



Segmentation to Label: Automatic Coronary Artery Labeling from Mask Parcellation

Zhuowei Li¹, Qing Xia¹(✉), Wenji Wang¹, Zhenan Yan², Ruohan Yin³, Changjie Pan³, and Dimitris Metaxas⁴

¹ SenseTime Research, Beijing, China
xiaqing@sensetime.com

² SenseBrain Technology Limited LLC, San Jose, USA

³ Changzhou No.2 People's Hospital, Changzhou, China

⁴ Rutgers University, New Jersey, USA

Abstract. Automatic and accurate coronary artery labeling technique from CCTA can greatly reduce clinician's manual efforts and benefit large-scale data analysis. Current line of research falls into two general categories: knowledge-based methods and learning-based techniques. However, no matter in which fashion it is developed, the formation of problem finally attributes to tree-structured centerline classification and requires hand-crafted features. Here, instead we present a new concise, effective and flexible framework for automatic coronary artery labeling by modeling the task as coronary artery parsing task. An intact pipeline is proposed and two paralleled sub-modules are further designed to consume volumetric image and unordered point cloud correspondingly. Finally, a self-contained loss is proposed to supervise labeling process. At experiment section, we conduct comprehensive experiments on collected 526 CCTA scans and exhibit stable and promising results.

Keywords: Coronary artery labeling · Parsing · Point cloud · Deep learning.

1 Introduction

Cardiovascular disease has long been a leading death reason worldwide. Coronary computed tomography angiography (CCTA) as a non-invasive imaging technique for diagnosis of coronary artery diseases has been widely used by physicians and radiologists. In a standard workflow, one of the most crucial and fundamental step is to correctly label branches by their anatomical names. Automatic and accurate coronary artery labeling technique can greatly reduce clinician's manual efforts and benefit large-scale data analysis.

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-59861-7_14) contains supplementary material, which is available to authorized users.

Among multiple concerns in automatic coronary artery labeling, the main challenge related is large individual variances among subjects. Current line of research falls into two general categories: knowledge-based methods and learning-based techniques. However, no matter in which fashion it is developed, the formation of problem finally attribute to tree-structured centerline classification. Previous methods are either matching target centerlines with pre-defined reference model [1, 3, 4, 15] or do segment-level classification using hand-crafted features [12, 13]. Most recently, Dan wu et al. [13] employed bi-directional tree-lstm to fully exploit topological information contained within the tree-structure using some hand-crafted features. Now the question we proposed here is: Does tree-structured centerline contains all useful information? Especially when the input features are manually collected. Clinically, the anatomical naming of a coronary artery relies on it’s functional effect that is for which cardiac field it supplies. Therefore, we conjecture that the scope of current methods may not contain all informative information and thus may be sub-optimal. Considering the fact that centerlines are commonly extracted using minimal-path or skeletonization algorithm from pre-segmentated masks or similar pixel-level representative space [5, 6, 8, 12, 16], so instead of dealing with centerlines, we here take a step back and reverse the order. We model the task as the coronary artery parsing problem and directly operate on coronary artery masks and then map back to centerlines for subsequent utilization. By doing so, we encode not only topological informations, but also morphological information and geometrical information as a whole. And the training ROIs are extensively enlarged comparing with pure centerline points. Moreover, the new pipeline can been implemented in an end-to-end deep-learning fashion without hand-crafted engineering.

In this paper, we demonstrate a new pipeline for automatic coronary artery labeling by constructing the task as parsing issue. Specifically, two sub-frameworks are proposed to consume volumetric images and point clouds correspondingly. A universal loss and voting strategy is designed for both sub-frameworks. Finally, we evaluate our proposed ideas and frameworks on collected CCTA scans and demonstrate promising results.

