

LA-Net: A Multi-Task Deep Network for the Segmentation of the Left Atrium

Fatmatülzehra Uslu¹, Member, IEEE, Marta Varela, Member, IEEE, Georgia Boniface², Thakshayene Mahenthiran, Henry Chubb, and Anil A. Bharath³, Member, IEEE

Abstract—Although atrial fibrillation (AF) is the most common sustained atrial arrhythmia, treatment success for this condition remains suboptimal. Information from magnetic resonance imaging (MRI) has the potential to improve treatment efficacy, but there are currently few automatic tools for the segmentation of the atria in MR images. In the study, we propose a LA-Net, a multi-task network optimised to simultaneously generate left atrial segmentation and edge masks from MRI. LA-Net includes cross attention modules (CAMs) and enhanced decoder modules (EDMs) to purposefully select the most meaningful edge information for segmentation and smoothly incorporate it into segmentation masks at multiple-scales. We evaluate the performance of LA-Net on two MR sequences: late gadolinium enhanced (LGE) atrial MRI and atrial short axis balanced steady state free precession (bSSFP) MRI. LA-Net gives Hausdorff distances of 12.43 mm and Dice scores of 0.92 on the LGE (STACOM 2018) dataset and Hausdorff distances of 17.41 mm and Dice scores of 0.90 on the bSSFP (in-house) dataset without any post-processing, surpassing previously proposed segmentation networks, including U-Net and SEGANet. Our method allows automatic extrac-

Manuscript received August 23, 2021; accepted September 23, 2021. Date of publication October 4, 2021; date of current version February 2, 2022. This work was supported by the British Heart Foundation under Grant RE/18/4/34215. The work of Anil A. Bharath was supported in part by the Rosetrees Trust through the Interdisciplinary Award “Atrial Fibrillation: A Major Clinical Challenge.” (Corresponding author: Fatmatülzehra Uslu.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of St Thomas’ Hospital (IRASID: 171620,191164), and performed in line with the Declaration of Helsinki.

Fatmatülzehra Uslu is with the Department of Electrical-Electronics Engineering, Bursa Technical University, 16310 Bursa, Turkey (e-mail: fatmatulzehra.uslu@btu.edu.tr).

Marta Varela is with the National Heart and Lung Institute, Imperial College London, London W12 0NN, U.K. (e-mail: marta.varela@imperial.ac.uk).

Georgia Boniface was with the Division of Biomedical Engineering and Imaging Sciences, King’s College London, London SE1 7EH, U.K. He is now with the Department of General Medicine, Darent Valley Hospital, Kent DA2 8DA, U.K. (e-mail: mckinleygeorgia@gmail.com).

Thakshayene Mahenthiran was with the Division of Biomedical Engineering and Imaging Sciences, King’s College London, London SE1 7EH, U.K. He is now with the Department of Cardiology, King’s College London, London WC2R 2LS, U.K. (e-mail: thakshayene.mahenthiran@nhs.net).

Henry Chubb was with the Division of Biomedical Engineering and Imaging Sciences, King’s College London, London SE1 7EH, U.K. He is now with the Department of Pediatrics—Cardiology, Stanford Hospital and Clinics, Lucile Packard Children’s Hospital, Palo Alto, CA 94304 USA (e-mail: henry.chubb@kcl.ac.uk).

Anil A. Bharath is with the Bioengineering Department, Imperial College London, London SW7 2AZ, U.K. (e-mail: a.bharath@imperial.ac.uk).

Digital Object Identifier 10.1109/TMI.2021.3117495

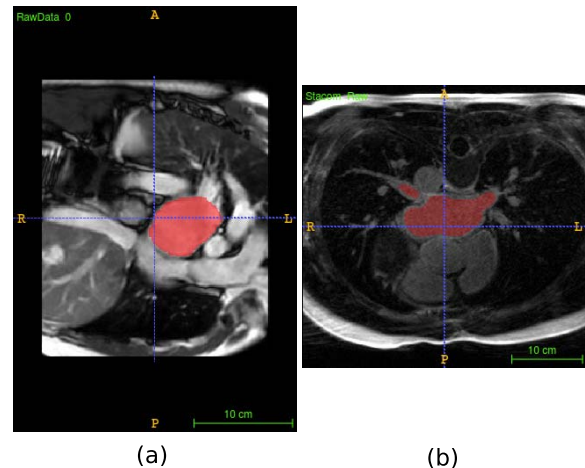


Fig. 1. Typical examples of: (a) a bSSFP image and (b) an LGE image, which are overlaid by segmentation masks of the LA, in red. (Best viewed in color.)

tion of information about the LA from MR images, which can play an important role in the management of AF patients.

Index Terms—Squeeze-excitation networks, edge detection, U-Net, à trous convolution, image segmentation, cardiac MRI.

I. INTRODUCTION

IN 2010, 32.5 million people were diagnosed with atrial fibrillation (AF) and 3% of the population older than 20 years is expected to have the disease by 2030 in Western Europe [1]. Permanence in AF leads to progressive changes in atrial tissue constitution, shape and size through AF-induced remodelling. AF is often accompanied by substantial decreases in quality of life and a high rate of hospitalizations [2]. Catheter ablation is arguably the most successful treatment for AF, but only a subset of patients respond well to it. To select patients suitable for this treatment, one may use biomarkers such as left atrial (LA) volume or diameter [3], alterations in the shape of the LA [4], changes in the symmetry of the LA shape [5] or the amount of LA fibrosis [6]. This characterisation of the LA typically relies on time-consuming manual segmentation [3], [4], [7] or delineations of LA contours in 2D views. The clinical characterisation of the LA in MRI would be greatly aided by a reliable automatic segmentation protocol, especially one that would work across different types of MR images. This would simultaneously improve the accuracy of the LA characterisation and greatly decrease the time that clinical experts require for this.

