Enhanced Automatic Coronary Sinus Vein Segmentation Model Based on Hausdorff Distance Loss Function and Multi-Architecture Neural Network Ensemble

Chiara Arduino[1], Stepan Zubarev[1,2], Margarita Budanova[1,2], Sergei Rud[2], Mikhail Chmelevsky[1,2], Aleksandr Sinitca[1]

[1] XSpline S.p.A., Bolzano, Italy
[2] Almazov National Medical Research Centre, Saint-Petersburg, Russia

Abstract

The key point of the study consists of performing automatic coronary sinus (CS) vein segmentation based on Hausdorff distance loss function and the multi-architecture neural network ensemble. A proprietary cardiac computed tomography protocol is implemented to provide adequate data quality. The dataset consists of two subsets: 133 patients selected for cardiac resynchronization therapy (CRT subset) and 127 candidates recruited for surgical interventions due to ventricular arrhythmias (Extra subset). To test the method, 39 of the 133 cases in the CRT subset were randomly selected. The remaining patients of the CRT subset and all cases of the extra subset are used to train neural networks. The ensemble neural network achieves a Dice score of 0.77, 30.8 mm in terms of Hausdorff distance, and 1.04 mm for mean surface distance in the test subset. In the study, an improvement in automatic CS anatomic reconstruction is presented, which could be helpful for pacemaker implantation and follow-up procedures.

1. Introduction

During CRT procedures, a crucial step consists of inserting a pacing lead into the coronary vein system. However, the anatomical structure of the CS is extremely variable; therefore, cardiologists must analyze and evaluate the characteristics of the CS structure during the operation. Certain complications arise due to the fact that the tactic decision must be made prior to endovascular implantation of the left ventricular electrode, which is sometimes impossible due to the peculiarities of the CS structure. Anatomical assessment before implantation can help clinicians and ensure efficient close placement of the left ventricular electrode in the most suitable area [1].

Having a look on literature, S. Shah [2] shows how difficult but important it is to perform CS imaging before CRT implantation to have a clear understanding of CS anatomy.

Thus, there are various attempts to solve the task. For example, S. Lu et al. in the paper [3] provides a method based on extraction of the center line for CS vein segmentation; however, this approach is strictly limited to the main trunk of CS. B. Bouraoui et. al [4] proposes a 3D segmentation of coronary arteries based on morphology techniques. A. Yoshida et al. in their study [5] proposed a U-Net-based approach for full 3D segmentation of the heart, but the scope of the work has been limited to the internal cavities. L. Baskaran et al. [6] and M. Habijan et al. [7] performed a CS segmentation approach using deep learning methods, but this is only related to the largest part of the CS system, processing 2D images. Also, G. Wang et al. [8] developed a novel U-shaped transformer network for CS segmentation using 2D images produced from coronary angiography. Following, C. Jahnke et al. [9] suggested a framework for surface mesh reconstruction, but their approach is also limited to the proximal part of the CS structure. Finally, our previous work [10] is based on the implementation of a 5-fold cross-validation ensemble. This research is concerned with segmentation of thin tributaries of CS.

This research aims at improving automatic segmentation of CS veins by employing the conventional mean ensemble of segmentation neural networks. A proprietary cardiac computed tomography (CT) protocol was applied to enhance NNs input data.

2. Materials and Methods

2.1. Dataset

Full CT dataset includes: 133 cases scheduled for CRT implantation (*CRT subset*) and 127 candidates recruited for heart interventions due to ventricular arrhythmias (*Extra subset*). CT data were collected on available scanners at the clinic (Somatom Definition 128 and Force, Siemens AG, Germany). Imaging acquisition steps implemented in

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.019

this study were based on the CRT-DRIVE clinical study CT scan protocol (ClinicalTrials.gov ID: NCT05327062). A total of 260 patients were included. The age at enrollment was 65 (37; 81) – median (min; max) years. 230 of the examined subjects had sinus rhythm, 26 atrial fibrillation without high-frequency heart rate during the study, and 4 having permanent cardiac pacing. A written informed consent was obtained from each patient prior to the procedure. The local Ethics Committee of the Almazov National Medical Research Centre approved the study (reference ID 203 from November 19, 2018).

Next, each DICOM file was manually segmented by a radiologist and further independently validated by two cardiologists. To test the method, 39 cases were randomly selected from the *CRT subset*. The remaining cases from the CRT subset and all cases from the extra subset were used to train the neural networks.

