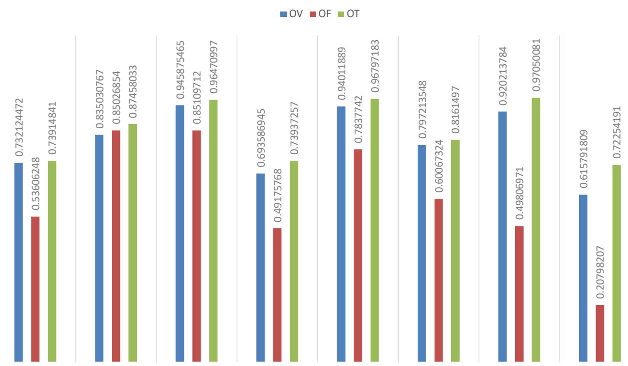

Figure 5. Average OV, OF, and OT of each dataset trained with CTCA08 dataset and CTCA13 dataset both without radius information

D. Discussions

The average OV, OF and OT over all testing datasets using the above three types of augmentations are shown in Figure 6. It is obvious that for each evaluation metrics, the training with the small dataset with centerlines and full radius information, and augmented with the additional partially labelled larger data achieved the best performance, with nearly 10% or more increases over the original method on the limited data. With more data, even with partial labels, we can still see 3% improvement in one of the metrics while maintain the similar performance for the other two evaluation measures, compared to the model using a small dataset with full labels.

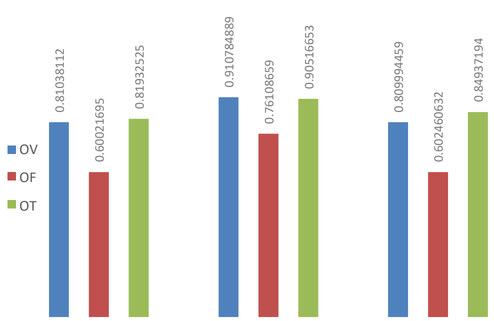

Figure 6. Result Comparison on models trained with different datasets and augmentations

VI. Conclusion

In this paper, we proposed to augment with partially labeled data for centerline tracking. Fully labeled dataset has both centerline and radius information. Partially labeled dataset has only centerline. Two radii were proposed here for data preparation, one is for the direction estimation and the other smaller radius was proposed to sample additional valid patches for training.

The first experiment used a fully labeled dataset with direction and radius information. The second training used a fully labeled dataset augmented with a partially labeled dataset. To compare the accuracy of centerline extraction with and without radius information, the third training used both datasets but without the exact radius information.

Experiments showed that the model trained on the limited fully labeled dataset and additional data with partial labeling (a simple CorNN) has much higher prediction performance, compared to the original (CorNN) method using the limited fully labeled dataset. While the model trained on the two datasets but with only partial labeling can still outperform the first model which was trained on the limited fully labeled data.

References

[1] A. H. A. Mawali, “Non-Communicable Diseases: Shining a Light on Cardiovascular Disease, Oman’s Biggest Killer,” *Oman Medical Journal*, vol. 30, pp. 227-228, Jul. 2015.

[2] I. Kawachi, “Stress and the Heart: Psychosocial Pathways to Coronary Heart Disease,” *British Medical Journal*, vol. 324, pp. 304, Jan. 2002.

[3] H. Tek, M. A. Gülsün, S. Laguitton, L. Grady, D. Lesage and G. FunkaLea, ‘‘Automatic coronary tree modeling,’’ *Insight J.*, pp. 1-8, Aug. 2008.

[4] H. A. Kirişliab, M. Schapa, C. T. Metza, et al, “Standardized evaluation framework for evaluating coronary artery stenosis detection, stenosis quantification and lumen segmentation algorithms in computed tomography angiography,” *Medical Image Analysis*, vol. 17, pp. 859-876, Dec. 2013.

[5] G. Yang, P. Kitslaar, M. Frenay, A. Broersen, M. J. Boogers, J. J. Bax, J. H. C. Reiber and J. Dijkstra, “Automatic centerline extraction of coronary arteries in coronary computed tomographic angiography,” *Int J of Cardiovascular Imaging*, vol. 28, pp. 921-933, Jun. 2011.

[6] J. M. Wolterink, R. W. van Hamersvelt, M. A. Viergever, T. Leiner and I. Išgum, “Coronary artery centerline extraction in cardiac CT angiography using a CNN-based orientation classifier,” *Medical Image Analysis*, vol. 51, pp. 46-60, Jan. 2019.

[7] O. Wink, A. F. Frangi, B. Verdonck, M. A. Viergever and W. J. Niessen, “3D MRA coronary axis determination using a minimum cost path approach,” *Magnetic Resonance in Medicine*, vol. 47, pp. 1169-1175, 2002.

[8] H. Cui, D. Wang, M. Wan, J. Zhang, X. Zhao, R. Tan, W. Huang, W. Xiong, Y. Duan, J. Zhou, T. Luo, G. S. Kassab and L. Zhong, “Fast Marching and Runge–Kutta Based Method for Centreline Extraction of Right Coronary Artery in Human Patients,” *Cardiovascular Engineering and Technology*, vol. 7, pp.159-169, Jun. 2016

[9] W. Huang, L. Huang, Z. Lin, S. Huang, Y. Chi, J. Zhou, J. Zhang, R. Tan and L. Zhong, “Coronary Artery Segmentation by Deep Learning Neural Networks on Computed Tomographic Coronary Angiographic Images,” *IEEE MEBC’18*, pp. 608-611, Jul. 2018.

[10] S. Jadon, “A survey of loss functions for semantic segmentation,” arXiv preprint arXiv:2006.14822, Jun. 2020.

[11] M. Schaap, C. T. Metz, T. van Walsum, A. G. van der Giessen, A. C. Weustink, N. R. Mol-let, C. Bauer, H. Bogunovic, C. Castro, X. Deng, et al, “Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms,” *Medical Image Analysis*, vol. 13, pp. 701-714, Oct. 2009.