2 Methods

2.1 Mask Space and Input

Under mask space, all inputs have been readily divided and regrouped, thus the complexity of the learning process is largely downgraded. Since acquisition of coronary artery mask is also a prerequisite for extracting centerlines under common circumstances, it is an straightforward and labor-saving choice to build our framework on mask space instead of original space. In order to encode functional information of coronary artery, the mask of whole cardiac has also been extracted and merged with coronary artery mask. Then an isotropic patch with fixed size is center-cropped at mass center of coronary artery mask as the final input.

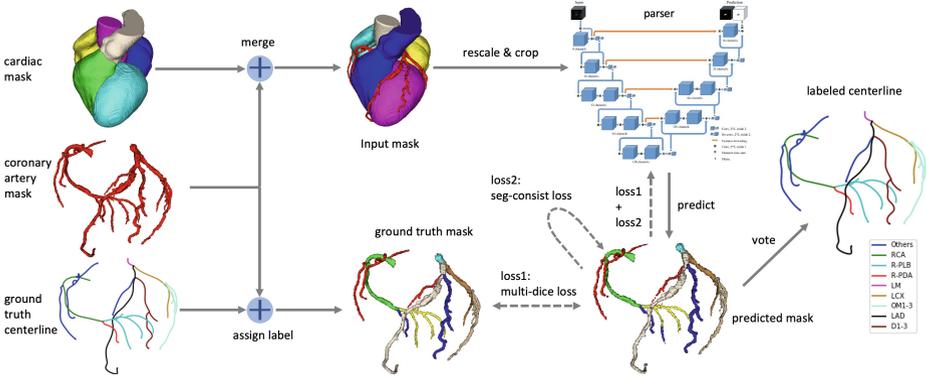


Fig. 1. General workflow for coronary artery parsing.

2.2 Volumetric Coronary Artery Parsing

Unlike objects in natural images, coronary artery branches in CTA image are revealed as tenuous and twisty tubular structure and they are all also tightly connected. This character makes anchor-based two-stage parsing techniques a failure in our case. Considering the fact that coronary arteries contain relative consistent components, here we model the task as multi-label segmentation task. In order to successfully extract structural information and stay robust against large individual variances, the most essential insight is to keep the view of the whole environment, that is the receptive field should contains whole input. So instead of cutting input volume into multiple cubes, here we down-sample the input to fit the memory. Due to the simplicity of mask space, large scale down-sampling operation is acceptable. Figure 1 elaborates the overall workflow of our proposed framework. As showed in Fig. 1, we employed a modified version of V-net [14] as our parser. Any other effective architecture can also substitute this backbone. During training, ground truth coronary artery mask is generated as follows: we first annotate tree-structured centerlines in segment-level, then each voxel in coronary artery mask is assigned with the same label as it’s nearest centerline point. Since the problem formation has been set as multi-label segmentation (pixel-level classification), multi-categorical dice loss is deployed to supervise the learning process. Even though dice-loss provides a good supervision on categorical segmentation, it neglects geometrical and semantic informations. One of the most challenging issue under coronary artery mask parsing is the chaotic predictions within a branch. Especially at bifurcation area where multiple branches are bordered. In order to address this issue, a self-contained loss called seg-consist loss is proposed. The intuition behind the loss design is based on a prior cognition: *points’ labels within the same coronary artery segment should be consist*. Formally, the loss is designed as:

$$Loss_{seg_entropy} = \sum_{n=1}^N \lambda_n \left(\frac{\sum_{a \in A_n} -p(a|\mathbf{S}_n) \log_2 p(a|\mathbf{S}_n)}{\log_2 |A_n|} \right) \quad (1)$$

where A_n is the predicted label space for segment S_n and $p(a|S_n)$ is the probability of predicting label a within segment S_n . In general, normalized entropy of predicted labels within each segment is calculated, then entropies of all segments are weighted summed. We set $\lambda_n = \frac{1}{N}$ across our experiments to provide equal attentions for all segments regardless of their sizes. In general, our final loss function is defined as:

$$Loss = \alpha Loss_{multi_dice} + (1 - \alpha) Loss_{seg_entropy} \quad (2)$$

α in Eq. (2) is experimentally set to 0.6.

