

Segmentation and visualization of left atrium through a unified deep learning framework

**Xiuquan Du, Susu Yin, Renjun Tang,
Yueguo Liu, Yuhui Song, Yanping
Zhang, Heng Liu & Shuo Li**

**International Journal of Computer
Assisted Radiology and Surgery**

A journal for interdisciplinary research,
development and applications of image
guided diagnosis and therapy

ISSN 1861-6410

Int J CARS

DOI 10.1007/s11548-020-02128-9



Your article is protected by copyright and all rights are held exclusively by CARS. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".



Segmentation and visualization of left atrium through a unified deep learning framework

Xiuquan Du^{1,2} · Susu Yin² · Renjun Tang² · Yueguo Liu² · Yuhui Song² · Yanping Zhang^{1,2} · Heng Liu³ · Shuo Li⁴

Received: 27 September 2019 / Accepted: 17 February 2020
© CARS 2020

Abstract

Purpose Left atrium segmentation and visualization serve as a fundamental and crucial role in clinical analysis and understanding of atrial fibrillation. However, most of the existing methods are directly transmitting information, which may cause redundant information to be passed to affect segmentation performance. Moreover, they did not further consider atrial visualization after segmentation, which leads to a lack of understanding of the essential atrial anatomy.

Methods We propose a novel unified deep learning framework for left atrium segmentation and visualization simultaneously. At first, a novel dual-path module is used to enhance the expressiveness of cardiac image representation. Then a multi-scale context-aware module is designed to effectively handle complex appearance and shape variations of the left atrium and associated pulmonary veins. The generated multi-scale features are feed to gated bidirectional message passing module to remove irrelevant information and extract discriminative features. Finally, the features after message passing are efficiently combined via a deep supervision mechanism to produce the final segmentation result and reconstruct 3D volumes.

Results Our approach primarily against the 2018 left atrium segmentation challenge dataset, which consists of 100 3D gadolinium-enhanced magnetic resonance images. Our method achieves an average dice of 0.936 in segmenting the left atrium via fivefold cross-validation, which outperforms state-of-the-art methods.

Conclusions The performance demonstrates the effectiveness and advantages of our network for the left atrium segmentation and visualization. Therefore, our proposed network could potentially improve the clinical diagnosis and treatment of atrial fibrillation.

Keywords Left atrium · Gadolinium-enhanced magnetic resonance image · Segmentation · Visualization

Introduction

Atrial fibrillation (AF) is the most common type of cardiac disease, which is caused by impaired electrical activity around the left atrium (LA) [1]. This can make blood more likely to form heart-threatening clots, which are capable of

limiting the blood supply to vital organs and further result in a stroke and heart failure. 3D gadolinium-enhanced magnetic resonance images (GE-MRIs) have been confirmed to improve the visibility of the patient's atrial structure. Therefore, GE-MRIs can provide guidance for ablation therapy of AF [2]. LA segmentation is an essential step during the diagnosis and treatment of patients with AF. Unfortunately, due to the lack of a basic understanding of atrial anatomy, the current clinical treatment effect is poor [3]. LA visualization is desired for analyzing atrial 3D geometry structures and providing reliable information for clinical treatment of AF after segmentation. However, most existing methods tend to focus only on the atrial segmentation and neglect the atrial visualization. Therefore, there is an urgent need for a unified framework to automatically segment LA from 3D GE-MRIs and accurately visualize atrial geometry for clinical analysis.

It is a very complicated task to assist the medical management of patients with AF based on LA geometric analysis.

✉ Xiuquan Du
dxqllp@163.com

¹ Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Anhui University, Hefei, China

² School of Computer Science and Technology, Anhui University, Hefei, China

³ Department of Gastroenterology, The First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China

⁴ Department of Medical Imaging, Western University, London, Canada

Many researchers have proposed a series of segmentation methods to solve it. For example, non-rigid registration [4] and deformable model [5] are typical methods for atrial segmentation. Tobongomez et al. [6] compared the performance of nine different approaches for LA segmentation from MRIs/computed tomography, and the results show that combining statistical models with regional growth methods is the most effective method. However, it is difficult to apply these methods directly to the GE-MRIs, because atrial structures are weaker under the effect of contrast agents. Therefore, the improved approaches [7, 8] were proposed and used to segment LA from GE-MRIs. Despite a lot of efforts, most LA structural analysis studies are still based on traditional methods and lack generalization capabilities for invisible cases, such as LA with the rare number of proximal pulmonary veins (PPVs). Moreover, they do not particularly emphasize the visualization work after the LA segmentation.

In recent years, deep learning is the most advanced method to process computer vision problems because of its efficiency and effectiveness. Moreover, the full convolutional neural (FCN) network architecture [9–11], such as U-Net, has been proven to handle effectively the medical images and achieve better performance. Therefore, many similar techniques have been presented for the LA segmentation; for instance, Yang et al. [12] closely combined 3D FCN, transfer learning, and deep supervision mechanisms to extract 3D context information to help segment the LA and other cardiac structures, and hybrid loss functions were designed to guide the training process by treating all classes equally. Subsequently, Yang et al. [13] further improved the method and applied to LA segmentation from GE-MRIs. Vesal et al. [14] presented a modified version of the 3D U-Net, using dilated convolutions in the lowest layer of the encoder branch, to segment the LA directly from GE-MRIs. Chen et al. [15] proposed a multi-task deep U-Net that could segment LA of the subject and detect whether the subject is pre- or post-ablation. Due to computer memory limitations and insufficient 3D data, multi-view FCN networks were proposed in Mortazi et al. [16] and Chen et al. [17].

Although these methods achieve competitive results, there are still some problems. (1) The FCN-based segmentation models sequentially stack single-scale convolutions and max-pooling layers to extract image features. This makes it impossible to accurately detect the LA and PPVs with various shapes and locations because of the limited receptive fields. (2) The afore-mentioned methods employ element-wise or concatenation operation to incorporate features from different levels. Since not all hierarchical features are positive for our goal, these incorporation fashions can cause information superfluous. (3) LA visualization as a critical step after segmentation can improve understanding of atrial anatomy and help develop AF treatment plan. However, most of the

above methods emphasize the segmentation of LA, and lack of analysis of atrial visualization.

In this paper, our solutions to the problems described above and contributions are as follows:

1. We propose a unified atrial segmentation and visualization framework for improving the diagnosis of AF and understanding of LA anatomy, which is a very crucial procedure toward subject-specific treatment.
2. We design a new dual-path structure for enhancing the expressiveness of GE-MRIs representation, and then a multi-scale context-aware feature extraction module is used to learn more contextual from hierarchical features to tackle the huge complexity of the LA with PPVs.
3. We present a gated bidirectional message passing module, which aims to adaptively filter redundant information and retain useful information. Integrated useful information is beneficial for LA segmentation and visualization.

Materials and methods

The overall workflow is shown in Fig. 1. Our framework consists of four main components: dual-path module (DPM), multi-scale context-aware module (MSCM), gated bidirectional message passing module (GBMPM), and deep supervision (DS) module. The details of each module are introduced in the following.

