Two-Stage 2D CNN for Automatic Atrial Segmentation from LGE-MRIs

Kevin Jamart[1(✉)], Zhaohan Xiong[1], Gonzalo Maso Talou[1], Martin K. Stiles[2], and Jichao Zhao[1]

[1] Auckland Bioengineering Institute, Auckland, New Zealand
kjam268@aucklanduni.ac.nz
[2] Waikato Clinical School, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand

Abstract. Atrial fibrillation (AF) is the most common sustained heart rhythm disturbance and a leading cause of hospitalization, heart failure and stroke. In the current medical practice, atrial segmentation from medical images for clinical diagnosis and treatment, is a labor-intensive and error-prone manual process. The atrial segmentation challenge held in conjunction with the 2018 the Medical Image Computing and Computer Assisted Intervention Society (MICCAI) conference and Statistical Atlases and Computational Modelling of the Heart (STACOM), offered the opportunity to develop reliable approaches to automatically annotate and perform segmentation of the left atrial (LA) chamber using the largest available 3D late gadolinium-enhanced MRI (LGE-MRI) dataset with 154 3D LGE-MRIs and labels. For this challenge, 11 out the 27 contestants achieved more than 90% Dice score accuracy, however, a critical question remains as which is the optimal approach for LA segmentation. In this paper, we propose a two-stage 2D fully convolutional neural network with extensive data augmentation and achieves a superior segmentation accuracy with a Dice score of 93.7% using the same dataset and conditions as for the atrial segmentation challenge. Thus, our approach outperforms the methods proposed in the atrial segmentation challenge while employing less computational resources than the challenge winning method.

Keywords: Automatic cardiac segmentation · LGE-MRI · Atrial fibrillation

1 Introduction

Atrial fibrillation (AF), is the most common sustained heart rhythm disturbance, with nearly 33 millions of people affected worldwide. The current overall prevalence of AF is 2% to 5% of the general population worldwide and is projected to more than double in the following couple of decades, becoming a global epidemic [1]. Currents treatments remain sub-optimal [2] and recent clinical studies, using late gadolinium enhancement MRI (LGE-MRI), suggest that this is probably due to the lack of understanding of the underlying left atrial (LA) structures which sustain AF. Unfortunately, most studies that use LGE-MRIs have relied on labor-intensive and error-prone manual segmentation methods [3], and therefore cannot reach beyond research

M. Pop et al. (Eds.): STACOM 2019, LNCS 12009, pp. 81–89, 2020.
https://doi.org/10.1007/978-3-030-39074-7_9

studies to be implemented in clinical practice. Alas, initial attempts for automatic segmentation using conventional approaches or some early machine learning strategies have also achieved limited efficacy [5, 6].

However, in 2018, new approaches were proposed for the atrial segmentation challenge held by the Statistical Atlases and Computational Modeling of the Heart (STACOM) workshop [7], during the 2018 edition of the Medical Image Computing and Computer-Assisted Intervention (MICCAI) conference. The challenge was a success with 18 teams attending the conference and proposing diverse approaches, with the winning team, Xia et al. [8], reaching 93.2% Dice score with a fully convolutional neural network (CNN). Nevertheless, the best score of the challenge was obtained using a 3D approach to render accurately the volume of the LA, we argue that such a small 3D dataset (100 3D LGE-MRIs) might not provide enough learning material for the CNN to reach the maximum score possible. Moreover, 3D segmentation approaches usually require more computational resources and are less efficient than CNN using 2D images. Furthermore, data augmentation, proven to be an effective method to extend and enrich the dataset, would remain of limited efficacy for 3D images, as the dataset can only be so much extended, and therefore we argue that a 2D approach is more appropriate to exploit the full potential of the dataset.

Our study addresses these problems. Firstly, we built a two-stage 2D convolutional neural network using extensive data augmentation to fully exploit the dataset potential. Secondly, we investigated the impact of the main transformations employed for data augmentation in medical segmentation tasks. Finally, we analyzed the volumetric prediction yielded from the aggregated 2D predictions.

