

V-FCNN: Volumetric Fully Convolution Neural Network For Automatic Atrial Segmentation

Nicoló Savioli¹, Giovanni Montana^{1,2}, Pablo Lamata¹

¹ Department of Biomedical Engineering, Kings College London, SE1 7EH, UK
{nicolo.l.savioli,pablo.lamata,giovanni.montana}@kcl.ac.uk

² WMG, University of Warwick, Coventry, CV4 7IAL
g.montana@warwick.ac.uk

Abstract. Atrial Fibrillation (AF) is a common electro-physiological cardiac disorder that causes changes in the anatomy of the atria. A better characterization of these changes is desirable for the definition of clinical biomarkers, and thus there is a need of its fully automatic segmentation from clinical images. In this work we present an architecture based in 3D-convolution kernels, a Volumetric Fully Convolution Neural Network (V-FCNN), able to segment the entire volume in one-shot, and consequently integrate the implicit spatial redundancy present in high resolution images. A loss function based on the mixture of both Mean Square Error (MSE) and Dice Loss (DL) is used, in an attempt to combine the ability to capture the bulk shape and the reduction of local errors products by over segmentation. Results demonstrate a reasonable performance in the middle region of the atria, and the impact of the challenges of capturing the variability of the pulmonary veins or the identification of the valve plane that separates the atria to the ventricle.

Keywords: Cardiac Imaging, V-FCNN, Atrial Fibrillation (AF), Dice Loss, MSE, Atrial, Segmentation, Shape, Pulmonary, Veins, Clinical, Biomarkers, Anatomy, Deep Learning.

1 Introduction

Atrial Fibrillation (AF) is a common electro-physiological cardiac disorder with a large prevalence worldwide [1] that causes changes in the anatomy of the atria. A better characterization of these changes, and the substrate that causes and sustains fibrillation, is desirable for the definition of clinical biomarkers. These biomarkers can be directly started from the image (i.e. the shape of the atria [3], or the fibrotic burden from late gadolinium enhanced (LGE) magnetic resonance imaging (MRI) [4]) or from mechanistic simulations of the function (i.e. computation of the risk of arrhythmia perpetuation [2]).

There is thus a need of a fully automatic segmentation of the atria from clinical images, specially in LGE studies. Current state of the art is based in tedious and error prone manual procedures, and the main difficulty is the lack of contrast from atrial tissue and surrounding background. Fully automated

solutions are desirable to speed-up the process and remove inter- and intra-observer variability. In this direction, a combination of multi-atlas registration within 3D level-set has been proposed, reporting a reasonable performance in the main atrial body and pulmonary vein regions [7]. The large computational burden of this multi-atlas approach can be alleviated by the use of convolutional neural networks (CNNs), as it has been illustrated in [8] for the analysis of 2D MRI slices.

The first idea followed in this work is the use of 3D CNNs as the effective solution for the segmentation of the 3D LGE datasets. The starting point is the v-net [9], which takes in consideration the spatial redundancy naturally present on the entire volumetric stack with 3D-kernels, showing good benefits in different cardiac segmentation problems [11,12]. This architecture is modified to reduce the memory burden and speed-up its training. The last idea explored in this work is the choice of a sensible loss function, the joint combination of both the mean squared error (MSE) and Dice Loss, which has been reported to be beneficial as the MSE minimizes global image details while the Dice Loss reduces local over-segmentation errors [9].