Dual-path module

In this paper, we design a new dual-path module (DPM) that aims to gradually enhance image representation. In fact, the proposed DPM provides the advantage that extracts semantic features and spatial details. Besides, the DPM capable of increasing the depth of traditional convolutional neural networks without distinctly increasing the number of parameters. In Fig. 2, we can find that each path has five stages/levels, each stage outputs h_{ij} , where i denotes the paths and j denotes stages, so

$$h_{2j} = \text{Conv}(h_{1j} + h_{2(j-1)}) \quad (1)$$

where $\text{Conv}(*; \theta)$ is a traditional convolution layer with parameter $\theta = \{W, b\}$. h_{2j} is also the j th block output. ‘+’ represents the element-wise addition which used to achieve the dual-path connection. h_{10} denotes the input and $h_{20} = 0$. The DPM can capture the context and structural information by adopting the connection of the block at different levels.

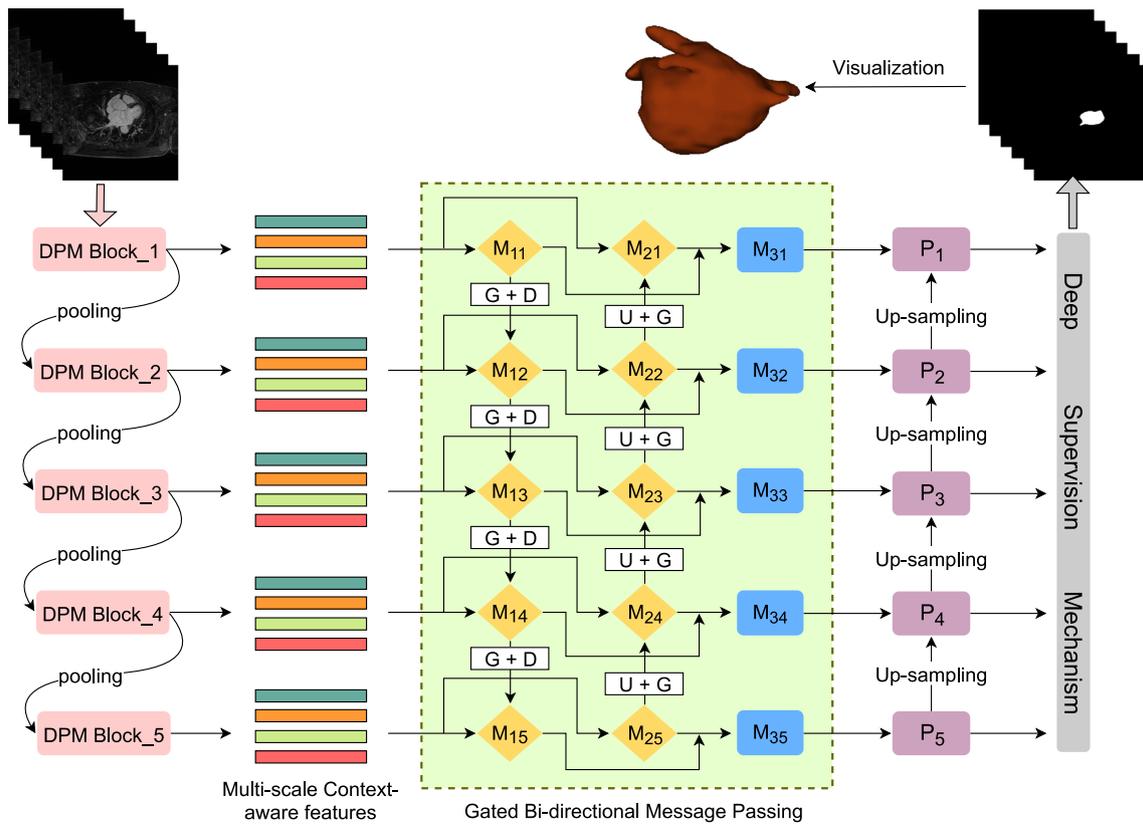


Fig. 1 The workflow of our proposed method. ‘G + D’ represent gate and down-sampling, and ‘U + G’ represent up-sampling and gate. In order to ensure consistent representation and easy understanding, we use m_i^1, m_i^2 and m_i^3 to represent the feature maps in M_{1i}, M_{2i} and M_{3i} , respectively, here $i = 1, \dots, 5$

Multi-scale context-aware module

Existing FCN-based methods, which consist of a series of convolution and max-pooling layers, cannot effectively handle these complex appearance variables. Therefore, Bian et al. [18] obtain the final LA segmentation results after extracting multi-scale features by adopting pyramid pooling. Due to the large stride of pooling, the image details will be easily missed. A recent work [19] used stack sequential blocks containing several dilated convolutions to capture different scale context information. Inspired by it, we put forward a multi-scale context-aware module (MSCM) consisting of four dilated convolutions with dilation ratios 1, 3, 5, and 7 to address large variations of LA and PPVs in shape and appearance. The main reason is that the traditional convolution filters with different kernel sizes produce redundant information and increase computation. Figure 3 shows the details of the MSCM, for the input image I , we first utilized the DPM to extract features at different levels, which are denoted as $F = \{f_i, i = 1, \dots, 5\}$. Then, four dilated convolution layers with various receptive fields are used to capture features f_i context information at a different scale. Finally, multi-level contextual features $F^s = \{f_i^s, i = 1, \dots, 5\}$ are

produced by concatenating the output of the four dilated convolution layers along the channel axis.

The gated bidirectional message passing module

Zeng et al. [20] proposed a bidirectional structure to pass information between context areas of the bounding box for object detection. Inspired by the work, we introduce and improve the gated bidirectional message passing module (GBMPM) to effectively and adaptively integrate the different level features. The difference from the bidirectional structure proposed in [20] is that our GBMPM is established among the different level output of the DPM. In the structure, the extracted multi-level features $F^s = \{f_i^s, i = 1, \dots, 5\}$ as input and $m_i^3, i = 1, \dots, 5$ represent the output features. Our GBMPM involves two directions of information passing among features with different spatial resolutions. We use the m to represent the output feature maps of each stage in different directions. The superscript numbers (1 and 2) of the m indicate the direction, and the subscript numbers indicate the different stages. Using m can not only clearly show the specific information passing process of the GBMPM but also lead to better understanding of the flow of information in the

