ORIGINAL PAPER

A deep learning approach with temporal consistency for automatic myocardial segmentation of quantitative myocardial contrast echocardiography

Mingqi Li[1,2] · Dewen Zeng[3] · Qiu Xie[1] · Ruixue Xu[1] · Yu Wang[1,2] · Dunliang Ma[1] · Yiyu Shi[3] · Xiaowei Xu[4] · Meiping Huang[5] · Hongwen Fei[1]

Abstract

Quantitative myocardial contrast echocardiography (MCE) has been proved to be valuable in detecting myocardial ischemia. During quantitative MCE analysis, myocardial segmentation is a critical step in determining accurate region of interests (ROIs). However, traditional myocardial segmentation mainly relies on manual tracing of myocardial contours, which is time-consuming and laborious. To solve this problem, we propose a fully automatic myocardial segmentation framework that can segment myocardial regions in MCE accurately without human intervention. A total of 100 patients' MCE sequences were divided into a training set and a test set according to a 7: 3 proportion for analysis. We proposed a bi-directional training schema, which incorporated temporal information of forward and backward direction among frames in MCE sequences to ensure temporal consistency by combining convolutional neural network with recurrent neural network. Experiment results demonstrated that compared with a traditional segmentation model (U-net) and the model considering only forward temporal information (U-net + forward), our framework achieved the highest segmentation precision in Dice coefficient (U-net vs U-net + forward vs our framework: 0.78 ± 0.07 vs 0.79 ± 0.07 vs 0.81 ± 0.07, $p < 0.01$), Intersection over Union (0.65 ± 0.09 vs 0.66 ± 0.09 vs 0.68 ± 0.09, $p < 0.01$), and lowest Hausdorff Distance (32.68 ± 14.6 vs 28.69 ± 13.18 vs 27.59 ± 12.82 pixel point, $p < 0.01$). In the visual grading study, the performance of our framework was the best among these three models (52.47 ± 4.29 vs 54.53 ± 5.10 vs 57.30 ± 4.73, $p < 0.01$). A case report on a randomly selected subject for perfusion analysis showed that the perfusion parameters generated by using myocardial segmentation of our proposed framework were similar to that of the expert annotation. The proposed framework could generate more precise myocardial segmentation when compared with traditional methods. The perfusion parameters generated by these myocardial segmentations have a good similarity to that of manual annotation, suggesting that it has the potential to be utilized in routine clinical practice.

Keywords Myocardial contrast echocardiography · Deep neural network · Image segmentation · Myocardial perfusion parameters

Introduction

Coronary artery disease (CAD) remains the leading cause of cardiovascular mortality and morbidity in China. Non-invasive imaging as the initial screening test for symptomatic patients suspected of myocardial ischemia is the Class I recommendation according to the European Society of Cardiology [1].

As a cost-effective examination, myocardial contrast echocardiography (MCE) is an imaging tool for the assessment of myocardial perfusion. According to the ultrasonic enhancing agents (UEAs) guideline by the American Society of Echocardiography [2], after an intravenous infusion of

Mingqi Li, Dewen Zeng and Qiu Xie have contributed equally to this work

✉ Xiaowei Xu
  xiao.wei.xu@foxmail.com

✉ Meiping Huang
  huangmeiping@126.com

✉ Hongwen Fei
  floyd2001@sina.com

Extended author information available on the last page of the article

UEAs which contain the microbubbles, the contrast imaging could be generated by retaining the microbubble signal in the coronary artery and eliminating the myocardial tissue signal. By the "flash" impulse (transient high mechanical (MI) index impulses), which leads to microbubbles destruction and then replenishment, clinicians qualitatively assess myocardial perfusion by observing the intensity of myocardial enhancement and counting the refill time. In terms of quantitative analysis, the localized time-intensity analysis of microvascular reentry of microbubbles can be used to assess the rate and extent of microbubble signal replenishment, reflecting microvascular flux rate and microvascular blood volume, respectively.

Meta-analysis has shown that quantitative stress-rest MCE has good diagnostic performance in detecting CAD [3]. However, the clinical application of quantitative MCE is limited. Accurate myocardial segmentation is a prerequisite for the output of reliable perfusion parameters but requires complex and time-consuming operation. Therefore, if myocardial segmentation can be done automatically and efficiently, quantitative MCE may be popularized and become an important tool for detecting myocardial ischemia in clinic.

Myocardial segmentation of MCE sequences has the following challenges. First, the concentration of enhancing agents in the myocardium and left ventricular changes over time because of the replenishment of microbubbles after "flash" impulses [2, 4]. Second, the shape and pose of the myocardium vary with heart motion, individual differences, and scan setting [5]. Also, myocardial structure features differ in different chamber views. Thirdly, misleading structures such as papillary muscles have the same intensity and grayscale information as myocardium, which makes it hard to extract accurate boundaries. Fourth, after the "flash" impulse associated with the destruction of microbubbles, a large number of signal sources are lost, resulting in a significant reduction in the quality of the first few frames, as shown in supplementary Fig. 1 (Online Resource 1).

To solve the above problems, we proposed an encoder-decoder architecture that combines convolutional neural networks [6, 7] and recurrent neural networks as shown in Fig. 1. The encoder servers as a feature extractor, which was able to extract a set of features from MCE sequences with different resolutions. The decoder used a pyramid of convolutional Long Short-Term Memory (LSTM) to incorporate temporal information between MCE frames to ensure segmentation consistency along time. The main contributions of our work were:

(1) We proposed an encoder-decoder architecture for myocardial segmentation of MCE sequences. Our model was able to incorporate temporal information between frames in MCE sequences.

(2) To further exploit the temporal information among sequences, we proposed a bi-directional training approach which can reduce segmentation error introduced by the first few sequences in the training process.

(3) We collected an MCE dataset which consists of three views from 100 subjects. Experimental results showed that our framework outperformed traditional works in terms of segmentation accuracy.

Materials and methods

Dataset

The database comprised of 100 patients who received MCE from 2019/01/01 to 2019/10/30, who were referred to Guangdong Provincial People's Hospital. This observational study was approved by the institutional review boards of Guangdong Provincial People's Hospital, and the informed consent of patients was obtained. The flow and the process for importing data are shown in Fig. 2.

For all subjects, we decomposed every MCE sequence into several frames, the next frame after the "flash" impulse is manually selected as the first frame, and the sampling rule is set to take one frame every other frame. We collected 30 frames for each MCE sequence. There are 100 (subjects) $\times$ 3 (chamber-views sequences per subject as shown as supplementary Fig. 2 (Online Resource 2)) $\times$ 30 (frames per sequences) = 9000 frames, with 3000 frames for each chamber view and 90 frames for each subject. Then we randomly split our data set into the training set and the testing set by a ratio of 7:3. Therefore, there were 70 subjects (210 MCE sequences) in the training set and 30 subjects (90 MCE sequences) in the testing set. Our segmentation model was used for each chamber view separately.

