DeepLA: Automated Segmentation of Left Atrium from Interventional 3D Rotational Angiography Using CNN

Kobe Bamps[1(✉)], Stijn De Buck[2,3,4(✉)], Jeroen Bertels[4(✉)], Rik Willems[3(✉)], Christophe Garweg[3(✉)], Peter Haemers[3(✉)], and Joris Ector[1,3(✉)]

[1] Department of Cardiovascular Sciences, KU Leuven, Leuven, Belgium
kobe.bamps@kuleuven.be
[2] Radiology UZ Leuven, Leuven, Belgium
stijn.debuck@uzleuven.be
[3] Cardiology UZ Leuven, Leuven, Belgium
{rik.willems,christophe.garweg,peter.haeners,joris.ector}@uzleuven.be
[4] Department of Electrical Engineering, KU Leuven, Leuven, Belgium
jeroen.bertels@kuleuven.be

Abstract. Accurate segmentation of the shape of the left atrium (LA) is important for treatment of atrial fibrillation (AF) by catheter ablation. Interventional 3D rotational angiography (3DRA) can be used to obtain 3D images during the intervention. Low dose 3DRA poses segmentation challenges due to high image noise. There is a significant amount of research focusing on the automatic segmentation from 3DRA images, all based on an active shape or atlas-based approaches.

We present an algorithm based on a 3D deep convolutional neural network (CNN) for automated segmentation of 3DRA images to predict the shape of the LA. The CNN is based on the U-Net architecture and consists of an encoder and a decoder part. It is designed to be trained end-to-end from scratch on interactive semi-automated 3DRA images, which include the body of the LA and the proximal pulmonary veins up to the first branching vessel.

The CNN is trained and validated using 5-fold cross-validation on 20 3DRA images by computing the Dice score (0.959 ± 0.015), recall (0.962 ± 0.026), precision (0.957 ± 0.021) and mean surface distance (0.716 ± 0.276 mm). We further validated the algorithm on an additional data set of 5 images. The algorithm achieved a Dice score and mean surface distance of 0.937 ± 0.016 and 1.500 ± 0.368 respectively.

Keywords: Segmentation · Atrial fibrillation · Left atrium · CNNs

1 Introduction

Atrial fibrillation (AF) is one of the most common cardiac arrhythmias [13]. The irregular heart beats originate from abnormal electrical discharges in the

M. Pop et al. (Eds.): STACOM 2019, LNCS 12009, pp. 138–146, 2020.
https://doi.org/10.1007/978-3-030-39074-7_15

LA. AF is treated by performing catheter ablation to isolate the pulmonary veins from the atrial body [1]. The ablation procedure is typically guided by X-ray fluoroscopy and electroanatomical mapping (EAM). Although a 3D surface of the LA can be generated by a roving catheter. A lot of centers use an image-based anatomical representation because of its superior anatomical detail. This image-based representation can be obtained from CT, MRI or interventional 3DRA and can be integrated in the EAM system. Furthermore, it improves patient outcome and reduces radiation dose from fluoroscopy [8]. Thus, a correct segmentation of the LA is important for the success of the ablation procedure.

The use of preoperative CT or MRI has already been proven useful to facilitate the guidance during interventions [4]. However, preoperative CT and MRI imaging is often performed days or weeks before the actual intervention and possibly under different cardiac loading conditions. As a consequence, the anatomical structures of the patient can differ between the preoperative acquisition and the intervention. Besides resulting in a more complicated workflow, CT or MRI may be contra-indicated and often associated with an additional financial cost [12]. Interventional 3DRA can address some of these shortcomings.

Previously, automated LA segmentation methods were primarily based on model-based approaches. These approaches make use of a prior shape of the LA and intensity values of a stack of known LA 3DRA images. Model-based methods have been implemented for CT and MR [2,11]. The success of these methods relies on both the excellent image quality of CT and MRI acquisitions and sufficiently large data sets. However, model-based methods can be less robust on low dose 3DRA due to the limited image quality in combination with a rather small data set [11]. The limited image quality is caused by the variable contrast density, catheter streak artefacts, low X-ray dose and possibly a reduced frame rate [3]. Recently, CNNs were used to segment the LA from MRI [7]. They proposed a multi-view 2D CNN with an adaptive fusion strategy to segment the pulmonary veins (PVs) and LA. However, automated LA segmentation based on CNN is nonexistent for 3DRA.