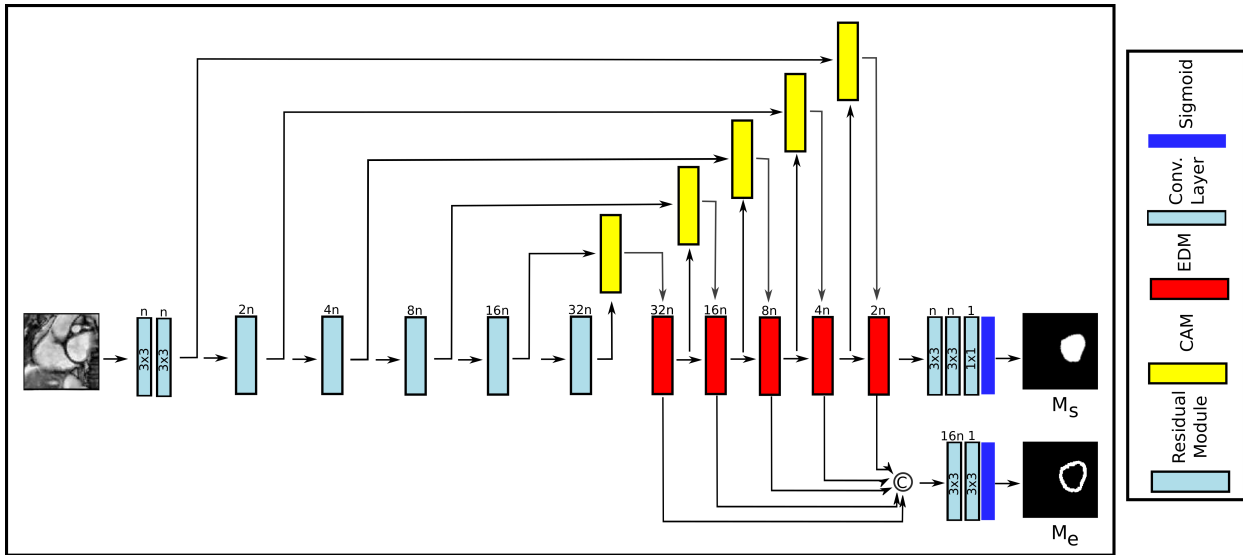


Fig. 2. LA-Net’s architecture. Output segmentation/edge masks are respectively labelled M_s and M_e . Detailed illustrations of the residual modules, CAM and EDM are shown, for n input features, in Fig. 3, 4 and 5. Numbers over module symbols show the counts of their output features. (Best viewed in color.)

Data-driven methods, for example, region growing and graph-cuts, as well as atlas-based methods, have been used for the segmentation of the LA in MRI [8]. However, their performance is limited by the low contrast of some images. Variations in the LA anatomy can be challenging for segmentation methods. These variations can emerge as a consequence of AF remodelling, but are more likely to come from topological variations of proximal sections of the pulmonary veins or the varying size or shape of the LA appendage across the population [8]. Recently, there has been an increasing interest in the use of deep neural networks for the segmentation of cardiac images [9]–[11]. However, in contrast to other cardiac structures such as the left ventricle, there has been comparably less interest in the LA for both clinical and data availability reasons [8]–[10]. Many studies on the segmentation of LA trained the U-Net [12] with dedicated modifications to improve its performance for the LA [13]–[16].

Cardiac MRI protocols rely on several different acquisition sequences, leading to very variable image contrast [17]. The most commonly used protocols for the evaluation of the LA are balanced steady state free precession (bSSFP, also known as trueFISP) and late gadolinium enhanced (LGE) MRI (see Section II-E). When cardiac image analysis literature is examined, one can see a rough grouping of methods targeting a particular MR sequence, either LGE [13], [14], [14], [18], [19] or bSSFP [15], [16], [20]. As far as we are aware, no single algorithm or network has been evaluated on both MR sequences to date.

In the medical image analysis literature, there are mainly two criteria to compare the performance of segmentation methods, regardless of imaging modality: overlap region size and surface fitness. The former one has been given more importance than the other [11] and generally quantified by Dice or Jaccard indices. The second criterion is usually measured with Hausdorff distance (HD) and average surface distance metrics (ASSD). The LA has a complicated shape with large variations in the topology of appendage and pulmonary vein insertions across patients. Successful LA

segmentation, therefore, needs to score highly on both overlap and surface distance metrics [11]. To achieve this, Yang *et al.* proposed the integration of the overlap loss and focal positive loss to encourage better delineation of LA contours [14]. However, their method did not produce a large improvement on distance metrics (see Table II). On the other hand, performing the auxiliary task of edge detection or extracting edge information through the network along with segmentation leads to improvements in the resolution of segmentation masks [21]. There are also studies incorporating model based approaches such as active contour models to deep networks, which showed better delineation of anatomical structures [22], [23].

We propose LA-Net, a multi-task network which, along with a segmentation mask, generates an edge mask as an auxiliary task to provide better delineation of LA borders. LA-Net thus explicitly penalises mistakes on the prediction of LA boundary as well as region overlap. Sharing the same path for overall segmentation and edge detection through the network leads to a low parameter count. For this work, we segment the LA from both LGE and bSSFP sequences as example MR sequences, which have very distinctive characteristics in terms of slice thickness and contrast.

II. METHOD

LA-Net is a multi-task network trained to simultaneously produce accurate LA segmentation maps and LA edges (Fig. 2). We introduce two new modules: a cross attention module (CAM) and an enhanced decoder module (EDM). A CAM makes encoder features sparse by modulating them with the guidance of decoder features (See Figure 11). An EDM smoothly incorporates segmentation features – guided by segmentation masks – and difference features – guided by edge masks – into the decoder path; therefore, this module also establishes a common path for the synthesis of both segmentation and edge masks. The shared path allows the network to have a small parameter count (approximately one million parameters) while benefiting from both target masks, allowing successful training with relatively small datasets.

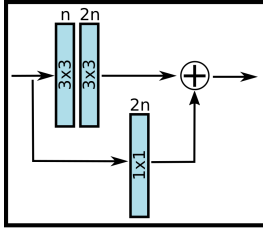


Fig. 3. A residual encoder module, REM. Note that the number of input features is n . (Best viewed in color.)

In further detail, as shown in Fig. 2, our network includes an encoder path with five *residual encoder modules (REM)*, shown in blue. There are also five *cross attention modules (CAM)*, in yellow, which are located at skip connections. The decoding path of our network contains five *enhanced decoder modules (EDM)*, in red. In contrast to the well-known U-Net [12], the network has extra depth with additional encoder-decoder modules.

There is an upsampling layer in each EDM, which doubles the resolution of the input channels using bilinear interpolation. Because MR images acquired in different scanners typically possess different characteristics, which may result in a range of segmentation performances by the same model, we use instance normalisation in the encoder path of our network and group normalisation for the rest of the network to tackle the issue [24]. Unless stated otherwise, each convolutional layer is followed by a normalisation layer and the parametric ReLU activation function [25]. The layers generating segmentation and edge masks are followed by the sigmoid function.

We minimise the loss function described with Equation (1) during training:

$$L = L_c(G_s, M_s) + \alpha L_c(G_e, M_e) \quad (1)$$

where $L_c(\cdot)$ shows the cross entropy loss. G_s and M_s respectively denote the ground truth segmentation mask and its prediction. Similarly, G_e and M_e are the ground truth edge mask and its prediction. α is a constant, adjusting the contribution of edge loss to the total loss.

Each of the proposed modules in the network is now described in more detail.

A. Residual Encoder Module, REM

Residual layers have been shown to improve the gradient propagation of deep networks by learning incremental changes over input features during training [26]. As shown in Fig. 3, an REM contains two 3×3 convolutional layers in the main path, where the first layer has a stride of 2 to reduce the resolution of input features by a factor of 2 and the second layer doubles the number of input channels, similar to the U-Net [12]. The 1×1 convolutional layer in the side path increases the number of input channels in a similar way to the main path. Encoder features generated at each path are added to produce output features of the module.