2.3 Back to Centerline

Herein, a two-step voting strategy is developed naturally to map coronary artery masks back to centerlines. Specifically, the label for the target segment is defined as:

$$S_i = mode\{mode(D_{ij}), mode(D_{ij+1}) \dots mode(D_{ij+n})\} \quad (3)$$

D_{ij} is the neighbor space of point j in segment i . In our framework, the neighborhood is set as a $3 \times 3 \times 3$ cube. To sum up, we perform a point-level voting following by a segment-level voting. By doing so, the final labeling result is impressively robust against noises and segmentation corruption. As long as the majority group remains correct, the final label will be sound. After assigning labels for all segments, segments with the same label will be extracted to compose final branches. In order to keep branch from breaking apart under exceptional situation, one step post-processing is performed to connect two separate segments if they have the same label and the segment between them is marked with different label.

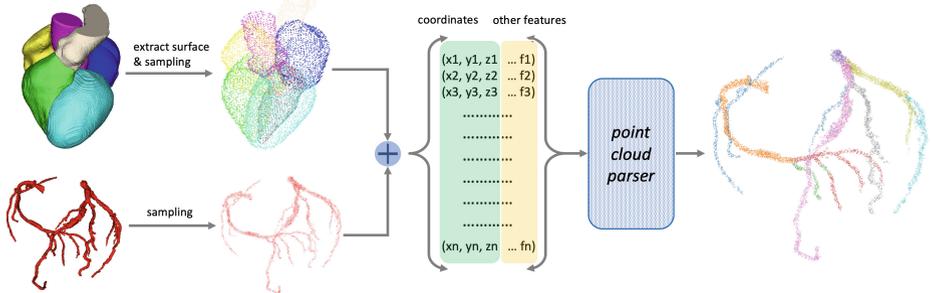


Fig. 2. Framework with point cloud input, voting step and loss calculation is omitted for the simplicity.

2.4 Point Cloud Extension

Despite the conciseness of proposed framework, it is in fact an computing and memory hungry implementation. Down-sampling and crop operations are needed to fit volumetric input into the GPU. The bottleneck causing this issue is the existence of vast less-informative background voxels. So here we also formulate parsing process from another perspective, instead of viewing input as volumetric image, we treat foreground masks as points cloud and get rid of backgrounds. Formally, we are aiming to assign k scores of k candidate categories for each point in a disordered 3D points set $\{P_i | i = 1, \dots, n\}$. By doing so, down-sample and crop operations are no longer needed. Figure 2 displays point cloud version of our framework. Here we deploy PointNet/PointNet++ [9, 10] as our point cloud parser. Since cardiac surface has already contains enough information, we keep only surface of the cardiac mask in order to save memories and speed up learning process. And above proposed seg-consist loss and two-step voting steps can be swiftly transferred to point cloud version without modification. Detailed comparisons between volumetric version and point cloud version are demonstrated in experiment section.

3 Experiments

This section will be organized in 3 subsections. We first introduce our dataset and metric used, then we present the configuration of our experiments and compare our results with other existing methods. Finally we substantiate our proposed ideas and heuristics by conducting ablation studies.

3.1 Dataset

We collect 526 multi-modality CCTA scans from multiple clinical institutes. Masks for 280 collected images are manually annotated by radiologists using self-developed annotation platform. Masks of remaining images are predicted by the segmentation model trained with manually annotated masks. Then all predicted masks are further scrutinized and revised by radiologists using annotation platform. Tree-structured centerlines are extracted using modified version of TEASAR [11]. Finally all centerlines are manually annotated in segment level. Specifically, we annotated RCA, PLB(R/L), PDA(R/L), LM, LAD, LCX, D1–2, OM1–2. All left-over segments are annotated as Others. Ground truth coronary artery masks are then generated as described in Sect. 2.2. Besides coronary artery masks, masks for whole cardiac are also collected, a model is trained using annotations from MM-WHS challenge [17]. Then masks of all 526 images are acquired increasingly using the similar strategy as mentioned above. The dataset is further split into 200 training set and 326 test set.