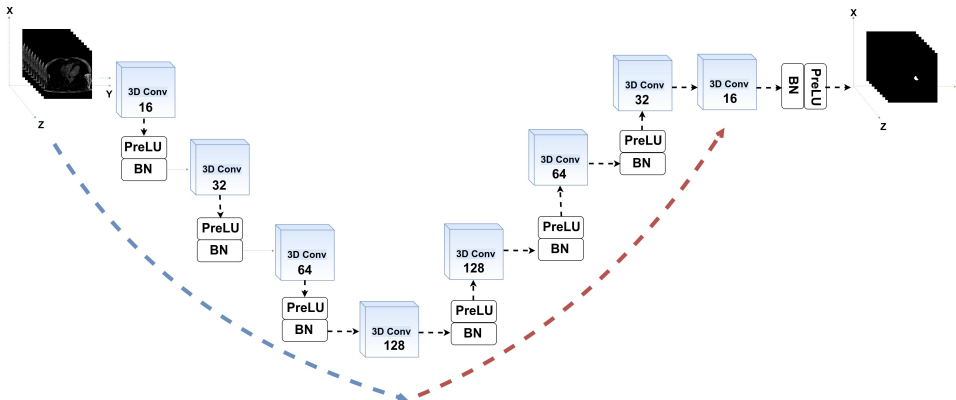


Fig. 1. V-FCNN architecture. Input is the (xyz) 3D MRI volume of size $(127 \times 127 \times 88)$, and is passed through the down-sampling path (blue arrow), represented by a 3D kernels Convolution Neural Network (CNN) able to progressively reduce the input volume slices. Then, the hidden features, at the end of it, are restored within 3D up-sampling kernels (red-arrow), ending in an output being a 3D mask of size $(127 \times 127 \times 88)$. Both down-sampling and up-sampling paths consist of four 3D-convolutions blocks (blue boxes) followed by PreLU and 3D-Batch Normalisation. The number of feature maps for each convolution layers are 16, 32, 64, 128 both in down and up-sampling.

2 Atrial Datasets

A population of 100 3D GE-MRIs, data and masks, were made available through the [2018 Atrial Segmentation Challenge] and used in this work. The scanner used for this clinical study is a whole-body MRI scanner, within a image acquisition resolution of $0.625 \times 0.625 \times 0.625 \text{ mm}^3$.

3 Method

A volumetric fully convolutional network, V-FCNN, is designed with two main paths (see Fig.1): the volumetric down-sampling path and the volumetric up-sampling path. Particularly, the volumetric down-sampling path has four 3-D convolutions blocks, each following by PreLU and 3-D Batch Normalisation (BN) layers. This path takes as input the entire (xyz) volume and progressively reduce the size of each slice (xy) and the number of stack spatial slices (z). In this phase the volume is compressed and presents both (xy) and (z) reduction. In a complementary fashion, the volumetric up-sampling path restores the compressed volume to its initial size, with every 3-D up-sampling convolution blocks being followed by PreLU and 3-D Batch Normalisation (BN). Each sampling paths of V-FCNN contains four blocks within 16,32,64,128 -3D kernels respectively (with the size fix to 3×3).

Image down-sampling during a segmentation task presents the problem of feature map reduction, followed by a strong spatial information loss. This problem has been addressed in v-net [9,11] by adding skips layers between down-sampling and up-sampling layers in order to fuse low level features within high level features. Our work explores an alternative approach to the skip path connections. In particular, we boost our model to capture fine details in two ways. Firstly, the use of max-pooling operations is avoided in order to prevent the loss of spatial resolution (i.e if pools do not overlap well, pooling operation loses appreciable information where the objects are located in the image). Secondly, with the loss proposed, we correct local pixels details through the Dice Loss which works as a local corrective loss.

The size of each MRI slice is reduced to 127×127 pixels using down-sampling bi-cubic interpolation, allowing the network to be faster in training stage and enabling a solution without expensive GPU hardware (i.e. large images need more GPU global memory). An up-sampling bi-cubic interpolation is finally used for restoring the mask to the size of 500×500 pixels.