B. Cross Attention Module, CAM

The module is inspired by squeeze-excitation (SE) networks [27] and the work of Yang *et al.* [28]. In SE networks, global contexts of features are used to recalibrate the

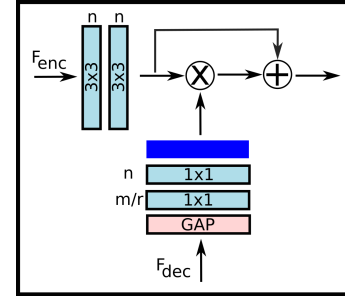


Fig. 4. The proposed cross attention module, CAM (Best viewed in color.) The numbers of encoder, F_{enc} , and decoder features, F_{dec} , are respectively n and m . \otimes refers to the Hadamard product.

importance of the same features. In contrast, Yang *et al.* used external class information to modulate features generated by a segmentation network for a video segmentation problem [28]. Our module combines the feature recalibration property of SE, which does not use class information directly, with the method of Yang *et al.*, where feature calibration is performed with an external signal. Although encoder features are used to increase resolution of decoder features, there is no guarantee that these features will point to the same structure as the decoder features. In the light of prior studies [27], [28], to allow encoder features to be refined with class information obtained with the decoder features, we scale encoder features with the global contexts of decoder features, which are obtained with global average pooling (GAP). CAMs are located at skip connections and modulate encoder features prior to sending them to the decoder path. See Figure 11 in Appendix A for features before and after a CAM.

After processing decoder features with GAP in the CAM, they are passed to a bottleneck block with two 1×1 convolutional layers. The first convolutional layer in the bottleneck block divides the number of channels by a factor of r and the second one brings it up to the number of encoder features. On the other hand, encoder features are passed through two 3×3 convolutional layers then multiplied by the context information obtained from decoder features. To increase the stability of training, we use a residual connection in the module, as illustrated in Fig. 4.

C. Enhanced Decoder Module, EDM

The local intensity of the LA may vary due to imaging noise and the flow characteristics of the blood pool. This may lead to incorrect segmentation of the LA and, furthermore, inaccurate delineation of LA borders. In the literature, it has been shown that performing the auxiliary task of edge detection or extracting edge information through the network along with segmentation, improves the accuracy of segmentation masks [21]. By following a similar reasoning, we design an EDM, demonstrated in Fig. 5. Inspired by ResNet [26], which learns incremental changes through the network, our EDM smoothly adds edge related features to the decoder path, on top of segmentation features. Our design of multitask learning differs from previous work in terms of using two different labels, where the auxiliary task of edge detection directly optimizes edge features. Because edge and segmentation features share

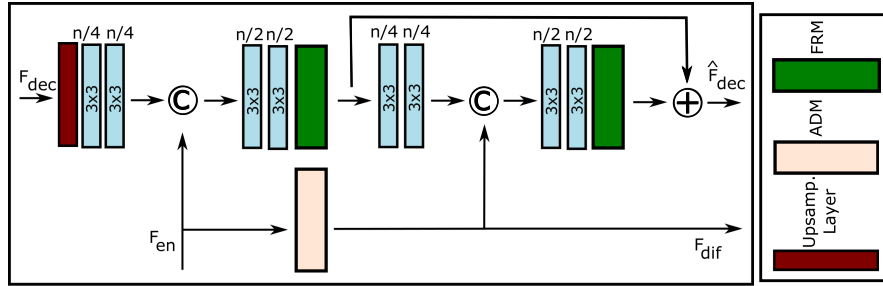


Fig. 5. The proposed enhanced decoder module, EDM. The top input line, F_{dec} , corresponds to n decoder features and the bottom one, F_{en} , represents n encoder features. \hat{F}_{dec} shows modified decoder features and F_{dif} denotes difference features. (Best viewed in color.)

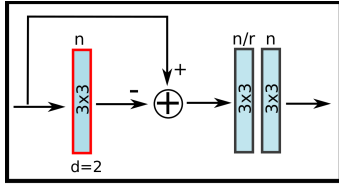


Fig. 6. The proposed à trous difference module (ADM) with depthwise separable à trous convolution layers of 3×3 . Red and black framed blue boxes respectively represent à trous and traditional convolutional layers. (Best viewed in color.)

the same path, the gradient propagation of either task supports the training of the other one.

In the first part of the module, decoder features from the previous layer are simply upsampled and combined with corresponding encoder features, similar to other encoder-decoder networks [12]. In its second part, the decoder features are joined with edge information – difference features – calculated with our à trous difference module (ADM). With the use of a residual connection [26] in the module, the second part learns only incremental edge information on top of decoder features generated in the first part. To keep only relevant information required for segmentation through the module, we integrate two feature refining modules (FRM) into an EDM. Because FRM applies GAP to input features, we use ReLU activation function in the double convolutional layers before FRM. We will explain ADMs and FRMs in more detail in the following section.

1) *À Trous Difference Module, ADM*: The proposed module is illustrated in Fig. 6. ADM intuitively works similar to unsharp masking, which is used for image enhancement [29]. Roughly, the module computes the difference between the original feature set and its blurred version using a subtraction operation. The original feature set is learned through training. Its blurred version is obtained by 3×3 à trous (dilated) convolutions with a dilation rate of d . Thanks to the use of à trous convolution, both the original and blurred feature sets have the same size. We use depth-wise separable à trous convolutions, where each convolutional layer processes a single input feature. Therefore, the module performs a one to one subtraction operation as formulated with equation (2).

$$\hat{F}_k^i = F_k^i - A^d(F_k^i) \quad (2)$$

where \hat{F}_k^i and F_k^i respectively denote the difference feature i in layer k and the input feature i in the same layer. $A^d(\cdot)$ shows an à trous convolution with a dilation rate of d .

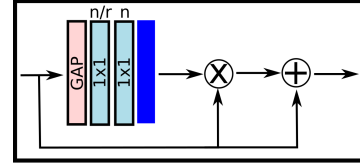


Fig. 7. Our feature refining module (FRM). (Best viewed in color.)

Although the difference features are useful for extracting details in encoder features, they may also add noise to the decoder features. To make the difference features more focused on chamber borders, they are further processed with a bottleneck layer with two 3×3 convolutional layers; the former layer is with a channel reduction ratio of r (see Fig. 6).

2) *Feature Refining Module, FRM*: Because our EDM contains two decoder modules inside, it is important to keep only informative features before passing them to its second part. We use SE networks [27] in our FRMs to calibrate decoder features by using their global contexts. Fig. 7 shows our FRM. The module firstly applies GAP to input features, which is followed by a bottleneck layer with two 1×1 convolutional layers. The former layer has a channel reduction ratio of r , similar to an ADM. The latter one is followed by the sigmoid function.

To adapt the difference features to edge detection during training, we combine the features generated at various resolution levels after upsampling them with the bilinear interpolation. Later, we process them with two 3×3 convolutional layers, where the latter one is followed by the sigmoid function (see Fig. 2) to yield an approximate edge mask, M_e .