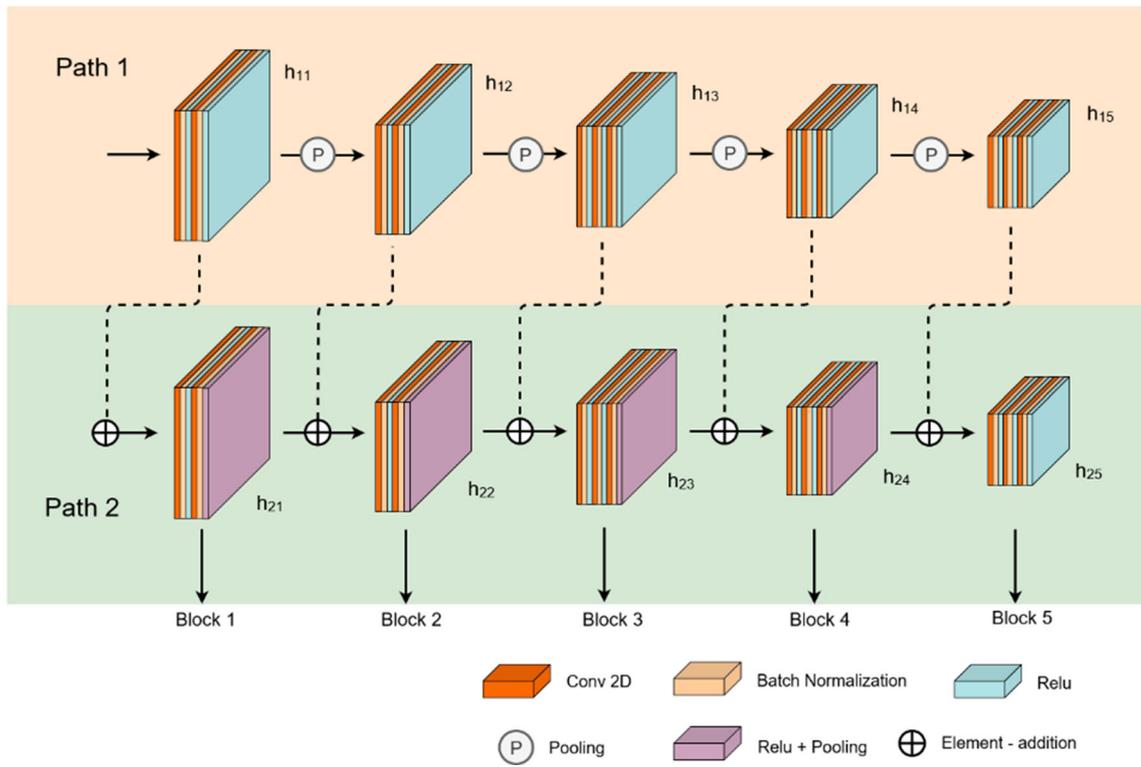
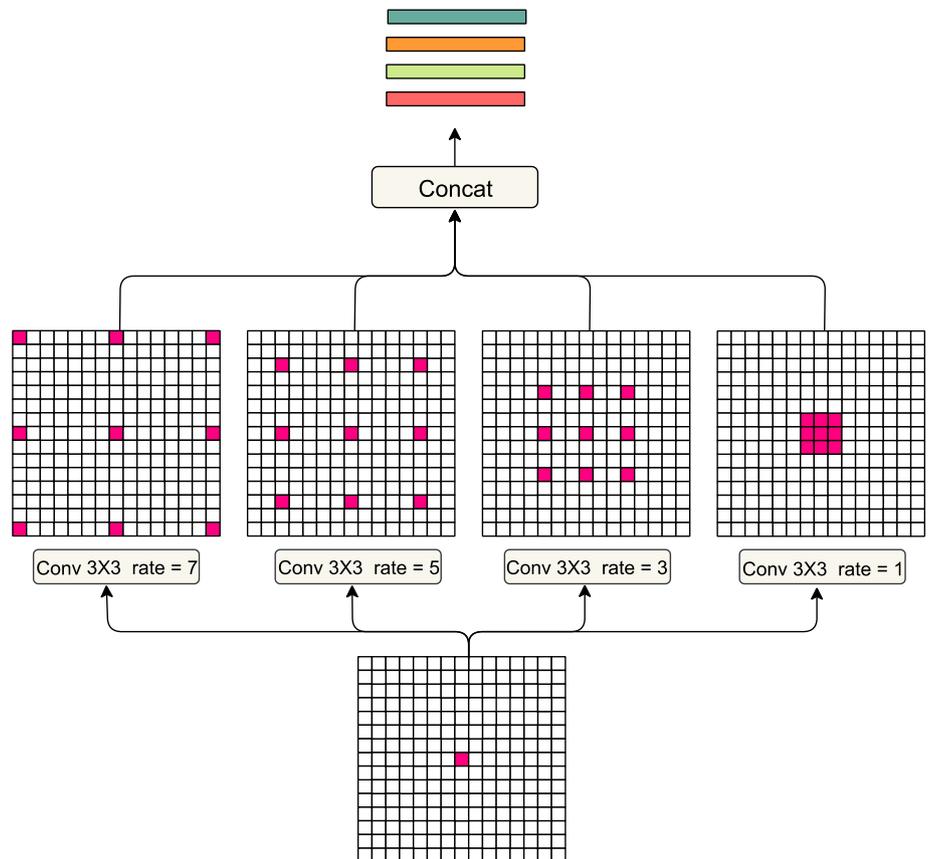


Fig. 2 Illustration of the dual-path module (DPM)

Fig. 3 Example of dilated convolution for the same input feature map. And the number of each dilated convolution channels is 64



GBMPM. Suppose $m_i^0 = f_i^s$, $m_0^1 = 0$, and $m_6^2 = 0$, the process of information transmitting from the shallow layers to deep layers as follow.

$$\begin{aligned}
 m_i^1 &= \text{Down} \left(G \left(m_{i-1}^0; \theta_{i-1,i}^{g^1} \right) \otimes \emptyset \left(\text{Conv} \left(m_{i-1}^1; \theta_{i-1,i}^1 \right) \right) \right) \\
 &\quad + \emptyset \left(\text{Conv} \left(m_i^0; \theta_i^1 \right) \right)
 \end{aligned} \tag{2}$$

And the message passing from the opposite direction is:

$$\begin{aligned}
 m_i^2 &= \text{Up} \left(G \left(m_{i+1}^0; \theta_{i,i+1}^{g^2} \right) \otimes \emptyset \left(\text{Conv} \left(m_{i+1}^2; \theta_{i,i+1}^2 \right) \right) \right) \\
 &\quad + \emptyset \left(\text{Conv} \left(m_i^0; \theta_i^2 \right) \right)
 \end{aligned} \tag{3}$$

where $\text{Down}()$ and $\text{Up}()$ represent the downsample and upsample the feature map, respectively. And $\emptyset()$ denotes ReLU activation function. The g in the superscript of the parameter θ indicates the parameter in the $G()$ gate function. As for the numbers 1 and 2 after g , they represent, respectively, the two directions of information passing in the GBMPM. \otimes denotes element-wise product. During the process of information passing along the bidirection, it should be decided whether the information on the current level is helpful for the next level. A gate function consisting of a convolutional layer with a sigmoid activation function, which generates a message rate in the range of $[0,1]$, is employed to control the information transmitting. After filtering by the gate function, the pixel value of the region containing the redundant information on the feature maps is less than 0.5 and close to 0, and the pixel value of the useful information region is greater than 0.5 and close to 1. Finally, the above feature maps and the feature maps by the previous convolution operation and the activation function is subjected to element-wise product operation to generate feature maps, and the produced feature maps pay more attention to the information of the useful area and discard other redundant information. Moreover, if we use convolution operations instead of the element-wise product, the network will not be able to achieve the effect of highlighting the object region. The $G(*; \theta^g)$ denotes the gate function which is defined as.

$$G(x; \theta^g) = \text{Sigmoid}(\text{Conv}(x; \theta^g)) \tag{4}$$

Here $\text{Conv}(*; \theta^g)$ represents a 3×3 convolution layer that the number of channels is equal to x , this indicates that each channel whose gate function is x learns a different gating filter. If $G(x; \theta^g) = 0$, the message of x will be prevented from passing to other levels.

After information passing, the features in m_i^1 could obtain more fine spatial details from low-level features, and features in m_i^2 obtain semantic information from high-level features.

As a result, we integrate the features from bidirections at multi-level context features as follows

$$m_i^3 = \emptyset \left(\text{Conv} \left(\text{Cat}(m_i^1, m_i^2); \theta_i^3 \right) \right). \tag{5}$$

where $\text{Cat}()$ denotes the concatenation operation among channel axis. m_i^3 contains both spatial details and semantic information. Hence, $m_i^3, i = 1, \dots, 5$ will be useful for LA segmentation and visualization. Taking the feature maps m_i^3 and the prediction P_{i+1} as input, the incorporate process is summarized as follows:

$$P_i = \text{Conv} \left(m_i^3; \theta_i^f \right) + \text{Up}(P_{i+1}), i < 5 \tag{6}$$

$$P_i = \text{Conv} \left(m_i^3; \theta_i^f \right), i = 5 \tag{7}$$

where $\text{Conv}(*; \theta^f)$ is a 1×1 convolutional layer. Using Eq. 6, predictions from deep layers are gradually transmitted to shallow ones. Finally, for the convenience of calculation, we set the number of channels in the convolution layer of GBMPM and P_i to 64.

