Enhanced Automatic Segmentation of Epi-Endocardial Ventricular Anatomy Using Multi-Architecture Neural Network Ensemble

Chiara Arduino[1], Stepan Zubarev[1,2], Margarita Budanova[1,2], Mikhail Chmelevsky[1,2], Sergei Rud[2], Danila Sudas[2], Daniil Kriukov[2], Gennady Trufanov[2], Aleksandr Sinitca[1]

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

Abstract

The modern development of cardiology as a branch of medicine is strongly linked to the improvement of neural networks for automatic heart segmentation. Our approach is based on three state-of-the-art neural networks for medical image segmentation: SwinUnet, UNet++ and SegResNet. The dataset consists of two subsets: 133 patients selected for cardiac resynchronization therapy (CRT subset) and 127 candidates recruited for heart interventions due to ventricular arrhythmias (Extra subset). Each computed tomography scan was manually segmented by a radiologist and independently validated by two cardiologists. Three classes were introduced: ventricular myocardium (wall), left, and right cavities. To test the method, 39 of the 133 cases in the CRT subset were randomly selected. The remaining patients in the CRT subset and all cases in the extra subset were used for training. The results of the study indicate that the model achieved a quality of 0.80 in terms of Dice score and 1.2 mm in terms of mean surface distance. The quality of automatic ventricular segmentation obtained demonstrates that this solution can be applied in real clinical practice.

1. Introduction

Recently, deep learning-based approaches have replaced conventional methods in the field of cardiac image segmentation. To address such tasks, Machine Learning and Deep Learning techniques have been widely used, with a strong predominance of Convolutional Neural Network-based solutions for volume segmentation. Specifically, U-Net, as mentioned by Ronneberger et al. [1], which is fast and provides great quality performances. In addition, there are several tasks-specific solutions like the one proposed by Matheus A.O. Ribeiro et al. [2]; or Tuan Anh Ngo et al.[3] that propose an automated segmentation using magnetic resonance imaging. The only limitation is that they are limited to the anatomy of the LV. Moreover, segmentation of the ventricle structure is complicated by the restriction of the imaging protocol. This restriction (late-phase imaging) leads to a lack of contrast in the ventricle cavities. In addition, to obtain epicardial and endocardial models, it is necessary to segment the very poorly visible wall between the ventricles.

This study focuses on improving automatic segmentation of epi-endocardial ventricular anatomy using the proprietary cardiac computer tomography (CT) protocol. The challenge of this research is the RV internal surface, which was not previously automatically reconstructed using CT images.

The used approach is based on a state-of-the-art neural networks (NN) for medical image segmentation namely SwinUnet, UNet++ and SegResNet in combination with a large and consistent dataset.

2. Materials and Methods

2.1. Dataset

For the current study, we used a dataset consisting of two cardiac CT subsets, specifically: 133 cases selected for cardiac resynchronization therapy (*CRT subset*) and 127 patients scheduled for heart surgical interventions due to ventricular arrhythmias (*Extra subset*). Two CT scanners (Somatom Definition 128 and Force, Siemens AG, Germany) were used. The acquisition steps implemented in this study were based on the CT scan protocol of the CRT-DRIVE clinical study (ClinicalTrials.gov ID: NCT05327062).

In total, 260 patients were included. The age at enrollment was 65 (37; 81) – median (min; max) years. 230 subjects examined had sinus rhythm, 26 atrial fibrillation without high-frequency heart rate during the study, and 4 with permanent cardiac pacing. 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 image was manually segmented by a radiologist and further independently validated by two cardiologists.

To test the method, 39 of the 133 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 NNs during a 5-fold cross-validation procedure.

2.2. Segmentation model

The proposed segmentation model consists in ensembling approach, based on 3 different architectures. In particular, as base models, the following architectures have been utilized: SwinUNETR [4], UNET++ [5], and SegResNet [6].

The 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: 4 (Background, Left Ventricle Cavity, Right Ventricle Cavity, Ventricular myocardium or Wall).
- Size of feature space: 48.