Contrast echocardiography

An ultrasonography system with a contrast specific multi-pulse amplitude modulation imaging algorithm (Philips 7C or iE ELITE, Philips Medical Systems, Best, Netherlands) equipped with a broadband (1 to 5 MHz) transducer was used, while SonoVue (Bracco Research SA, Geneva, Switzerland) as UEA. SonoVue was diluted with a normal saline to 10 ml in total followed by 5 mL of a slow normal saline flush. The intravenous continues infusion rate was adjusted to obtain maximal opacification of the myocardium with minimal attenuation throughout the examination. The mean infusion.

rate was 2.5–3 ml/min. MCE was performed in the apical two-chamber (A2C), three-chamber (A3C) and four-chamber (A4C) views using power-modulation MCE with a mechanical index of 0.10 to 0.17. A transient high MI
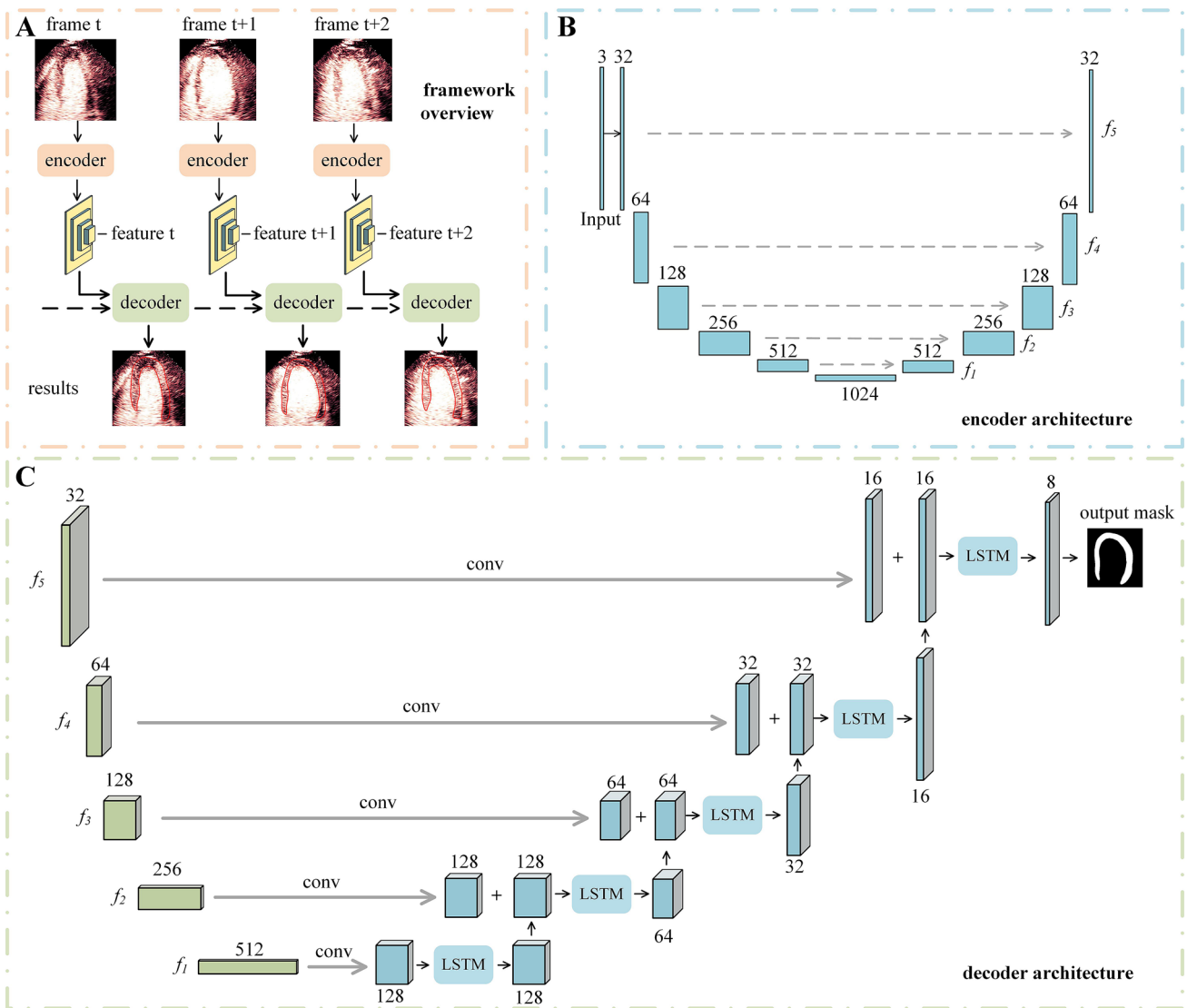
Fig. 1 a The proposed myocardial segmentation architecture, which contains an encoder (U-Net) and a decoder (ConvLSTM). **b** Network structure of the U-net based encoder. The rectangle denotes the feature maps at this layer, the number above the rectangle is channel size. The first rectangle is the input image, which has 3 channels as a RGB image. **c** Network structure of the proposed decoder. The input of our decoder is the features extracted from our encoder. The decoder consists of hierarchical ConvLSTMs and is able to incorporate temporal information between MCE frames

(1.20–1.35) flash (10 frames) was used to clear myocardial microbubbles; the subsequent replenishment imaging was digitally captured (10 to 15 cardiac cycles) in each view with a frame rate of 20 to 30 Hz. The focus was set at the mitral valve level. Once the UEA entered the left ventricular cavity, MCE acquisition began. As shown in supplementary Fig. 2 (Online Resource 2), the myocardium is divided into 17 segments which are ascribed to coronary territories according to the American Heart Association [8]. In perfusion analysis, we manually select only end-systolic frames in order to reduce the effect of signals from large blood vessels in the myocardium on the information of microvascular blood flow velocity [9, 10].

Segmentation method

We proposed an encoder-decoder architecture for myocardial segmentation [11–13] of MCE sequences as shown in Fig. 1a. The input consisted of a set of RGB frames from an MCE sequence. The output was a sequence of myocardial segmentation of the input. The manual annotations of myocardial region of interests (ROIs) were performed by an experienced echocardiographer, and the time for labeling each image was around 1 min.