Deep supervision mechanism

Since gradient-vanishing problems usually happen in the shallow layers of the network, we add the batch normalization and ReLU activation functions after the convolution operation to effectively avoid the gradient-vanishing problems in the shallow layers. In addition, we use deep supervision to train hidden layers in the network. Specifically, we firstly use additional upsampling operations to expand the feature maps at each level; we then use the soft-max function on these full-scale feature maps and obtain additional dense predictions. For these branch prediction results, we calculate their classification errors for ground truth. These auxiliary losses are combined with the auxiliary loss of the last feature integration layer to stimulate the backpropagation of the gradient in order to update the parameters efficiently. In each training iteration, the input of the network is large capacity data, and the errors of these different loss components are back-propagated at the same time. In a word, the auxiliary losses are fused to stimulate the backpropagation of the gradient so that the network can update the parameters efficiently in each iteration. Therefore, it can assist shallow layers to avoid gradient-vanishing. Furthermore, deep supervision in our network can also alleviate the loss of some details caused by sampling operations. As illustrated in Fig. 4, we add auxiliary side paths as well as the specific side with multi-level integrated features and

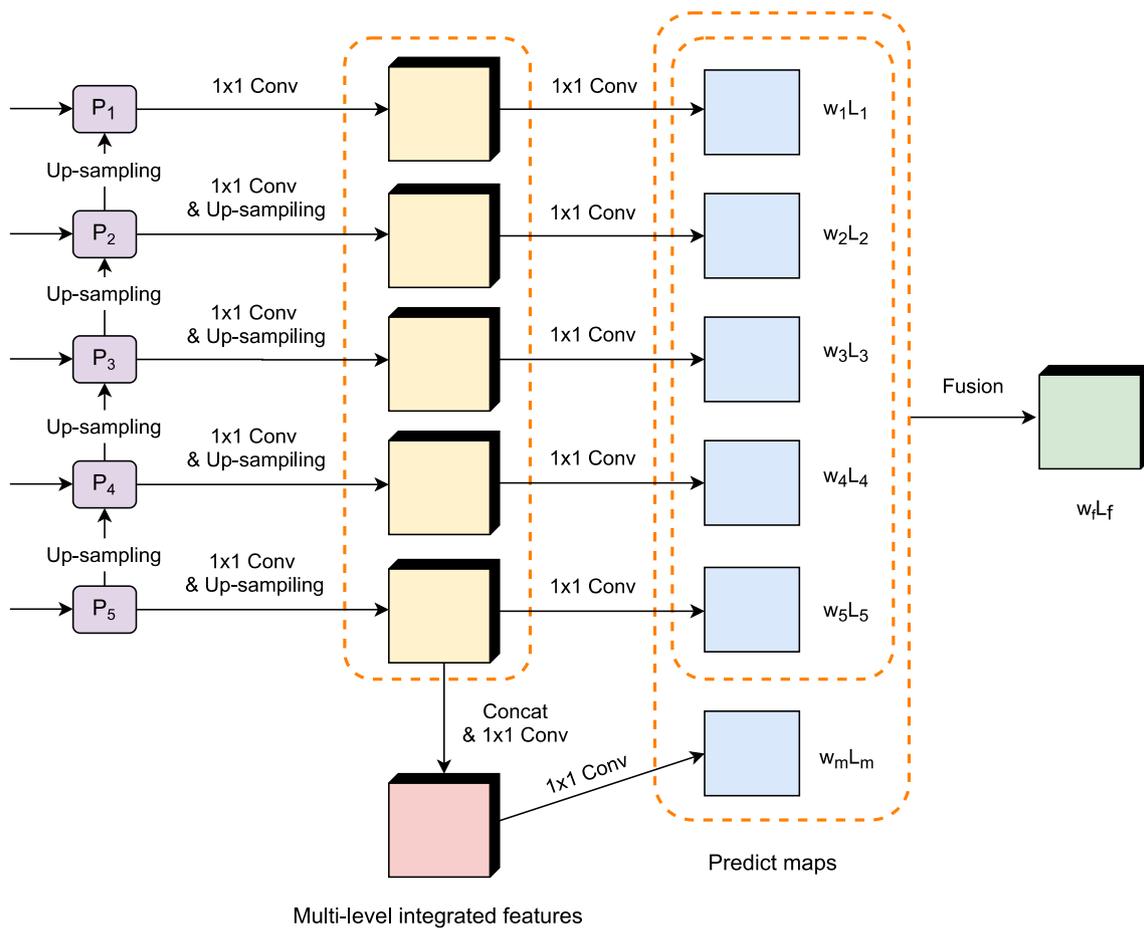


Fig. 4 Details of the deep supervision mechanism

thus expose each level to extra supervisions. As a result, the overall loss function $L_{overall}$ is elaborated by:

$$L_{overall} = \sum_i w_i L_i + w_m L_m + w_f L_f + \alpha L_2 \quad i = 1, \dots, 5 \quad (8)$$

here w and L represent the weight and loss, respectively, and the value of α is 0.0005. Note that we empirically set all the weights w_i , w_m and w_f are 1. In this work, it contains six auxiliary losses and a main loss in the overall network. Each loss function L_t in the network is calculated by:

$$L_t = L_{ce} + L_{dice} + L_{mae} + L_{overlap}, \quad t \in \{i, m, f\} \quad (9)$$

And the first, we use the cross-entropy loss L_{ce} as a basic component for main loss and all auxiliary losses to optimize the network.

$$L_{ce} = - \sum_{x,y} l_{x,y} \log(P_{x,y}) + (1 - l_{x,y}) \log(1 - P_{x,y}) \quad (10)$$

where $l_{x,y} \in \{0, 1\}$ is the label for each pixel of the image (x, y) , and $P_{x,y}$ is the probability for each pixel of the image (x, y) . The dice coefficient loss function can solve the problem of class imbalance which causes the traditional loss function to bias and makes the network ignores the minor classes. The generalized dice loss [21] L_{dice} is adopted as follows:

$$L_{dice} = 1 - 2 \left[\frac{(w_g \sum_{x,y} S_{x,y} G_{x,y} + w_s \sum_{x,y} (1 - S_{x,y})(1 - G_{x,y}))}{w_g \sum_{x,y} (S_{x,y} + G_{x,y}) + w_s \sum_{x,y} (2 - S_{x,y} - G_{x,y})} \right] \quad (11)$$

where $w_g = (\sum_{x,y} G_{x,y})^{-1}$, $w_s = (\sum_{x,y} (1 - G_{x,y}))^{-1}$ G is ground truth and S is predicted map. We use the modified mean absolute errors L_{mae} penalize misclassified LA region pixels and misclassified non-LA region pixels:

$$L_{mae} = \sum_{x,y} \ln(1 + e^{|G_{x,y} - S_{x,y}|}) \quad (12)$$

Note that the soft-plus function is adopted in L_{mae} make it easy to optimize. The boundary pixels in the LA region

are more likely to be classified incorrectly, so that boundary ambiguity makes the classification of the LA region and the non-LA region more difficult. Therefore, we introduce the overlap loss [22] L_{overlap} to enlarge the gap between the LA region and the non-LA region to minimize the misclassification of pixels.

$$L_{\text{overlap}} = \sum_{x,y} \ln(1 + e^{|S_{x,y} * (1 - S_{x,y})|}) \quad (13)$$

where $*$ is basic multiplication and L_{overlap} also use the soft-plus function to assist optimize.