2 Methods

2.1 Dataset

For this study, we employed the 3D LGE-MRI dataset used in the 2018 atrial segmentation challenge in conjunction with the 2018 MICCAI and STACOM workshop. The dataset was acquired and labeled by experts' consensus at the University of Utah and consists of 154 original 3D LGE-MRIs with a spatial resolution of $0.625 \times 0.625 \times 0.625$ mm^3 and dimensions of $640 \times 640 \times 88$ and $576 \times 576 \times 88$ assorted with their respective manual segmentation of the LA cavity and used as *ground truth* (labels).

For the challenge, the dataset was divided into a training set (100 3D LGE-MRIs) and a testing set (54 3D LGE-MRIs) yielding a grand total of 8800 2D images MRIs and labels for training and 4752 MRIs and labels for testing, respectively. For our approach development, and fair comparison with other approaches published in the 2018 STACOM proceedings [7], we split the original training dataset into 80 3D LGE-MRIs for training and 20 3D LGE-MRIs for validation. Finally, our approach was evaluated using the other 54 3D LGE-MRIs kept unseen during training and fine-tuning of the network hyper-parameters, to replicate the challenge testing conditions.

2.2 Network Architectures

For each 3D LGE-MRI, the atrial cavity represents only a small fraction of the image volume, therefore creates a severe class imbalance (∼0.7% cavity pixels versus ∼99.3% of background). As class imbalance is a recurrent and serious problem in segmentation tasks [9] the first stage of our two-stage 2D CNN approach was dedicated to reduce the background predominance by extracting the region of interest (ROI). To this end, the first network (*Vnet 1*), using an image-based regression approach, was employed to precisely locate the LA by determining the coordinate of the LA cavity center of mass on each 2D image. Once the LA was localized, the images were cropped to an optimized size (240 × 240 pixels), increasing speed and accuracy of the training process. The second network (*Vnet 2*) was dedicated to accurately segment the LA from the cropped image. The final prediction image was reconstructed to the original image input size as illustrated in Fig. 1.

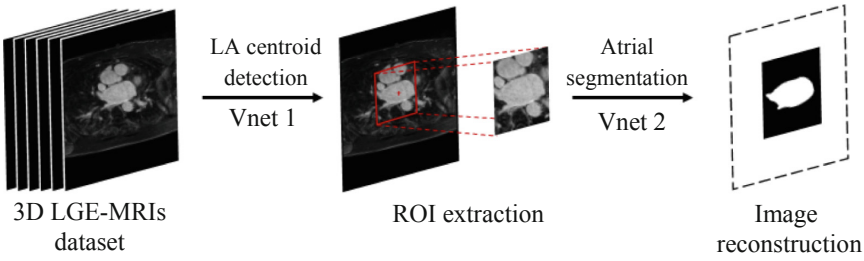

Fig. 1. Framework of our proposed two-stage 2D convolutional neural network (CNN) approach. The first network (Vnet 1) was used to determinate the center of the region of interest (ROI), i.e., left atrium (LA) cavity, on each MRI, then the images were cropped to an optimized region with a dimension of 240 × 240 pixels. The second network (Vnet 2) was used to accurately segment LA cavity.

For this study, both of the CNNs we used were based on the V-net [10] architecture as depicted in Fig. 2. In order to make the best of the dataset, we implemented a 2D version of V-net which allowed us to process 8800 2D LGE-MRIs. Our approach used a fully convolutional neural network in which the convolution operations were used to extract information, reduce the image resolution and reconstruct the image for the final output (prediction). The architecture of our V-net can be described in two parts: an initial encoder part in which the image information is extracted in a local-to-global manner, and a subsequent decoder part mirroring the encoding part and used to reconstruct the predicted segmentation. In our approach, we used 5 encoding and 5 decoding blocks where each block consisted of a batch normalization [11] followed by a succession of 5 × 5 of padded 2D convolutions keeping constant image size, and an increasing number of feature maps as the network goes deeper (respectively 8/16/32/64/128 features maps). Each block was followed by a strided 2D convolution layer for the encoding part allowing image down-sampling and global features extraction, and a strided 2D deconvolution layer for the decoding part permitting image

