



Edge-Enhanced Feature Guided Joint Segmentation of Left Atrial and Scars in LGE MRI Images

Siping Zhou[✉], Kai-Ni Wang[✉], and Guang-Quan Zhou[✉]

School of Biological Science and Medical Engineering, Southeast University,
Nanjing, China

guangquan.zhou@seu.edu.cn

Abstract. Automatic segmentation of the left atrial (LA) cavity and atrial scars in late gadolinium enhancement magnetic resonance imaging has significant clinical relevance to diagnosing atrial fibrillation (AF). Nevertheless, automatic segmentation remains challenging because of the poor image quality, the shape variability of LA, and the small size of scars. Therefore, this study proposes a multi-task learning model in a coarse-to-fine framework, among which the fine model simultaneously segmenting the LA cavity and scars. Specially, we develop an edge-enhanced feature-guided module (EFGM) to exploit the spatial relationship between LA and scars using a 3D central difference convolution, exploring the feature dependence from multi-task learning. Also, a dilated inception module (DIM) is plugged in to learn multi-scale representation, further improving the joint segmentation considering the shape difference between the LA cavity and scar. We evaluate our model on the LAScarQS 2022 validation set. The average Dice scores of the LA cavity and scar are 0.875 and 0.631. Also, the Average Surface Distance (ASD) and Hausdorff Distance (HD) of the LA cavity are 2.233 mm and 24.731 mm, respectively. The accuracy, specificity, sensitivity, and generalized Dice score of LA scar are 0.999, 0.999, 0.603, and 0.916, respectively.

Keywords: Deep learning · Cardiac Segmentation · Joint optimization · Difference convolution

1 Introduction

Atrial fibrillation (AF) is the most common cardiac arrhythmia observed in clinical practice, occurring in up to 2% of the population and rising fast with advancing age [10]. Recently, late gadolinium enhancement magnetic resonance imaging (LGE MRI) has been considered as a promising and reliable technique to visualize and quantify left atrial scars [11]. The segmentation or quantification of LA and scars provides important information for the clinical diagnosis and the treatment of AF patients. Since manual delineations of LA and scars are time-consuming and subjective, it is crucial to develop techniques for automatic segmentation of the LA cavity and scar for LGE MRI.

However, the poor image quality in LGE MRI, various shapes of LA, the surrounding enhanced noise, and the complex patterns of scars make it challenging to automatically and accurately segment LA and scars. Li et al. reviewed algorithms proposed to perform the LA cavity and scar segmentation or quantification from medical images in [2]. Among them, deep learning-based methods are dominant in these two tasks and they achieved promising results [1, 15, 16, 19–22]. Nevertheless, most of the methods mentioned in [2] normally solved the two tasks independently and ignored the intrinsic spatial relationship between LA and scars which are located on the LA wall, as Fig. 1 shows. The performance of segmenting the LA cavity and scar may be bottlenecked by the failure in exploiting the correlation between these two tasks. Multi-task learning has been shown to outperform methods considering related tasks separately by leveraging the relationship between different tasks. Recently, Li et al. [1] developed a novel framework where LA segmentation, scar projection onto the LA surface, and scar quantification are performed simultaneously in an end-to-end style. The relationship between LA segmentation and scar quantification was explicitly explored and has shown significant performance improvements for both tasks in their work.

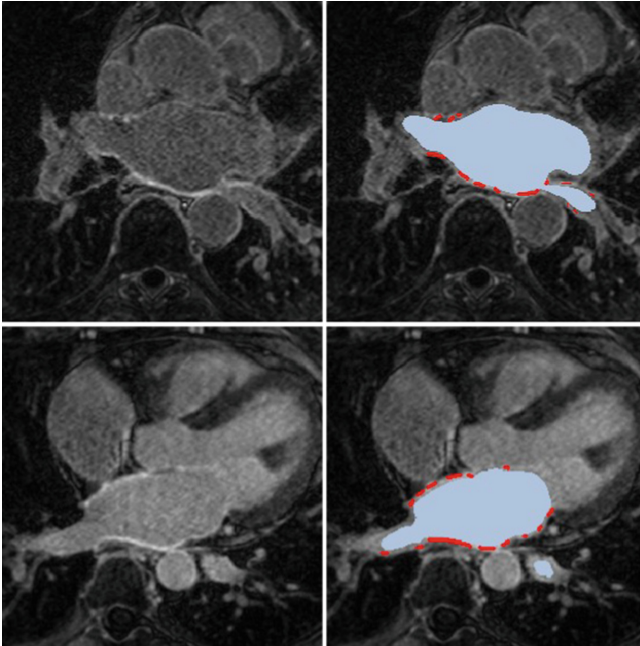


Fig. 1. Examples of axial views from two cases in the LAScarQS2022 dataset. The LA cavity and scar are highlighted in blue and red, respectively. One can see that scars are located on the LA wall. (Color figure online)

