Weakly Supervised Deformation Network for 3D Echocardiography Segmentation on Left Ventricle

Suyu Dong[1], Gongning Luo [1], Naren Wulan [1], Shaodong Cao[2], Kuanquan Wang[1], Henggui Zhang[1,3]

[1]Harbin Institute of Technology, Harbin, China
[2]The Fourth Hospital of Harbin Medical University, Harbin, China
[3]University of Manchester, Manchester, UK

Abstract

The automated 3D echocardiography segmentation on left ventricle (LV) is very important for clinical evaluation of LV function. However, the segmentation is difficult due to the 3D echocardiography's challenges, such as the low signal-to-noise ratio, indistinguishable boundaries between LV and other heart substructures, and limited annotation data. This paper aims to propose a novel method to achieve accurate 3D echocardiography segmentation on LV, based on a weakly supervised deformable network. The deformation network was optimized by generative adversarial constraint and volume similarity constraint. The proposed framework was trained and validated on 3D echocardiography datasets which including 70 patients (35 train patients and 35 test patients). The results demonstrated the proposed method is relatively accurate and has potential for further research and application.

1.   Introduction

Left ventricle (LV) function is important for the diagnosis and treatment of cardiac disease. Echocardiography is the most widely used imaging modality for assessing the cardiac function in clinical application. Hence, there are many researchers pay attention to echocardiography segmentation on LV, especially for 3D echocardiography, which contains more information comparing with 2D echocardiography.

However, 3D echocardiography segmentation on LV is still a challenging task because of the inherent problems of 3D echocardiography: (1) low signal-to-noise ratio, which makes the low contrast between LV boundaries and other areas; (2) indistinguishable boundaries between LV and other heart substructures, which makes achieving no leakage segmentation difficult; (3) limited annotation data, which makes achieving accurate segmentation difficult.

Up to now, researchers have proposed many methods to segment the LV from echocardiography. There are many traditional segmentation methods, such as deformable methods, statistical methods, machine learning methods. Deformable methods are the most widely used methods, which achieves segmentation by optimizing energy function. These methods can be classified as snake-based models and level set-based methods. [1] proposed a fast adaptive B-spline snake, which combines external forces, adaptive node insertion, and multiresolution strategy, and avoids the expensive computation. [2] used an improved level set method to achieve segmentation. The proposed method was robust to attenuation artifact by using local phase and local orientation information. They also used Cauchy kernels to extract features. [3] proposed pSnakes, which used the 1D Hilbert transform as external energy, to segment left ventricle. The commonly used statistical methods are active shape model and active appearance model, which all need a large number of labelled information from experts. Active shape model is based on the statistical shape model, and active appearance model also integrates texture model into the statistical shape model. [4] combined active shape model and active contour model to achieve the left ventricle segmentation. They used PCA to achieve shape representation of left ventricle so that there was no need of the difficult determination of corresponding landmarks. [5] proposed a fast and fully automatic LV segmentation of the whole heart cycle. The proposed method combined B-spline explicit active surfaces framework and statistical shape models to achieve accurate segmentation. [6] used active appearance model, which integrated local edge detector to the active appearance model, to segment left ventricle on multi-view and multi-frame. Machine learning methods are based on a large number of handcrafted features and also used for LV segmentation. [7] combined structured random forests and active shape model to achieve segmentation of left ventricle. [8] used random forest, which integrated a statistical shape model to provide prior knowledge to guide the random forest, to achieve myocardial segmentation. The deformable methods depend on initialization seriously. The statistical methods and machine learning methods all need a great deal of
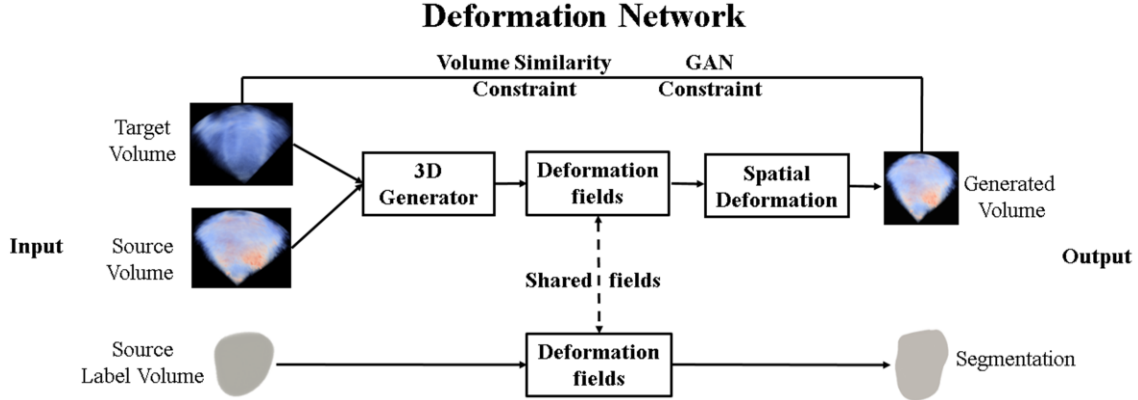
Deformation Network