Experiment setup

Datasets

The 100 3D GE-MRIs with ground truth annotation were provided from the STACOM 2018 challenge dataset (<http://atriaseg2018.cardiacatlas.org/>). The size of the 3D GE-MRIs is $88 \times 640 \times 640$ and $88 \times 576 \times 576$ with the unified spacing $1.0 \times 1.0 \times 1.0$ mm. Because of the information on the edge of the GE-MRIs is useless, we crop each volume center to a size of $88 \times 400 \times 400$, and each 3D volume is divided into multiple slices along the first dimension and each slice further resizes to 256×256 . All the samples are normalized as zero mean and unit variance before the network started running. We randomly split the dataset and validate the overall network via fivefold cross-validation, that is, 80 subjects were used for training and 20 subjects for testing. We implemented our network by TensorFlow on NVidia Tesla P100 GPU (16 GB GPU memory). For the dilated convolutional layers in the MSCM, we initialize the weights using the truncated normal method, and the parameters of the other convolutional layers are initialized by [23]. We use the Adam method (batchsize = 4, learning rate = $1e-6$) as the optimization algorithm to train our proposed network.

Evaluation criteria

We first evaluate the performance of our method and other approaches using four metrics, including dice score, conform coefficient and Jaccard coefficient, which are defined, respectively, by

$$\text{dice} = 2 * \frac{S \cap G}{S + G} \quad (14)$$

$$\text{conform} = \frac{3 * \text{dice} - 2}{\text{dice}} \quad (15)$$

$$\text{jaccard} = \frac{S \cap G}{S \cup G} \quad (16)$$

where G is ground truth and S is predicted map. These metrics focus on the overlap between ground truth area and predicted area and they are in the range of $0 \sim 1$. Furthermore, higher values denote predicted contour closer to manual contour. Besides, we also employ Hausdorff distance (HD) [24] to measure shape variations. It is written as

$$\text{HD}(G, S) = \max_{g \in G} \left\{ \min_{s \in S} \left\{ \sqrt{g^2 - s^2} \right\} \right\} \quad (17)$$

A smaller HD represents a better match between predicted and manual contours.

Results and analysis

Analysis of the segmentation results

Our method delivers accurate segmentation for LA with PPVs. The proposed network achieves average dice scores of 0.936 as well as an average HD of 11.889 mm. Figure 5 shows segmentation results for the whole subjects in our dataset. Among these results, the maximum of dice score in our dataset is 0.95, and the minimum of dice score is 0.893 which was potentially due to the PPVs of the patient are relatively small and thin compared to other patients and thus it more difficult to predict. Figure 6 shows the comparison between the segmentation contours by our method and ground truth counterparts manually obtained by human experts. From Fig. 6, we can clearly see that the predicted contour is very close to the manual contour. For the shape and size of LA and PPVs vary widely among different subjects, our method is capable of highlighting boundary positions of the LA region with PPVs. Therefore, the proposed method not only handles successfully the complex variation pattern of LA and associated PPVs but also achieves accurate segmentation for all the patients from GE-MRIs, revealing its great potential for identification and diagnosis of AF in clinical practice.

Analysis of the visualization results

Atrial visualization can provide reliable information for clinical treatment and optimize the therapy plan. Figure 7 shows the evaluation of the visualization quality of the proposed approach. In Fig. 7, we can better see the overall prediction performance of the LA of each patient from a 3D perspective. Compared with the ground truth, our method also reconstructs some spatial details well, such as the small branch of the PPVs and the fine-grained details of its ends. To better verify the effects of the visualization results, we quantify the 3D volumes of one of the clinical indicators. In Fig. 8, the blue dot around the red line indicates that the LA body and asso-