Fig. 2 The flow and the process for importing data. *MCE* myocardial contrast echocardiography. *There are 3 chamber-view sequences per patient, and 30 frames per sequence
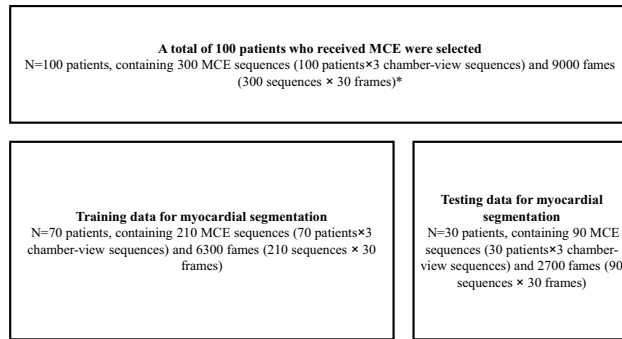
Encoder

The encoder was based on U-net [14–17] which was an effective method for medical image segmentation tasks [18]. The network structure of the encoder was shown in Fig. 1b. We extracted five features at different resolution levels from the up-sampling layers as the output of our encoder. The features were denoted as f1, f2, f3, f4, and f5, each of which had 32, 64, 128, 256, and 512 feature maps, respectively. These features were then fit into the decoder for segmentation. The features at different resolution levels have different receptive filed, which enables the network to learn general and fine-grained information of myocardial at the same time. Thus, the segmentation accuracy can be improved.

Decoder

The network structure of the decoder is shown in Fig. 1c. It contained a hierarchical Convolutional LSTMs [19] which was able to incorporate the input features at different resolutions. The output of the decoder was the segmentation of the myocardium. Inspired by the task of object tracking [20], we incorporated the temporal information of the myocardium in the previous frame as additional features to enhance the segmentation of the current frame. The dash arrows in Fig. 1a depict the temporal recurrence in the decoder. We depicted the features extracted by the encoder for frame t at the level k as $f_{t,k}$ and the output of the k-th ConvLSTM layer for frame t as $y_{t,k}$. $y_{t,k}$ depends on three variables: (1) the output of the previous ConvLSTM layer $y_{t,k-1}$; (2) the extracted features from the encoder $f_{t,k}$; (3) the hidden state of the same ConvLSTM layer for the previous frame $y_{t-1,k}$.

$$h_{input} = \left[Conv\left(f_{t,k}\right) | y_{t,k-1} \right], \tag{1}$$

$$h_{state} = y_{t-1,k}, \tag{2}$$

$$y_{t,k} = ConvLSTM_k\left(h_{input}, h_{state}\right). \tag{3}$$

In Eq. (3), h_{input} is the input of ConvLSTM, and h_{state} is the hidden state input of ConvLSTM. $[Conv(f_{t,k})|y_{t,k-1}]$ is the concatenation operation for matrix $Conv(f_{t,k})$ and $y_{t,k-1}$. $Conv(f_{t,k})$ is a convolutional layer for projecting the feature $f_{t,k}$ into a smaller dimension. For the first frame of an MCE sequence, h_{state} is a zero matrix, which means no prior information is known.

Bi-directional training

We noticed that the prediction results of the myocardium in MCE images is highly dependent on the image quality of the frame. Supplementary Fig. 1 (Online Resource 1) shows some sample frames of MCE of different chamber views. We can see that the image quality in the first frames is usually worse than that of the others. Consequently, the prediction error of these frames may propagate to the rest of the MCE sequences.

In order to alleviate this problem, we introduced a bi-directional training approach. Specifically, after our model was trained from frame 1 to T, we further trained the model from frame T to 1, where T is the total number of frames. Figure 3 presents the workflow of the proposed bi-directional training approach.
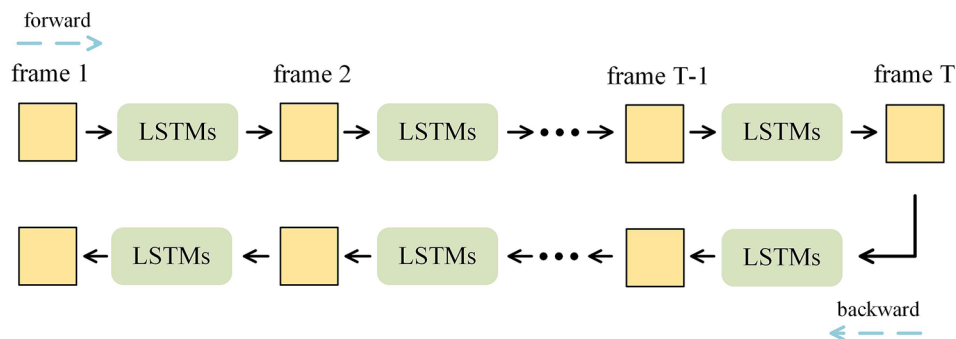
Perfusion quantification

According to the guideline of UEA by the American Society of Echocardiography [2] and [21], time-intensity replenishment curves showed in supplementary Fig. 3 (Online Resource 3) were generated and fitted to the following exponential equation:

$$Y = A \times \left(1 - e^{-\beta t}\right), \tag{4}$$

where Y is the intensity of myocardium at time t after the "flash" impulse, A is the plateau myocardial contrast intensity, and the slope of the replenishment curve depicts mean microbubble velocity β. The myocardial blood flow (*MBF*) can be described as:

Fig. 3 Illustration of our proposed bi-direction training approach. T refers to the total number of frames

$$MBF = A \times \beta. \tag{5}$$

Experiment setup

In this section, we evaluated the performance of our proposed encoder-decoder architecture for the myocardial segmentation task using our collected MCE dataset. We implemented two versions of the proposed framework, (a) training with forward temporal information (denoted as U-net + f-ConvLSTM), (b) training with temporal information from both directions (denoted as U-net + bi-ConvL-STM). U-net was used as our baseline. U-net and ConvL-STM were implemented using Pytorch based on [22] and [19], separately.

The MCE images were centrally cropped into 512×512. For data augmentation during training, we randomly scaled all images by $[0.8, 1.2]$ and rotate them by $[-30°, 30°]$ respectively. We also shifted the image brightness, contrast, and saturation by $[-0.1, 0.1]$. During testing, we did not employ any augmentations. For each iteration, a complete MCE sequence containing 30 frames of a subject was used for training. The batch size was set to 5, which meant in each iteration, 5 MCE sequences containing 150 frames were fed for training. This is the largest batch size in order to train our network on a single GPU. Our experimental results show that increasing the batch size to 10 and training the network on 2 GPUs does not improve the accuracy.

We trained our network for 30 epochs, and the learning rate for the encoder and decoder was set to 0.0001 for the first 15 epochs and 0.00001 for the rest. All the experiments run on an Nvidia GTX 1080Ti GPU with 11 GB memory.