up-sampling and prediction reconstruction. Each convolution or deconvolution layer was followed by a leaky rectifier linear unit (Leaky ReLU) activation function (using $\alpha = 0.1$) to ensure non-linearity while limiting vanishing gradient problems. In addition, we applied 25% dropout on every layer to prevent overfitting. Moreover, we used skip connections to keep proper information forwarding as usually used in residual block architectures [12], by merging (element-wise sum) the first layer input and the final layer output of a residual block before each strided convolution or deconvolution layer. Furthermore, our network utilized horizontal features map forwarding between same level residual blocks from the encoding path, to the decoding path to avoid network singularities [13].

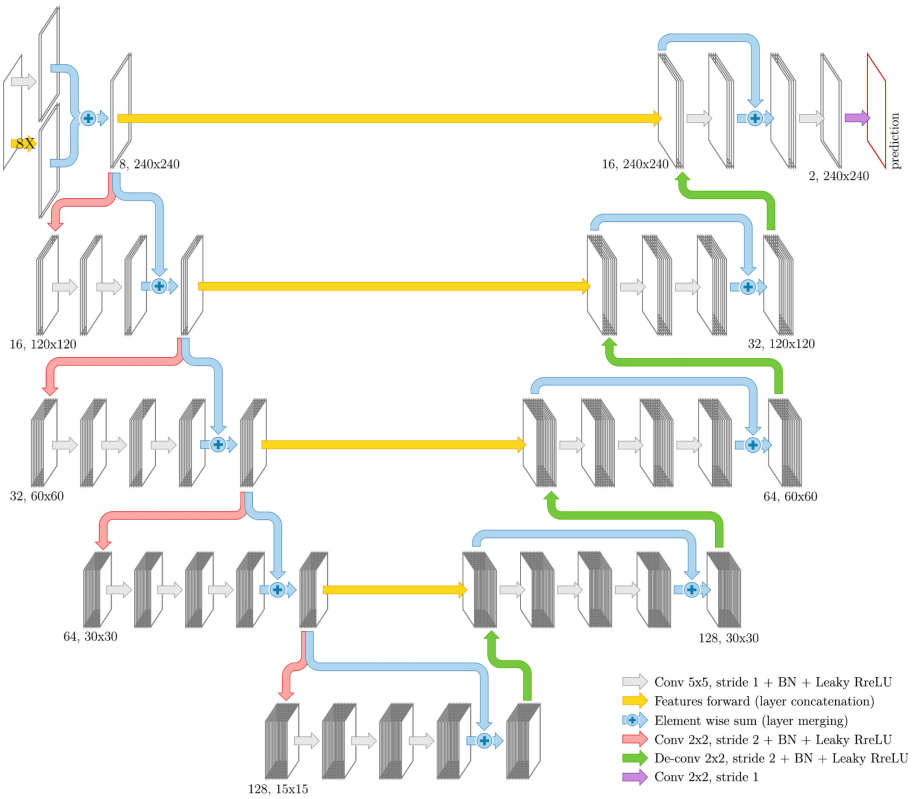

Fig. 2. Our proposed 2D V-net architecture consists of an encoder part extracting features and a decoder part reconstructing the predicted image.

The two networks differ in the loss function that we used and the activation function for the final layer of the network. For the first network (*Vnet 1*) we used mean squared error loss function and sigmoid activation function in the regression approach to determine the coordinates of the centroid of the LA. Whereas for the segmentation

task of the second network (*Vnet 2*), we used Dice loss as loss function and softmax activation function to distinguish background information from LA cavity information.