2.2. Segmentation model

The approach applies the conventional mean ensemble, based on 3 different architectures. In particular, SwinUNETR [11], UNET++ [12], and SegResNet [13] have been used as base models networks.

The base models were built with the following common configuration and hyperparameters:

- Input and output shape: $96 \times 96 \times 96$.
- Number of input channels: 1.
- Number of output channels: 2 (background, coronary sinus).
- Batch size: 4.

Due to the heterogeneous spatial resolution of the data set, and the necessity to normalize pixels' intensity, we used the following pre-processing pipeline:

1. Cast image orientation to the next 3D orientation: (Left, Right), (Posterior, Anterior), (Inferior, Superior).
2. Resample 3D image to resolution $0.5 \times 0.5 \times 1.0$ mm. The resolution was chosen according to the target sizes of anatomical structure.
3. The scale intensity ranges from $[0, 800]$ Hounsfield units to $[0, 1]$ with values clipping.

Additionally, because of the limited possible input size of the neural networks, a sliding-window approach was applied. For quality estimations, an overlap of 0.5 was applied using a Gaussian window, which gives less weight to predictions on the edges of window with $\sigma = 0.125$. During the training procedure, a random cropping approach with balance based on a central pixel class was applied.

For better generalization, a well-known training-time data augmentation approach was set. Random intensity shift with 0.10 offset was implemented.

In order to evaluate each network performances, a 5-fold cross-validation was performed, which is allowing to provide an estimation of base model quality without the use of test subset.

Next, for each model architecture, the best cross-validation checkpoint was selected as a base network weighs.

For the NNs training, AdamW optimizer [14] was used, with an initial learning rate of $1.5e-4$ and a weight decay of $1e-5$.

As a loss function, a combination of TverskyLoss (T) [15] and Hausdorff Distance (HD) loss [16] was implemented (equation 1).

$$Loss = \zeta HD(\bullet) + \eta T(\bullet) \tag{1}$$

ζ and η are equation's balancing coefficients.

The Hausdorff distance loss is employed to add penalty for mistakes in a small branch segmentation. Such tributaries are a small fraction of the CS volume, and errors in these portions cannot be excessively penalized with other conventional losses. In the scope of the present work, we used $\zeta = 1.0$ and $\eta = 0.1$ to reduce the effect of Hausdorff loss at the beginning of the training procedure. This loss has an extremum-related nature and can be unstable, which can negatively affect the training process.

TverskyLoss function (T) shown in equation 2 has the following parameters: $alpha = 0.3$ and $beta = 0.7$ to balance an extremely unbalanced dataset. The parameters were chosen on the basis of recommendations from the original paper.

$$T(\alpha, \beta) = \frac{\sum_{i=1}^{N} p_{0i} g_{0i}}{\sum_{i=1}^{N} p_{0i} g_{0i} + \alpha \sum_{i=1}^{N} p_{0i} g_{1i} + \beta \sum_{i=1}^{N} p_{1i} g_{0i}} \tag{2}$$

where p_{ji} is the output of the Softmax prediction layer, p_{0i} is the probability that voxel i is a background. p_{1i} is the probability that the i'th voxel is a foreground. Also, $g_{0i} = 1$ and $g_{0i} = 0$ for foreground and background voxels accordingly, and vice versa for g_{1i}.

The bidirectional Hausdorff Distance loss, which is utilized as the loss function, is shown in equation 3.

$$HD_{Loss}(q, p) = \frac{1}{|\Omega|} \sum_{\Omega} ((p - q)^2 \circ (d_p^{\alpha} + d_q^{\alpha})) \tag{3}$$

where q and p are the predicted scores, α is penalization ratio (for this purpose $\alpha = 2$), d is distance. This function is based on Hausdorff Distance (4), with the penalization of larger segmentation errors (squared distance is used).

$$HD(Y, X) = \max(hd(X, Y), hd(Y, X)) \tag{4}$$

where $hd(X, Y)$ and $hd(Y, X)$ are one side Hausdorff Distances as defined in 5 and 6, respectively.

$$hd(X,Y) = \max_{x \in X} \min_{y \in Y} ||x - y||_2 \quad (5)$$

$$hd(Y,X) = \max_{y \in Y} \min_{x \in X} ||x - y||_2 \quad (6)$$

2.3. Quality estimation

In this study, the quality of the segmentation models was estimated in the following terms.

- Dice Score (DSC):

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \quad (7)$$

where X is the predicted mask and Y the ground truth.