Evaluation

Dice coefficient and Intersection over Union (IoU) were used to evaluate the segmentation performance of our framework, which were defined as:

$$Dice(P, T) = 2 \times \frac{\sum_{(n=1)}^{N} (P_n \times T_n)}{\sum_{(n=1)}^{N} (P_n + T_n)}, \tag{6}$$

$$IoU(P, T) = \frac{\sum_{n=1}^{N} (P_n \times T_n)}{\sum_{n=1}^{N} (P_n + T_n - P_n \times T_n)}, \tag{7}$$

in which P and T refer to the prediction of our encoder-decoder model and ground truth (manual annotation method) mask, respectively. n is the index of pixel space N. Hausdorff distance (HD) was also used for measuring the surface distance between prediction and ground truth boundaries.

An observation study was conducted to further show the segmentation improvement. An independent and experienced echocardiographer was asked to grade results of the testing frames. For each chamber view of each patient, we randomly selected 5 frames for grading, resulting in $3 \times 30 \times 5$ frames in total. A semi-quantitative scoring system was defined based on the segmentation boundary on the output frame whose precision is classified into 4 levels: Level 4, very good; Level 3, with a little deviation; Level 2, existing unregulated shapes; Level 1, with severe disconnection and mistake. The visualization of segmentation results at each level can be seen in supplementary Fig. 4 (Online Resource 4).

Intra-observer and inter-observer variability

The variability of the annotation of myocardial ROI was assessed in 8 randomly selected frames, by comparing the parameters Dice, IoU, HD, A, β, and MBF. The intra-observer variability was assessed in the same frames 3 months apart by one echocardiographer. The inter-observer variability was assessed among the frames by two independent echocardiographers.

Statistical methods

Data were analyzed using R (http://www.R-project.org) and Empower Stats software (http://www.empower.stats.com,

X&Y solutions, Inc., Boston, MA, USA). Kruskal–Wallis rank sum test was used to compare the Dice, IoU, HD, and scores by expert between three models, and if it achieved significant difference, multiple comparisons were carried out. Two-sided p values with $p \leq 0.05$ were considered statistically significant.

The intraclass correlation coefficient (ICC) of a perfect correlation between two variables would equal to 1, and an ICC above 0.8 shows a very strong correlation and between 0.6 and 0.8 a strong correlation [23].

Results

Dice coefficient, IoU, and HD

Table 1 shows the results of myocardial segmentation for A2C, A3C and A4C chamber views. We can notice that U-net + bi-ConvLSTM achieves the higher Dice coefficient, IoU, and lower HD in each chamber view when compared with U-net + f-ConvLSTM method and U-net. When the parameters of all the chamber views are summarized, their differences remain significant (Dice coefficient: 0.81 ± 0.07 vs 0.79 ± 0.07 vs 0.78 ± 0.07, $p < 0.01$; IoU: 0.68 ± 0.09 vs 0.66 ± 0.09 vs 0.65 ± 0.09, $p < 0.01$, HD: 27.59 ± 12.82 vs 28.69 ± 13.18 vs 32.68 ± 14.61 pixels point, $p < 0.01$).

Observation study

As shown in Table 2, there was a statistical difference in the scores among the three models on A2C and A3C. When comparing U-net + bi-ConvLSTM with U-net and U-net + f-ConvLSTM, respectively, the statistical differences remained significant. In terms of total score, U-net + bi-ConvLSTM was statistically higher than that of U-net + f-ConvLSTM and U-net (54.63 ± 4.72 vs 52.50 ± 5.12 vs 50.80 ± 4.32, $p < 0.01$). However, when comparing U-net + f-ConvLSTM and U-net, no statistical differences were found.

Intra-observer and inter-observer variability

For intra-observer variability, the Dice, IoU, and HD of all chamber views were 0.87, 0.77, 11.65 pixels. The correlation coefficients of A, β, and MBF were 0.99 (95% CI 0.92 to 1.00), 0.98 (95% CI 0.90 to 1.00), and 0.99 (95% CI 0.96 to 1.00), respectively. For interobserver variability, Dice, IoU, and HD were 0.88, 0.78, 20.66 pixels, respectively. The correlation coefficients of A, β, and MBF were 0.99 (95% CI 0.94 to 1.00), 0.93 (95% CI 0.65 to 0.99), and 0.94 (95% CI 0.67 to 0.99), respectively.

Table 1 Comparison of segmentation results of different models

	Model 1	Model 2	Model 3	P value	P value*	P value**	P value***
Dice							
A2C	0.80 ± 0.08	0.80 ± 0.08	0.81 ± 0.08	**< 0.01**	0.69	**< 0.01**	**< 0.01**
A3C	0.75 ± 0.11	0.76 ± 0.12	0.78 ± 0.11	**< 0.01**	**0.06**	**< 0.01**	**< 0.01**
A4C	0.80 ± 0.10	0.81 ± 0.09	0.82 ± 0.09	**< 0.01**	**< 0.01**	**< 0.01**	**< 0.01**
Total	0.78 ± 0.07	0.79 ± 0.07	0.81 ± 0.07	**< 0.01**	**0.01**	**< 0.01**	**< 0.01**
IoU							
A2C	0.67 ± 0.10	0.67 ± 0.10	0.69 ± 0.10	**< 0.01**	0.65	**< 0.01**	**< 0.01**
A3C	0.61 ± 0.13	0.62 ± 0.14	0.65 ± 0.13	**< 0.01**	**0.01**	**< 0.01**	**< 0.01**
A4C	0.67 ± 0.12	0.69 ± 0.11	0.71 ± 0.12	**< 0.01**	**< 0.01**	**< 0.01**	**< 0.01**
Total	0.65 ± 0.09	0.66 ± 0.09	0.68 ± 0.09	**< 0.01**	**< 0.01**	**< 0.01**	**< 0.01**
HD (pixel point)							
A2C	28.46 ± 20.36	26.98 ± 19.44	25.79 ± 19.94	**0.02**	0.25	**0.01**	0.41
A3C	42.54 ± 27.90	34.90 ± 24.55	33.01 ± 24.17	**< 0.01**	**< 0.01**	**< 0.01**	0.26
A4C	27.03 ± 18.92	24.19 ± 16.38	23.96 ± 17.73	**< 0.01**	**< 0.01**	**< 0.01**	0.97
Total	32.68 ± 14.61	28.69 ± 13.18	27.59 ± 12.82	**< 0.01**	**< 0.01**	**< 0.01**	0.20

Data are expressed as mean ± SD. Model 1: U-net; Model 2: U-net + f-ConvLSTM; Model 3: U-net + bi-ConvLSTM. The U-net + f-ConvLSTM refers to training our model forwardly from frame 0 to frame 30 while U-net + bi-ConvLSTM refers to training in bi-direction