Our final results are evaluated in two ways, firstly a branch-level precision are judged manually by experts. Then a segment-level statistic result is evaluated using precision, recall and f1 score.

Table 1. Results of previous work comparing with ours, numbers here are the same as those collected by Wu et al [13]. P represents precision and R stands for recall. Our results reported here are in branch level, Ours-v is the volumetric version of our methods and Ours-pc is the point cloud version.

Study	Akiyemi et al. [1]	Yang et al. [15]	Gülsün et al. [4]	Cao et al. [2]	Wu et al. [13]	Ours-pc	Ours-v
Subject	52	58	37	83	44 (each fold)	326	326
Metric	R	P	P	P	P	P	P
LM	100.0	99.3	100.0	100.0	99.1	92.6	95.7
LAD(p/m/d)	97.4	93.4/86.8/93.4	99.2/97.1/100	93.6/85.8/95.4	96.9	99.7	99.1
LCX(p/m/d)	91.7	84.6/80.3	99.2/99.2/83.9	87.3/83.2	93.5	96.9	95.1
RCA(p/m/d)	98.9	97.8/94.1/92.7	100	85.1/82.3/92.5	96.0	96.9	97.6
D(1-2)	80.0	100/86.8	91.2/83.2	93.5/82.2	91.0	85.0	86.5
OM(1-2)	78.9	86.1/78.8	90.5/83.2	90.4/79.7	85.2	85.1	82.9
PLB(R/L)	86.5/ -	88.3/ -	91.2/89.7	89.8/85.7	82.7/65.9	94.0	95.5
R-PDA	65.0	94.1	94.8	96.6	79.8	87.5	89.6

3.2 Coronary Artery Mask Parsing

Volumetric input: As described in Sect. 2.1, we isotropically downsample masks to spacing 0.8mm. Then a patch with size $144 \times 144 \times 144$ is cropped at ROI. Then extracted patches are feed into the parser. Seg-entropy loss and multi-dice loss co-supervise the training process. **Point cloud input:** Surfaces of cardiac masks are extracted using marching cubes [7]. Points in coronary artery mask are all saved. During training, 16384 points are sampled from each case, Among these points, half are randomly sampled from the coronary artery mask and others are sampled proportionally from the cardiac surfaces. All coordinates are transferred to relative coordinates according to coronary artery’s mass center and input masks’ labels are collected as extra features. Architectures for pointnet and pointnet++ remain the same as in original work except that seeds generated for pointnet++ are doubled due to the complexity of coronary artery mask. Then the point cloud parser is trained under joint supervision of NLL loss and seg-consist loss.

Table 1 shows results of our proposed framework comparing with other existing methods. Unlike other methods which are evaluated on relative small datasets or tested with many folds, we here train on 200 images and inference on 326 unseen data. A stable and promising result is exhibited.

3.3 Ablation Studies

At this section, we report segment-level result to eliminate compensatory effect of post-processing and potential manual deviation. Experiments are organized as follows: **(1) org-img:** We firstly parse coronary artery directly from raw image instead of mask space in order to demonstrate the complexity of the task in original space. **(2) cor:** Secondly, we input on coronary artery mask without cardiac masks to clarify the contribution of cardiac atlas information. **(3) cor+car:** Here merge two coronary artery mask and cardiac mask as final input. **(4) cc+loss:**

Table 2. Results of our ablation experiments. All metrics here are measured in segment level. P represents precision and R stands for recall.