D. Ablation Studies

We performed ablation studies to reveal how much each of the proposed modules contributes to the segmentation performance of the LA-Net. In these experiments, we removed one of: (i) the additional encoder-decoder modules (to match the depth of the original U-Net), (ii) EDMs or (iii) CAMs. For the experiment removing EDMs, we only kept the first part of EDMs and removed ADMs and FRMs. Because ADMs were not used in the network anymore for this experiment, there was no need to generate edge masks. For the other experiments, we simply removed the components from the network to see their effects on segmentation performance. All ablation experiments were conducted on the LGE-MRI (STACOM 2018) and bSFFP (in-house) datasets.

E. Cardiac MR Images

The performance of our network is evaluated on short axis MR images acquired with different MR sequences, namely T1-weighted LGE-MRI and bSSFP CINE MRI.

bSSFP gives high signal to noise ratio bright-blood images and is the sequence of choice for dynamic cardiac MRI (CINE MRI). Full coverage multi-slice CINE data is typically acquired in a short axis geometry with slice thicknesses of 8 – 10 mm.

On the other hand, LGE MRI requires the administration of a gadolinium based contrast agent several minutes before imaging. Atrial LGE MRI typically is acquired in a short-axis geometry using T_1 -weighted fast gradient echo (FLASH) sequences. This sequence is mostly used for the identification of atrial scar or fibrosis in the LA and healthy myocardium is thus typically dark in them [6], [30]. Fig. 1 shows the variable contrast of images of the LA captured by bSSFP and LGE MR sequences.

III. MATERIAL

We assessed the performance of the proposed neural network on two distinct types of atrial short-axis MRI: 1) LGE-MRI, obtained from the STACOM 2018 Atrial Segmentation Challenge dataset¹ and 2) CINE (dynamic) MRI acquired in house.

We used 100 LGE-MR images from the STACOM 2018 dataset, acquired in AF patients across multiple sites and reconstructed with an isotropic resolution of $0.625 \text{ mm} \times 0.625 \text{ mm} \times 0.625 \text{ mm}$. The images were acquired using a 3D inversion recovery prepared gradient echo pulse sequence with fat suppression. Each MRI volume contains 88 slices. The ground truth consists of LA blood pool segmentation masks performed manually by experts, which include the LA appendage, mitral valve and the antrum region of pulmonary veins. We assigned 70 volumetric images for training, 10 for validation and 20 for evaluation, as performed by [31].

The second dataset consists of 47 short-axis CINE MRI volumes acquired in AF patients at St Thomas' Hospital, London, UK, under ethical approval, which was granted by the Ethics Committee of St Thomas' Hospital (IRASID:171620, 191164). Images were acquired with a 2D bSSFP sequence, reconstructed with a spatial resolution of $1 \text{ mm} \times 1 \text{ mm} \times 10.00 \text{ mm}$ and for 20 phases of the cardiac cycle. There is no gap between slices. In contrast to STACOM 2018, the bSSFP dataset covers the entire heart and typically does not include the LA in more than half of the MR slices acquired in each patient. There are 14-20 slices in each MRI volume, of which 4-8 slices cover the LA. All slices were used during training and performance evaluation. The LA was manually segmented by experts at three cardiac phases: atrial systole (AS), end ventricular systole (ES) and end of ventricle diastole (ED), which increased the total number of MRI volume to 47×3 . These segmentation masks were used as ground truth, which contain LA blood pool and LA appendage but not pulmonary vein insertions. Our training/validation/test set contains 24/3/20 patients.

¹<http://atriaseg2018.cardiacatlas.org>

TABLE I
PERFORMANCE OF LA-NET ON STACOM 2018 DEPENDING ON
VARIOUS α IN EQ.(1)

α	HD (mm)	ASSD (mm)	Jaccard	Dice
0.25	17.66 \pm 17.39	0.96 \pm 0.31	0.85 \pm 0.03	0.92 \pm 0.02
0.5	11.84 \pm 5.35	0.88 \pm 0.18	0.85 \pm 0.02	0.92 \pm 0.01
1.	12.43 \pm 7.20	0.86 \pm 0.24	0.86 \pm 0.03	0.92 \pm 0.02

IV. PERFORMANCE METRICS

We used the medical image processing library² in Python to compute: Hausdorff distance (HD), Jaccard similarity, Dice score and Average Symmetric Surface Distance (ASSD) metrics for each MR volume after thresholding a volume with Otsu's threshold. It should be noted that there is no post-processing applied in any experiments. We performed the Wilcoxon signed-rank test, without any family-wise error correction, to assess the statistical significance of performance differences between different tested networks.

V. EXPERIMENTAL SETTINGS

We generated edge masks of the LA to be used as ground truth edge masks for M_e in Fig. 2. We firstly applied the Sobel edge detector to segmentation masks, followed by a dilation operation with a square structuring element of 3×3 pixels. We set the number of channels to $n = 4$ in the LA-Net (see Fig. 2) and assign n in other modules (see Fig. 3-7) according to the number of features entering the modules, in the main network. The dilation rate d and channel reduction ratio r were set to 2.

We optimise the parameters of our network with the Adam algorithm by minimising the loss function, L , presented in equation (1). We conducted an experiment on STACOM 2018 dataset to decide on α , whose results are given in Table I. Because $\alpha = 1$ produces a good balance between region and distance based metrics, this value was thus used in all experiments.

We normalised the intensity of the signal in each MR slice, to restrict it to values between 0 and 1 [32]. We additionally cropped input images at their centers to make their size 256×256 pixels for the STACOM 2018 dataset and 192×192 pixels for our in-house dataset. Because the sizes of the datasets used are small, we applied data augmentation on-the-fly to increase the variety of training samples. Data augmentation for both datasets included translating each input image by at most (± 10 , ± 10) pixels from its center, rescaling image size with a factor sampled from [0.8, 1.2] for STACOM 2018 and [0.5, 1.5] for our in-house dataset, and contrast augmentation using gamma correction [19]. For our-in-house dataset, we also applied vertical and horizontal flipping, rotation by 90° and -90° , each with a probability of 0.1.