This paper, inspired by [1], proposes a coarse-to-fine framework to achieve joint segmentation of LA and scars. In the coarse stage, a vanilla 3D U-Net [14] is trained to coarsely segment LA and crop a region of interest (ROI) that contains the whole LA. In the fine stage, a modified dual-task learning 3D U-Net consisting of two decoders for LA and scars segmentation respectively, is proposed to segment LA and scars simultaneously. We also introduce an edge-enhanced feature-guided module (EFGM) at the skip connection between the shared encoder and the decoder layers for scar segmentation. It includes a difference convolution submodule based on 3D central difference convolution (CDC) [7], followed by a spatial attention submodule. We argue that it can help pass the edge-enhanced features to guide the localization and segmentation of scars as they are located at the LA wall while utilizing the spatial relationship between LA and scars. In addition, a dilated inception module (DIM) to extract multi-scale features is plugged in at the bottleneck of the modified 3D U-Net.

2 Methods

Figure 2 shows the pipeline of our coarse-to-fine joint segmentation framework. In our work, we develop a two-stage strategy to perform coarse-to-fine joint segmentation of the LA cavity and scar. In the coarse stage, a vanilla 3D U-Net is first trained to segment the ROI which contains the whole LA from the entire 3D volume of each MRI. After the ROIs are detected, they are all cropped out with a fixed size from the processed MRIs and then fed into the proposed modified multi-task learning 3D U-Net to obtain segmentation results of LA cavity and scars simultaneously in the fine stage.

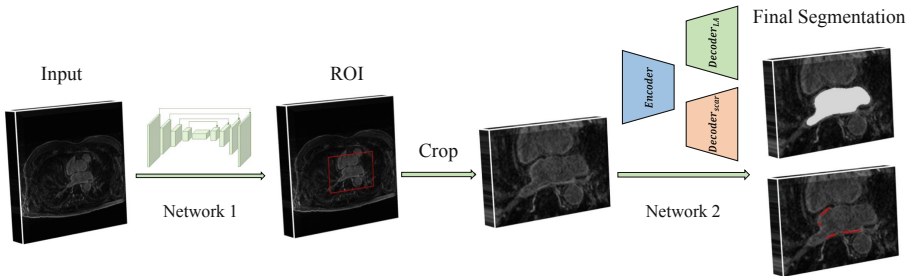


Fig. 2. The overall pipeline of our coarse-to-fine joint segmentation of left atrial and scars framework. The network 1 is a vanilla 3D U-Net to segment the ROI coarsely. The network 2 is a modified 3D U-Net consisting of two decoders for LA and scar segmentation respectively to get more accurate segmentation results.

2.1 Coarse Segmentation of ROIs

As shown in Fig. 1, the regions of the LA cavity and scar are only part of the whole volume, especially for a scar of such a small size. Therefore, we first employ a coarse segmentation stage to segment the ROI containing the LA cavity and scar, aiming at alleviating the class imbalance problem and discarding redundant or irrelevant surrounding voxels. We choose the vanilla 3D U-Net as our coarse segmentation network for its effectiveness in various medical image segmentation tasks without any complex design.