The dataset presents a heterogeneous spatial resolution, and it is necessary to normalize the pixels' intensity. To make the input data uniform the data, the following preprocessing pipeline was used:

1. Cast image orientation to the next 3D orientation: (Left, Right), (Posterior, Anterior), (Inferior, Superior).
2. Resample 3D image to resolution $1.5 \times 1.5 \times 1.5$. The resolution was chosen respectively to the target sizes of anatomical structure.
3. The scale intensity ranges from $[0, 550]$ Hounsfield unit intensity to $[0, 1]$ with values clipping.

Furthermore, because of the relatively limited resolution of the NN, a sliding window method was implemented. For quality assessment, an overlap of 0.5 was employed using the Gaussian window. This allows to assign a lower weight to the predictions located at the edges of the window (having $\sigma = 0.125$). In the training phase, we applied a random cropping technique with balancing dependent on the central pixel class.

For better generalization, a well-known training-time data augmentation approach was applied. For this purpose, in particular the random intensity shift was used, with a 0.10 offset.

To evaluate performance during the training step, each network was trained using 5-fold cross-validation. This approach allows to provide an estimation of base model quality without test subset utilization.

Following, for each model architecture, during the 5-fold cross-validation 5 different checkpoint files were produced. The best was chosen as a base network weighs.

For every NN training, AdamW optimizer [7] was used with an initial learning rate of $5e-4$ and a weight decay of $1e-5$.

To balance extremely unbalanced classes, TverskyLoss (T) [8] was employed as a loss function (equation 1) with the following parameters: $alpha = 0.3$ and $beta = 0.7$.

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

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}.

2.3. Quality estimation

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

- Dice Score (DSC):

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \tag{2}$$

where X is the predicted mask and Y the ground truth mask. This metric is evaluated for each separate class (left ventricle cavity, right ventricle cavity, and wall).

- Bidirectional Hausdorff distance (HD) as defined in Equation 3

$$HD(Y, X) = \max(hd(X, Y), hd(Y, X)) \tag{3}$$

where $hd(X, Y)$ and $hd(Y, X)$ is one side Hausdorff distances as defined in 4 and 5, respectively.

$$hd(X, Y) = \max_{x \in X} \min_{y \in Y} ||x - y||_2 \tag{4}$$

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

- Mean Surface Distance (MSD) estimated as follows.

$$\text{MSD} = \frac{1}{n_S + n_{S'}} \left(\sum_{p=1}^{n_S} d(p, S') + \sum_{p'=1}^{n_{S'}} d(p', S) \right) \tag{6}$$

3. Results

For each network, different NNs performances have been evaluated in order to choose the best fold to be included in the ensemble.

As shown in Table 1, the cross-validation scores for all NNs used range between 0.65 and 0.66 in terms of the Dice Score and have a limited relative standard deviation.

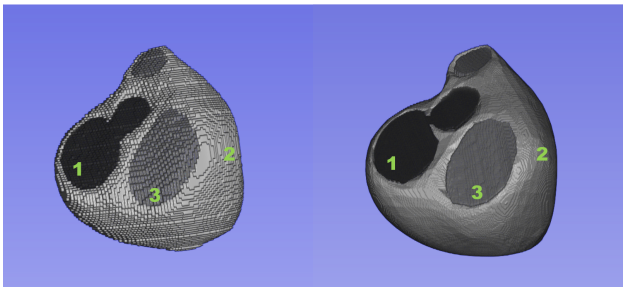

Figure 1. Resulted segmentation from ensemble classifier (left), manual segmentation of the cavities and wall, used as a ground truth (right). It is possible to identify different regions by grey scale color labels (1: Left ventricle cavity, 2: Wall, 3: Right ventricle cavity)

In Table 2, the quality metrics related to the entire ensemble classifier are presented. The dice score is calculated separately for the three classes (left ventricle cavity, right ventricle cavity, and wall), and it ranges between 0.70 and 0.90. These scores have a small standard deviation, which justifies the reliability of quality assessment during validation. Hence, the Hausdorff distance, related to the similarity between segmentation results and ground truth, is relatively low, more specifically it is about 23 mm. The mean surface distance results in 1.2 mm, this parameter is related to the average distances between the boundaries of the two surfaces.

As shown, the Dice Score was significantly improved by the ensemble network. This demonstrates a more coherent segmentation made by the ensemble classifier compared to the use of a single-network classification. The results are more similar and less distant from the ground truth segmentation. An example of the comparison between the manual and automatic segmentation performed by the NN is shown in Figure 1.