To avoid a potential over-fitting during training with the small size of our in-house dataset, we used transfer learning where the weights of the networks trained with the STACOM 2018 dataset were used for the weight initialisation of the same networks. The learning rate was set to be 0.0005

²<https://pyip.org/project/MedPy/>

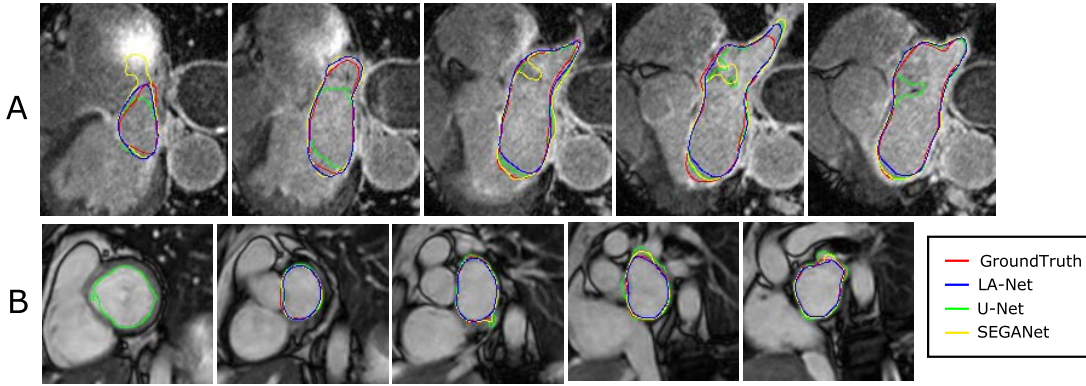


Fig. 8. The contours of the LA are shown overlaying LGE images in (A) and bSSFP images in (B). (Best viewed in color.)

for the STACOM 2018 dataset and 0.0001 for the bSSFP dataset, which was reduced by a factor of 0.9 exponentially. To avoid over-fitting, we set the weight decay to 0.001 for the former dataset and 0.01 for the latter dataset. Training continued for 34 epochs for the STACOM 2018 dataset and 30 epochs for our in-house dataset, which was determined with early-stopping. We used reflection padding for 3×3 convolution [19]. Each mini-batch contained 8 slices.

For performance comparison purposes, we trained the U-Net [12], with two dropout layers with a probability of 0.5 to avoid over-fitting and SEGANet, [16] using almost the same training settings as in LA-Net, where we also adjusted weight decay to avoid any over-fitting. We also used transfer learning for both networks when training them with the bSSFP dataset. To observe the contribution of training with edge detection in other networks, we also trained a U-Net with the additional task of edge detection, which will be called U-Net-e in following experiments.

VI. RESULTS

LA-Net was able to successfully segment the LA in both bSSFP MRI and LGE MRI, by producing smaller distance and larger overlapping based metrics.

A. Visual Assessment of Estimated LA Contours

When compared with LGE images, CINE bSSFP images show a larger contrast between the LA and the surrounding structures. As seen in Fig. 8, while SEGANet and the U-Net struggle to correctly delineate the LA in LGE slices in the top row, LA-Net shows the best fit. We observe a different problem in the bottom row of the same figure where the U-Net confidently label the left ventricle as the LA, and the U-Net and SEGANet occasionally deviate from the correct LA boundaries in the other slices. Only LA-Net consistently predicts accurate LA boundaries.

B. Performance Assessment on the LGE-MRI STACOM 2018 Dataset

Table II tabulates the performance of our network and that of previous studies on the LGE-MRI STACOM 2018 dataset. LA-Net outperforms previous methods on distance metrics, with a mean HD of 12.43 mm and a mean ASSD of 0.86 mm

TABLE II
PERFORMANCE COMPARISON OVER THE LEFT ATRIAL
LGE-MRI 2018 CHALLENGE DATASET

Method	HD (mm)	ASSD (mm)	Jaccard	Dice
LA-Net	12.43 \pm 7.20	0.86 \pm 0.24	0.86 \pm 0.03	0.92 \pm 0.02
U-Net	16.48 \pm 6.15	0.99 \pm 0.27	0.84 \pm 0.04	0.91 \pm 0.02
U-Net-e	21.23 \pm 11.64	0.99 \pm 0.23	0.85 \pm 0.03	0.92 \pm 0.02
SEGANet	15.47 \pm 5.26	1.00 \pm 0.21	0.84 \pm 0.03	0.91 \pm 0.02
Bian <i>et al.</i> [31]	17.89		0.87	0.93
Chen <i>et al.</i> [19]	14.23 \pm 4.83	1.04 \pm 0.32	0.82 \pm 0.06	0.90 \pm 0.03
Yang <i>et al.</i> [14]	18.29		0.86	0.92
Jia <i>et al.</i> [33]	19.66 \pm 4.52			0.92 \pm 0.03
Li <i>et al.</i> [34]				0.92

while producing comparable or better overlap performance metrics, with a mean Dice score of 0.92 and a Jaccard index of 0.86. We found that the superiority of LA-Net over the U-Net and SEGANet on the segmentation of the dataset is significant ($p < 0.01$ on Dice, Jaccard and ASSD metrics for the U-Net and $p < 0.01$ on the four metrics for SEGANet.). LA-Net significantly outperforms U-Net-e on distance metrics of ASSD and HD ($p < 0.01$). LA-Net also outperforms other studies given in Table II on HD and ASSD and show better or comparable performance on Dice and Jaccard indices. The only exception is the results of Bian *et al.* [31], where their method produces slightly larger Dice and Jaccard scores. Of previous studies in Table II, only Yang *et al.*, Li *et al.* and Jia *et al.* used 3D networks. The others trained 2D architectures. It should be noted that Bian *et al.*, Chen *et al.*, Yang *et al.* and Jia *et al.* reported to remove isolated structures in final segmentation masks [14], [19], [31], [33], which is not the case for LA-Net, U-Net, U-Net-e and SEGANet.

Fig. 9 shows some segmentation examples by LA-Net, U-Net and SEGANet on STACOM 2018 dataset. The segmentation masks from the three networks show some discontinuities across slices, because they both operate on 2D slices without the knowledge of the 3D structure. Furthermore, the very low contrast of the LA wall increases the labelling uncertainty of pixels in close proximity to the periphery of the LA. However, when compared with the U-Net and SEGANet, LA-Net yields a more accurate estimation of LA surface. It preserves the overall shape of the LA better, without adding false structures or missing existing ones (examine locations pointed by arrows in Fig. 9).

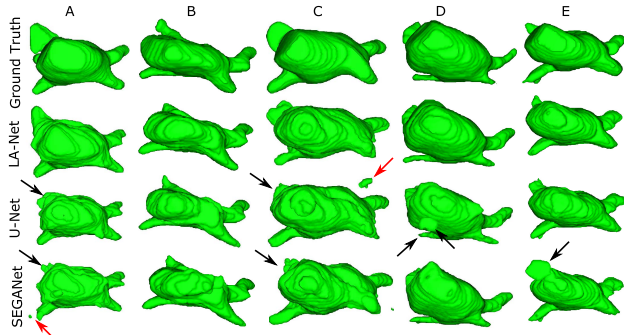


Fig. 9. Some segmentation results from the STACOM 2018 dataset. Red arrows show extra structures labelled by U-Net and SEGANet. Black arrows point to structures under or over segmented by the same networks.