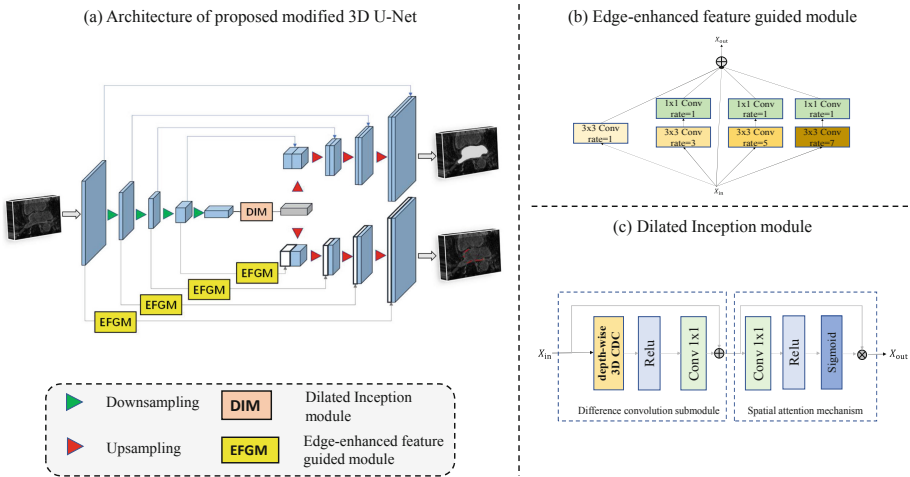


Fig. 3. (a)An overview of our proposed modified 3D U-Net with two decoders for LA and scar segmentation, respectively. (b)Edge-enhanced feature-guided module using 3D central difference convolution. (c)Dilated Inception module using dilated convolutions with different rates and shortcut connections.

2.2 Fine and Joint Segmentation of LA and Scars

Most of the automatic scar segmentation or quantification methods require an accurate initial LA segmentation considering the prior knowledge that atrial scars are located on the LA wall. Additionally, previous methods usually solved the two tasks independently and ignored the intrinsic spatial relationship between LA and scars [2]. Therefore, we propose a modified 3D U-Net consisting of two decoders for LA and scar segmentation and train it in a multi-task learning manner in the fine stage. Figure 3 (a) provides an overview of the proposed dual-task learning network architecture. First, an edge-enhanced feature guided module (EFGM) is introduced at the skip connection between the shared encoder and decoder layers for scar segmentation. Different from the original skip connection, the EFGM, which can serve as an edge detector, helps

preserve differential or edge-related information via extracting edge-enhanced features and passing them to the corresponding layers at the scar segmentation decoder. In addition, a dilated inception module (DIM) is introduced at the end of the original encoder. With the equipment of DIM, the modified 3D U-Net can capture deep multi-scale semantic features, which is beneficial to the joint segmentation of LA cavities and scars as they are totally different in size. The details of the EFGM and the DIM are described below.

Edge-enhanced Feature Guided Module. Difference Convolution, which explicitly calculates pixel differences during convolution to aggregate local gradient information, has been gradually used in computer vision tasks such as edge detection [6], face recognition [5], gesture recognition [7], and so on in recent years. By contrast, vanilla convolution aggregates intensity-level information [6]. As a result, although modern CNNs based on vanilla convolution are powerful enough to learn rich and hierarchical image representations, it is still hard for them to focus on edge-related features due to the lack of explicit encoding for gradient information [5]. The formulations of vanilla convolution and difference convolution can be written as (take 2D convolution as an example):

$$y = \sum_{i=1}^{k \times k} w_i \cdot x_i \quad (\text{vanilla convolution}) \quad (1)$$

$$y = \sum_{x_i, x_j \in S} w_i \cdot (x_i - x_j) \quad (\text{difference convolution}) \quad (2)$$

where, x_i and x_j are the input pixels, w_i is the weight in the $k \times k$ convolution kernel. S is the local receptive field over the feature map.

As mentioned before, scars are located on the LA wall, so we intuitively argue that the edge information of the LA cavity is important to the localization and further segmentation of scars. In the vanilla U-Net [13], long skip connections were introduced to pass features from the encoder path to the decoder path to recover spatial information lost during downsampling. However, original low-level features which are simply passed through the skip connections to fuse with high-level features may contain substantial redundant location or spatial information. Motivated by these assumptions, we propose the EFGM equipping difference convolution at the skip connection between the encoder and the decoder for scar segmentation in the modified 3D U-Net only to pass the edge-enhanced features containing rich edge-related information. With the implementation of the EFGM, our model can learn to suppress irrelevant regions and highlight salient regions (edge of LA cavity) useful for more precise localization and segmentation of scars due to the ability of difference convolution to extract local differential information from feature maps. Moreover, the edge-enhanced features can also be regarded as localization guidance for decoding high-level semantic features in the decoder path in scar segmentation, which benefits the segmentation of scars located on the LA wall.