3.1 Loss layer

The loss layer used is a combination of MSE_{loss} and $DICE_{loss}$: the $DICE_{loss}$ term searches for local details in the volume data, and the MSE_{loss} can be seen as a regularization that instead focuses on global features on the MRI volume. This bimodal loss also prevents the V-FCNN to fall in a local minima, especially within small atrial regions. Specifically, the loss is defined as:

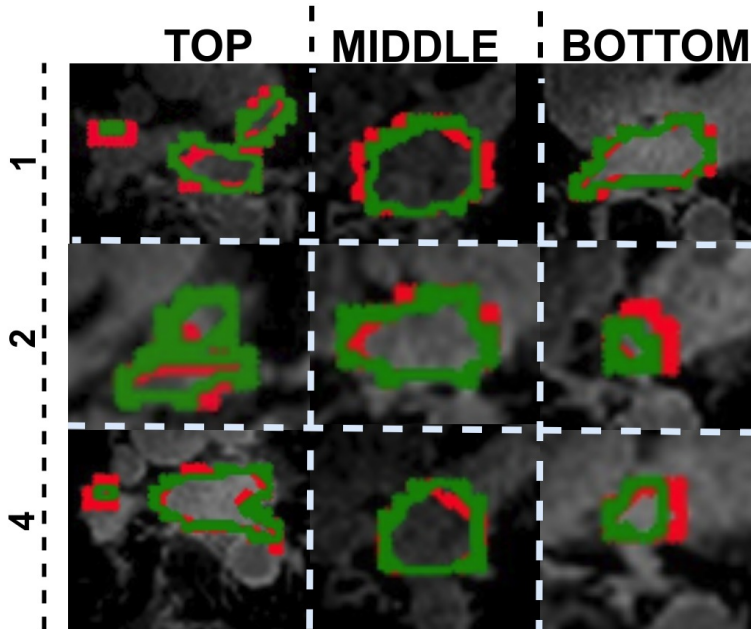


Fig. 2. Visual comparison of the segmentations obtained from V-FCNN (green line) vs clinical ground truth (red line) between three different test patients (represented by an univocal number). The comparison is made at three different sections of the atrium: top, middle and bottom. As can be seen, the V-FCNN is able to segment not only visually simpler slices (middle section) but also more complex cases (top and bottom sections).

$$\begin{aligned}
 Loss = MSE_{loss} + \lambda \cdot DICE_{loss} = & \frac{1}{Z} \sum_{s=0}^Z (y[s] - \hat{y}[s])^2 \\
 & + \lambda \cdot \frac{1}{Z} \sum_{s=0}^Z \frac{2 \sum_i^N y[s]_i \hat{y}[s]_i}{\sum_i^N y[s]_i + \sum_i^N \hat{y}[s]_i},
 \end{aligned} \tag{1}$$

where $y[s]$ are the Ground Truth (GT) binary slices, and $\hat{y}[s]$ are the correspondents prediction masks. s is an index through the spatial slices in the Z (z-axis). Then, the sums for i in $DICE_{loss}$ runs over all N pixels of the prediction masks $\hat{y}[s]$ and the $y[s]$ GT masks. Finally, λ controls the amount of $DICE_{loss}$ during the optimisation training process (set to $1e-3$).

3.2 Implementation details

For our experiments we train the network with a number of epochs of 1000 up to convergence. The Stochastic Gradient Descent (SDG) was used with a learning

rate of $1e-4$, while the momentum and weight decay are 0.9, $1e-5$ respectively. Furthermore, to increase the generalization of the network, data argumentation was used, finding particularly effective the random vertical and horizontal flip in close combination with plane translation. Input sequences were equalised in grayscale intensity with CLAHE (Contrast Limited Adaptive Histogram Equalization) [10], and noise was minimised with a combination of High-Pass Filters and Gaussian blurring filters.