In this paper, we present DeepLA, an automated LA segmentation method for low dose 3DRA. Our method is based on a modified version of the U-Net architecture [10]. In order to train the CNN, 3D patches were sampled from low dose 3DRA images. Furthermore, we implemented a preprocessing step. It reduces redundant information to improve the learning process. The overall segmentation process for a given 3DRA takes on average 19.6 ± 5.2 s on a GPU (Geforce GTX 1080 Ti) at test time.

2 Method

The method is composed of two parts: preprocessing and training. First, we discuss the preprocessing step. This is important to optimize the training process. Then, we elucidate the architecture and training process.

2.1 Preprocessing

Preprocessing aims at making the data more meaningful to the CNN. This is accomplished by removing redundant information and ensuring that the intensity value range matches across images of the entire data set. The latter avoids an initial bias in the CNN at test time. The preprocessing is realized in two steps. First, we construct a LA region that consists of all voxels with an intensity value higher than zero. The region outside the LA is set to zero. In addition, the gaps in the LA region are filled with morphological closing operations. Secondly, we normalize every patient's image by subtracting its mean and dividing by its standard deviation.

2.2 Network Architecture

In this paper, we propose a modified 3D U-Net architecture. The network consists of a contracting path and an expanding path. The contracting path contains successive convolutions and pooling layers to extract semantic features at the expense of spatial information. Two max pooling layers are used to reduce the dimensions of the image, first by $3 \times 3 \times 3$ and then by $3 \times 3 \times 1$. Before every max-pooling layer, we find two successive convolutional layers with a filter size of $3 \times 3 \times 3$ and $3 \times 3 \times 1$, respectively. In addition, we used batch normalization and a Parametric Rectified Linear Unit (PReLU) activation after every convolution. The PReLU activation function is a function with a learnable parameter. It follows: $f(x) = \alpha * x$ for $x < 0$, $f(x) = x$ for $x >= 0$, where α is learned [5].

However, in order to perform a correct segmentation, both spatial and semantic information is needed. The U-Net recovers the spatial information by shortcut connections from the feature maps in the contracting path of the same resolution. First the feature maps of the expanding path are cropped and then concatenated with the feature maps in the expanding path of the same resolution. Furthermore, the semantic information is passed on from bottom to top by up-sampling. The convolutional layers in the expanding path have the same configurations as in the contracting path. Due to the valid padding of the CNNs, the output size ($38 \times 38 \times 13$) of the U-Net is smaller than the input. In order to allow for sharper boundaries between voxels, the output of the U-Net is connected to two successive convolutional layers with 32 filters of size $1 \times 1 \times 1$. In front of each of the final convolutional layers, we added dropout with a probability of 0.25 to improve the generalization of the network. The final layer of the network includes a convolution with a filter size of 1 and a sigmoid activation function for generating a probability score for each voxel. Eventually, the LA shapes are produced by thresholding the probabilities of the output at 0.5. The algorithm is implemented in Keras (Python). A summary of the CNN is presented in Fig. 1.

2.3 Training Procedure

The training procedure is an important step to learn the right features in order to perform a correct segmentation of the LA. The whole training procedure was

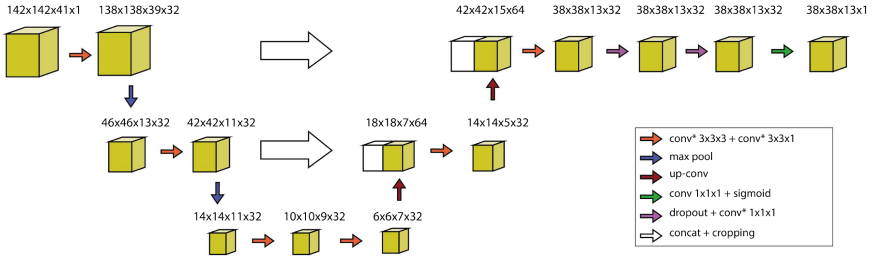


Fig. 1. Details of the U-Net architecture. Conv* refers to a convolutional, a batch normalization and PReLU activation layer.

designed using the DeepVoxNet [9]. This is a framework that enables a fast implementation of neural networks on medical images. DeepVoxNet provides components to handle the memory requirements, the various image modalities and data augmentation. The proposed network architecture is trained with input-patches of size $142 \times 142 \times 41$ and corresponding output-patches of size $38 \times 38 \times 13$. These patches are sampled via a class weighted sampler to compensate the class imbalance between the background and foreground. We define one epoch as 128 batches, with one batch consisting of 8 samples. The training is terminated early if there is no improvement during 70 epochs in the validation Dice coefficient. The weights of the network are optimized by the ADAM optimizer with an initial learning rate of 10^{-3}. The learning rate is divided by a factor of 2 when the validation Dice coefficient did not improved over the last 20 epochs. Furthermore, the network is regularized by L1 and L2 weight delay, both with a weight of 10^{-5}. Lastly, we use the binary cross-entropy loss function to train our network.