Figure 1. The proposed weakly supervised deformation network for 3D echocardiography segmentation on left ventricle.

manpower and material resources, which is difficult for medical researches. Different from these traditional methods, deep learning methods, which can extract features automatically, have been successfully used for cardiac research [9]. [10] used a dynamic convolutional neural networks to segment fetal LV. The proposed method also used block matching and line matching method to separate the left ventricle and left atrium. [11] combined convolutional neural network and snake to segment LV. They used convolutional neural network to achieve coarse segmentation and then used snake to refine the segmentation results. However, deep learning methods depend on a large number of training data, which is difficult for medical image, especially for 3D echocardiography.

In this paper, we proposed a novel and weakly supervised deformable network to achieve accurate 3D echocardiography segmentation on LV. The proposed method is three-fold: (1) For the first time, we used weakly supervised deformation network to achieve 3D echocardiography segmentation on LV. (2) We designed a generative adversarial network (GAN) to achieve weakly supervised segmentation. (3) Volume similarity constraint and GAN constraint were jointed to optimize the proposed frame.

2.   Methods

The proposed method used weakly supervised deformation network to achieve 3D echocardiography segmentation on LV. In this paper, the weakly supervised deformation network means that there is no need of labels to train the deformation network. In the training process, we only discriminate the consistency of the target volume and the generated volume. There is no need of any annotation. As shown in Figure 1, the proposed method was based on a generative adversarial network (GAN). It was jointly optimized by GAN constraint and volume similarity constraint.

2.1.   Deformation network architecture

In this paper, we used 3D conditional generative adversarial network [12] as the base architecture of the deformable network. The generator and discriminator networks can be other network structure. The 3D generator was used for regression of deformation fields. In the training process, the generated volume can be acquired by spatial deformation basing on the deformation fields. And the generator is optimized by the volume similarity constraint and GAN constraint. In the testing process, the segmentation can be acquired by the spatial deformation basing the shared deformation fields, which is common field between the source volume and target volume, as well as the source label and segmentation.

In this paper, the 3D generator was an encoder-decoder structure, and had been shown in Figure.2. It uses a single volume as the input of the network. The single volume was constructed by concatenating the source volume and target volume. In the encoder phase, every 3D convolution layer was also followed by a residual module and a max pooling layer. The receptive field sizes of max pooling layer were 2*2*2 voxels, and the strides was 2. The decoder structure was constructed by 3D convolution layers and 3D deconvolution layers. The receptive field sizes of every deconvolution layer was 2*2*2. Every convolution layer and deconvolution layer were followed by a BN layer and Relu activation layer.

The discriminator was a fully convolutional network. The inputs of the discriminator are target volume and generated volume. The structure of the network had five 3D convolution layers. Except the last one convolution layer, every convolution layer is also followed by a BN layer and Relu activation layer. The last convolution layer was followed by a sigmoid activation layer. The receptive field sizes of every convolution layer and max pooling layer were also 3*3*3 voxels and 2*2*2 voxels, and the stride is 2. And the discriminator output the judgement for the input.

Spatial deformation was based on spatial transformer networks [13]. In this paper, we used the spatial transformation function in [14] and linear interpolation to
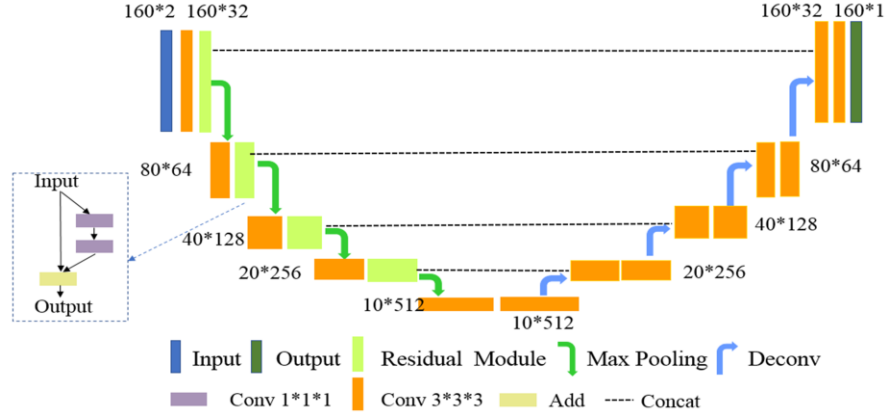
Figure 2. The network of 3D generator.

achieve volume's deformation. The spatial transformation was differentiable. In the testing process, the generated deformation fields were used for the segmentation based on deformation of source label volume.