	Metric	org-img	cor	cor+car	cc+loss	pc	pc+loss	pc!	pc!+loss
LM	P	76.4	89.8	89.9	92.8	83.1	87.9	88.4	89.6
	R	97.5	98.1	98.8	99.1	95.4	94.4	98.4	98.5
	F1	85.7	93.8	94.1	95.8	88.8	91.1	93.1	93.8
LAD	P	86.9	82.7	87.9	85.9	79.1	78.8	80.4	79.7
	R	95.3	97.5	96.8	97.1	91.6	91.2	94.8	94.2
	F1	90.9	89.5	92.1	91.2	84.9	84.6	87.0	86.3
LCX	P	86.3	83.9	84.8	89.6	88.6	87.3	90.7	91.8
	R	77.3	89.0	90.9	91.1	86.5	88.0	86.7	85.7
	F1	81.6	86.3	87.8	90.3	87.5	87.7	88.7	88.7
RCA	P	85.4	94.2	94.0	94.6	90.5	91.9	91.8	91.4
	R	94.1	95.5	95.9	95.5	89.0	92.5	92.5	92.6
	F1	89.5	94.8	95.0	95.1	89.8	92.2	92.2	92.0
D	P	88.3	84.1	88.1	87.1	81.9	88.3	88.7	87.3
	R	72.4	92.3	91.2	93.1	85.3	78.2	88.2	90.4
	F1	79.6	88.0	89.6	90.0	83.6	83.0	88.4	88.8
OM	P	78.2	82.9	84.7	87.2	88.2	86.8	84.7	85.5
	R	47.6	84.4	80.9	84.5	68.9	71.0	79.8	84.0
	F1	59.2	83.7	82.8	85.8	77.4	78.1	82.2	84.7
PLB	P	91.4	91.3	92.0	93.9	81.2	89.6	90.1	90.3
	R	90.5	93.8	92.9	92.7	92.6	88.4	88.7	91.1
	F1	91.0	92.5	92.4	93.9	86.5	89.0	89.4	90.7
PDA	P	85.3	82.9	82.2	84.4	81.2	72.8	76.9	79.9
	R	72.8	80.6	81.8	84.7	68.3	86.3	83.4	83.8
	F1	78.5	81.7	82.0	84.7	74.2	79.0	80.0	81.8
Avg.	P	84.8	86.5	88.0	89.4	84.2	85.4	86.5	86.9
	R	80.9	91.4	91.2	92.3	84.7	86.3	89.1	90.0
	F1	82.0	88.8	89.5	90.8	84.1	85.6	87.6	88.4

We finally add our proposed seg-consist loss to reach final proposed volumetric version of framework. As for point cloud extension, we transfer experiment set (3) to point cloud version using pointnet and pointnet++ noted as (5) **pc** and (6) **pc!** correspondingly. Then we integrate seg-consist to form (5) **pc+loss** and (5) **pc!+loss**.

As showed in Table 2, task is more challenging in original space than in mask space. Since gap between (2) and (3) is relative small. We conjecture that even though cardiac mask is informative, coronary artery mask itself has already contained most valuable information So even when the cardiac mask is

Table 3. Efficiency analysis between volumetric input and point cloud input. Train time calculates costs for training 1000 epochs on 200 training set

Object	GPU	Train Time (GPU hrs)	Infer Memory (MB)	Infer Time (s)
V-net	1060	197	3736	1.0 ± 0.1
PointNet	1060	8	938	0.03 ± 0.01
PointNet++	1060	42	3082	1.2 ± 0.1

not obtainable, our proposed framework will remain effective. Comparing with volumetric input, point cloud version achieves slightly inferior but compatible results. And no matter in which formation of parsing process, seg+consist loss is beneficial.

We further analyze efficiency in terms of both time and space for two formations. As listed in Table 3, point cloud version is much more space and time efficient. It serves as a good choice under limited memories or facing abundant training data.

4 Conclusion

In this paper, we present a new concise, effective and flexible framework for automatic coronary artery labeling by modeling the task as coronary artery parsing task. An intact pipeline is proposed and two paralleled sub-modules are further designed to consume volumetric image and unordered point cloud correspondingly. Finally, a self-contained loss is proposed to supervise labeling process. At experiment section, we conducted comprehensive experiments on collected 526 CCTA scans. Stable and promising results are exhibited.