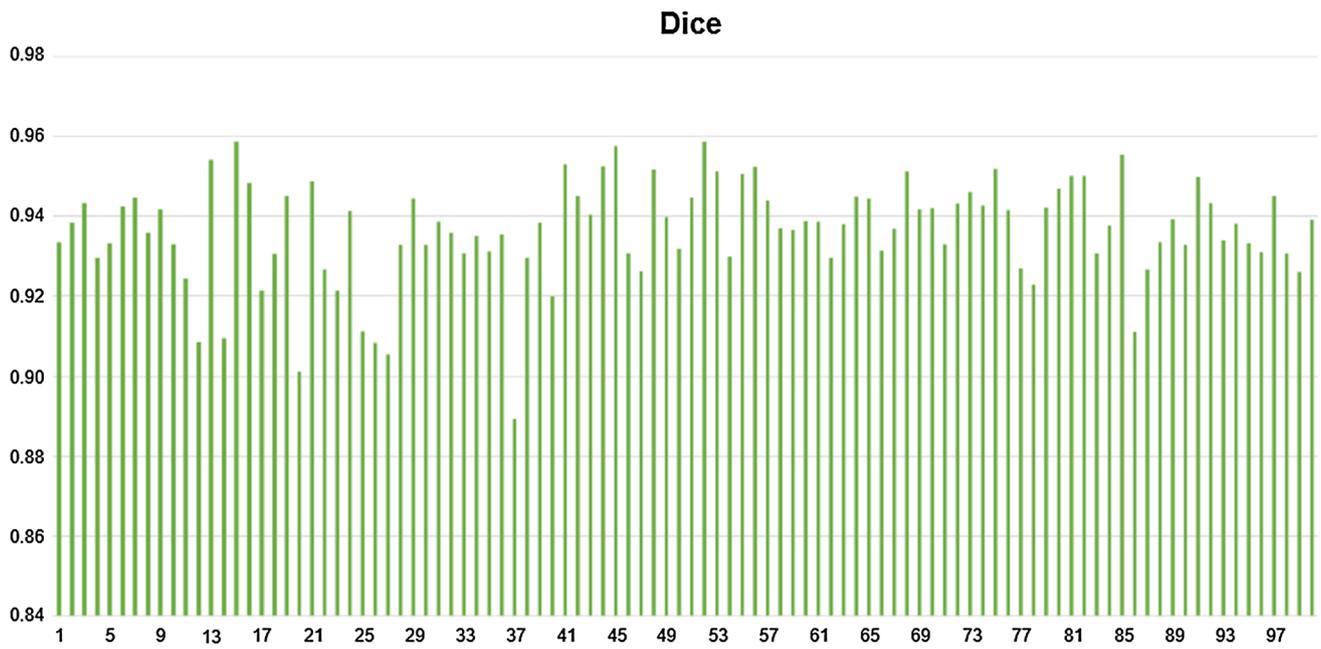


Fig. 5 Dice score for all the subjects based on our proposed network

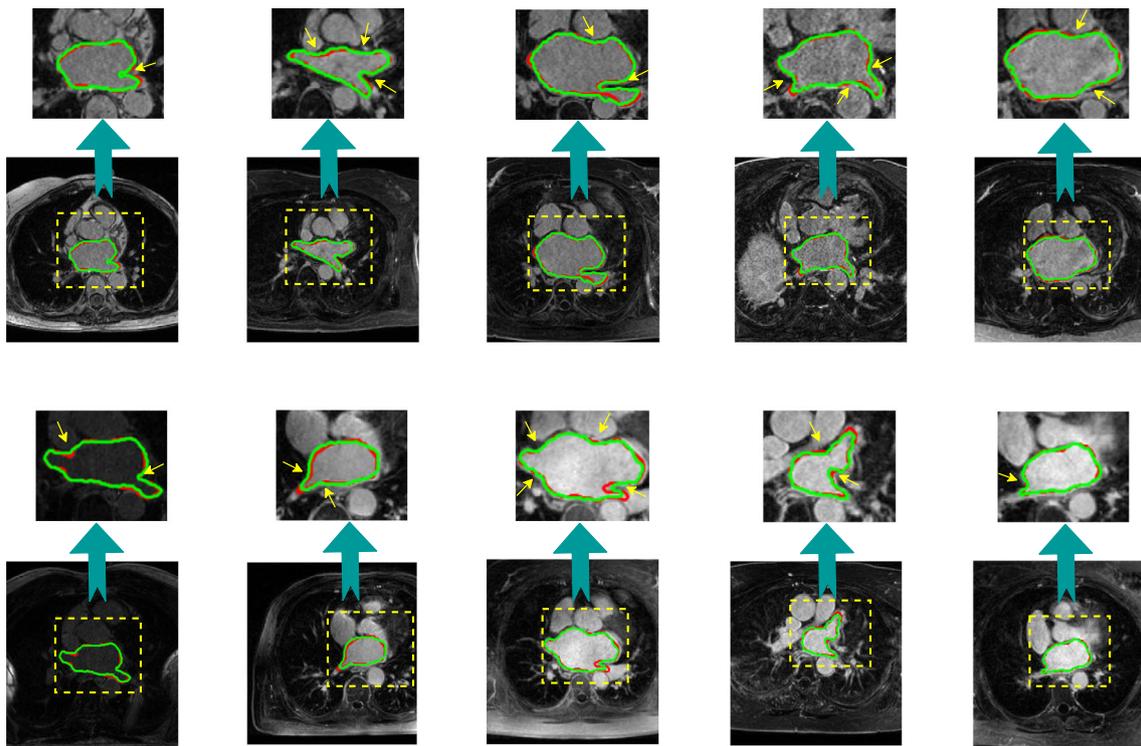


Fig. 6 Segmentation results of the LA from the proposed method compared to the ground truth for representative slices on the same 3D GE-MRIs. The red line represents the manual definition, and the green

line represents the predicted contours. Our method can well segment areas pointed by the arrow that is difficult to segment

ciated PPVs are completely reconstructed by our approach from the corresponding LA with PPVs. From Fig. 8, we can obviously observe that the volumes predicted by the proposed method is very close to the truth 3D volumes, and the

results display that it has a good correlation between our automatic method and ground truth. Our method yielded accurate visualization results, demonstrating its advantages in dealing with high structural variability in the LA with PPVs. More-

Fig. 7 3D surface visualization of the LA for the ground truth and the output generated by the proposed method, respectively

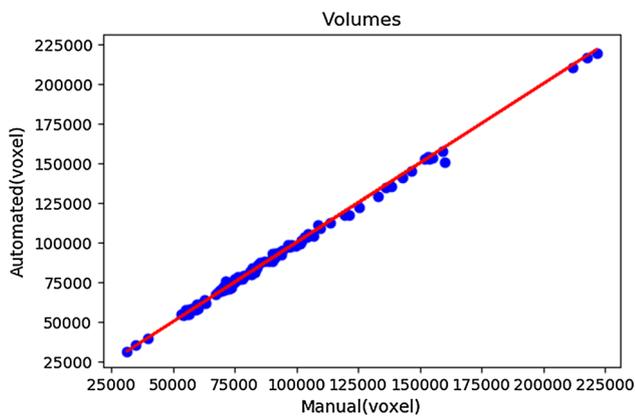
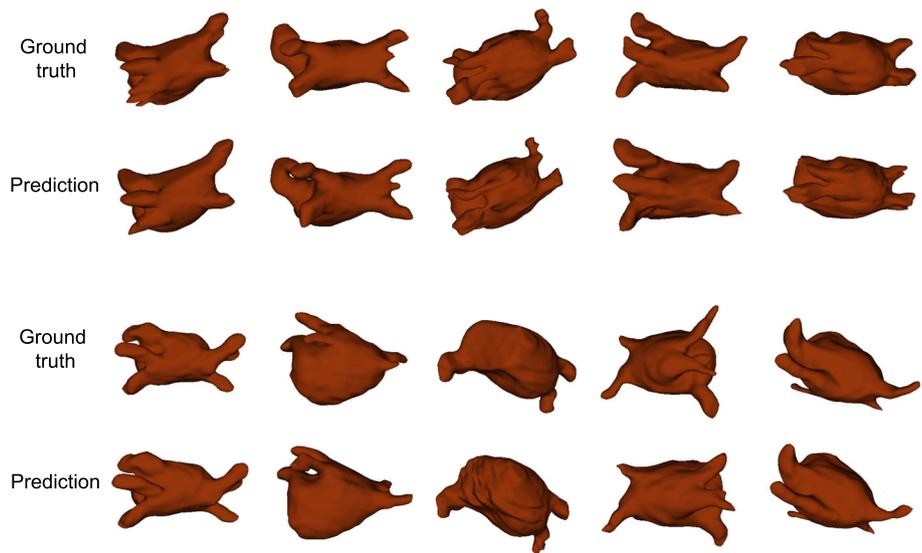


Fig. 8 Comparison between visualization volumes by the automated method and truth volumes

over, our network is capable of automatically segmenting the 2D slices of GE-MRIs for accurate 3D visualization of atrial geometry. This is beneficial to clinical patient-specific diagnostics and treatment.

Ablation analysis

Our approach is composed of four modules, including DPM, MSCM, GBMPM, and DS module. To investigate whether each module has a positive effect on the final segmentation and visualization results, several settings are used to demonstrate the four components' contributions to the network. From Table 1 and Fig. 9, we can observe that the results are effective in LA segmentation and visualization, respectively. We conduct experiments to highlight the advantages of our proposed network which consisting of four modules. The experiments demonstrate that the results improve with each addition of a proposed module. Consequently, the four

Table 1 Quantitative comparisons of different settings in our network. The value in parentheses indicates the standard deviation