Figure 3 (b) illustrates an edge-enhanced feature-guided module. Each module mainly includes a difference convolution submodule where we utilize a 3D central difference convolution (CDC) [7], which is formulated as follows:

$$y = \sum_{i \in C} w_i \cdot (x_i - x_0) \quad (3)$$

3D convolution with kernel size $3 \times 3 \times 3$ and dilation 1 is used for demonstration. The local receptive field cube for the 3D convolution is $C = (-1, -1, -1), (-1, -1, 0), \dots, (0, 1, 1), (1, 1, 1)$.

The same as [6], we use the separable depth-wise convolutional structure with a shortcut for fast inference and easy training. The residual path in this module includes a depth-wise convolutional layer, a ReLU layer, and a point-wise convolutional layer sequentially. To further highlight the edge-related features and filter background noise, we apply the spatial attention mechanism at the end of the difference convolution submodule.

Dilated Inception Module. Motivated by the Inception-ResNet-V2 module [8] and Atrous Spatial Pyramid Pooling (ASPP) [9], we propose the DIM to encode deep multi-scale features for both LA and scar segmentation. As shown in Fig. 3 (c), the DIM has four parallel paths with dilated convolutions with different dilation rates followed by one 1×1 convolution. At last, we directly add the original features with the other four multi-scale features to make a shortcut mechanism. Different dilation rates can increase the receptive field sizes of parallel convolution paths by adding zeros between kernel elements without incrementing parameters. As a result, the proposed DIM can capture features of objects of various sizes, such as LA cavities and scars, due to the combination of the dilation convolutions with different dilation rates.

2.3 Loss Function

For the coarse stage only regarding the segmentation of the LA cavity, our loss function is Dice Loss.

For the fine and joint segmentation stage, our loss function is the sum of the loss function of segmenting the LA cavity and the loss function of segmenting scar, as shown in Eq. 4:

$$Loss_{total} = Loss_{LA} + Loss_{scar} \quad (4)$$

The loss function used in LA segmentation is the sum of the Dice Loss and the Cross-Entropy Loss, as shown in Eq. 5:

$$Loss_{LA} = Loss_{ce} + Loss_{dice} \quad (5)$$

For scar segmentation, as the scar only takes up a small fraction of the whole volume, which can cause a severe class-imbalance problem, the loss function is the sum of the Dice Loss and the Weighted Cross Entropy Loss, as shown in Eq. 6:

$$Loss_{scar} = Loss_{wce} + Loss_{dice} \quad (6)$$

3 Experiments

3.1 Dataset and Data Preprocessing

MICCAI 2022-LAScarQS2022 (Left Atrial and Scar Quantification & Segmentation Challenge) [1–3] provides 194 LGE MRIs acquired in real clinical environment from patients suffering atrial fibrillation (AF) and is composed of two tasks: 1. LA Scar Quantification 2. Left Atrial Segmentation from Multi-Center LGE MRIs. In this study, we focus on task 1.

The training dataset provided for task 1 of the LAScarQS 2022 challenge [1–3] consists of 60 LGE-MRIs with segmentation annotations of LA cavities and scars. In our experiments, the images and masks were first resampled to the isotropic resolution of $1 \times 1 \times 1 \text{mm}^3$. And then, all the volumes were cropped and zero-pad to the uniform size of $576 \times 576 \times 96$. Then we used a 3D version of contrast limited adaptive histogram localization (CLAHE) [4] to enhance the contrast of LGE-MRIs, and finally applied sample-wise normalization.

3.2 Implementation Details

Our experiments were run on NVIDIA GeForce RTX 3090 GPU with 24 GB RAM. We firstly down-sampled the input for the coarse segmentation from $576 \times 576 \times 96$ to $144 \times 144 \times 48$ due to memory restriction. The first network was trained for 100 epochs using the Adam optimizer with a fixed learning rate of 0.001. The batch size is 4. We randomly chose 48 out of the 60 MRIs as training data; the rest 12 are validation data. After the training procedure was completed, the model with the best dice scores on validation data was saved for ROI detection. For the fine and joint segmentation, we first computed the barycenter of the ground truth and cropped a region of size $288 \times 192 \times 96$ centered with the barycenter from the original data. Then the cropped ROIs were fed into the second network. The second network was trained for 100 epochs using the Adam optimizer with an initial learning rate of 0.001. The learning rate was reduced by 0.1 every 1000 iterations and the batch size is 2. We randomly split the data into training (48 subjects) and testing (12 subjects) subsets for the fine stage.

