A Semi-Automatic Method To Segment The Left Atrium in MR Volumes With Varying Slice Numbers

Fatmatülzehra Uslu [1], Marta Varela [2] and Anil A. Bharath [3]

***Abstract*— Atrial fibrillation (AF) is the most common sustained arrhythmia and is associated with dramatic increases in mortality and morbidity. Atrial cine MR images are increasingly used in the management of this condition, but there are few specific tools to aid in the segmentation of such data. Some characteristics of atrial cine MR (thick slices, variable number of slices in a volume) preclude the direct use of traditional segmentation tools. When combined with scarcity of labelled data and similarity of the intensity and texture of the left atrium (LA) to other cardiac structures, the segmentation of the LA in CINE MRI becomes a difficult task. To deal with these challenges, we propose a semi-automatic method to segment the left atrium (LA) in MR images, which requires an initial user click per volume. The manually given location information is used to generate a chamber location map to roughly locate the LA, which is then used as an input to a deep network with slightly over 0.5 million parameters. A tracking method is introduced to pass the location information across a volume and to remove unwanted structures in segmentation maps. According to the results of our experiments conducted in an in-house MRI dataset, the proposed method outperforms the U-Net [1] with a margin of 20 mm on Hausdorff distance and 0.17 on Dice score, with limited manual interaction.**

I. Introduction

Atrial fibrillation (AF) affected 32.5 million people worldwide in 2010 and by 2030 its incidence in Western Europe is expected to rise to 3% [2]. AF is independently associated with a two-fold increase in all-cause mortality and increased morbidity, particularly stroke, heart failure and cognitive impairment [2]. Outcomes of AF have been linked to remodelling of the left atrium, altering its shape and size [3].

CINE bSSFP MRI is an imaging modality used to efficiently analyse the function of the heart, enabling the calculation of maximal volumes, stroke volume and ejection fractions of the left atrium (LA) [4]. However, the estimation of these biomarkers requires the segmentation of the LA from CINE MRI. In clinics, the segmentation of the LA can be performed manually by human experts or via commercial software based on semi-automatic methods [5], the latter requiring some user input to delineate the wall of the organ [6]. In both cases, manual input may add bias to the outputs and the inter-variation of the manual segmentations performed by human experts may not negligible [7].

Until recently, model based and atlas based methods were found to outperform their counterparts [8]. Recently some deep learning based methods have been proposed for the segmentation of the cardiac structures [8], [9]. However, very few of them have been developed for the segmentation of the LA [10], [11], [12], [13].

MRI scans in typical clinical settings may have lower spatial resolution and thicker slices than required for optimal measurement. Also, the number of slices in the z axis may dramatically vary from patient to patient, as demonstrated in Figure 4 for our in-house dataset. Given the varying number of slices, the generation of a shape priori per slice, as proposed in [14] for the ventricles, may not be possible. Furthermore, in CINE-MRI, slices are typically acquired in different breath-holds, which may lead to inconsistencies in the location of the atrium in adjacent slices, preventing the use of $3D$ consistency as a guide for segmentation.

In order to deal with the challenges in the segmentation of the LA, we propose a semi-automatic method using an encoder-decoder deep network. The network is provided by a chamber location map, which softly constrains the location of the LA. The generation of the map requires only a mouse click per image stack by a medical expert to locate the LA. In contrast to semi-automatic single-click region growing methods, working on pixel-wise similarity to perform segmentation, and usually requiring strong edge information to stop growing, we use the mouse-click provided location for abstract level-object detection [15]. A simple tracking method is incorporated into the proposed method to sequentially pass the location information provided by the expert to the next slice. The results of our experiments conducted on an in-house dataset show that the proposed chamber location map significantly improves the segmentation performance of a deep network when compared with that of fully automatic segmentation by the same network.

II. Method

The proposed method consists of a deep network, performing the segmentation of the LA given its rough location by a human operator or a simple tracking method in an MR slice. The tracking algorithm also automatically removes unwanted structures in a segmentation map.