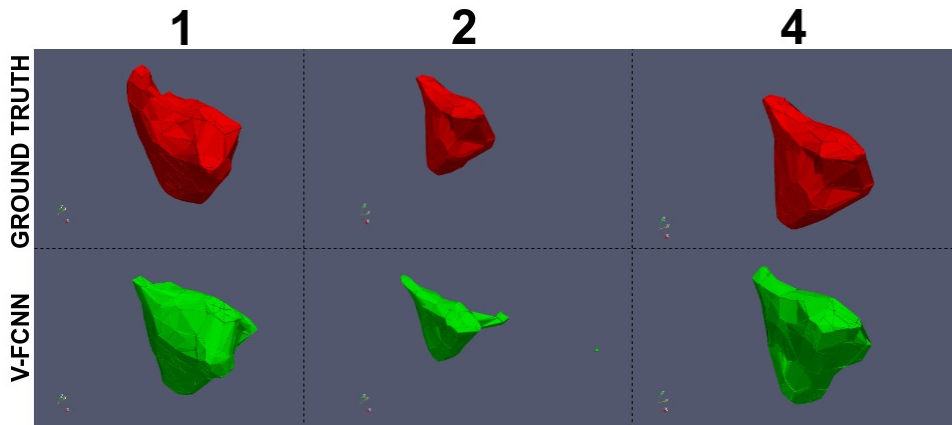


Fig. 3. Visual comparison between ground truth (red) compared with those obtained by proposed V-FCNN (green). Note that, the mesh coarse resolution is related to the low number of triangles used.

4 Experiments

We extracted five patients for validation, five for test-set, and the rest (90) is used for training. The accuracy of segmentation was measured through the Dice Metric (DM) [16], which was subdivided in 3 regions of the atria: top (including the pulmonary veins), middle and bottom (including the valve plane that divides the atria and ventricle). Besides, the surface Hausdorff Distance (HD) [17] is computed for the entire atrial anatomy.

Proposed V-FCNN achieved an average in Dice of 69.6 ± 16.1 , 82.1 ± 1.4 and 78.0 ± 6.0 in top, middle and bottom parts, respectively (see more details in Table 1, and an illustrative example in Fig. 2). Table 1 also reports the surface HD for the five testing cases, which ranges from 0.31 to 0.86, and provide a geometric view of the resemblance between the ground truth’s mesh is that generated by V-FCNN model. Particularly, given a point p and a surface S , the distance $\varepsilon(p, S)$ is defined as:

$$\varepsilon(p, S) = \min_{p' \in S} d(p, p') \quad (2)$$

whereas the $d(\cdot)$ is the euclidean distance between two points in a euclidean space. Then, the HD mesh surfaces distance between $S_{GroundTruth}$ and S_{vfcnn} is given as:

$$HD(S_{GroundTruth}, S_{vfcnn}) = \max_{p \in S_{GroundTruth}} \varepsilon(p, S_{vfcnn}) \quad (3)$$

Visual inspection of the contours reveals how a loss of accuracy is seen on the first up slices for two of the cases, creating an artificial flattening of the shape (see patients 1 and 2 of Fig. 3), and on the top of the atria with the more variable anatomy of the pulmonary veins. However, as a preliminary overall DM result, we have 76.58(7.87) within a HD mean surface of 0.59.

	<i>TOP</i>	<i>MID</i>	<i>BOTTOM</i>	<i>3D-MESH</i>
Patient number	DM (%)	DM (%)	DM (%)	HD
1	77.74 (8.46)	84.62 (2.40)	76.99 (10.89)	0.86
2	72.26 (9.50)	81.25 (1.57)	54.18 (33.70)	0.66
3	74.10 (8.92)	77.75 (1.36)	68.27 (14.75)	0.31
4	81.60 (0.74)	81.54 (1.20)	72.32 (14.32)	0.55
5	84.48 (2.58)	85.36 (0.66)	76.33 (7.05)	0.58

Table 1. The table shows the Dice Metric (DM) [16] results for all five testing patients. The results are divided in three different atrium sections: top, middle and bottom, correspondingly respectively in average and standard deviation. Finally, in the last column is show the total mean of mesh surface Hausdorff Distance (HD) [17] between geometric difference V-FCNN volumes and those provide by manual segmentation.