To reduce the risk of over-fitting and further improve the generalization ability of our framework, we also apply data augmentation including random flipping and rotation in both networks training.

At the inference stage, each MRI volume from the testing subset was firstly down-sampled to $144 \times 144 \times 48$ and fed into the first network. The network would output the predicted binary mask used to locate the ROI. We computed the barycenter of the predicted mask, cropped a region of size $288 \times 192 \times 96$ centered with this barycenter, and then fed it into the second network. The second network output the predicted masks of the LA cavity and scar simultaneously inside the target region and mapped them back to the original size volume, which finished the inference. The end-to-end segmentation process takes approximately 9s for each case.

4 Results and Discussions

4.1 Ablation Experiments

We run a number of ablation experiments to evaluate the effectiveness of multi-task learning and the two proposed modules in our modified 3D U-Net. All the experiments were run in the coarse-to-fine framework mentioned above, sharing the same coarse stage and we only performed different models in the fine stage to conduct ablation experiments. Here, U-Net_{LA} denotes the vanilla 3D U-Net architecture for LA segmentation individually. U-Net_{scar} denotes the vanilla 3D U-Net architecture for scar segmentation individually. U-Net_{LA and scar} denotes the multi-task learning 3D U-Net consists of a shared encoder and two decoders for joint segmentation of LA and scars, which is also our baseline model. Besides, we successively tested the performance of the baseline model incorporating the DIM, the baseline model incorporating the EFGM, and the baseline model incorporating both the EFGM and the DIM. All these experiments were conducted using the same aforementioned training configurations and loss functions.

All the models were evaluated through the validation platform provided by the LAScarQS2022 organizer. As shown in Table 1, the segmentation performance of LA was evaluated by the Dice score, average surface distance (ASD) and Hausdorff distance (HD). The scar’s quantification performance was evaluated via first projecting the segmentation result onto the manually segmented LA surface. Then, the Accuracy, Specificity and Sensitivity measurement of the two areas in the projected surface, Dice score (*Dice*) and generalized Dice score (*Dice_g*) were used as indicators of the accuracy of scar quantification [1]. *Dice_g* is a weighted Dice score by evaluating the segmentation of all labels [17, 18], which is formulated as follow [1]:

$$Dice_g = \frac{2 \sum_{k=0}^{N_k-1} |S_k^{auto} \cap S_k^{manual}|}{\sum_{k=0}^{N_k-1} (|S_k^{auto} + S_k^{manual}|)} \quad (7)$$

where S_k^{auto} auto and S_k^{manual} indicate the segmentation results of label k from the automatic method and manual delineation, respectively, and N_k is the number of labels.

Table 1. Summary of the quantitative evaluation results of LA segmentation and scar quantification on the LAScarQS 2022 validation set in ablation experiments. EFGM denotes the proposed edge-enhanced feature-guided module discussed in Sect. 2.2, and DIM denotes the proposed dilation inception module discussed in Sect. 2.2.

Method	LA			Scar				
	Dice	ASD(mm)	HD (mm)	Accuracy	Specificity	Sensitivity	Dice	<i>Dice_g</i>
U-Net _{LA}	0.860	2.625	30.028	N\ A	N\ A	N\ A	N\ A	N\ A
U-Net _{scar}	N\ A	N\ A	N\ A	0.999	0.999	0.534	0.580	0.909
Base(U-Net _{LA and scar})	0.869	2.318	27.050	0.999	0.999	0.549	0.596	0.913
Base+EFGM	0.867	2.489	26.170	0.999	0.999	0.588	0.621	0.914
Base+DIM	0.871	2.355	25.636	0.999	0.999	0.605	0.617	0.914
Base+EFGM+DIM	0.875	2.233	24.731	0.999	0.999	0.603	0.631	0.916

Table 1 presents the quantitative results for LA segmentation and scar quantification. It demonstrates that our baseline model outperforms U-Net_{LA} and U-Net_{scar} which consider these two related tasks separately, verifying the superiority of multi-task learning. The relationship between LA segmentation and scar segmentation is exploited implicitly through multi-task learning. Figure 4 and Fig. 5 illustrate the segmentation results of the LA cavity and scar, respectively, from the mentioned ablation experiments. One can see that the boundary of segmentation results of U-Net_{LA} is far from the boundary of the ground-truth and U-Net_{scar} tends to make mistakes on non-LA wall regions and under-segment scars, while the baseline model results are closer to the ground truth.