2.2. Optimization

As shown in Figure 1, GAN constraint and volume similarity constraint were used to jointly optimize the proposed framework using stochastic gradient descent algorithm. The final optimization object is

$$L = \alpha L_{volume} + \beta L_{image-GAN} \qquad (1)$$

Volume similarity constraint was used to measure the similarity of generated volume and target volume. In this paper, we used the normalized mutual information to model the volume similarity constraint, which has high robustness for noisy and high intensity variability. Specifically, the volume similarity constraint is

$$L_{volume} = -\big(H(X) + H(X')\big)/H(X, X') \qquad (2)$$

where X and X' denote the target volume and generated volume, H(X) and H(X') denote the marginal entropies of X and X' respectively, H(X, X') denotes the joint entropies of X and X'.

GAN constraint was based on image-conditional GAN, which extracts high frequency features to penalize any inconformity between generated volume and target volume. It means that GAN constraint takes advantage of high frequency features to further refine the generated volume. Specially, the GAN constraint is

$$\underset{G}{MIN}\ \underset{D}{MAX}\ \mathrm{L}(G, D)$$

$$L(G, D) = E_X[log D(X)] + E_{X'}\Big[log\big(1 - D\big(G(X)\big)\big)\Big] (3)$$

where G denotes generator and D denotes the discriminator.

2.3. Segmentation

As the atlas method, which is a segmentation method

based on registration, we used this strategy to achieve our segmentation. In this paper, we used the source volume as the atlas volume and the label of source volume as atlas label. Hence, the deformation fields of target volume and source volume were also the deformation fields of the segmentation of target volume and label of source volume. In the testing process, the deformation fields can be regressed by trained deformation network. Then, based on these shared deformation fields and spatial deformation strategy, we achieve segmentation by deforming label of source volume.

3. Results

3.1. Datasets

In this paper, we used 35 training subjects and 35 testing subjects. Every subject contains end-systole volume (ES) and end-diastole volume (ED), and all these volumes were labeled. The datasets were acquired by Philips iE33 using X7-2T. Each volume was normalized into 160*160*160.

3.2. Metrics

In this paper, we used the evaluation metrics in [15] to evaluate the proposed method to compare accuracy of the proposed method and ground truth. Specially, the evaluation metrics contain mean surface distance (dM), hausdorff surface distance (dH), modified dice similarity index (D*).

3.3. Experiments results and Discussion

Table 1 showed the segmentation accuracy comparison between the proposed method and ablation methods on mean surface distance, hausdorff surface distance and modified dice similarity index. The statistical results demonstrate that the proposed method achieved good

segmentation. Comparing with the ablation methods, the segmentation result of the proposed method demonstrated that the proposed volume similarity constraint and GAN constraint was appropriate. Volume similarity was mainly used to measure the intensity similarity, and it focused on

Table 1. The segmentation accuracy comparison between the proposed method (PM) and ablation methods on mean surface distance, hausdorff surface distance and modified dice similarity index. (PM-GAN denotes the proposed method without GAN constraint, PM-VS denotes the proposed method without volume similarity constraint.)

	dM (mm)	dH (mm)	D* (val)
PM-GAN	2.50±0.37	8.95±5.2	0.181
PM-VS	2.35±0.40	8.67±4.7	0.106
PM	1.87±0.39	7.8±4.9	0.094

voxel-level similarity. GAN constraint was used to measure anatomic structure similarity and intensity similarity, it focused on the global similarity and high frequency information. Combining volume similarity constraint and GAN constraint was better for measuring the similarity between the target volume and generated volume from different levels.

4.   Conclusion

In this paper, we proposed a weakly supervised deformable network to achieve 3D echocardiography segmentation of LV. We used generative adversarial network to construct the deformation network. The deformation network was jointly optimized by volume similarity constraint and GAN constraint. The results demonstrated the proposed method is relatively accurate and has potential for further research and application.