TABLE III

PERFORMANCE COMPARISON OVER OUR IN-HOUSE LEFT ATRIAL CINE-MRI DATASET

Method	HD (mm)	ASSD (mm)	Jaccard	Dice
LA-Net	17.50 \pm 15.49	1.47 \pm 1.33	0.82 \pm 0.10	0.90 \pm 0.06
U-Net	38.43 \pm 33.06	2.58 \pm 2.32	0.75 \pm 0.09	0.86 \pm 0.06
U-Net-e	17.48 \pm 14.99	1.57 \pm 1.32	0.81 \pm 0.09	0.89 \pm 0.06
SEGANet	39.16 \pm 37.55	2.07 \pm 1.89	0.80 \pm 0.09	0.89 \pm 0.06

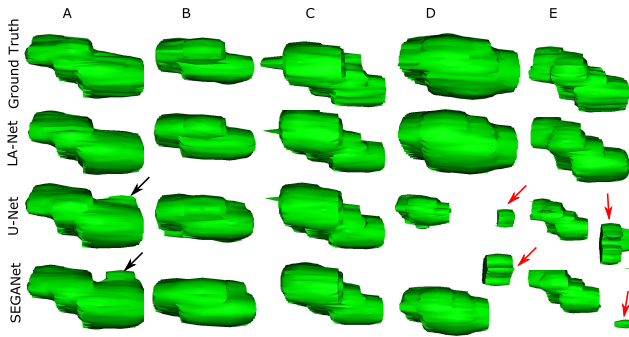


Fig. 10. Some segmentation results from our in-house dataset. Red and black arrows respectively show isolated and connected false positives. In (D) and (E), we down-scale the segmentation masks of the U-Net and SEGANet to fit them into the figure.

C. Performance Assessment on the bSSFP MRI In-House Dataset

Table III compares the performance of our network to that of U-Net on our in-house bSSFP MRI dataset. As before, LA-Net significantly outperforms the U-Net and SEGANet on all four performance metrics with $p < 0.05$ on the four performance metrics for two networks (see Table III). As in the LGE dataset, we observe a striking enhancement on both distance metrics when compared to those obtained by the U-Net and SEGANet. U-Net-e generates very close performance to LA-Net in this dataset.

Fig. 10 shows representative segmentation results by LA-Net, U-Net and SEGANet on our in-house dataset. As understood from the figure, the three networks show similar fit to the surface of the LA, however, the latter two networks suffers from extra labelled structures.

We also compare the performance of LA-Net with that of U-Net and SEGANet, across the three cardiac phases, in Table IV. LA-Net consistently is observed to outperform the

TABLE IV

PERFORMANCE COMPARISON ACROSS THE THREE CARDIAC PHASES OF AS, ED AND ES, OVER OUR IN-HOUSE LEFT ATRIAL CINE-MRI DATASET

Method	Cardiac Phase	HD (mm)	ASSD (mm)	Jaccard	Dice
LA-Net	AS	17.41 \pm 17.00	1.43 \pm 1.17	0.83 \pm 0.08	0.91 \pm 0.05
	ES	13.63 \pm 7.26	1.28 \pm 1.01	0.83 \pm 0.07	0.91 \pm 0.04
	ED	21.46 \pm 18.71	1.68 \pm 1.68	0.79 \pm 0.13	0.88 \pm 0.09
U-Net	AS	34.12 \pm 31.37	3.00 \pm 3.06	0.76 \pm 0.08	0.86 \pm 0.05
	ES	36.57 \pm 31.63	1.81 \pm 1.44	0.78 \pm 0.07	0.87 \pm 0.05
	ED	44.37 \pm 35.10	2.94 \pm 1.99	0.71 \pm 0.09	0.83 \pm 0.07
SEGANet	AS	32.73 \pm 33.88	1.66 \pm 1.25	0.83 \pm 0.07	0.91 \pm 0.04
	ES	44.11 \pm 39.79	2.02 \pm 1.56	0.81 \pm 0.06	0.89 \pm 0.04
	ED	40.33 \pm 37.69	2.51 \pm 2.51	0.77 \pm 0.11	0.87 \pm 0.08

TABLE V

ABLATION STUDIES ON STACOM 2018. FOR E-D: Extra Depth, EDM AND CAM SEE THE MAIN TEXT. (THE PERFORMANCE METRICS FOR THE FULL LA-NET ARE SHOWN IN THE TOP ROW)

E-D	EDM	CAM	HD (mm)	ASSD (mm)	Jaccard	Dice
✓	✓	✓	12.429 \pm 7.198	0.862 \pm 0.237	0.855 \pm 0.013	0.921 \pm 0.018
✗	✓	✓	15.278 \pm 6.771	1.011 \pm 0.302	0.841 \pm 0.030	0.913 \pm 0.018
✓	✗	✓	16.086 \pm 11.182	0.965 \pm 0.242	0.842 \pm 0.029	0.914 \pm 0.029
✓	✓	✗	15.045 \pm 9.862	0.965 \pm 0.299	0.845 \pm 0.036	0.916 \pm 0.021

TABLE VI

ABLATION STUDIES ON OUR IN-HOUSE DATASET. FOR E-D: Extra Depth, EDM, CAM AND TL: Transfer Learning SEE THE MAIN TEXT. (THE PERFORMANCE METRICS FOR THE FULL LA-NET ARE SHOWN IN THE TOP ROW)

E-D	EDM	CAM	TL	HD (mm)	ASSD (mm)	Jaccard	Dice
✓	✓	✓	✓	17.500 \pm 15.489	1.466 \pm 1.332	0.819 \pm 0.096	0.897 \pm 0.063
✗	✓	✓	✓	14.400 \pm 5.219	1.725 \pm 1.120	0.790 \pm 0.087	0.880 \pm 0.056
✓	✗	✓	✓	26.595 \pm 28.390	2.020 \pm 2.138	0.800 \pm 0.093	0.886 \pm 0.061
✓	✓	✗	✓	15.610 \pm 6.813	1.445 \pm 1.065	0.810 \pm 0.085	0.893 \pm 0.054
✓	✓	✓	✗	19.615 \pm 22.075	1.711 \pm 1.305	0.798 \pm 0.097	0.884 \pm 0.064

other networks for the four performance metrics. There was no significant difference in segmentation performance across the cardiac phases of AS, ED and ES for LA-Net and SEGANet; however, we found a significant difference between ED and ES phases of the U-Net with $p < 0.05$ for the four performance metrics.

D. Ablation Studies

The results of our ablation experiments are given in Table V for STACOM 2018 dataset and in Table VI for in-house dataset. We found that EDM, the extra depth and CAM are significantly important for STACOM 2018 dataset with $p < 0.05$ for the four performance metrics, excluding CAM on HD. On the other hand, in our in-house dataset, the presence of the extra depth significantly improved the performance of LA-Net on Dice and Jaccard scores with $p < 0.05$ while that of EDMs showed significant improvement for only HD metric ($p < 0.05$). We did not see any significant improvement due to the use of transfer learning in our in-house dataset.

VII. DISCUSSION

We present LA-Net, a novel left atrial segmentation network that also minimises discrepancies in LA edge masks. LA-Net outperforms previous LA segmentation methods, particularly on distance based metrics, on two different MR sequences: LGE (STACOM 2018) and bSSFP (our in-house) datasets.