5 Discussion and Conclusions

The exploitation of spatial coherence across different volumetric slices is a important resource for improving the accuracy of fully automated segmentation in clinical scenario. Proposed V-FCNN achieves a good segmentation performance, mostly in the middle atrial section. Limitations occur in the top and bottom sections, caused by the presence of the pulmonary veins and the difficulty to identify where the atria and ventricle split. Comparison with other methods will be the result of the 2018 Atrial Segmentation Challenge, but in an early preliminary comparison V-FCNN presents a Dice metric between 82.05(3.43) (top case) and 69.23(14.92) (worst case) within a global average (for 5 test patients) of 76.58 (7.87).

A cardiac imaging study has a huge level of correlation in space, and many segmentation algorithms have exploited the spatial coherence within different deep-learning techniques [9,11,12,13,14]. Some of them have used 3D convolutional kernels [9,11,12] that have the advantage of directly capturing the spatial information on in each convolutional layer without adding extra parameters (i.e. the addition of a recurring network). This is the main reason that has driven

us to use them in our optimised V-FCNN. In fact, proposed V-FCNN is a simplification of the v-net [9]. The constitutive advantage of the skip-paths is to increase the localization, but this comes at the cost of a considerable slow down of the speed of the network (i.e GPU global memory leak) while propagating the gradient, at every iteration, forward and back through those paths [18].

The choice of the loss function is an important consideration in any CNN, and the joint minimization of MSE and Dice Loss has been adopted in our solution. The interpretation is that the MSE looks at global volumetric features, while the Dice Loss (DL) regularize it trying to fit local details. This choice allowed the learning to avoid local minima and the corresponding slow convergence of using only MSE. The optimal weight between them, and the inclusion of further criteria such as a L1/L2 loss or a statistical distance to a library of existing cases, should be addressed in future extensions of this work.

An alternative strategy to 3D kernels is the use of recurrent units [13,14], where recurrence is used to capture the redundancy between adjacent slices. The use of recurrence is a much more memory efficient approach (i.e less parameters to capture recurring partners), also reporting improvements at the challenging apical slices of the left ventricle [13]. But these solutions are limited by the vanishing gradient that unfortunately occurs within long sequences, what can be partially avoided by imposing an upper bound constraints on the backwards gradient (i.e. gradient clipping) or with a regularisation term [15]. Future work is thus still needed to maximise the synergies between 3D kernels and recurrent units.

Acknowledgements

This work was supported by the Wellcome/EPSRC Centre for Medical Engineering at Kings College London [g.a. 203148/Z/16/Z]. PL holds a Wellcome Trust Senior Research Fellowship [g.a. 209450/Z/17/Z].

References

1. Prystowsky EN, Benson DW, Fuster V, et al. Management of patients with atrial fibrillation?: A statement for healthcare professionals from the Subcommittee on Electrocardiography and Electrophysiology, American Heart Association. *Circulation* 1996;93:1262-1277.
2. Boyle, Patrick M. and Zahid, Sohail and Trayanova, Natalia A. Towards personalized computational modelling of the fibrotic substrate for atrial arrhythmia, *EP Europace*, vol. 18, p 136-145, 2016.
3. Varela, Marta and Bisbal, Felipe and Zacur, Ernesto and Berruezo, Antonio and Aslanidi, Oleg V. and Mont, Lluís and Lamata, Pablo. Novel Computational Analysis of Left Atrial Anatomy Improves Prediction of Atrial Fibrillation Recurrence after Ablation, *Frontiers in Physiology*, vol. 8, p. 68, 2017
4. Kim RJ, Wu E, Rafael A, et al. The use of contrast-enhanced magnetic resonance imaging to identify reversible myocardial dysfunction. *N Engl J Med* 2000;343:1445-1453.