"P value" represents the comparison among all the three models

A2C apical two-chamber; *A3C* apical three-chamber, *A4C*: apical four-chamber, *HD* hausdorff distance, *IoU* intersection over union

*Represents the comparison between Model 1 and Model 2

**Represents the comparison between Model 1 and Model 3

***Represents the comparison between Model 2 and Model 3

Table 2 Scores of different models on testing data for three chamber views

	Model 1	Model 2	Model 3	P value	P value*	P value**	P value***
A4C (n = 30)	17.93 ± 2.10	18.17 ± 1.76	18.33 ± 1.95	0.73	–	–	–
A2C (n = 30)	17.30 ± 1.84	17.60 ± 1.79	18.67 ± 1.45	**< 0.01**	0.77	**< 0.01**	**0.05**
A3C (n = 30)	15.60 ± 3.15	16.73 ± 3.39	17.63 ± 3.03	**0.05**	0.36	**0.04**	0.52
Total scores (n = 30)	50.80 ± 4.32	52.50 ± 5.12	54.63 ± 4.72	**< 0.01**	0.34	**< 0.01**	0.19

Data are expressed as mean ± SD. The score was obtained by adding the grades of 5 randomly selected frames of each subject. The full score of each frame and whole fames are 4 and 20, respectively

Model 1: U-net; Model 2: U-net + f-ConvLSTM; Model 3: U-net + bi-ConvLSTM

A2C apical two-chamber; *A3C* apical three-chamber, *A4C* apical four-chamber

*Represents the comparison between Model 1 and Model 2

**Represents the comparison between Model 1 and Model 3

***Represents the comparison between Model 2 and Model 3

Case report

Myocardial segmentation and perfusion parameter acquisition were performed on the end-systolic frames of a randomly selected subject in A4C. The end-systolic frames of 8 consecutive cardiac cycles were manually chosen for segmentation. It took 0.8 s for our model to complete the whole segmentation of these frames. Figure 4 shows the visualization of segmentation results by U-net + bi-ConvLSTM and the ground truth. It can be seen that our model can generate myocardial segmentation close to human annotations. Figure 5 shows the time-intensity curve and perfusion parameters using the myocardial segmentation generated by U-net + bi-ConvLSTM and by the ground truth, respectively. The ICC of A, β, and MBF were 0.991 (95%CI 0.946 to 0.998), 1 (0.999 to 1.000), and 0.999 (0.996 to 1.000). We can see that there is little difference between the perfusion parameters generated by using the myocardial segmentation of our model and that of the expert, suggesting that our proposed model has the potential to replace human in myocardial segmentation tasks.

Discussion

To summarize this work, we proposed a framework with temporal consistency for automatic myocardial segmentation. It can generate accurate myocardial segmentation and reduce the manual labor of creating and adjusting ROIs. The experimental results on our collected MCE dataset showed that by fusing bi-directional temporal information into convolutional neural network, we can improve the segmentation performance.

To assess the precision of our model, we evaluated the segmentation results from multiple aspects. In terms of segmentation metrics, it is objectively proved that our model performs well on Dice, IoU, and HD. When compared with the other two baseline models on the observational level, the visual grading results show that there is no statistical difference in the scores of the three models on A4C. This might be because the fewer image artifacts on A4C (reducing the difficulty of boundary recognition) narrowed the difference among models' predictions, which needs further research and confirmation. And in total score, the myocardial segmentation contour generated by our framework is closest to that of experts' expectation.

Methodology comparison with current software

Take Qlab of Philips as an example, it uses a single ROI created on any end-systolic frame in the MCE sequence to cover all end-systolic phase frames, assuming that the shape and position of the heart remain the same at the same phase. However, the structure of the myocardium changes over time in the real world, it is impossible to meet this situation perfectly. In contrast, ROI is generated for each frame by our framework, so the errors of segmentation and parameters caused by the variation of myocardial shape and position could be avoided. In a word, if the dynamic changes of the position of the myocardium are not considered, misjudgment and inaccuracy of myocardial segmentation may occur.

Related work

There are two approaches among existing works towards myocardium segmentation. The first approach is based on point distribution models (PDMs) [24]. A good example is the active shape model (ASM) [25] or active appearance model (AAM) [26]. The main idea of ASM is to learn patterns of variability from a training set of correctly annotated images. ASM uses principal component analysis (PCA) to build a statistical shape model from a set of training shapes and then fits an image in a way that is most similar to the