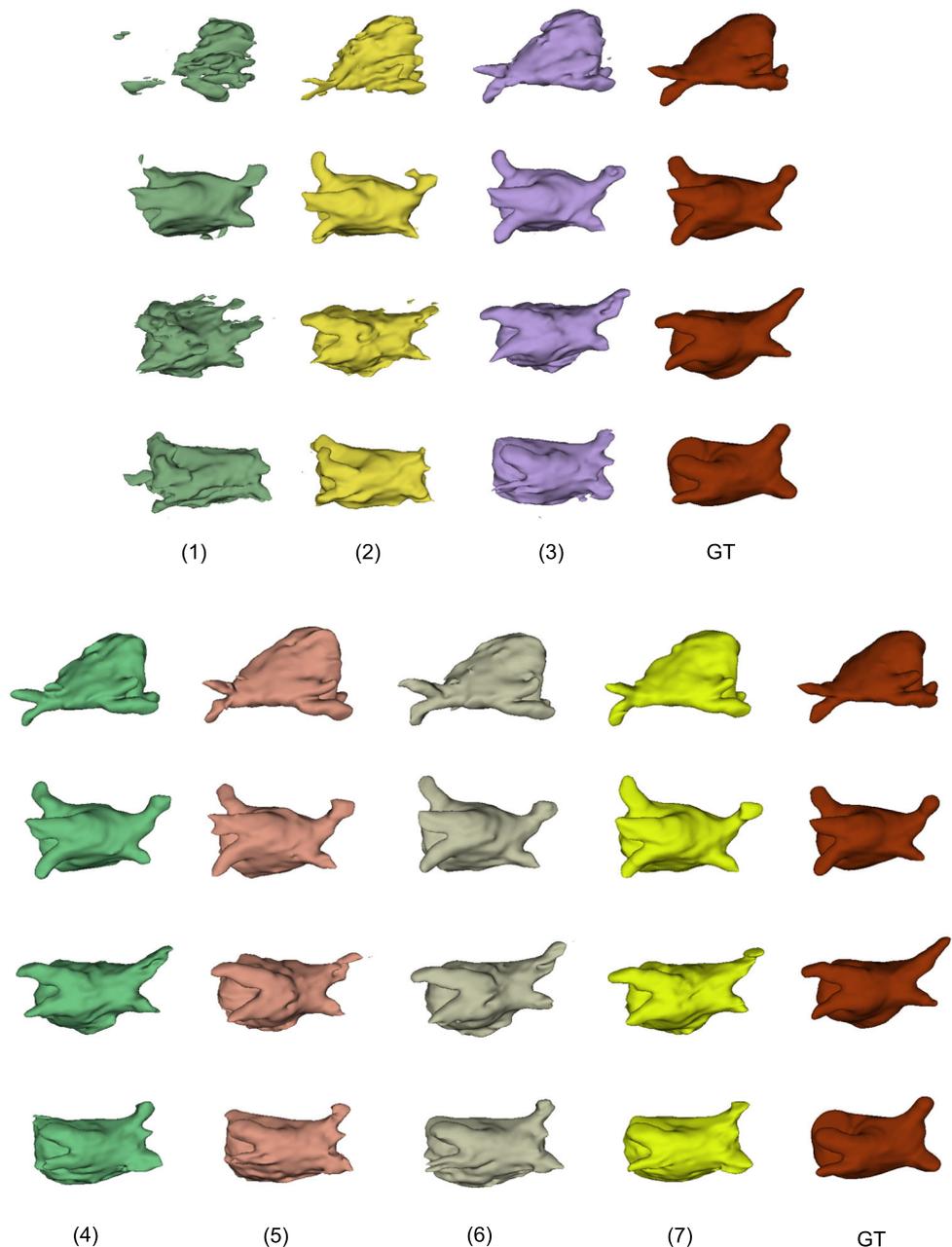
Model setting	Dice (%)
Basic	86.148(0.984)
Basic + DPM	90.299(0.685)
Basic + DPM + MSCM	92.840(0.450)
Basic + DPM + MSCM + GBMPM (downsampling stream)	93.503(0.522)
Basic + DPM + MSCM + GBMPM (upsampling stream)	93.402(0.553)
Basic + DPM + MSCM + GBMPM (bidirectional structure)	93.583(0.522)
Basic + DPM + MSCM + GBMPM + DS	93.640(0.536)

modules can assist the network to predict LA regions more accurately and reduce the impact of adjacent tissues and others. The comparison results prove that the combination of the four components in our model contributes to the final segmentation and visualization result because the performance will decrease when using one of the four components alone. Therefore, the experiments demonstrate that the combination of DPM, MSCM, GBMPM, and DS module makes our method an efficient and reliable solution for the diagnoses and understanding of AF.

Performance comparison with state-of-the-art methods

In order to further verify the effectiveness of our proposed method, we compared with several state-of-the-art algorithms, including the LA segmentation methods and other typical image segmentation techniques. Table 2 reports the results of different approaches under different measurements.

Fig. 9 Comparison visualization results of the LA for the ablation analysis. (1) Basic network; (2) Basic + DPM; (3) Basic + DPM + MSCM; (4) Basic + DPM + MSCM + GBMPM (downsampling stream); (5) Basic + DPM + MSCM + GBMPM (upsampling stream); (6) Basic + DPM + MSCM + GBMPM (bidirectional structure); (7) Basic + DPM + MSCM + GBMPM + DS



When compared with existing typical segmentation networks, our method significantly outperforms the U-Net-2D network by 4% average dice score and SegNet network by 2.8% average dice score. Besides, our method is obviously superior to the recent approaches proposed on this dataset. Among them, the average dice score obtained by our method is 8.8% higher than the Vesal et al. [14] and 0.8% higher than Bian et al. [18]. Our method overcomes complex shape variance and size of LV and PPVs and achieves promising performance. It proves the effectiveness and advantages of our network is beneficial to LA with PPVs segmentation. Therefore, automated segmentation and reconstruct of LA

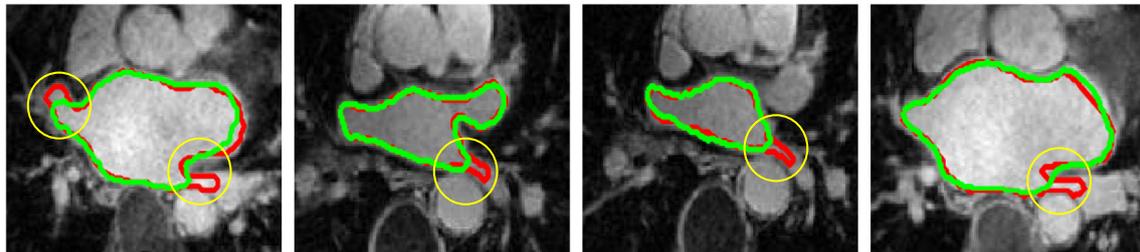
with PPVs is extremely useful in providing reliable and objective diagnoses and treatment of AF and relieving clinicians from laborious workloads.

Analysis of HD metric

The HD refers to the Hausdorff distance between the manual contours and predicted contours, which better reflects the degree of matching between manual and predicted contours. As shown in Fig. 10, it can be clearly seen that the manual contours are very close to the predicted contours for the LA body, and large HD values often occur around the

Table 2 Performance comparison between our method and state-of-the-art methods for LA segmentation

Methods	Dice (%)	Conform (%)	Jaccard (%)	HD (mm)
PSPNet [25]	87.904	72.218	78.532	23.319
U-Net-2D	89.638	76.485	81.404	22.711
U-Net-3D	87.020	68.033	77.785	25.243
DeepLabV3 [26]	88.456	72.449	79.846	23.099
SegNet [27]	90.762	78.912	83.422	23.362
Yang et al. [13]	92.244	83.120	85.644	18.293
Vesal et al. [14]	84.800	–	77.000	–
Chen et al. [15]	90.100	–	82.200	14.230
Bian et al. [18]	92.834	84.445	86.694	17.891
Our method	93.640	86.377	88.068	11.889

**Fig. 10** Some slices of GE-MRIs with poor matching, yellow circles indicate the PPVs with higher HD values

PPVs, which indicates a poor match between the predicted contours and the ground truth. It is caused by the shape and size of the LA with PPVs of different subjects. The final HD is the average result of combining the LA body and the PPVs. The HD of the PPVs is higher, which will increase the final average result. It is worth noting that not all PPVs have poor matching. All in all, the average HD is reasonable and has a great competitive advantage compared with the result of other methods. However, we also admit that our method does not perfectly solve the segmentation problem of PPVs with different shapes, sizes, and lengths. This makes the predicted structure of the LA and PPVs different from the ground truth, which may make the atrial fibrillation ablation plans can be biased and affect treatment effects. Therefore, we are working hard to explore new solutions to solve this problem so that the LA and PPVs have a high degree of matching with ground truth.