Meanwhile, Table 1 illustrates the effectiveness of each proposed module, suggesting the advantage of the EFGM and the DIM. Compared to the baseline model, incorporating the DIM reduces about 2 mm in HD in segmenting LA and improves the Dice in segmenting scar by around 2%. This observation implies the need for learning deep multi-scale features when coping with segmenting targets of different sizes since scars are quite small compared with the LA cavity. Note that incorporating the EFGM into the baseline model improves the Dice in segmenting scar by around 2.5% compared to the baseline model and outperforms the baseline model only equipped with the DIM. As shown in Fig. 5, introducing the EFGM can alleviate the problem of under-segmenting scars observed in other models. Furthermore, it indicates that the edge-related information can effectively guide the segmentation of scars while encoding the prior spatial knowledge that scars are located at the LA wall into the framework, thus utilizing the spatial relationship between LA and scars more explicitly. However, the performance of the model which only incorporates the EFGM even degrades a little in Dice and ASD of segmenting LA compared to the baseline model. We argue that this is because the EFGM is mainly designed for the scar segmentation task, which is much more challenging than LA segmentation, so it may not improve the segmentation performance of LA.

The highest performance gain (about 1.5% in Dice of LA segmentation and 5% in Dice of scar segmentation compared to U-Net_{LA} and U-Net_{scar}) is observed when incorporating both the DIM and the EFGM. Moreover, the model equipped with both two modules achieves the best segmentation performance in almost all metrics in both tasks. Figure 4 also demonstrates that the boundary of LA segmentation results is the most consistent with the ground truth among all the experiments, while Fig. 5 illustrates that our final model can detect and segment scars more precisely than any other model in our ablation experiments. It shows that the combination of two modules can further improve the performance of the framework. Note that the model incorporating both the DIM and the EFGM outperforms the model only incorporating the DIM in LA cavity segmentation, but the introduction of the EFGM cannot improve the segmentation of the LA cavity as mentioned above. This finding is probably attributed to the explanation that relatively good performance in the scar segmentation task can boost the LA segmentation task during the simultaneous optimization process in multi-task learning.

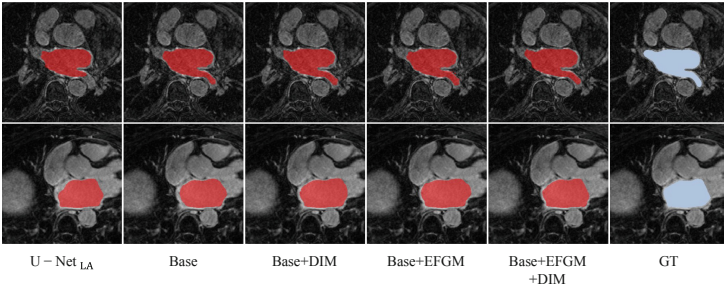


Fig. 4. Visualization of the LA cavity segmentation results on the LAScarQS 2022 validation set by using different training combinations.

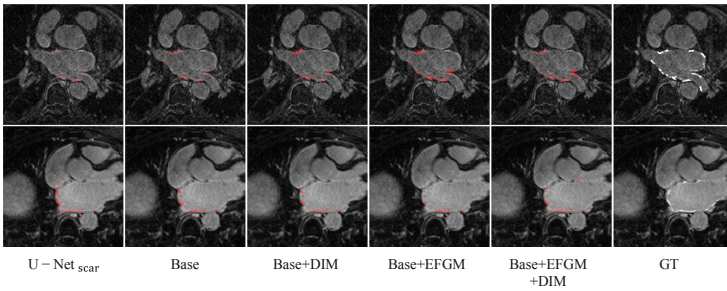


Fig. 5. Visualization of the scar segmentation results on the LAScarQS 2022 validation set by using different training combinations.

4.2 Comparison Experiments

We implemented U-Net with different loss functions to conduct comparison experiments for both LA segmentation and scar segmentation. We used the same hyper-parameters in these experiments for consistency.

Table 2 tabulates the quantitative comparison results for LA segmentation and scar quantification. For LA segmentation, our method achieves 0.875 in Dice, demonstrating its advantage in segmenting the LA cavity more accurately. Meanwhile, the proposed coarse-to-fine joint segmentation framework obtains the smallest HD and ASD, which means it can identify the correct boundaries of LA cavities despite their various shape. Figure 6 also proves that our proposed model can achieve better segmentation compared to other methods.