3 Validation

The 3DRA images were obtained through a floor-mounted Siemens Axiom Artis dBC biplane fluoroscopy system (Siemens). During joint rapid atrial and ventricular pacing at 250 ms, contrast was injected directly into the left atrium. Iomeron 350 contrast (Bracco) was diluted up to 50% using normal saline. Ninety millilitres of diluted contrast was injected at a rate of 20 mL/s, starting 4 s before the actual start of the C-arm rotation. The X-ray tube rotates around the patient over a course of 200°. During the rotation, fluoroscopic images were sequentially acquired with target detector entrance doses of 0.24 μGy/frame. In total, 62 frames were acquired. These 62 frames were used to reconstruct a 3D volume. This 3D volume is also called a 3DRA image. Further information about the image acquisition can be found in [3].

Twenty subjects were interactively semi-automated annotated by an expert. The annotated images include the body of the LA and the proximal pulmonary veins up to the first branching vessel. The 20 3DRA images were randomly divided into 5 subsets of 4 images. The learning process was repeated for 5 times while each time a different validation set of 4 images was used. In addition,

Table 1. Mean, median and standard deviation of the metrics for LA segmentation

	Mean	STD
DICE	0.959	0.015
sDICE	0.932	0.020
PRE	0.957	0.021
REC	0.962	0.026
MSD	0.716	0.276
HD	11.088	4.770
HD95	2.451	1.686

(a) Validation set

	Mean	STD
DICE	0.937	0.016
sDICE	0.915	0.024
PRE	0.961	0.035
REC	0.917	0.038
MSD	1.500	0.368
HD	32.390	13.390
HD95	6.888	2.749

(b) Test set

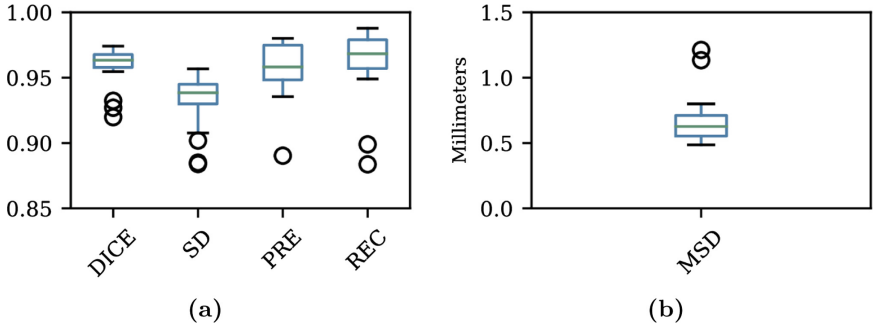
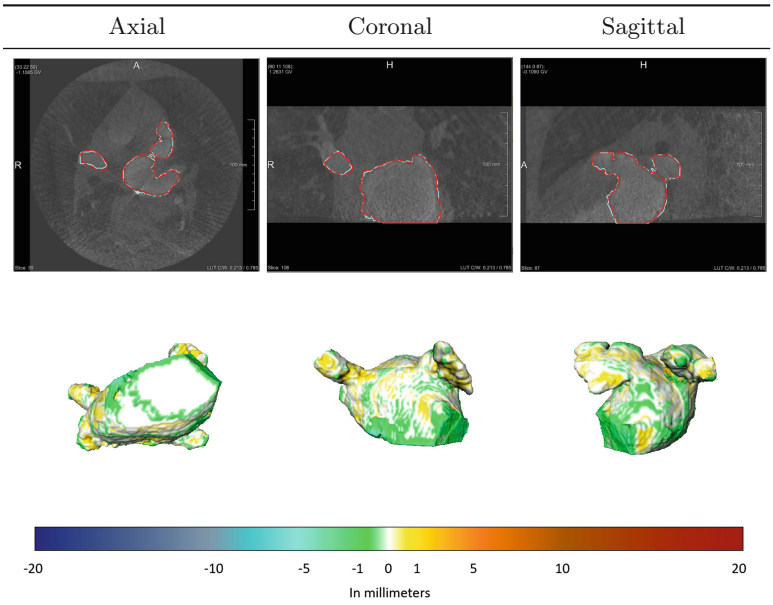


Fig. 2. Box plots for Dice, soft Dice, precision, recall and mean-surface-distance in the validation set