A. The Proposed Architecture

In medical image analysis, encoder-decoder networks, such as U-Net [1], have been widely used for image seg-

[1] F. Uslu is with Bursa Technical University, Department of Electrical-Electronics Engineering, Turkey, fatmatulzehra.uslu@btu.edu.tr

[2] M. Varela was with the School of Biomedical Engineering Imaging Sciences, King's College London, UK. She is now with the National Heart and Lung Institute and Bioengineering Department, Imperial College London,UK, marta.varela@imperial.ac.uk

[3] A.A. Bharath is with Imperial College London, Department of Bioengineering, UK, a.bharath@imperial.ac.uk

mentation [9], [8]. In this work, we use a shallow encoder-decoder network with 12 convolutional, two max pooling and two upsampling layers containing 3 skip connections, as illustrated in Figure 1.In contrast to common architectures, we incorporate features in the first layer for the second time to decoding path in order to refine edges in segmentation maps. The total parameter count of the proposed network is around $600,000$. We use batch normalisation right after the convolutional layers, followed by a ReLU activation function. The final layer uses the sigmoid function to generate the output masks.

The network takes three types of images as input (see Figure 2): (a) the current slice to be segmented, (b) an edge emphasised version of the current slice and (c) a corresponding chamber location map, explained below. We emphasise edge information (see Figure 2 (b)) by subtracting the median of intensities from images , and then taking the square of the result.

B. A Chamber Location Map

As shown in Figure 2(a) and Figure 5, the blood pool inside the LA has very similar intensity and texture to that of other cardiac structures. Also, because the LA has very thin walls and partial volume effect is severe due to thick slices, edge information in the figures is not strong enough (even after emphasising edges as in Figure 2(b)) for a segmentation method to delineate them. When there is not a sufficient amount of training data, a deep network may not well localise the LA given its challenging appearance. In order to guide the network to segment the right structure, we introduce a way to generate chamber location maps. Figure 2(c) demonstrates some examples of our chamber location maps for corresponding images.

A chamber location map $P_l(x, y)$ can be generated by using a Gaussian-like radial weighting function, given by (1):

$$P_l(x, y) = e^{-\frac{(x-\mu_x)^2+(y-\mu_y)^2}{\sigma^2}}, \quad (1)$$

where μ_x, μ_y are the centroid location of an estimate of the LA, which can be manually input by a medical expert with a mouse click on the current slice or can be estimated from a previously segmented adjacent slice. Also, a region detector may be used to estimate μ_x, μ_y [12], [13] when there is adequate amount of training data available. $\sigma_x = \sigma_y = \sigma$ are the spread of the function along x and y axis, which are manually set. Because the size and orientation of the LA may abruptly change in subsequent slices, we fix σ for the complete dataset and encourage the network to learn the actual location of the LA by using our denoising method as follows.

a) A Denoising Method For Training With Prior Information: When a deep network is fed with a chamber location map, it may simply associate a probability level of the map with segmentation masks regardless of any structural information in the other input images. In order to avoid this situation, we purposefully add two types of noise during training: (i) noisy centroid locations $\tilde{\mu}_x, \tilde{\mu}_y$ and (ii) noisy spread parameters $\tilde{\sigma}$.

The first one imitates a weighted centroid location of the imperfectly segmented structure in an adjacent slice. This is realised by moving centroid locations calculated in the previous slice up to 5 pixels (5 mm) in any direction with a probability of 0.5. The second noise type changes the size of the LA relative to the predetermined σ of the location map. This is performed by rescaling input images and segmentation masks during data augmentation while maintaining the size of the location map, as shown in Figure 3.

C. A Tracking Method

After segmentation of the LA in an MR slice, some unwanted structures may also be labelled as LA. We propose a simple tracking method (see Algorithm 1 in Appendix) to (i) automatically remove the unwanted structures and (ii) to optionally pass prior location information across slices through an MR volume. When the tracking method performs both tasks, we call this operation "Mode 1". In contrast, if it only performs task (i), it is called "Mode 2".