2.3 Data Augmentation

Data augmentation is a worthy tool for many application in machine learning and image processing, but every dataset presents their specificities and therefore requires tailored data augmentation. As the utilized dataset was of limited size, we applied online data augmentation, a more efficient approach than the alternative offline strategy, to train the network over a wider range of biological variations to obtain the best coverage of the human heart shape variability. To this end, we investigated the impact of four different image augmentation, two transformations – rotation and left/right flip, and two shape deformation– scaling and perspective alteration. Rotation and left/right flip addressed the relative position of the heart within the image, whereas scaling and perspective alteration varied the cavity volume and the contours of the LA, respectively. For our approach we used a rotation angle randomly selected between −25° and 25°, a scaling coefficient randomly selected between 0.5 and 1.5 for x and y, and a perspective factor ranging from 0.05 and 0.1. Moreover, we also investigated the effects of two image histogram augmentations, "add" and "gamma" addressing the contrast and brightness variations generally encountered in LGE-MRIs. Add consisted of adding selected values to each pixel values on the image (between −40 and 40), and gamma adjusted the contrast of the image by scaling each pixel value using $255 \times (I_{ij}/255)^{\gamma}$ (with γ between 0.3 and 1.7).

As image transformation can generate artifacts on the LGE-MRIs, it is important to control the emergence of these aberrant features, in order to prevent the network to learn them. Moreover, using multiple image augmentation at the same time amplifies the risk of artifacts appearance and therefore can impair the learning process. To avoid these effects, we only applied data augmentation on 50% of the dataset and only one type of image transformation and one type of image histogram augmentation to each image.

2.4 Metrics

In order to evaluate our results, several commonly used metrics were utilized to represent different aspects of the predictions. We used Dice score to evaluate the similarity between the ground truth and the predictions. We also employed Jacquard index (intersection over union) more sensible and severe upon small variation than Dice score. Moreover, we also included surface distance metrics, such as mean symmetric surface distance and Hausdorff distance, which are more representative of shape and contour accuracy of the LA than the Dice score. Finally, we added antero-posterior diameter error and volume error calculations used clinically to assess the medical relevance of the predicted reconstructed LA volumes.

3 Results

We implemented our approach using TFLearn, a high-level API of TensorFlow, and ran all experiments on Nvidia Tesla V100-PCIe with cuDNN. Our final results were obtained after training our network for 300 epochs with a learning rate of 0.001 using Adam optimizer and a maximum batch size up to 44 for speed and performance. Weights were initialized once using He normalization [14], saved, and re-employed for each experiments in order to avoid weights bias from initialization.

Performance of the Two-Stage 2D CNN Approach. Our two-stage 2D CNN achieved 93.72% Dice score accuracy, outperforming the proposed approaches for the 2018 atrial segmentation challenge. Moreover, our framework also obtained better accuracy for most of the other metrics (Jacquard index, Mean surface distance, LA diameter error and volume error) as shown in Table 1, only, the Hausdorff distance appeared larger than for the other approaches compared. We believe that the superiority of our approach relies on its two-stage architecture, the total exploitation of the dataset using 2D images and the optimized data augmentation employed. Furthermore, our approach alleviated the class imbalance issue by using small patch size images centered on the ROI, and improved the ROI learning process, providing by centering the image on the centroid of the LA cavity allowing to obtain a better Dice score. Finally, we employed carefully selected image augmentation to improve the learning process and provide an enlarged shape variability database increasing the segmentation accuracy further.

Table 1. Comparison between our approach (2D V-net), the top participants [7] of the 2018 atrial segmentation challenge and Unet 2D [15] using various metrics (Dice score, Jacquard index, Mean square distance (MSD), Hausdorff distance, Diameter error and Volume error).