- Bidirectional Hausdorff distance (HD) as defined in Equation 3.
- Mean Surface Distance (MSD).

3. Results

As shown in Table 1, cross-validation scores for networks are about 0.75 in terms of the Dice score. These values have a small standard deviation, which justifies the reliability of quality assessment during validation. Hence, the Hausdorff distance, which is related to the similarity between segmentation results and ground truth, is homogeneous and ranges, respectively, from almost 40 to 74 mm. These results highlight the comparability and similarity between the performance of the base networks.

Table 1. 5-Fold cross-validation scores of CS segmentation, ($Mean \pm std$, fold level).

Model	DSC	HD_{Loss}	Best Fold
SwinUNETR	0.75 ± 0.01	0.50 ± 0.24	5
UNET++	0.75 ± 0.01	0.74 ± 0.05	1
SegResNet	0.74 ± 0.02	0.39 ± 0.21	1

Next, in Table 2 an estimated score is given for the best fold chosen during the cross-validation process using the test subset. As shown, despite a comparable Dice Score, the Hausdorff distance metric was significantly improved by the ensemble network. At the same time, there is a slight improvement in terms of mean surface distance. This demonstrates a more coherent segmentation of the ensemble classifier in comparison with the use of a single-network classification. The segmentations are more similar and less distant from the corresponding ground truth.

An example of the comparison between manual and automatic segmentation performed by a neural network is shown in figure 1.

Table 2. Test scores of CS segmentation for base models and the ensemble, ($Mean \pm std$, sample level).

Model	DSC	HD, mm	MSD, mm
SwinUNETR	0.75 ± 0.07	73.6 ± 35.3	1.20 ± 0.58
UNET++	0.76 ± 0.06	97.4 ± 28.8	1.29 ± 0.46
SegResNet	0.76 ± 0.07	92.0 ± 46.1	1.21 ± 0.60
Ensemble	0.77 ± 0.07	30.8 ± 15.2	1.04 ± 0.46

Figure 2 represents a violin and a boxplot diagram, regarding the statistical distribution of the Dice score. These data refer to the dice score that belongs to the sample sets. The major part of the distribution ranges from just under 0.75 to about 0.85. The resulting dice scores are homogeneous, indicating that the network performances are stable and reliable. There are few outliers, probably because of the complexity of the CS geometry or the features of the CT image.

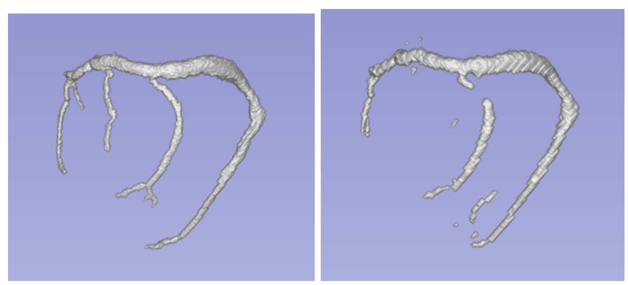

Figure 1. Manual segmentation of CS veins performed by experts, used as a ground truth (left); resulted segmentation from ensemble classifier (right)

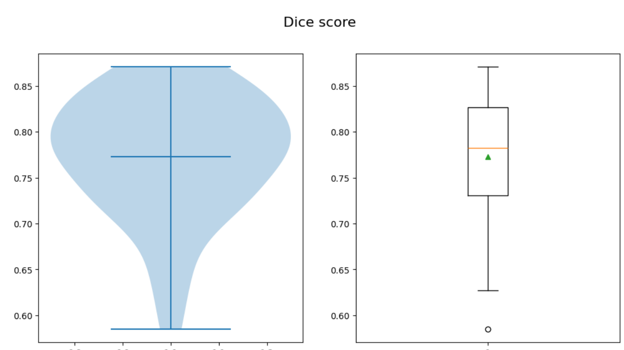

Figure 2. Violin diagram of Dice Scores distribution related to test set (left); boxplot diagram of Dice Scores distribution related to test set results (right)

4. Conclusion

The results of the study indicate that the ensemble neural network achieved a quality of 0.77 (Dice Score), $30.8\ mm$ in terms of Hausdorff distance and $1.04\ mm$ in terms of mean surface distance. All these parameters are calculated based on test subset results. Despite a minor improvement in terms of the dice score, this represents a significant improvement in terms of Hausdorff distance over our previous published results. This means that the proposed approach reduces the number of segmentation gaps, which is one of the main obstacles for this kind of task.