Fig. 4 Visualization of myo-cardial segmentation results of a randomly selected subject on A4C. The different colors in the output represent different segments of the myocardium. In the third column, the blue and green boundaries refer to the ground truth and prediction respectively. Each row refers to the frame at the end-systolic phase of consecutive cardiac cycles after "flash" impulse
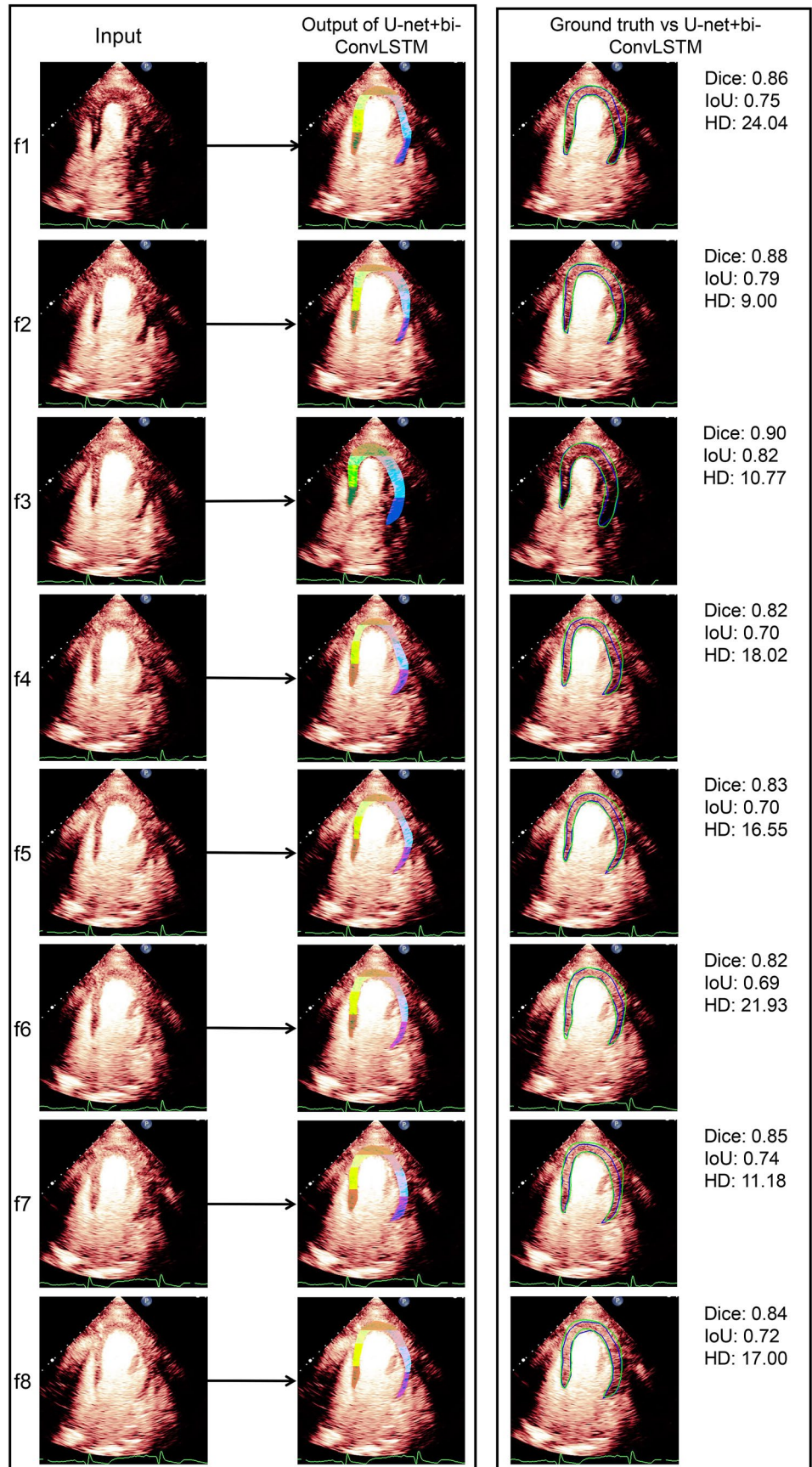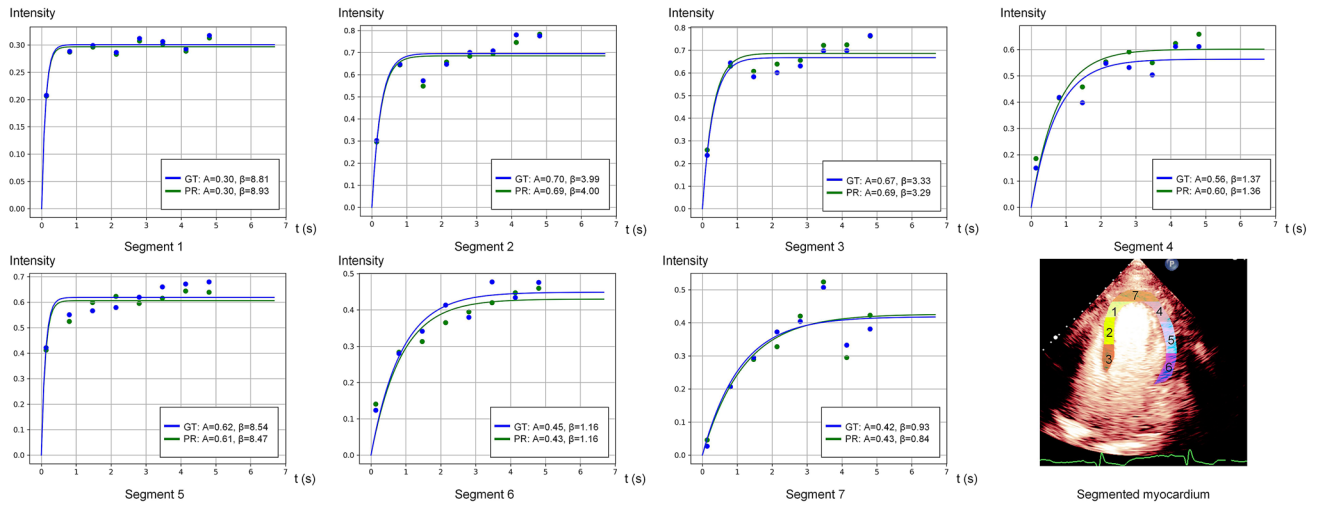
Fig. 5 Measured perfusion parameters by U-net + bi-ConvLSTM against ground truth. **a** the plateau myocardial contrast intensity; β: mean microbubble velocity

statistical shape in the training set [27]. Proposed a real-time algorithm for myocardial and left ventricular segmentation in three-dimensional (3D) cardiac ultrasound based on 3D ASM [28]. Coupled the advantages of ASM and B-spline explicit active surfaces framework and present a fast and automatic left ventricular segmentation and tracking framework [29]. Improved the performance of ASM in four-chamber contrast echocardiography by introducing a gradient vector flow field [30]. Pointed out the difficulty of manual landmarking in 3D Ultrasound images and proposed simulated ultrasound to obtain the training set. The experiment results in vivo data demonstrated the feasibility of the approach. The prediction results of ASM must be constrained into certain shape variations so that the shape of the segmentation result does not go too far from the regular myocardium shape, which is very important when artifacts in the ultrasound image make the myocardium boundary unclear and hard to recognize. However, ASM is based on linear intensity information in the image, which is insufficient to model the appearance of MCE data with huge intensity variations and large artifacts. Besides, ASM requires a manual initialization shape and the final segmentation result is very sensitive to the shape and position of this initialization. Thus, a fully automatic and non-linear model is needed.

The second approach adopts machine learning and deep learning techniques to perform image segmentation. For example [31], used fully convolutional networks for anatomical structure detection and segmentation from ultrasound images. And [32] used structured random forest algorithm to segment myocardium and left ventricular on heterogeneous clinical data [33]. Proposed an ASM based random forest for myocardial segmentation in contrast echocardiography. Although these above methods showed great improvements in the segmentation performance compared to ASM or AAM, they do not take into account the temporal consistency when treating each frame, which may lead to inaccurate segmentation results or the lack of coherence in some sequences.

Notice that work [34] dealt with the same problem and they used Shape Model guided Random Forests to segment myocardium in MCE images. They achieved a Dice of 0.81 ± 0.10 on A4C chamber view at the end-systolic phase of their dataset. Despite the difference in dataset, our model also obtains a Dice of 0.822 ± 0.09 on the whole cardiac cycle.

Clinical perspective

The new framework of automatic myocardial segmentation has several potential clinical applications, such as improving the quantitative analysis workflow of myocardial perfusion, guiding inexperienced users to the expert-level myocardial segmentation with minimal effort and time. On the other hand, the new framework can make the widespread use of MCE possible.