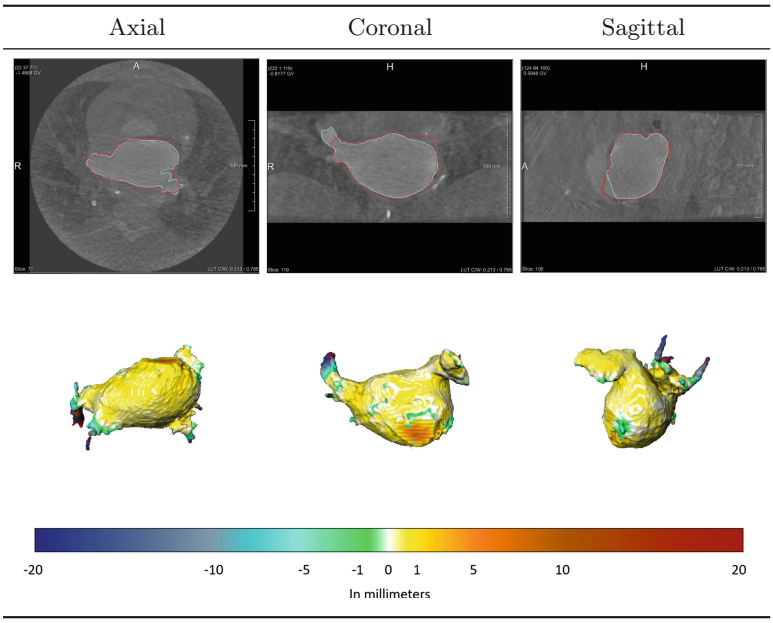
Table 2. Clinical usefulness

Case	1	2	3	4	5	Mean
Overall segmentation of the LA and PVs	4	4	4	4	4	**4.0**
Ridge between the LPVs and LAA	3	3	4	4	5	**3.8**
Position of the Mitral valve	4	3	4	4	5	**4.0**
Number of PVs correctly segmented	4	3	4	4	4	**3.8**

5 unseen 3DRA images were independently collected. These 5 3DRA images were semi-automated annotated by a different reviewer. This resulted in slightly different ground truth masks compared to the other 20 3DRA images. The automatic LA segmentation on the test data were generated by the model that yielded the highest Dice score on the validation set. For evaluation of the proposed method, we computed the Dice coefficient (DICE), soft Dice coefficient (sDICE), recall (REC) and precision (PRE) for both validation and test set. In addition, we assessed the mean-surface-distance (MSD) maximal Hausdorff distance (HD) and 95 percentile Hausdorff distance (HD95).



(a)



(b)

Fig. 3. Visualization of segmentation in validation (a) and test set (b). The first row of each table presents 3DRA slices from axial, coronal and sagittal views. The white contour is the ground truth. The red contour was generated by the proposed method. The second row visualizes the difference between surface of the ground truth and the surface of the output. (Color figure online)

Using 5-fold cross-validation on the training and validation set, we obtained a DICE (0.959 ± 0.015) and MSD (0.716 ± 0.276 mm). Similar, in almost all evaluation metrics in the test set, the algorithm achieved promising performance (DICE 0.937 ± 0.016 and MSD 1.500 ± 0.368 mm). More details about the performance and the distributions of the evaluation metrics are presented in Table 1 and Fig. 2. Furthermore, an additional test is performed to assess the clinical usefulness. The generated shapes in the test set were reviewed by a physician. Every shape is scored on the base of three criteria, namely the overall segmentation of the LA and PVs, the ridge between the Left PVs (LPVs) and the Left Atrial Appendage (LAA) and the position of the Mitral valve. The score per criterion ranges from 0 (very poor) to 5 (excellent). Moreover, the number of correctly segmented PV ostia were counted. The results of the review are presented in Table 2. It can be noticed that on average the proposed method scored a 3.9 out of 5 on the three criteria. In order to perform this qualitative evaluation, we used surface rendering of the predictions in the test set and compared them with the ground truth (Fig. 3).

Figure 3 shows an example of a segmentation in the validation and test set. The first row of the figure shows 3DRA slices from axial, coronal and sagittal views. The white contour represents the contour of the ground truth and the red contour was generated by the proposed method. The second row visualizes the difference between the surface of the ground truth and the surface of the predicted segmentation. The differences between the ground truth and the predicted segmentation of the proposed method is around 1.00 mm.