References

1. Akinyemi, A., Murphy, S., Poole, I., Roberts, C.: Automatic labelling of coronary arteries. In: 2009 17th European Signal Processing Conference, pp. 1562–1566 (2009)
2. Cao, Q., et al.: Automatic identification of coronary tree anatomy in coronary computed tomography angiography. *Int. J. Cardiovasc. Imaging* **33**(11), 1809–1819 (2017). <https://doi.org/10.1007/s10554-017-1169-0>
3. Chalopin, C., Finet, G., Magnin, I.E.: Modeling the 3d coronary tree for labeling purposes. *Med. Image Anal.* **5**(4), 301–315 (2001). [https://doi.org/10.1016/S1361-8415\(01\)00047-0](https://doi.org/10.1016/S1361-8415(01)00047-0)
4. Gülsün, M.A., Funka-Lea, G., Zheng, Y., Eckert, M.: CTA coronary labeling through efficient geodesics between trees using anatomy priors. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) *MICCAI 2014*. LNCS, vol. 8674, pp. 521–528. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10470-6_65

5. Guo, Z., et al.: DeepCenterline: a multi-task fully convolutional network for centerline extraction. In: Chung, A.C.S., Gee, J.C., Yushkevich, P.A., Bao, S. (eds.) IPMI 2019. LNCS, vol. 11492, pp. 441–453. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-20351-1_34
6. Jin, D., Iyer, K.S., Chen, C., Hoffman, E.A., Saha, P.K.: A robust and efficient curve skeletonization algorithm for tree-like objects using minimum cost paths. *Pattern Recogn. Lett.* **76**, 32–40 (2016)
7. Lorensen, W.E., Cline, H.E.: Marching cubes: a high resolution 3d surface construction algorithm. In: Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques. SIGGRAPH '87, Association for Computing Machinery, New York, NY, USA, pp. 163–169 (1987), <https://doi.org/10.1145/37401.37422>
8. Metz, C.T., Schaap, M., Weustink, A.C., Mollet, N.R., van Walsum, T., Niessen, W.J.: Coronary centerline extraction from ct coronary angiography images using a minimum cost path approach. *Med. Phys.* **36**(12), 5568–5579 (2009)
9. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 652–660. IEEE (2017)
10. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: deep hierarchical feature learning on point sets in a metric space. arXiv preprint, (2017) [arXiv:1706.02413](https://arxiv.org/abs/1706.02413)
11. Sato, M., Bitter, I., Bender, M.A., Kaufman, A.E., Nakajima, M.: Teasar: tree-structure extraction algorithm for accurate and robust skeletons. In: Proceedings the Eighth Pacific Conference on Computer Graphics and Applications, pp. 281–449 (2000)
12. Wu, D., et al.: A learning based deformable template matching method for automatic rib centerline extraction and labeling in ct images. In 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 980–987. IEEE (2012)
13. Wu, D., et al.: Automated anatomical labeling of coronary arteries via bidirectional tree LSTMs. *Int. J. Comput. Assist. Radiol. Surg.* **14**(2), 271–280 (2018). <https://doi.org/10.1007/s11548-018-1884-6>
14. Xia, Q., Yao, Y., Hu, Z., Hao, A.: Automatic 3D atrial segmentation from GE-MRIs using volumetric fully convolutional networks. In: Pop, M., et al. (eds.) STACOM 2018. LNCS, vol. 11395, pp. 211–220. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-12029-0_23
15. Yang, G., et al.: Automatic coronary artery tree labeling in coronary computed tomographic angiography datasets. *Computing in Cardiology*, pp. 109–112. IEEE (2011)
16. Zheng, Y., Tek, H., Funka-Lea, G.: Robust and accurate coronary artery centerline extraction in CTA by combining model-driven and data-driven approaches. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. LNCS, vol. 8151, pp. 74–81. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40760-4_10
17. Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. *Med. Image Anal.* **31**, 77–87 (2016). <https://doi.org/10.1016/j.media.2016.02.006>. <http://www.sciencedirect.com/science/article/pii/S1361841516000219>