We found that explicitly minimising the loss on LA edge masks generated by our EDM leads to more accurate LA contours and also improves overlap based performance metrics. LA-Net outperforms a previous method to improve the delineation of LA contours with a large margin of roughly 6 mm on HD [14] (see Table II). We also found that our method is more effective at reducing errors in boundary estimation of the LA than previous methods combining various loss functions for the same aim [14], [33], without any post-processing. We also observed that generating edge masks per se do not necessarily improve distance based metrics (see Table II for our result on U-Net-e). These are important findings based on the argument of Jamart *et al* [11] that boundary distance based performance metrics are more important than volume overlap based scores to compare the performance of segmentation methods for anatomical structures with complicated shape, such as the LA.

LA-Net’s design is robust enough to outperform dedicated LA segmentation networks on two different types of MR images: bSSFP and LGE. Although the network faces different challenges in these datasets such as the abundance of non-LA slices in the former dataset [35] and very low contrast in the latter one, LA-Net copes well with these difficulties. Of the two MR images analysed in this study, LA-Net has a better performance on the LGE dataset. However, we believe that this is not a sign of sequence-sensitive performance, but the result of having a small training dataset for the bSSFP sequence, which is almost half the size of the LGE dataset regarding patient number. To avoid any potential overfitting to the bSSFP dataset, we used transfer learning by initializing the parameters of a network with those trained with LGE images.

VIII. CONCLUSION

As far as we are aware, there has been no study investigating the automatic segmentation of the LA using the same neural network in MR sequences with different contrasts. Future work will explore if the proposed network can easily be extended to include other MR sequences and other cardiac chambers such as the right atrium and the aortic root. We will also investigate whether our network can also benefit from smart training strategies to further improve segmentation performance, such as curriculum learning [19]. Moreover, we will assess the performance of LA-Net on larger cardiac MR datasets, preferably collected from multiple sites, to evaluate the use of our network in the clinical setting.

In short, we present LA-Net, a simultaneous segmentation and edge-matching network which outperforms current state-of-the-art segmentation networks on two different types of MR images. We expect this network to contribute to a better clinical characterisation of LA, greatly reducing the burden of manual processing tasks and improving the accuracy of LA imaging biomarkers. This could be particularly useful for the clinical management of atrial fibrillation patients.

APPENDIX A

THE TRANSFORMATION OF FEATURES WITH CAM

Figure 11 shows features generated for the bSSFP dataset before and after CAM. Note the sparsity and refined features after the CAM. The network puts more emphasis on

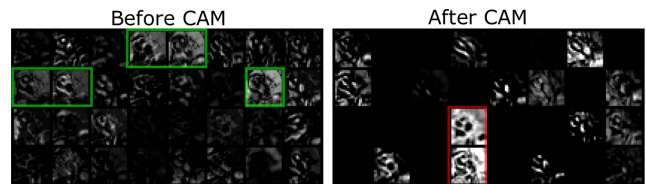


Fig. 11. Features before and after a CAM. Feature size is 24×24 pixels. See text for more details. (Best viewed in color.)

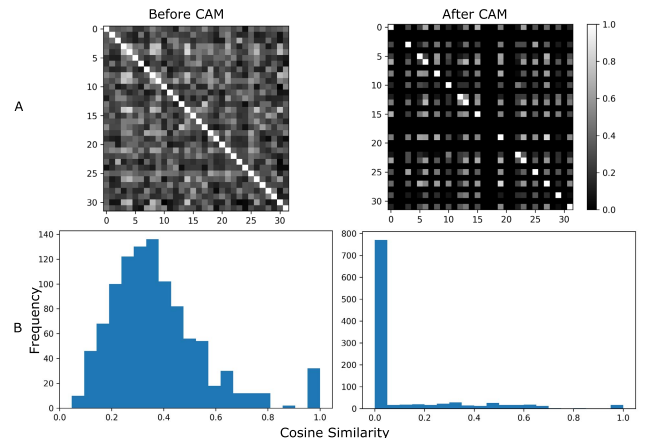


Fig. 12. Heat maps (A) and histograms (B) of pairwise cosine similarity of features before and after CAM.

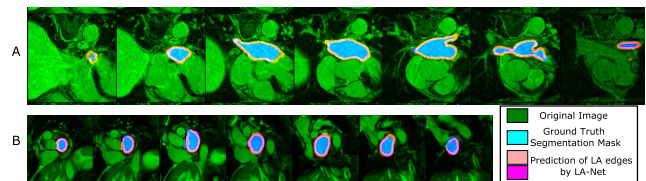


Fig. 13. Edge prediction by LA-Net on LGE, (A), and bSSFP, (B), datasets.

anatomical topology with the use of CAMs, which is indicated with clearly visible cardiac chambers, framed in red. In contrast, the approximately correspondent features before CAM, framed in green, suggest a greater weighting on the distinction between cardiac chambers.

To see if there is any decrease in feature redundancy after the use of CAM, we calculated pairwise cosine similarities of features [36]. The more features are dissimilar, the closer their cosine similarity to 0. Figure 12 displays the heat maps and histograms of the calculated similarities for features before and after CAM. After CAM, the histogram of feature similarities is mostly dominated with closer-to-zero values, indicating greatly reduced feature redundancy.

APPENDIX B

EDGE PREDICTION BY LA-Net

Figure 13 shows estimated edges by LA-Net, where our network well approximates LA boundaries.

APPENDIX C

LA VOLUME COMPARISON

Figure 14 shows predicted volumes by LA-Net. Because of different characteristics of STACOM 2018 dataset and our in-house dataset, LA volume varies between datasets. For STACOM 2018 dataset, volumes predicted by LA-Net closely

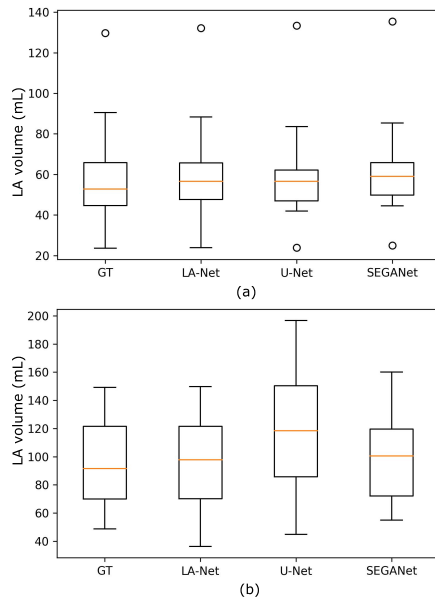


Fig. 14. Volume prediction by LA-Net, U-Net and SEGANet on LGE in (a) and bSSFP datasets in (b).

match those obtained from manual segmentation masks, while U-Net and SEGANet overestimate small volumes, which may be explained with falsely labelled structures by these networks. On the other hand, for bSSFP dataset, LA-Net underestimates very small volumes, which may be indication of missed slices for these volumes whereas U-Net and SEGANet overestimates almost all volumes, which may relate to extra labelled structures, as shown in Figure 10.