Conclusion

This paper proposes a unified approach that can be used to perform automatic LA segmentation and accurate LA visualization from GE-MRIs. We have studied a robust model from GE-MRIs in four modules: (1) a novel dual-path structure

for improving the expressiveness of GE-MRIs representation at different levels. (2) MSCM, which contains four dilated convolution layers with different receptive fields, to capture the context information for LA feature learning at multiple scales. (3) GBMPM to incorporate features from different levels adaptively transmit information to each other. (4) the DS mechanism is used to generate score maps on different levels and utilize several loss functions to make effective the training process. In the end, we evaluated the effectiveness of our designed network on STACOM 2018 challenge dataset via fivefold cross-validation and our method achieves the desired results. It can accelerate the development of a more accurate segmentation and visualization of the atrial geometry method, which can possibly be assisting in improving clinical diagnosis and clinical guidance during ablation treatment for subjects with AF. In the future, we will look to expand our method and improve the existing network framework to deal with other problems of the organ segmentation and visualization.

Acknowledgements We thank all the anonymous reviewers for their valuable comments and constructive suggestions, which were helpful for improving the quality of the paper. This work was supported in part by the National Science Foundation of China under Grant 61673020, the Anhui Provincial Natural Science Foundation under Grand 1708085QF143.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in our study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee.

References

- Peng P, Lekadir K, Gooya A, Shao L, Petersen SE, Frangi AF (2016) A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging. *Magma* 29(2):155–195
- Christopher MG, Nazem A, Amit P, Eugene K, Patricia R, Kavitha D, Brent W, Josh C, Alexis H, Ravi R (2014) Atrial fibrillation ablation outcome is predicted by left atrial remodeling on MRI. *Circ Arrhythm Electrophysiol* 7(1):23
- Zhao J, Butters TD, Zhang H, Pullan AJ, Legrice IJ, Sands GB, Smaill BH (2012) An image-based model of atrial muscular architecture: effects of structural anisotropy on electrical activation. *Circ Arrhythm Electrophysiol* 5(2):361–370
- Zhuang X, Rhode KS, Razavi R, Hawkes DJ, Ourselin S (2010) A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI. *IEEE Trans Med Imaging* 29(9):1612–1625
- Zheng Y, Yang D, John M, Comaniciu D (2014) Multi-part modeling and segmentation of left atrium in C-arm CT for image-guided ablation of atrial fibrillation. *IEEE Trans Med Imaging* 33(2):318–331
- Tobongomez C, Geers AJ, Peters J, Weese J, Pinto K, Karim R, Ammar M, Daoudi A, Margeta J, Sandoval Z (2015) Benchmark for algorithms segmenting the left atrium from 3D CT and MRI datasets. *IEEE Trans Med Imaging* 34(7):1460–1473
- Zhu L, Gao Y, Yezzi AJ, Tannenbaum AR (2013) Automatic segmentation of the left atrium from MR images via variational region growing with a moments-based shape prior. *IEEE Trans Image Process* 22(12):5111–5122
- Tao Q, Ipek EG, Shahzad R, Berendsen FF, Nazarian S, Der Geest RJV (2016) Fully automatic segmentation of left atrium and pulmonary veins in late gadolinium-enhanced MRI: towards objective atrial scar assessment. *J Magn Reson Imaging* 44(2):346–354
- Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. In: *Medical image computing and computer assisted intervention*. Lecture notes in computer science, vol 9351. Springer, Cham, pp 234–241
- Milletari F, Navab N, Ahmadi S (2016) V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: *International conference on 3D vision*. pp 565–571
- Tran PV (2016) A fully convolutional neural network for cardiac segmentation in short-axis MRI. [arXiv:1604.00494](https://arxiv.org/abs/1604.00494)
- Yang X, Bian C, Yu L, Ni D, Heng P-A (2017) Hybrid loss guided convolutional networks for whole heart parsing. In: *International workshop on statistical atlases and computational models of the heart*. Springer, Cham, pp 215–223
- Yang X, Wang N, Wang Y, Wang X, Nezafat R, Ni D, Heng P (2018) Combating uncertainty with novel losses for automatic left atrium segmentation. In: *International workshop on statistical atlases and computational models of the heart*, pp 246–254
- Vesal S, Ravikumar N, Maier AK (2018) Dilated convolutions in neural networks for left atrial segmentation in 3D gadolinium enhanced-MRI. In: *International workshop on statistical atlases and computational models of the heart*. Springer, Cham, pp 319–328
- Chen C, Bai W, Rueckert D (2018) Multi-task learning for left atrial segmentation on GE-MRI. In: *International workshop on statistical atlases and computational models of the heart*. Springer, Cham, pp 292–301
- Mortazi A, Karim R, Rhode KS, Burt J, Bagci U (2017) Cardiac-NET: Segmentation of left atrium and proximal pulmonary veins from MRI using multi-view CNN. In: *Medical image computing and computer assisted intervention*. Lecture notes in computer science, vol 10434, pp 377–385
- Chen J, Yang G, Gao Z, Ni H, Angelini ED, Mohiaddin RH, Wong T, Zhang Y, Du X, Zhang H (2018) Multiview two-task recursive attention model for left atrium and atrial scars segmentation. *Medical image computing and computer assisted intervention*. Lecture notes in computer science, vol 11071. Springer, Cham, pp 455–463
- Bian C, Yang X, Ma J, Zheng S, Liu Y, Nezafat R, Heng P, Zheng Y (2018) Pyramid network with online hard example mining for accurate left atrium segmentation. *International workshop on statistical atlases and computational models of the heart*. Lecture notes in computer science, vol 11395. Springer, Cham, pp 237–245
- Li D, Chen X, Zhang Z, Huang K (2017) Learning deep context-aware features over body and latent parts for person re-identification. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 384–393
- Zeng X, Ouyang W, Yang B, Yan J, Wang X (2016) Gated Bidirectional CNN for Object Detection. In: *European conference on computer vision*. pp 354–369
- Liu T, Yuan Z, Sun J, Wang J, Zheng N, Tang X, Shum H (2011) Learning to detect a salient object. *IEEE Trans Pattern Anal Mach Intell* 33(2):353–367
- Wang Z, Liu C, Cheng D, Wang L, Yang X, Cheng K (2018) Automated detection of clinically significant prostate cancer in mp-MRI images based on an end-to-end deep neural network. *IEEE Trans Med Imaging* 37(5):1127–1139
- He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: surpassing human-level performance on Imagenet classification. In: *International conference on computer vision*. pp 1026–1034
- Huttenlocher DP, Klanderman GA, Rucklidge W (1993) Comparing images using the Hausdorff distance. *IEEE Trans Pattern Anal Mach Intell* 15(9):850–863
- Zhao H, Shi J, Qi X, Wang X, Jia J (2017) Pyramid scene parsing network. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 6230–6239
- Chen L, Papandreou G, Schroff F, Adam H (2017) Rethinking atrous convolution for semantic image segmentation. [arXiv:1706.05587](https://arxiv.org/abs/1706.05587)
- Badrinarayanan V, Kendall A, Cipolla R (2017) SegNet: a deep convolutional encoder–decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell* 39(12):2481–2495

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.