4 Conclusion and Discussion

Accurate knowledge of the shape of the LA is important for treatment of AF by cardiac ablation. In this paper we present DeepLA, an automated LA segmentation tool for 3DRA images. DeepLA was trained and optimized on 20 3DRA images by 5-fold cross-validation. Additionally, DeepLA is tested on 5 3DRA images and assessed by a physician. To the best of our knowledge, we are first to present results of a CNN based approach to segment the LA from 3DRA images.

We were able to achieve promising performance in Dice coefficient and MSD on the training and test set through a deep learning algorithm. Despite overall promising results on the validation and test sets, some local discrepancies between the ground truth and predictions can be found in the test set as reflected by the MSD values (Table 1b). This can possibly be explained by the differences in annotations of the ground truth in the training and test set (Fig. 3b). In terms of clinical usefulness, the segmentation of the LA scores well (Table 2).

We cannot directly compare our approach with those reported in the literature due to the difference in imaging modalities used and datasets. In [7], the LA is segmented from cardiac MRI data sets provided by the STACOM 2013 challenge through a 2D multi-view CNN. Their method achieved a Dice score of 0.905. Furthermore, Zheng et. al [14] and Manzke et. al [6] proposed a model-based method to segment LA from 3DRA images and obtained a mean

segmentation error of 1.30 and 1.50 mm respectively. Even though they obtained a slightly smaller mean segmentation error than our method, DeepLA is trained with 3DRA images which were acquired with a lower radiation dose and consequently higher image noise.

We can conclude that we are able to *successfully* segment the LA and PVs as a whole. Our method achieves a mean accuracy of 1.5 ± 0.368 mm with a mean processing time of 19.6 ± 5.2 s. Hence, it may be easily implemented for online use during PVI, given an ablation lesion size ≥ 5 mm.

References

1. Bajpai, A., Savelieva, I., Camm, A.J.: Treatment of atrial fibrillation. Br. Med. Bull. **88**(1), 75–94 (2008)
2. von Berg, J., Lorenz, C.: Accurate left atrium segmentation in multislice CT images using a shape model. In: Fitzpatrick, J.M., Reinhardt, J.M. (eds.) Medical Imaging 2005: Image Processing, vol. 5747, p. 351. International Society for Optics and Photonics (2005)
3. De Buck, S., et al.: Cardiac three-dimensional rotational angiography can be performed with low radiation dose while preserving image quality. Europace **15**(12), 1718–1724 (2013)
4. De Buck, S., et al.: An augmented reality system for patient-specific guidance of cardiac catheter ablation procedures. IEEE Trans. Med. Imaging **24**(11), 1512–1524 (2005)
5. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In: Proceedings of the IEEE International Conference on Computer Vision 2015 International Conference on Computer Vision, ICCV 2015, pp. 1026–1034 (2015)
6. Manzke, R., et al.: Automatic segmentation of rotational X-ray images for anatomic intra-procedural surface generation in atrial fibrillation ablation procedures. IEEE Trans. Med. Imaging **29**(2), 260–272 (2010)
7. Mortazi, A., Karim, R., Rhode, K., Burt, J., Bagci, U.: *CardiacNET*: segmentation of left atrium and proximal pulmonary veins from MRI using multi-view CNN. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10434, pp. 377–385. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66185-8_43
8. Nakagawa, H., et al.: 2017 HRS/EHRA/ECAS/APHRS/SOLAECE expert consensus statement on catheter and surgical ablation of atrial fibrillation. Heart Rhythm **14**(10), e275–e444 (2017)
9. Robben, D., Bertels, J., Willems, S., Vandermeulen, D., Maes, F., Paul, S.: DeepVoxNet: voxel-wise prediction for 3D images (2018)
10. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
11. Stender, B., Blanck, O., Wang, B., Schlaefer, A.: Model-based segmentation of the left atrium in CT and MRI scans. In: Camara, O., Mansi, T., Pop, M., Rhode, K., Sermesant, M., Young, A. (eds.) STACOM 201. LNCS, vol. 8330, pp. 31–41. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-642-54268-8_4

12. Thiagalingam, A., et al.: Intraprocedural volume imaging of the left atrium and pulmonary veins with rotational X-ray angiography: implications for catheter ablation of atrial fibrillation. J. Cardiovasc. Electrophysiol. **19**(3), 293–300 (2008)
13. Wyndham, C.R.: Atrial fibrillation: the most common arrhythmia. Texas Heart Inst. J. **27**(3), 257–67 (2000)
14. Zheng, Y., Wang, T., John, M., Zhou, S.K., Boese, J., Comaniciu, D.: Multi-part left atrium modeling and segmentation in C-Arm CT volumes for atrial fibrillation ablation. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011. LNCS, vol. 6893, pp. 487–495. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23626-6_60