REFERENCES

- [1] P. Kirchhof *et al.*, “2016 ESC guidelines for the management of atrial fibrillation developed in collaboration with EACTS,” *Kardiologia Polska*, vol. 74, no. 12, pp. 1359–1469, Dec. 2016.
- [2] H. Calkins *et al.*, “2012 HRS/EHRA/ECAS expert consensus statement on catheter and surgical ablation of atrial fibrillation,” *Europace*, vol. 14, no. 4, pp. 528–606, 2012.
- [3] A. Berrueto *et al.*, “Pre-procedural predictors of atrial fibrillation recurrence after circumferential pulmonary vein ablation,” *Eur. Heart J.*, vol. 28, no. 7, pp. 836–841, Mar. 2007.
- [4] F. Bisbal *et al.*, “Left atrial sphericity: A new method to assess atrial remodeling. Impact on the outcome of atrial fibrillation ablation,” *J. Cardiovascular Electrophysiol.*, vol. 24, no. 7, pp. 752–759, Jul. 2013.
- [5] M. Varela *et al.*, “Novel computational analysis of left atrial anatomy improves prediction of atrial fibrillation recurrence after ablation,” *Front. Physiol.*, vol. 8, p. 68, Feb. 2017.
- [6] N. F. Marrouche *et al.*, “Association of atrial tissue fibrosis identified by delayed enhancement MRI and atrial fibrillation catheter ablation: The DECAAF study,” *JAMA*, vol. 311, no. 5, pp. 498–506, 2014.
- [7] R. Karim, H. Wang, S. Conjeti, H. Tang, E. Dehghan, and A. Karargyris, “Algorithms for left atrial wall segmentation and thickness—evaluation on an open-source CT and MRI image database,” *Med. Image Anal.*, vol. 50, pp. 36–53, Dec. 2018.
- [8] C. Tobon-Gomez *et al.*, “Benchmark for algorithms segmenting the left atrium from 3D CT and MRI datasets,” *IEEE Trans. Med. Imag.*, vol. 34, no. 7, pp. 1460–1473, Jul. 2015.
- [9] C. Chen *et al.*, “Deep learning for cardiac image segmentation: A review,” *Frontiers Cardiovascular Med.*, vol. 7, p. 25, Mar. 2020.
- [10] O. Bernard *et al.*, “Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?” *IEEE Trans. Med. Imag.*, vol. 37, no. 11, pp. 2514–2525, Nov. 2018.
- [11] K. Jamart, Z. Xiong, G. D. Maso Talou, M. K. Stiles, and J. Zhao, “Mini review: Deep learning for atrial segmentation from late gadolinium-enhanced MRIs,” *Frontiers Cardiovascular Med.*, vol. 7, p. 86, May 2020.
- [12] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in *Proc. MICCAI*. Munich, Germany: Springer, 2015, pp. 234–241.
- [13] J. Chen *et al.*, “Multiview two-task recursive attention model for left atrium and atrial scars segmentation,” in *Proc. MICCAI*. Granada, Spain: Springer, 2018, pp. 455–463.
- [14] X. Yang *et al.*, “Combating uncertainty with novel losses for automatic left atrium segmentation,” in *Proc. STACOM*. Granada, Spain: Springer, 2018, pp. 246–254.
- [15] A. Mortazi, R. Karim, K. Rhode, J. Burt, and U. Bagci, “CardiacNET: Segmentation of left atrium and proximal pulmonary veins from MRI using multi-view CNN,” in *Proc. MICCAI*. QC, Canada: Springer, 2017, pp. 377–385.
- [16] A. Lourenço *et al.*, “Left atrial ejection fraction estimation using SEGANet for fully automated segmentation of CINE MRI,” 2020, *arXiv:2008.13718*. [Online]. Available: <http://arxiv.org/abs/2008.13718>
- [17] Y. Jo *et al.*, “Guideline for cardiovascular magnetic resonance imaging from the Korean society of cardiovascular imaging—Part 1: Standardized protocol,” *Korean J. Radiol.*, vol. 20, no. 9, pp. 1313–1333, 2019.
- [18] 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 Trans. Med. Imag.*, vol. 38, no. 2, pp. 515–524, Feb. 2018.
- [19] C. Chen, W. Bai, and D. Rueckert, “Multi-task learning for left atrial segmentation on GE-MRI,” in *Proc. STACOM*. Granada, Spain: Springer, 2018, pp. 292–301.
- [20] C. Ma, G. Luo, and K. Wang, “A combined random forests and active contour model approach for fully automatic segmentation of the left atrium in volumetric MRI,” *BioMed Res. Int.*, vol. 2017, pp. 1–14, Feb. 2017.
- [21] W. Wang, S. Zhao, J. Shen, S. C. H. Hoi, and A. Borji, “Salient object detection with pyramid attention and salient edges,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1448–1457.
- [22] M. Zhang, B. Dong, and Q. Li, “Deep active contour network for medical image segmentation,” in *Proc. MICCAI*. Lima, Peru: Springer, 2020, pp. 321–331.
- [23] X. Chen, B. M. Williams, S. R. Vallabhaneni, G. Czanner, R. Williams, and Y. Zheng, “Learning active contour models for medical image segmentation,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11632–11640.
- [24] Y. Wu and K. He, “Group normalization,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 3–19.
- [25] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.
- [26] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [27] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
- [28] L. Yang, Y. Wang, X. Xiong, J. Yang, and A. K. Katsaggelos, “Efficient video object segmentation via network modulation,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6499–6507.
- [29] R. E. Woods and R. C. Gonzalez, *Digital Image Processing*, 3rd ed. New York, NY, USA: Pearson, 2021.
- [30] H. Chubb *et al.*, “The reproducibility of late gadolinium enhancement cardiovascular magnetic resonance imaging of post-ablation atrial scar: A cross-over study,” *J. Cardiovascular Magn. Reson.*, vol. 20, no. 1, p. 21, Dec. 2018.
- [31] C. Bian *et al.*, “Pyramid network with online hard example mining for accurate left atrium segmentation,” in *Proc. STACOM*. Granada, Spain: Springer, 2018, pp. 237–245.
- [32] H. M. Ali, “MRI medical image denoising by fundamental filters,” in *High-Resolution Neuroimaging-Basic Physical Principles and Clinical Applications*. Rijeka, Croatia: InTech, 2018, pp. 111–124.
- [33] S. Jia *et al.*, “Automatically segmenting the left atrium from cardiac images using successive 3D U-Nets and a contour loss,” in *Proc. STACOM*. Granada, Spain: Springer, 2018, pp. 221–229.
- [34] C. Li *et al.*, “Attention based hierarchical aggregation network for 3D left atrial segmentation,” in *Proc. STACOM*. Granada, Spain: Springer, 2018, pp. 255–264.
- [35] F. Uslu and M. Varela, “SA-Net: A sequence aware network for the segmentation of the left atrium in cine MRI datasets,” in *Proc. IEEE 18th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2021, pp. 766–769.
- [36] H. Schütze, C. D. Manning, and P. Raghavan, *Introduction to Information Retrieval*, vol. 39. Cambridge, U.K.: Cambridge Univ. Press, 2008.