We use Otsu's threshold to automatically binarize the output maps generated by our network, then take $3D$ consistency of the LA in a volume into account to keep a single segmented region. Our markers for $3D$ consistency are (i) the maximum overlapping region between a candidate region in the current slice and the region kept in a previous slice when tracking is in "Mode 1" and (ii) the minimum distance between the centroid of a candidate region in the current slice and an estimate of a centroid location of the LA in the same slice, provided by the user, when tracking is in "Mode 2". We assess the agreement of overlapping regions with the Dice score. When tracking is in "Mode 1", the centroid location of the segmentation mask is also calculated to generate a chamber location map for the next slice.

Algorithm 1: The Proposed Tracking Algorithm

```
Set tracking "Mode", 1 or 2 ;
Get output maps generated by the proposed method ;
Binarize the output maps with Otsu's threshold ;
Find independent regions in the binarized output maps with connected component analysis ;
for each region do
    if "Mode 1" then
        Calculate Dice score between the region and ground truth segmentation mask of
          the previous slice ;
    else
        Get the centroid location of the left atrium ;
        Calculate the distance between the centroid of the region and that of the ground truth
          segmentation mask of the current slice ;
    end
end
if "Mode 1" then
    Return the region with the maximum Dice score ;
    Calculate the centroid location of the region ;
    Generate a chamber location map for the next slice ;
else
    Return the region with the minimum distance ;
end
```

III. Material

47 image stacks from AF patients (31-72 years old, 75% male, all in sinus rhythm during scanning) were acquired at St Thomas' Hospital, London, UK, in a 1.5T Philips Ingenia scanner, under ethical approval and following written

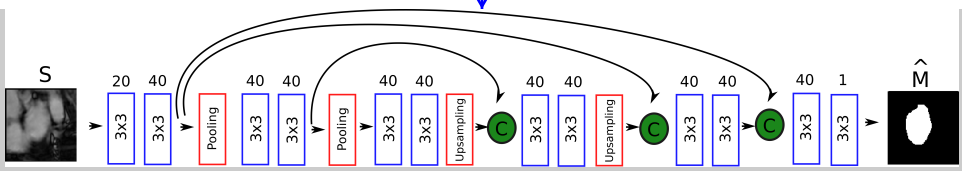

Fig. 1. The proposed network for the segmentation of the LA: upsampling modules use bilinear interpolation to upsample input features. S and $\hat{M}$ respectively denote the set of input images and the output of the network.

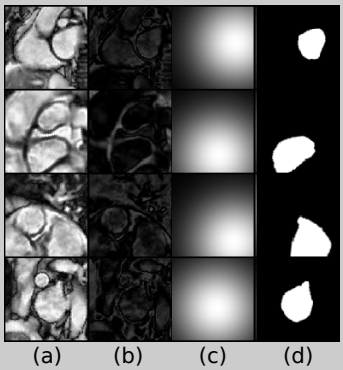

Fig. 2. Input images fed to the proposed network during training and target segmentation masks, which are cropped at the image center: (a) raw MR slices, (b) edge emphasised version of slices in (a), (c) proposed chamber location maps for $\sigma = 30$ and (d) target segmentation masks.