Metrics						
Network	Dice score (%)	Jacquard (%)	MSD (mm)	Hausdorff (mm)	Diam. Err (%)	Volume Err (%)
Xia et al.	93.2	87.4	0.748	8.892	4.0	4.9
Huang et al.	93.1	87.2	0.754	**8.495**	3.6	4.9
Bian et al.	92.6	86.9	0.759	9.213	3.9	4.4
Yang et al.	92.5	86.1	0.850	9.759	3.6	6.1
Unet 2D	92.5	86.2	0.842	15.88	3.1	4.9
2D V-net	**93.7**	**88.2**	**0.614**	10.60	**2.7**	**4.2**

Patch Size and Centroid Cropping. As the LA on a 3D LGE-MRI represented only a fraction of the image to segment, we first investigated the effects of cropping the image to different patch size on the Dice score with decreasing patch sizes (from 512 × 512 to 240 ×240 Fig. 3A). Using image centered cropping (blue bars Fig. 3A) we observed a minor increase on the Dice score using small patches (240 × 240 to

320 × 320) rather than large patches (400 × 400 to 512 × 512). Secondly, we evaluated the importance of the location of the LA within the selected region using LA centroid cropping (red bars Fig. 3A). To this end, the first network of our approach, dedicated to locate the centroid of the LA, was able to precisely determine the coordinates of the LA centroid with a 15 pixels mean precision (mean square error = 0.35). Applying LA centroid cropping, we observed a significant Dice score increase compared to image centered cropping (form 92% to 92.56%, *p-value* > 0.01) for all patch size (Fig. 3A). Moreover, applying LA centroid cropping we noticed a significant accuracy increase using small patches (92.86% versus 92.26% for large patches, *p*-value < 0.01). These results show the importance of controlling the background, by, for example removing irrelevant background (appropriate patch size) associated with pertinent centering (LA centroid cropping), to improve the learning process, and obtain a better prediction.

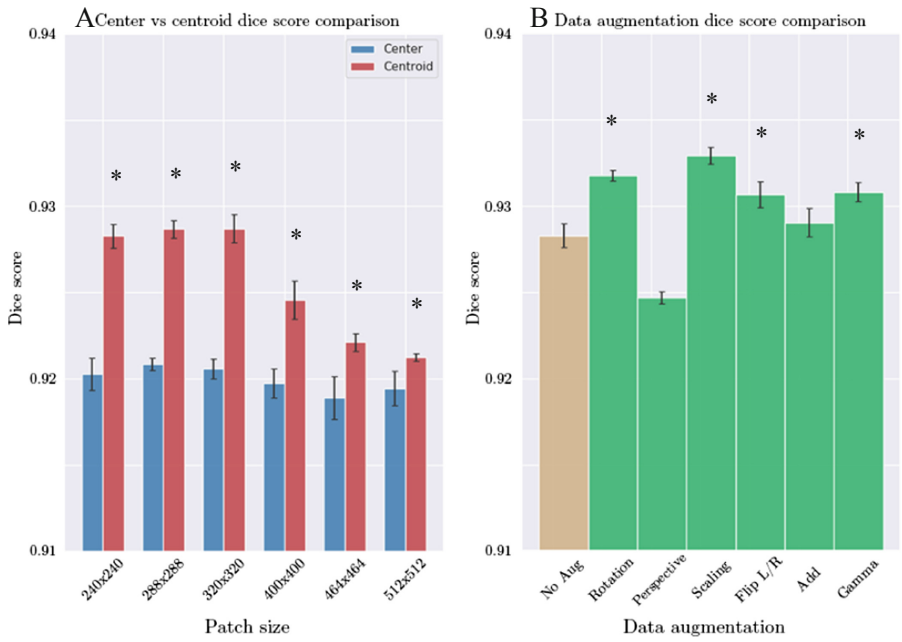

Fig. 3. Effects of patch size, region of interest extraction method, and data augmentation on Dice score. A: Comparison between the different patch size and cropping method. Using small patch size and LA centroid cropping yields better accuracy. B: Comparison of the effect on the Dice score of various data augmentation on 240 × 240 images using atrial centroid cropping.