Acknowledgements

References

[1] M. Marsousi, A. Eftekhari, A. Kocharian, and J. Alirezaie, "Endocardial Boundary Extraction in Left Ventricular Echocardiographic Images Using Fast and Adaptive B-Spline Snake Algorithm," International Journal Of Computer Assisted Radiology And Surgery, vol. 5, no. 5, pp. 501-513, Sep 2010.

[2] A. Belaid, D. Boukerroui, Y. Maingourd, and J. F. Lerallut, "Phase-Based Level Set Segmentation of Ultrasound Images," IEEE Transactions on Information Technology In Biomedicine, vol. 15, no. 1, pp. 138-147, Jan 2011.

[3] A. R. De Alexandria, P. C. Cortez, J. A. Bessa, J. H. D. Felix, J. S. de Abreu, and V. H. C. de Albuquerque, "PSnakes: A New Radial Active Contour Model and Its Application in the Segmentation of The Left Ventricle From Echocardiographic Images," Computer Methods And Programs In Biomedicine, vol. 116, no. 3, pp. 260-273, Oct 2014.

[4] G. Hamarneh and T. Gustavsson, "Combining Snakes and Active Shape Models for Segmenting the Human Left Ventricle in Echocardiographic Images," Computers In Cardiology 2000, Vol 27, vol. 27, pp. 115-118, 2000.

[5] J. Pedrosa et al., "Fast and Fully Automatic Left Ventricular Segmentation and Tracking in Echocardiography Using Shape-Based B-Spline Explicit Active Surfaces," IEEE Transactions on Medical Imaging, vol. 36, no. 11, pp. 2287-2296, Nov 2017.

[6] J. Hansegard, S. Urheim, K. Lunde, and S. I. Rabben, "Constrained Active Appearance Models for Segmentation of Triplane Echocardiograms," IEEE Trans Med Imaging, vol. 26, no. 10, pp. 1391-400, Oct 2007.

[7] F. Khellaf, S. Leclerc, J. D. Voorneveld, R. S. Bandaru, J. G. Bosch, and O. Bernard, "Left Ventricle Segmentation in 3D Ultrasound by Combining Structured Random Forests With Active Shape Models," Medical Imaging 2018: Image Processing, vol. 10574, 2018.

[8] Y. W. Li, C. P. Ho, M. Toulemonde, N. Chahal, R. Senior, and M. X. Tang, "Fully Automatic Myocardial Segmentation of Contrast Echocardiography Sequence Using Random Forests Guided by Shape Model," IEEE Transactions on Medical Imaging, vol. 37, no. 5, pp. 1081-1091, May 2018.

[9] G. N. Luo, S. Y. Dong, K. Q. Wang, W. M. Zuo, S. D. Cao, and H. G. Zhang, "Multi-Views Fusion CNN for Left Ventricular Volumes Estimation on Cardiac MR Images," IEEE Transactions on Biomedical Engineering, vol. 65, no. 9, pp. 1924-1934, Sep 2018.

[10] L. Yu, Y. Guo, Y. Y. Wang, J. H. Yu, and P. Chen, "Segmentation of Fetal Left Ventricle in Echocardiographic Sequences Based on Dynamic Convolutional Neural Networks," IEEE Transactions on Biomedical Engineering, vol. 64, no. 8, pp. 1886-1895, Aug 2017.

[11] S. Y. Dong, G. N. Luo, K. Q. Wang, S. D. Cao, Q. C. Li, and H. G. Zhang, "A Combined Fully Convolutional Networks and Deformable Model for Automatic Left Ventricle Segmentation Based on 3D Echocardiography," Biomed Research International, 2018.

[12] P. Isola, J. Y. Zhu, T. H. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," 30th IEEE Conference on Computer Vision And Pattern Recognition, pp. 5967-5976, 2017.

[13] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial Transformer Networks," Advances In Neural Information Processing Systems (NIPS 2015), vol. 28, 2015.

[14] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "VoxelMorph: A Learning Framework for Deformable Medical Image Registration," IEEE Transactions on Medical Imaging, vol. 38, no. 8, pp. 1788-1800, Aug 2019.

[15] O. Bernard et al., "Standardized Evaluation System for Left Ventricular Segmentation Algorithms in 3D Echocardiography," IEEE Trans Med Imaging, Nov 25

2015.

Address for correspondence.
Henggui Zhang
Henggui.zhang@manchester.ac.uk
Harbin Institute of Technology, Harbin, China