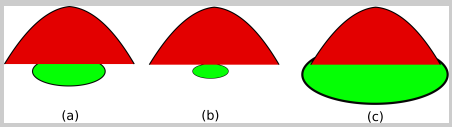

Fig. 3. Image scaling as a regularization to improve segmentation performance: Red bells show a $3D$ representation of a chamber location map whose standard deviations and mean locations are the same. Green ellipses show a target segmentation mask; (a) in its original size, (b) after scaling by a factor of 0.5 and (c) by a factor of 2 (Best viewed in color.).

informed consent. CINE MR images were acquired in a short-axis view, planned with the help of traditional 2- and 4-chamber scans. A 2D bSSFP protocol (flip angle: 60° TE/TR: 1.5/2.9 ms, SENSE factor 2). Images were acquired with ECG-based retrospective gating in a typical field of view of 385 x 310 x 150 mm, an acquisition matrix of 172 x 140 and a slice thickness of 10 mm. Images were reconstructed to a resolution of 1 x 1 x 10 mm and 50 cardiac phases with 70% view sharing. Depending on the size of the heart, the number of slices along the z axis varies between 2 to 7 across the dataset as demonstrated in Figure 4.

The images of the LA at atrial systole (AS), the end of ventricle systole (ES) and the end of ventricle diastole (ED) of a cardiac phase were manually labelled by a medical expert and then reviewed by one of the authors. Generally, the dataset suffers from partial volume effect due to thick slices and relatively low contrast (see Figure 5). Because the size of the dataset is very small – 47 image stacks – we divided it into 9 sub-sets, where we trained our network with 8 sub-sets and evaluated its performance on the remaining one.

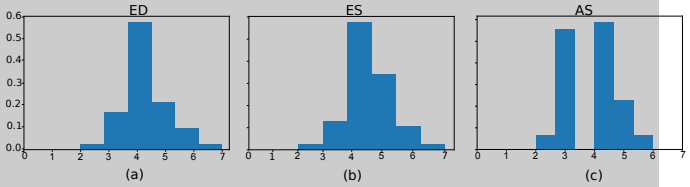

Fig. 4. The normalised histograms of slice numbers of MRIs for (a) ED phase (b) ES phase and (c) AS phase.

IV. Experimental Settings

Before normalising intensities of MR slices in the range of $[0, 1]$, we removed intensities at the 1^{st} and 100^{th} percentiles of the intensity histogram. Finally, a contrast limited adaptive histogram equalisation (CLAHE) operation [16] was used to improve the contrast of MRI images, with a tile grid size set to 8×8 pixels and a clip limit of 3.

During training, on-the-fly data augmentation was applied to input images, which included translating images up to 20 pixels (20 mm) in any direction, rotating by an angle between -45^0 and 45^0 and scaling, each with a probability of 0.2. Input images were cropped around their centers to be 96×96 pixels.

We optimised the parameters of our network with the Adam algorithm, until there was no change in performance metrics observed. Training usually lasted between 60 and 150 epochs. We set the initial learning rate to 0.00008 and exponentially decayed it by a factor of 0.92, with a mini batch size of 32 images. In order to prevent the network from over-fitting, a weight decay of 0.25 was used.

Parameters in the tracking algorithm and the chamber location map were tuned to increase Dice score on a validation dataset. We set the maximum size of the LA to 100 pixels (100 mm^2), the maximum distance between centroids as 8 pixels (8 mm) and the σ, required to calculate P_l, to 60 pixels (60 mm)(see Section II-B and II-C for details.).

V. Performance Criteria

We used Python libraries to calculate Hausdorff distance and Dice score[1] and the sensitivity and precision[2]. The Dice score, sensitivity and precision metrics were calculated for the complete test set. Hausdorff distances were computed on a slice-by-slice basis and the average Hausdorff distance is reported.

[1]https://pypi.org/project/MedPy/
[2]https://scikit-learn.org/stable/

VI. Results

We conducted three experiments to assess the performance of the proposed network on left atrium segmentation, with or without chamber location maps, and how frequently the use of these maps affected the segmentation performance of the proposed method. Table I tabulates our findings where average performance metrics over the three cardiac phases are presented.

a) Experiment 1: To compare the performance of the proposed network with a standard network, we trained a U-Net [1], with a similar parameter count to our network. In the experiment, the U-Net took only an MR slice as an input, in contrast to our network, also fed with the proposed edge enhanced MR slice. Both networks were trained without chamber location maps and then their segmentation maps were refined with our tracking method in "Mode 2". According to Table I, our network outperforms the U-Net on Dice score, Hausdorff distance and precision, with a large margin.