Study limitations

This study of deep learning applied to MCE data has several limitations. First, this study only performed rest MCE cross-verification with a limited number of patients in a single center. A study containing a larger population in external center should be performed. And a study to assess the efficacy of the deep learning algorithm in stress echocardiography is needed. Second, we did not do the backward

experiment, but we expected the performance would be between the forward model and the bi-directional model. Third, a diagnostic test is needed to verify the performance to identify myocardial ischemia by our framework. Forth, we can do the myocardial segmentation for any frame, but the end-systolic frames need to be extracted manually, because we have not developed a program to automatically identify them from the cardiac cycles. Fifth, the images in this study were only from Philips machines using SonoVue as UEA, so caution should be taken when applying this study to other ultrasound vendors and UEAs, due to different image quality. Besides, there was inter- or intra-observer variability that should be noticed.

Conclusions

In this study, we proposed an automatic myocardial segmentation framework for MCE sequences. Experimental results on our collected MCE dataset showed that our proposed framework can produce better myocardial segmentation compared to the baseline methods. Perfusion analysis demonstrated that the myocardial perfusion parameters generated by the myocardial segmentation of our model had a good correlation to the manual annotation by the experienced echocardiographer.

Data availability The data that support the findings of this study are available on request. The data are not publicly available due to the containing information that could compromise the privacy of research participants.

Code availability The codes of this study are available on request.

Compliance with ethical standards

Conflict of interest We declare that we have no significant competing financial, professional, or personal interests that might have influenced the performance or presentation of the work described in this manuscript.

Ethics approval and consent to participate We further confirm that any aspect of the data collection covered in this manuscript that has involved our patients has been conducted with the ethical approval of all relevant bodies and that such approvals are acknowledged within the manuscript.

Consent for publication We guarantee that the contribution to the work has not been previously published elsewhere, or that if it has been published in whole or in part, any permission necessary to publish it in the work has been obtained.

References

1. Knuuti J, Wijns W, Saraste A, Capodanno D, Barbato E, Funck-Brentano C, Prescott E, Storey RF, Deaton C, Cuisset T, Agewall S, Dickstein K, Edvardsen T, Escaned J, Gersh BJ, Svitil P, Gilard M, Hasdai D, Hatala R, Mahfoud F, Masip J, Muneretto C, Valgimigli M, Achenbach S, Bax JJ (2019) 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes. Eur Heart J. https://doi.org/10.1093/eurheartj/ehz425
2. Porter TR, Mulvagh SL, Abdelmoneim SS, Becher H, Belcik JT, Bierig M, Choy J, Gaibazzi N, Gillam LD, Janardhanan R, Kutty S, Leong-Poi H, Lindner JR, Main ML, Mathias W Jr, Park MM, Senior R, Villanueva F (2018) Clinical applications of ultrasonic enhancing agents in echocardiography: 2018 American Society of Echocardiography Guidelines Update. J Am Soc Echocardiogr: Off Publ Am Soc Echocardiogr 31(3):241–274. https://doi.org/10.1016/j.echo.2017.11.013
3. Abdelmoneim SS, Dhoble A, Bernier M, Erwin PJ, Korosoglou G, Senior R, Moir S, Kowatsch I, Xian-Hong S, Muro T, Dawson D, Vogel R, Wei K, West CP, Montori VM, Pellikka PA, Abdel-Kader SS, Mulvagh SL (2009) Quantitative myocardial contrast echocardiography during pharmacological stress for diagnosis of coronary artery disease: a systematic review and meta-analysis of diagnostic accuracy studies. Eur J Echocardiogr: J Work Group Echocardiogr Eur Soc Cardiol 10(7):813–825. https://doi.org/10.1093/ejechocard/jep084
4. Kaufmann BA, Wei K, Lindner JR (2007) Contrast echocardiography. Curr Probl Cardiol 32(2):51–96. https://doi.org/10.1016/j.cpcardiol.2006.10.004
5. Tang MX, Mulvana H, Gauthier T, Lim AK, Cosgrove DO, Eckersley RJ, Stride E (2011) Quantitative contrast-enhanced ultrasound imaging: a review of sources of variability. Interface focus 1(4):520–539. https://doi.org/10.1098/rsfs.2011.0026
6. Ding Y, Jiang W, Lou Q, Liu J, Xiong J, Hu XS, Xu X, Shi Y (2020) Hardware design and the competency awareness of a neural network. Nat Electron 3(9):514–523. https://doi.org/10.1038/s41928-020-00476-7
7. Xu X, Ding Y, Hu SX, Niemier M, Cong J, Hu Y, Shi Y (2018) Scaling for edge inference of deep neural networks. Nat Electron 1(4):216–222. https://doi.org/10.1038/s41928-018-0059-3
8. Cerqueira MD, Weissman NJ, Dilsizian V, Jacobs AK, Kaul S, Laskey WK, Pennell DJ, Rumberger JA, Ryan T, Verani MS (2002) Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart. A statement for healthcare professionals from the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association. Circulation 105(4):539–542. https://doi.org/10.1161/hc0402.102975
9. Leong-Poi H, Le E, Rim SJ, Sakuma T, Kaul S, Wei K (2001) Quantification of myocardial perfusion and determination of coronary stenosis severity during hyperemia using real-time myocardial contrast echocardiography. J Am Soc Echocardiogr: Off Publ

Am Soc Echocardiogr 14(12):1173–1182. https://doi.org/10.1067/mje.2001.115982

10. Leong-Poi H, Swales J, Jayaweera AR, Bin JP, Kaul S, Lindner JR (2005) Effect of microbubble exposure to ultrasound on quantitation of myocardial perfusion. Echocardiography (Mount Kisco, NY) 22(6):503–509. https://doi.org/10.1111/j.1540-8175.2005.40001.x

11. Zhang J, Zhang Y, Zhu S, Xu X (2020) Constrained multi-scale dense connections for accurate biomedical image segmentation. In: 2020 IEEE international conference on bioinformatics and biomedicine (BIBM), pp 877–884. https://doi.org/10.1109/BIBM49941.2020.9313254

12. Xu X, Wang T, Zhuang J, Yuan H, Huang M, Cen J, Jia Q, Dong Y, Shi Y (2020) ImageCHD: a 3D computed tomography image dataset for classification of congenital heart disease. In: Medical image computing and computer assisted intervention – MICCAI 2020. Springer, Cham, pp 77–87. https://doi.org/10.1007/978-3-030-59719-1_8

13. Wang T, Xu X, Xiong J, Jia Q, Yuan H, Huang M, Zhuang J, Shi Y (2020) ICA-UNet: ICA inspired statistical Unet for real-time 3D cardiac cine MRI segmentation. International conference on medical image computing and computer-assisted intervention, 2020, pp. 447–457. https://doi.org/10.1007/978-3-030-59725-2_43

