A Novel Spatio-Temporal Self-Supervised Framework to Improve the Generalization Ability for Left Ventricle Volume Quantification Based on CMR Data

Gongning Luo[1], Kuanquan Wang[1]*, Naren Wulan[1], Shaodong Cao[2], Qince Li[1], Yongfeng Yuan[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 quantification of the left ventricular (LV) volume on MRI is a crucial step for cardiac disease diagnosis. In recent years, deep learning (DL) technology has been widely used in the field of ventricle quantification and achieves relatively higher quantification accuracy. However, the LV volume quantification is still a challenging task, mainly because of limited labelled data. Hence, this study aims to propose an innovative approach to achieve accurate LV volume estimation based on limited labelled data.. The proposed method is three-fold: (1) For the first time, we proposed a self-supervised framework to model the significant spatio-temporal correlation information of adjacent slices from CMR images. (2) We designed a deep learning network based on spatio-temporal ranking loss to achieve self-supervised training utilizing large-scale unlabelled CMR data. (3) An iterative optimization strategy was developed to achieve efficient model optimization. The deep learning network was trained and validated on cardiac MRI datasets from MICCAI 2012 LV segmentation challenge including 100 patients (50 training patients and 50 test patients).

1. Introduction

The cardiac disease is the leading killer of human beings. The left ventricle (LV) function is critical to the heart. Hence, automated LV quantification task gets more and more attention in recent years [1-4]. What's more, considering the strong feature representation ability, most of the recent methods are based on deep learning technology.

The LV volume (LVV) is an important index for cardiac disease diagnosis. Most of the LV function quantification studies focus on addressing automated LVV estimation problem. Based on accurate LVV index, the eject fraction (EF) can be got naturally, which is crucial for accurate cardiac disease diagnosis.

Before 2014, the image segmentation technology is always used for addressing LVV estimation problem. The LV segmentation technology can be classified as four kinds: active contour model, atlas-based segmentation method, and the method based on deep learning [5]. However, segmentation-based technology is a two-stage volume estimation method. The accurate volume estimation relies on accuracy segmentation. But automated and accurate segmentation is still an unsolved problem. Hence, some researchers try to develop a fully new volume estimation method (i.e., direct volume estimation method).

After 2014, Zhen et al. proposed a direct ventricle volume estimation method based on regression forests method [6]. Inspired by the good feature representation ability of deep learning technology, Zhen et al further combined the multi-scale deep networks and regression forests into a unified framework to achieve direct ventricle volume estimation [7]. Wang et al proposed an direct ventricle volume framework using the Bayesian estimation method [8], which is the first direct volume estimation method based on traditional statistical learning method. Li et al. proposed four-chamber volume estimation method by multi-task regression method [9, 10]. Luo et al. present a new data input style based on multi-view combination deep convolutional network to achieve good ventricle volume estimation [2, 11]. [2] also shows the potential of LV volume estimation based on a single slice or small number of slices.

Though a large number of methods were proposed to address LVV estimation problem on CMR image, it is still an open challenge, due to the limited labelled CMR images. To our best knowledge, the state-of-the-art deep learning methods still rely on large numbers of labelled training data. Manually labelling large number of training data is time-consuming. Hence, how to achieve good quantification performance without relying on large

 ISSN: 2325-887X DOI: 10.22489/CinC.2019.146

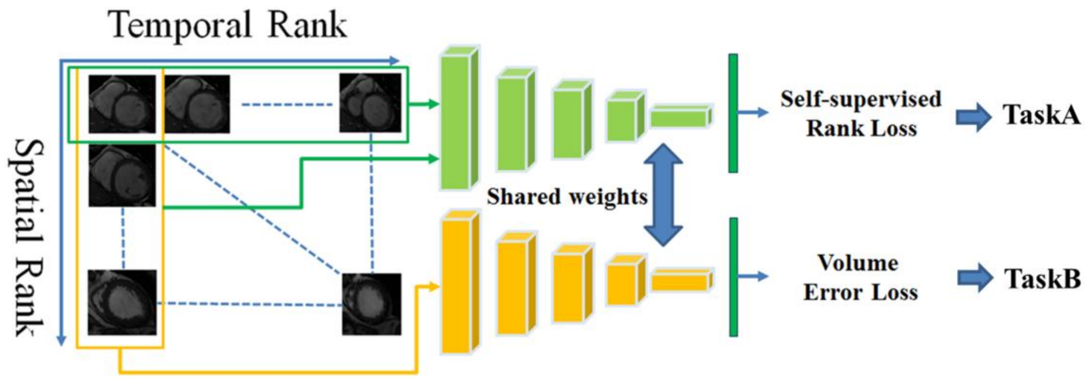

Figure 1. The framework of the proposed method

number of labelled training data is an important problem urgently to be solved. In this paper, utilizing the spatio-temporal correlation information of adjacent slices from CMR images, we proposed self-supervised deep learning framework to address automated LV volume estimation problem. This framework may be insightful and can be extended to other medical image analysis task. Besides, we conducted related experiments on the large scale CMR dataset (i.e., MICCAI 2012 LVSC) [12]. The results show that the proposed method has potential to be further studied to address deep learning task based on limited data.