Note that our method shows significant improvement in scar quantification results. As demonstrated in Fig. 7, the vanilla U-Net models tend to under-segment scars while our method alleviates this problem. With the help of the DIM and the EFGM, edge-enhanced low-level and multi-scale features are fused while more contextual semantic information and more precise spatial information are integrated, facilitating the segmentation of scars which are hard to recognize and locate due to their small size, complex patterns, and surrounding noise.

Overall, our method outperformed superiorly to other methods, implying its effectiveness. This could result from the two major contributions in our framework. First, the multi-task learning model can effectively exploit the relationship between LA and scars. Moreover, the EFGM and the DIM are introduced to further boost the multi-task learning process through providing spatial guidance for segmenting scars and learning multi-scale representation. Second, the two-stage coarse-to-fine framework can suppress the background pixels that dominate foreground pixels in the scar segmentation, thus significantly mitigate the class imbalance problem.

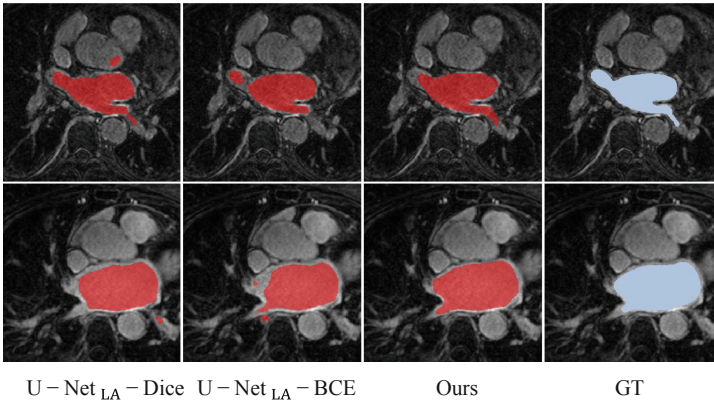


Fig. 6. Visualization of the LA cavity segmentation results on the LAScarQS 2022 validation set compared with other classic methods.

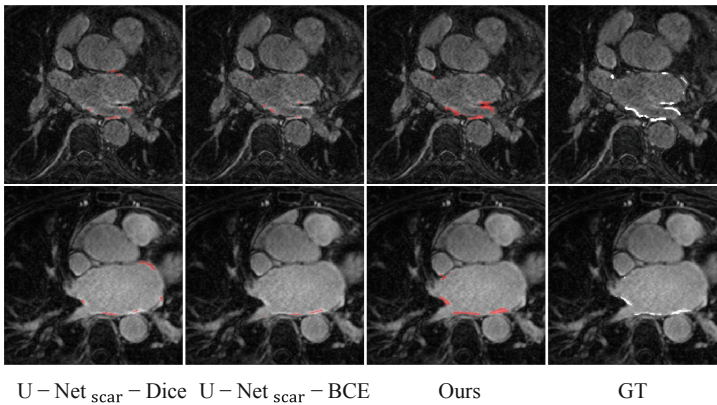


Fig. 7. Visualization of the scar segmentation results on the LAScarQS 2022 validation set compared with other classic methods.

Table 2. Summary of the quantitative evaluation results of LA segmentation and scar quantification on the LAScarQS 2022 validation set in comparison experiments.

Method	LA			Scar				
	Dice	ASD(mm)	HD (mm)	Accuracy	Specificity	Sensitivity	Dice	$Dice_g$
U-Net _{LA/scar} -BCE	0.849	2.980	41.934	0.681	0.999	0.362	0.466	0.890
U-Net _{LA/scar} -Dice	0.845	3.227	43.622	0.670	0.999	0.340	0.427	0.883
Ours	0.875	2.233	24.731	0.999	0.999	0.603	0.631	0.916

5 Conclusion

This paper proposes a coarse-to-fine framework for joint segmentation of LA and scars from LGE MRI. The coarse segment network is a vanilla 3D U-Net to extract ROI of the volume, and the fine segment network is a modified 3D U-Net consisting of two decoders for LA and scar segmentation, respectively, aiming at segmenting the LA cavity and scar simultaneously in a multi-task learning manner. In addition, we introduce an edge-enhanced feature-guided module using 3D central difference convolution to exploit the spatial relationship between LA and scars and a dilated inception module to learn multi-scale semantic features in our modified 3D U-Net. We evaluated our method on the LAScarQS 2022 validation dataset, and the convincing results suggest the effectiveness of the newly proposed coarse-to-fine framework, especially for scar segmentation or quantification.