b) Experiment 2: In the experiment, the user was asked to click a location only at the top slice of an image stack at a cardiac phase (e.g. ED). Chamber location maps for other slices were obtained with the use of our tracking method working in "Mode 1". During training and evaluation, we simulated user inputs by calculating centroid location of ground truth segmentation masks. According to Table I, there is a large improvement over all performance metrics when compared with those obtained for the first experiment such that Dice score increases from 0.82 to 0.92, Hausdorff Distance reduces from 16.26 mm to 8.64mm, precision rises from 0.92 to 0.94 and sensitivity increases from 0.75 to 0.90.

c) Experiment 3: In contrast to the second experiment, user input was required for each slice of an image stack. The tracking method was used in "Mode 2" to only refine the segmentation maps. Despite using more user inputs in the experiment, it appears that there is a slight improvement observed in Dice score, precision and sensitivity metrics, and a considerable decrease on Hausdorff distance when compared with the performance of the second experiment (see Table I).

In the same experiment, we also assess the effect of variation in user-selected centroid locations on segmentation performance. To set this scenario, we independently added ± 3 pixels to x and y positions of the entire set of the manually selected centroid locations. For each of the test sub-sets, we ran the experiment 5 times and averaged their results. As a result of the perturbations on centroid locations, we observe a slight increase of 1.1 mm in Hausdorff distance and a slight decrease of 0.01 in sensitivity (see Table I).

d) The Effect Of Varying Slice Numbers In Segmentation Performance: Figure 5 exemplifies the impact of varying slice number for our MR image dataset at segmentation performance. As observed in the figure, in contrast to case A, segmentation methods are more prone to failure in the delineation of the LA in case B and C. As understood from the figure, the use of chamber location maps improves the segmentation performance of the proposed method for the all image stacks, case A, B and C. We can also see that there is no disadvantage of using the proposed method in "Mode 2", when the size and shape of the LA changes smoothly, as in the case of A and C. However, in case B, the shape and size of the LA changes abruptly and also the wall of the LA is not distinguishable at the third row. This leads to a leak in the segmentation, when tracking in "Mode 2". A similar situation happens in case C when tracking in "Mode 1" as a result of vague contrast of the LA wall.

VII. Conclusion

In this study, we introduced a semi-automatic method to segment the LA in MR image volumes sampled with thick slices in small datasets. For such datasets, the number of slices in MRIs may also significantly vary. As a consequence, there may be a large variation observed in the appearance of the LA regarding its shape and size. To deal with these challenges, we proposed a chamber location map which is given as an input image to a deep network, to locate the LA. Our approach also includes a tracking method to refine segmentation maps. We showed a large and consistent improvement on segmentation performance on a small in-house dataset when a single slice per volume was provided with manually detected centroid locations. When there are sufficient quantitites of labelled training data available, user inputs could be replaced with an automatic region detector, such as provided by region proposal networks [17], [18]. In this work, we manually tuned parameters used in the generation of chamber location maps and tracking methods. Future work will investigate how the parameters can be automatically tuned according to image resolution and the actual size of the LA.

Acknowledgment

This research was supported by the Wellcome/EPSRC Centre for Medical Engineering [WT203148/Z/16/Z], the Medical Imaging Network (MedIAN) [EP/N026993/1] and the British Heart Foundation Centre of Research Excellence at Imperial College London [RE/18/4/34215].

References

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[2] P. Kirchhof, S. Benussi, D. Kotecha, A. Ahlsson, D. Atar, B. Casadei, M. Castella, H.-C. Diener, H. Heidbuchel, J. Hendriks *et al.*, "2016 esc guidelines for the management of atrial fibrillation developed in collaboration with eacts," *European journal of cardio-thoracic surgery*, vol. 50, no. 5, pp. e1–e88, 2016.