Data Augmentation. Then, we investigated individually the effect of various online data augmentation usually employed for segmentation tasks (Fig. 3B). We showed that 4 out of the 6 image modifications employed (rotation, scaling, left/right flip, and gamma) yielded significant Dice score improvement compared with no data augmentation. However, perspective alteration worsened the results (*p*-value < 0.01), and

"add" didn't improve the Dice score significantly. This shows that whole shape alterations (rotation, scale, left/right flip) can be beneficial for the learning process while local contours modification (perspective alteration) using Dice loss can impair the learning process. Moreover, gamma also improved the Dice score, this can indicate the importance of considering the contrast as part of the data augmentation. Thus, combined inline data augmentation (rotation, scale, left/right flip and gamma) with 240x240 centroid cropped images allowed us to rise the Dice score to 93.72% (Table 1).

Segmentation Error Analysis. By examining 3D predictions, we observed that the best segmentation results were obtained at the center of the atrial volume reaching a Dice score 98.6% and a lower accuracy, mostly due to over-prediction (false positive), on the atrial regions presenting the smallest surface area, corresponding to the upper region of the LA (LA roof, Dice score 28.2%), and LA lower region (mitral valve opening, Dice score 66.9%). This can be explained by the use of the Dice loss function which weighs more towards the volume rather than the boundary of the atrium, therefore smaller labeled regions (LA roof and valve) can become over-weighted, leading, *in fine*, to false positive prediction error and a lower Dice score.

4 Conclusion

In this paper, we have proposed and extensively validated a novel two-stage 2D CNN architecture using the same dataset and conditions as the 2018 atrial segmentation challenge. Our segmentation approach achieves a segmentation accuracy with a Dice score of 93.7% outperforming all previously proposed approaches. In this study, we showed the importance of controlling the background by reducing the class imbalance using appropriate patch size and relevant region of interest centering for the learning process, we also displayed the impact of selecting pertinent data augmentation for dataset enrichment, yielding, *in fine*, better accuracy.

References

1. Krijthe, B.P., Kunst, A., Benjamin, E.J., Lip, G.Y., Franco, O.H., Hofman, A., et al.: Projections on the number of individuals with atrial fibrillation in the European Union, from 2000 to 2060. Eur. Heart J. **34**(35), 2746–2751 (2013)
2. Brooks, A.G., Stiles, M.K., Laborderie, J., Lau, D.H., Kuklik, P., Shipp, N.J., et al.: Outcomes of long-standing persistent atrial fibrillation ablation: a systematic review. Heart Rhythm **7**(6), 835–846 (2010)
3. Oakes, R.S., Badger, T.J., Kholmovski, E.G., Akoum, N., Burgon, N.S., et al.: Detection and quantification of left atrial structural remodeling with delayed-enhancement magnetic resonance imaging in patients with atrial fibrillation. Circulation **119**(13), 1758–1767 (2009)
4. Mortazi, A., Karim, R., Kawal, R., Burt, J., Bagci, U.: CardiacNET: segmentation of left atrium and proximal pulmonary veins from MRI using multi-view CNN. arXiv:170506333 (2017)

5. Tobon-Gomez, C., Geers, A.J., Peters, J., Weese, J., Pinto, K., Karim, R., et al.: Benchmark for algorithms segmenting the left atrium from 3D CT and MRI datasets. IEEE Trans. Med. Imaging **34**(7), 1460–1473 (2015)
6. Xiong, Z., Zhao, J., Stiles, M.: Machine learning for fully automatic 3D atria segmentation and reconstruction from gadolinium enhanced MRIs. Heart Lung Circ. **26**, S33 (2017)
7. Pop, M., et al. (eds.): Statistical Atlases and Computational Models of the Heart: Atrial Segmentation and LV Quantification Challenges: 9th International Workshop, STACOM 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers. Springer, Heidelberg (2019). https://doi.org/10.1007/978-3-030-12029-0
8. 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
9. Buda, M., Maki, A., Mazurowski, M.: A systematic study of the class imbalance problem in convolutional neural networks. Neural Netw. **106**, 249–259 (2018)
10. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). IEEE (2016)
11. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv:1502.03167 (2015)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
13. Orhan, E., Pitkow, X.: Skip connections eliminate singularities. arXiv:1701.09175 (2017)
14. 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, pp. 1026–1034 (2015)
15. 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