In Figure 2, it is possible to understand the statistical distribution of the Dice Scores, represented in the three boxplot diagrams. These data refer to the target anatomical regions, belonging to the test subset elements. The interquartile range (IQR) of Walls segmentation quality, starts from just over 0.6 to end at about 0.8. The DSC distributions for the right and left ventricular cavities are

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

Model	DSC_{mean}	Best Fold
SwinUNETR	0.66 ± 0.05	3
UNET++	0.65 ± 0.05	5
SegResNet	0.66 ± 0.03	4

Table 2. Test scores of Ventricle segmentation ensemble, ($Mean \pm std$, sample level).

Metric	Value
DSC_{wall}	0.72 ± 0.13
DSC_{LV}	0.89 ± 0.07
DSC_{RV}	0.79 ± 0.15
DSC_{mean}	0.80 ± 0.12
HD, mm	23.57 ± 63.73
MSD, mm	1.2 ± 1.01

narrow; the median is above 0.8. Overall, the resulting Dice Scores are homogeneous, indicating that the network performances are stable and reliable. There are few outliers, regarding ventricle segmentation, especially related to RV. This is related to the radiodensity of the cavities and the complexity of the internal heart structure.

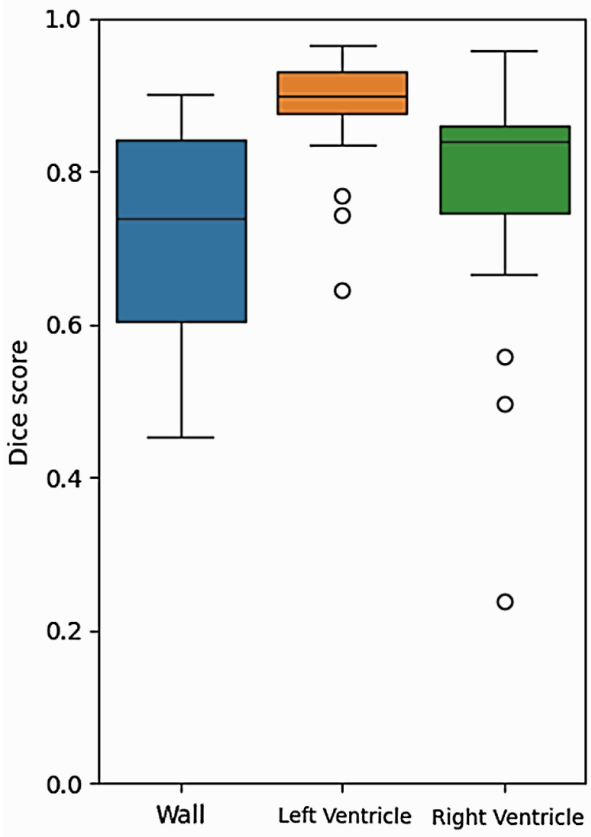

Figure 2. Boxplot diagrams of Dice Score related to Wall, Left ventricle cavity, Right ventricle cavity estimated on test subset.

4. Conclusion

The results of the study indicate that the ensemble neural network achieved a Dice Score quality of about 0.80, $23.57\ mm$ in terms of Hausdorff distance and $1.2\ mm$ in terms of mean surface distance in the test subset. Comparing the single neural network architectures with the ensemble, the Dice score has increased by about 15 percentage points. The ensemble quality in terms of Hausdorff distance and mean surface distance are low, which means that the volumes are similar and coherent to the ground truth. Moreover, the statistical distributions are homogeneous and there are only a few outliers.

The results can be improved by augmenting and perfecting the data set at disposal. Nevertheless, the quality obtained is already sufficient for implementation in clinical practice.

Acknowledgments

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

References

[1] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. volume 9351. ISBN 978-3-319-24573-7, 10 2015; 234–241.

[2] A.O. Ribeiro M, L.S. Nunes F. Left ventricle segmentation combining deep learning and deformable models with anatomical constraints. Journal of Biomedical Informatics 2023;142:104366. ISSN 1532-0464.

[3] Ngo TA, Lu Z, Carneiro G. Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance. Medical Image Analysis 2017;35:159–171. ISSN 1361-8415.

[4] 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.

[5] 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.

[6] 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.

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

[8] 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.

Address for correspondence:

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