References

1. Li, L., Zimmer, V.A., Schnabel, J.A., et al.: AtrialJSQnet: a new framework for joint segmentation and quantification of left atrium and scars incorporating spatial and shape information. *Med. Image Anal.* **76**, 102303 (2022)
2. Li, L., Zimmer, V.A., Schnabel, J.A., et al.: Medical image analysis on left atrial LGE MRI for atrial fibrillation studies: a review. *Medical Image Analysis*, 102360 (2022)
3. Li, L., Zimmer, V.A., Schnabel, J.A., Zhuang, X.: AtrialGeneral: domain generalization for left atrial segmentation of multi-center LGE MRIs. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12906, pp. 557–566. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87231-1_54
4. Pizer, S.M., Johnston, R.E., Ericksen, J.P., Yankaskas, B.C., Muller, K.E.: Contrast-limited adaptive histogram equalization: speed and effectiveness. In: *Proceedings of the First Conference on Visualization in Biomedical Computing*, Atlanta, Georgia, vol. 1, pp. 337–345 (1990)
5. Yu, Z., Zhao, C., Wang, Z., et al.: Searching central difference convolutional networks for face anti-spoofing. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5295–5305 (2020)
6. Su, Z., Liu, W., Yu, Z., et al.: Pixel difference networks for efficient edge detection. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 5117–5127 (2021)

7. Yu, Z., Zhou, B., Wan, J., et al.: Searching multi-rate and multi-modal temporal enhanced networks for gesture recognition. *IEEE Trans. Image Process.* **30**, 5626–5640 (2021)
8. Szegedy, C., Ioffe, S., Vanhoucke, V., et al.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Thirty-first AAAI Conference on Artificial Intelligence* (2017)
9. Chen, L.C., Papandreou, G., Kokkinos, I., et al.: DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(4), 834–848 (2017)
10. Chugh, S.S., Havmoeller, R., Narayanan, K., et al.: Worldwide epidemiology of atrial fibrillation: a global burden of disease 2010 study. *Circulation* **129**(8), 837–847 (2014)
11. Vergara, G.R., Marrouche, N.F.: Tailored management of atrial fibrillation using a LGE-MRI based model: from the clinic to the electrophysiology laboratory. *J. Cardiovasc. Electrophysiol.* **22**(4), 481–487 (2011)
12. Xiong, Z., Xia, Q., Hu, Z., et al.: A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging[J]. *Med. Image Anal.* **67**, 101832 (2021)
13. 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
14. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) *MICCAI 2016*. LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_49
15. Zhao, Z., Puybareau, E., Boutry, N., et al.: Do not treat boundaries and regions differently: an example on heart left atrial segmentation. In: *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 7447–7453. IEEE (2021)
16. Li, L., Wu, F., Yang, G., et al.: Atrial scar quantification via multi-scale CNN in the graph-cuts framework. *Med. Image Anal.* **60**, 101595 (2020)
17. Crum, W.R., Camara, O., Hill, D.L.G.: Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Trans. Med. Imaging* **25**(11), 1451–1461 (2006)
18. Zhuang, X.: Challenges and methodologies of fully automatic whole heart segmentation: a review. *J. Healthc. Eng.* **4**(3), 371–407 (2013)
19. Borra, D., Andalò, A., Paci, M., et al.: A fully automated left atrium segmentation approach from late gadolinium enhanced magnetic resonance imaging based on a convolutional neural network. *Quant. Imaging Med. Surgery* **10**(10), 1894 (2020)
20. Du, X., Yin, S., Tang, R., et al.: Segmentation and visualization of left atrium through a unified deep learning framework. *Int. J. Comput. Assist. Radiol. Surg.* **15**(4), 589–600 (2020)
21. Chen, J., et al.: Multiview two-task recursive attention model for left atrium and atrial scars segmentation. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) *MICCAI 2018*. LNCS, vol. 11071, pp. 455–463. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_51
22. Yang, G., Chen, J., Gao, Z., et al.: Simultaneous left atrium anatomy and scar segmentations via deep learning in multiview information with attention. *Future Gener. Comput. Syst.* **107**, 215–228 (2020)