[3] F. Bisbal, E. Guiu, N. Calvo, D. Marin, A. Berruezo, E. Arbelo, J. Ortiz-Pérez, T. M. de Caralt, J. M. Tolosana, R. Borràs *et al.*, "Left atrial sphericity: a new method to assess atrial remodeling. impact on the outcome of atrial fibrillation ablation," *Journal of cardiovascular electrophysiology*, vol. 24, no. 7, pp. 752–759, 2013.

[4] L. E. Hudsmith, S. E. Petersen, J. M. Francis, M. D. Robson, and S. Neubauer, "Normal human left and right ventricular and left atrial dimensions using steady state free precession magnetic resonance imaging," *Journal of cardiovascular magnetic resonance*, vol. 7, no. 5, pp. 775–782, 2005.

[5] C. Petitjean and J.-N. Dacher, "A review of segmentation methods in short axis cardiac mr images," *Medical image analysis*, vol. 15, no. 2, pp. 169–184, 2011.

TABLE I

PERFORMANCE COMPARISON OVER THE THREE CARDIAC PHASES.

Method	Experiment No	Prior Type	Dice	Hausdorff Distance (mm)	Precision	Sensitivity
U-net	1	-	0.75 ± 0.05	28.61 ± 1.10	0.73 ± 0.06	0.79 ± 0.06
Our Network	1	-	0.82 ± 0.03	16.26 ± 2.92	0.92 ± 0.03	0.75 ± 0.04
	2	Previous Slice	0.92 ± 0.01	8.64 ± 1.97	0.94 ± 0.02	0.90 ± 0.03
Our Network	3	Current Slice	0.93 ± 0.01	6.70 ± 1.33	0.95 ± 0.02	0.92 ± 0.02
	3	Current Slice *	0.93 ± 0.01	7.80 ± 1.22	0.95 ± 0.02	0.91 ± 0.02

* Random noise was added to calculated centroid locations.