5. McGann Christopher, Akoum Nazem, Patel, Amit, Kholmovski, Eugene Revelo Patricia, Damal Kavitha Wilson Brent, Cates Josh, Harrison Alexis, Ranjan Ravi, Burgon Nathan S., Greene Tom Kim Dan, DiBella Edward V.R., Parker Dennis, MacLeod Rob S., Marrouche Nassir F. Atrial Fibrillation Ablation Outcome Is Predicted by Left Atrial Remodeling on MRI Circulation: Arrhythmia and Electrophysiology Circ Arrhythm Electrophysiol, vol 7, pp 23-30, 2014
6. Brian J. Hansen, Jichao Zhao, Thomas A. Csepe, Brandon T. Moore, Ning Li, Laura A. Jayne, Anuradha Kalyanasundaram, Praise Lim, Anna Bratasz, Kimerly A. Powell, Orlando P. Simonetti, Robert S.D. Higgins, Ahmet Kilic, Peter J. Mohler, Paul M.L. Janssen, Raul Weiss, John D. Hummel, Vadim V. Fedorov; Atrial fibrillation driven by micro-anatomic intramural re-entry revealed by simultaneous sub-epicardial and sub-endocardial optical mapping in explanted human hearts, European Heart Journal, Volume 36, Issue 35, 14 September 2015, Pages 23902401,
7. Tao Qian, Shahzad Rahil, Ipek Esra Gucuk Berendsen Floris F, Nazarian, Saman, van der Geest, Rob J. Fully automated segmentation of left atrium and pulmonary veins in late gadolinium enhanced MRI, Journal of Cardiovascular Magnetic Resonance, vol 18, pp 08, 2016.
8. Aliasghar Mortazi, Rashed Karim, Kawal S. Rhode, Jeremy Burt, Ulas Bagci CardiacNET: Segmentation of Left Atrium and Proximal Pulmonary Veins from MRI Using Multi-View CNN, CoRR, vol abs/1705.06333, 2017.
9. Fausto Milletari, Nassir Navab, Seyed-Ahmad Ahmadi V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation, CoRR, vol abs/1606.04797, 2016.
10. H. Kaur, J. Rani, MRI brain image enhancement using Histogram Equalization techniques, International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2016.
11. F. Isensee, P. Jaeger, P. Full, I. Wolf, S. Engelhardt, and K. H. MaierHein, Automatic cardiac disease assessment on cine-mri via time-series segmentation and domain specific features, in Proc. STACOM-MICCAI, LNCS, volume 10663, 2017, pp. 120129.
12. Hinrich B. Winther, Christian Hundt, Bertil Schmidt, Christoph Czerner, Johann Bauersachs, Frank K. Wacker, and Jens Vogel-Claussen. v-net: Deep learning for generalized biventricular cardiac mass and function parameters. CoRR, abs/1706.04397, 2017.
13. Poudel Rudra P. K., Lamata Pablo, Montana Giovanni Recurrent Fully Convolutional Neural Networks for Multi-slice MRI Cardiac Segmentation. Reconstruction, Segmentation, and Analysis of Medical Images, pp 83-94, 2017.
14. Jianxu Chen and Lin Yang and Yizhe Zhang and Mark S. Alber and Danny Z. Chen Combining Fully Convolutional and Recurrent Neural Networks for 3D Biomedical Image Segmentation, CoRR, vol. abs/1609.01006, 2016.
15. Razvan Pascanu, Tomas Mikolov, Yoshua Bengio, Understanding the exploding gradient problem, CoRR, abs/1211.5063, 2012.
16. Avendi, M.R., Kheradvar, A., Jafarkhani, H., A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. Medical Image Analysis 30, 108119, 2016.
17. Cignoni, Paolo and Rocchini, Claudio and Scopigno, Roberto, Metro: Measuring Error on Simplified Surfaces, 1996, Centre National de la Recherche Scientifique, Paris, France, France.
18. Adam Paszke, Abhishek Chaurasia, Sangpil Kim, Eugenio Culurciello, ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation, CoRR, vol, abs/1606.02147, 2016