In summary, the developed approach improved the overall quality of automatic reconstruction of CS anatomy using an efficient combination of three different neural networks. The results obtained could lead to a greater understanding and characterization of the anatomy of the patient, which could be beneficial in enhancing the implantation and outcome of the CRT procedure.

Acknowledgments

Type of funding sources: Private company. Main funding source(s): XSpline S.p.A.

References

[1] Chmelevsky M, Zubarev S, Khamzin S, Dokuchaev A, Bazhutina A, Sinitca A, Budanova M, Auricchio A. Clinical evaluation of the new 12-lead ecg noninvasive epi-endocardial mapping technology. Europace 2023;25(Supplement_1):euad122–647.

[2] Shah SS, Teague SD, Lu JC, Dorfman AL, Kazerooni EA, Agarwal PP. Imaging of the coronary sinus: normal anatomy and congenital abnormalities. Radiographics 2012;32(4):991–1008.

[3] Lu S, Huang X, Wang Z, Zheng Y. Sparse appearance learning based automatic coronary sinus segmentation in cta. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014: 17th International Conference, Boston, MA, USA, September 14-18, 2014, Proceedings, Part I 17. Springer, 2014; 779–787.

[4] Bouraoui B, Ronse C, Baruthio J, Passat N, Germain P. 3d segmentation of coronary arteries based on advanced mathematical morphology techniques. Computerized medical imaging and graphics 2010;34(5):377–387.

[5] Yoshida A, Lee Y, Yoshimura N, Kuramoto T, Hasegawa A, Kanazawa T. Automated heart segmentation using u-net in pediatric cardiac ct. Measurement Sensors 2021; 18:100127.

[6] Baskaran L, Al'Aref SJ, Maliakal G, Lee BC, Xu Z, Choi JW, Lee SE, Sung JM, Lin FY, Dunham S, et al. Automatic segmentation of multiple cardiovascular structures from cardiac computed tomography angiography images using deep learning. PloS one 2020;15(5):e0232573.

[7] Habijan M, Galić I, Romić K, Leventić H. Ab-resunet+: Improving multiple cardiovascular structure segmentation from computed tomography angiography images. Applied Sciences 2022;12(6):3024.

[8] Wang G, Zhou P, Gao H, Qin Z, Wang S, Sun J, Yu H. Coronary vessel segmentation in coronary angiography with a multi-scale u-shaped transformer incorporating boundary aggregation and topology preservation. Physics in Medicine Biology 2024;69(2):025012.

[9] Jahnke C, Darma A, Lindemann F, Oebel S, Hilbert S, Bode K, Stehning C, Smink J, Paetsch I. Electrophysiological cardiovascular mr: procedure-ready mesh model generation for interventional guidance based on non-selective excitation compressed sensing whole heart imaging. Scientific Reports 2024;14(1):8974.

[10] Sinitca A, Chmelevsky M, Zubarev S, Shirshin A, Dokuchaev A, Budanova M, Arduino C, Khamzin S, Bazhutina A, Rud S, et al. A machine learning-based approach for automatic coronary sinus vein segmentation and anatomy reconstruction. In 2023 Computing in Cardiology (CinC), volume 50. IEEE, 2023; 1–4.

[11] Hatamizadeh A, Nath V, Tang Y, Yang D, Roth HR, Xu D. Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In Crimi A, Bakas S (eds.), Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Cham: Springer International Publishing, 2022; 272–284.

[12] Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. Unet++: A nested u-net architecture for medical image segmentation. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. Springer, 2018; 3–11.

[13] Myronenko A. 3d mri brain tumor segmentation using autoencoder regularization. In Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II 4. Springer, 2019; 311–320.

[14] Loshchilov I, Hutter F. Decoupled weight decay regularization. In International Conference on Learning Representations. 2017; .

[15] Salehi SSM, Erdogmus D, Gholipour A. Tversky loss function for image segmentation using 3d fully convolutional deep networks. In Wang Q, Shi Y, Suk HI, Suzuki K (eds.), Machine Learning in Medical Imaging. Springer International Publishing, 2017; 379–387.

[16] Karimi D, Salcudean SE. Reducing the hausdorff distance in medical image segmentation with convolutional neural networks. IEEE Transactions on Medical Imaging 2020; 39(2):499–513.

Address for correspondence:

Aleksandr Sinitca
XSpline S.p.A., Via Ressel 2F, 39100, Bolzano, Italy
sinitca@xspline.com