14. Liu Z, Li S, Chen Y-k, Liu T, Liu Q, Xu X, Shi Y, Wen W (2020) Orchestrating medical image compression and remote segmentation networks. In: Medical image computing and computer assisted intervention – MICCAI 2020. Springer, Cham, pp 406–416. https://doi.org/10.1007/978-3-030-59719-1_40

15. Ding Y, Liu J, Xu X, Huang M, Zhuang J, Xiong J, Shi Y (2020) Uncertainty-aware training of neural networks for selective medical image segmentation. In: Medical imaging with deep learning. PMLR, pp 156–173

16. Wang T, Xiong J, Xu X, Jiang M, Yuan H, Huang M, Zhuang J, Shi Y (2019) MSU-net: multiscale statistical U-net for real-time 3D cardiac MRI video segmentation. In: Medical image computing and computer assisted intervention – MICCAI 2019. Springer, Cham, pp 614–622. https://doi.org/10.1007/978-3-030-32245-8_68

17. Xu X, Wang T, Shi Y, Yuan H, Jia Q, Huang M, Zhuang J (2019) Whole heart and great vessel segmentation in congenital heart disease using deep neural networks and graph matching. In: Medical image computing and computer assisted intervention – MICCAI 2019. Springer, Cham, pp 477–485. https://doi.org/10.1007/978-3-030-32245-8_53

18. Ronneberger O, Fischer P, Brox T U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Cham, 2015. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer International Publishing, pp 234–241

19. Shi X, Chen Z, Wang H, Yeung D-Y, Wong WK, Woo W-c (2015) Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting.

20. Xu N, Yang L, Fan Y, Yang J, Yue D, Liang Y, Price B, Cohen S, Huang T Youtube-vos (2018) Sequence-to-sequence video object segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV) 585–601

21. Wei K, Jayaweera AR, Firoozan S, Linka A, Skyba DM, Kaul S (1998) Quantification of myocardial blood flow with ultrasound-induced destruction of microbubbles administered as a constant venous infusion. Circulation 97(5):473–483. https://doi.org/10.1161/01.cir.97.5.473

22. Isensee F, Petersen J, Klein A, Zimmerer D, Jaeger PF, Kohl S, Wasserthal J, Koehler G, Norajitra T, Wirkert S, Maier-Hein KH Abstract: nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation. In: Wiesbaden, 2019. Bildverarbeitung für die Medizin 2019. Springer Fachmedien Wiesbaden, pp 22–22

23. Hallgren KA (2012) Computing inter-rater reliability for observational data: an overview and tutorial. Tutorials Quantitat Methods Psychol 8(1):23–34. https://doi.org/10.20982/tqmp.08.1.p023

24. Tobon-Gomez C, Butakoff C, Aguade S, Sukno F, Moragas G, Frangi AF (2008) Automatic construction of 3D-ASM intensity models by simulating image acquisition: application to myocardial gated SPECT studies. IEEE Trans Med Imaging 27(11):1655–1667. https://doi.org/10.1109/tmi.2008.2004819

25. Cootes TF, Taylor CJ, Cooper DH, Graham J (1995) Active shape models-their training and application. Comp Vision Image Understand 61(1):38–59. https://doi.org/10.1006/cviu.1995.1004

26. Cootes TF, Edwards GJ, Taylor CJ (2001) Active appearance models. IEEE Trans Pattern Anal Machine Intelligence 23(6):681–685. https://doi.org/10.1109/34.927467

27. Hansegård J, Orderud F, Rabben SI (2007) Real-Time Active Shape Models for Segmentation of 3D Cardiac Ultrasound. Berlin, Heidelberg. Computer Analysis of Images and Patterns. Springer, Berlin Heidelberg, pp 157–164

28. Pedrosa J, Queiros S, Bernard O, Engvall J, Edvardsen T, Nagel E, D'Hooge J (2017) Fast and fully automatic left ventricular segmentation and tracking in echocardiography using shape-based B-spline explicit active surfaces. IEEE Transact Med Imaging 36(11):2287–2296. https://doi.org/10.1109/tmi.2017.2734959

29. Pickard JE, Hossack JA, Acton ST Shape model segmentation of long-axis contrast enhanced echocardiography. In: 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006, 6–9 April 2006 2006. pp 1112–1115. doi:https://doi.org/10.1109/ISBI.2006.1625117

30. Butakoff C, Balocco S, Ordas S (2007) Simulated 3D ultrasound LV cardiac images for active shape model training, vol 6512. SPIE, Medical Imaging

31. Chen H, Zheng Y, Park J-H, Heng P-A, Zhou SK Iterative Multidomain Regularized Deep Learning for Anatomical Structure Detection and Segmentation from Ultrasound Images. In, Cham, 2016. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. Springer International Publishing, pp 487–495

32. Leclerc S, Grenier T, Espinosa F, Bernard O A fully automatic and multi-structural segmentation of the left ventricle and the myocardium on highly heterogeneous 2D echocardiographic data. In: 2017 IEEE International Ultrasonics Symposium (IUS), 6–9 Sept. 2017 2017. pp 1–4. https://doi.org/https://doi.org/10.1109/ULTSYM.2017.8092797

33. Li Y, Ho CP, Toulemonde M, Chahal N, Senior R, Tang MX (2018) Fully automatic myocardial segmentation of contrast echocardiography sequence using random forests guided by shape model. IEEE Trans Med Imaging 37(5):1081–1091. https://doi.org/10.1109/tmi.2017.2747081

34. Li Y, Ho CP, Chahal N, Senior R, Tang M-X Myocardial Segmentation of Contrast Echocardiograms Using Random Forests Guided by Shape Model. In: Cham, 2016. Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016. Springer International Publishing, pp 158–165

Authors and Affiliations

Mingqi Li[1,2] · Dewen Zeng[3] · Qiu Xie[1] · Ruixue Xu[1] · Yu Wang[1,2] · Dunliang Ma[1] · Yiyu Shi[3] · Xiaowei Xu[4] · Meiping Huang[5] · Hongwen Fei[1]

[1]   Department of Cardiology, Guangdong Cardiovascular Institute, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China

[2]   Shantou University Medical College, Shantou, Guangdong, China

[3]   Department of Computer Science and Engineering, University of Notre Dame, South Bend, IN, USA

[4]   Guangdong Provincial Key Laboratory of South China Structural Heart Disease, Guangdong Cardiovascular Institute, Guangdong Provincial People's Hospital,

Guangdong Academy of Medical Sciences, Guangzhou, China

[5]   Department of Catheterization Lab, Guangdong Provincial Key Laboratory of South China Structural Heart Disease, Guangdong Cardiovascular Institute, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China