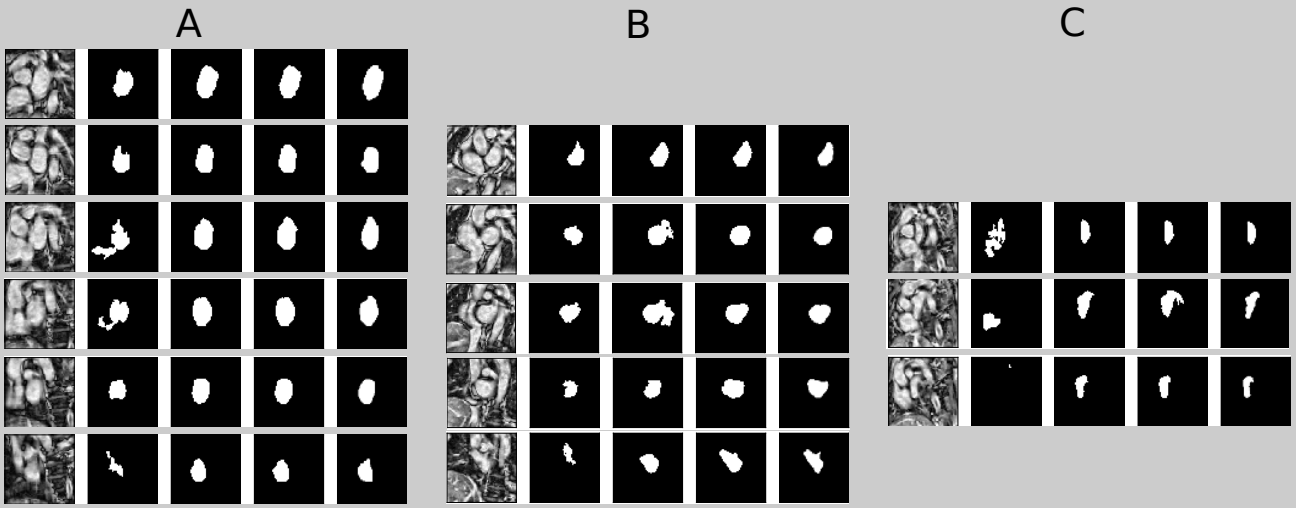

Fig. 5. Three stacks of MR images from the test dataset with varying slice numbers. From A to C, the image stacks respectively contain 6, 5 and 3 slices from ED phase. Columns from left to right show input MR images, segmentation masks generated by the U-net, those produced by our method when it is in "Mode 1", those synthesised by our method when it is in "Mode 2" and ground truth segmentation masks, consecutively.

[6] D. W. den Uijl, N. Cabanelas, E. M. Benito, R. Figueras, F. Alarcon, R. Borras, S. Prat, E. Guasch, R. Perea, M. Sitges *et al.*, "Impact of left atrial volume, sphericity, and fibrosis on the outcome of catheter ablation for atrial fibrillation," *Journal of cardiovascular electrophysiology*, vol. 29, no. 5, pp. 740–746, 2018.

[7] C. Tobon-Gomez, A. J. Geers, J. Peters, J. Weese, K. Pinto, R. Karim, M. Ammar, A. Daoudi, J. Margeta, Z. Sandoval *et al.*, "Benchmark for algorithms segmenting the left atrium from 3d ct and mri datasets," *IEEE transactions on medical imaging*, vol. 34, no. 7, pp. 1460–1473, 2015.

[8] C. Chen, C. Qin, H. Qiu, G. Tarroni, J. Duan, W. Bai, and D. Rueckert, "Deep learning for cardiac image segmentation: A review," *arXiv preprint arXiv:1911.03723*, 2019.

[9] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. G. Ballester *et al.*, "Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: Is the problem solved?" *IEEE transactions on medical imaging*, vol. 37, no. 11, pp. 2514–2525, 2018.

[10] C. Chen, W. Bai, and D. Rueckert, "Multi-task learning for left atrial segmentation on ge-mri," in *International Workshop on Statistical Atlases and Computational Models of the Heart*. Springer, 2018, pp. 292–301.

[11] Z. Xiong, V. V. Fedorov, X. Fu, E. Cheng, R. Macleod, and J. Zhao, "Fully automatic left atrium segmentation from late gadolinium enhanced magnetic resonance imaging using a dual fully convolutional neural network," *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 515–524, 2018.

[12] X. Yang, N. Wang, Y. Wang, X. Wang, R. Nezafat, D. Ni, and P.-A. Heng, "Combating uncertainty with novel losses for automatic left atrium segmentation," in *International Workshop on Statistical Atlases and Computational Models of the Heart*. Springer, 2018, pp. 246–254.

[13] S. Jia, A. Despinasse, Z. Wang, H. Delingette, X. Pennec, P. Jaïs, H. Cochet, and M. Sermesant, "Automatically segmenting the left atrium from cardiac images using successive 3d u-nets and a contour loss," in *International Workshop on Statistical Atlases and Computational Models of the Heart*. Springer, 2018, pp. 221–229.

[14] C. Zotti, Z. Luo, A. Lalande, and P.-M. Jodoin, "Convolutional neural network with shape prior applied to cardiac mri segmentation," *IEEE journal of biomedical and health informatics*, vol. 23, no. 3, pp. 1119–1128, 2018.

[15] Y. Gu, V. Kumar, L. O. Hall, D. B. Goldgof, C.-Y. Li, R. Korn, C. Bendtsen, E. R. Velazquez, A. Dekker, H. Aerts *et al.*, "Automated delineation of lung tumors from ct images using a single click ensemble segmentation approach," *Pattern recognition*, vol. 46, no. 3, pp. 692–702, 2013.

[16] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics gems IV*. Academic Press Professional, Inc., 1994, pp. 474–485.

[17] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[18] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.