2. Methods

2.1. Data Selection

We selected the short-axis CMR image series as the input of the proposed framework. The CMR image preprocessing procedure follows our previous research in [2]. The CMR image series include spatial image series and temporal image series. The spatial image series are shown as following:

$$S = \left[A_{1,N}, A_{2,N}, \bullet, \bullet, \bullet, A_{i,N}, \bullet, \bullet, A_{M,N} \right] \quad (1)$$

where $A_{i,N}$ denotes the spatial ranked slices on moment N, and i denotes spatial slice index. The temporal image series are shown as following:

$$T = \left[A_{M,1}, A_{M,2}, \bullet, \bullet, \bullet, A_{M,j}, \bullet, \bullet, A_{M,N} \right] \quad (2)$$

where $A_{M,j}$ denotes the temporal ranked slices on location M, and j denotes temporal slice index.

2.2. Network Structure

As shown in Figure 1, the proposed framework includes four parts: The data selection module, the shared encoder network (The detailed structure is similar to VGG16 [13]), the multi-task output module for ranking and LVV estimation, and the optimization module with spatio-temporal self-supervised ranking loss and volume estimation loss. Specifically, the proposed framework is two-path network based on weight-shared backbone network. One can think that such an shared backbone network as two-path network structure in the stage of forward propagation, and think that the weights of network are optimized jointly in the stage of backpropagation.

We divided the proposed framework into two tasks, i.e., ranking task (Task A) and volume estimation task (Task B). In the stage of forward propagation, the spatial image series S and the temporal image series T are disorganized and used as the input of task A, and the spatial image series S is used as the input of task B. The two kinds of loss functions (i.e., Self-supervised rank loss and volume error loss) are used to optimize the shared weights based on Task A and Task B respectively.

Inspired by the fact that a subject's CMR image sequences are ordered naturally, we proposed a novel spatio-temporal self-supervised ranking loss. In space, the CMR image sequences are ranked from base to apex on a ventricle geometry model. In the time order, the CMR image sequences are ranked from end-diastole (ED) to end-systole (ES). Hence, the self-supervised ranking loss is modeled as:

$$SRL = \left\| S_e - S_t \right\| + \left\| T_e - T_t \right\| \quad (3)$$

where SRL denotes the self-supervised ranking loss, S_e denotes the estimated ranking vector in space order, S_t denotes the ground truth ranking vector in space order, T_e denotes the estimated ranking vector in the time order, and T_t denotes the ground truth ranking vector in the time order. The volume estimation loss is modeled as:

$$VEL = \left\| V_e - V_t \right\| \quad (4)$$

where VEL denotes the volume estimation loss, V_e denotes the estimated volume, and V_t denotes the ground truth ground truth volume.

2.3. Model Optimization

We proposed the two optimization strategies (i.e., iterative optimization pattern (IOP) and joint optimization pattern (JOP)) to achieve effective model optimization. As shown in Algorithm 1, the IOP is achieved through iteratively computing the *SRL* and *VEL* based on series <T,S> and series <S> respectively. As shown in Algorithm 2, the JOP is achieved through jointly computing the *SRL* and *VEL* based on series <T,S> and series <S> respectively. Finally, the shared weights of network are updated through backpropagation algorithm.

Algorithm 1:IOP

```
Input: CMR image series T and S;
Initializing parameters including N, MaxIterations
For i=1 to N do:
  If M< MaxIterations then:
     Computing the SRL based on image series T and S
     Update network through backpropagation algorithm
  End
  Else
     Termination of training
  End
  M=M+1
End
M=0
For i=1 to N do:
  If M< MaxIterations then:
     Computing the VEL based on image series S
     Update network through backpropagation algorithm
  End
  Else
     Termination of training
  End
  M=M+1
End
```

Algorithm 2:JOP

```
Input: CMR image series T and S;
Initializing parameters including N, MaxIterations
For i=1 to N do:
  If M< MaxIterations then:
     Computing the SRL based on image series T and S
     Computing the VEL based on image series S
     Update network through backpropagation algorithm
  End
  Else
     Termination of training
  End
  M=M+1
End
```

3. Results and Discussion

3.1. Metrics

The proposed method was evaluated by the widely used criterion in the LVV estimation field, including correlation coefficient (R), mean absolute errors (MAE) ± the standard deviation (SD) [2].

3.2. Environment of Experiments

We conducted the experiments in Keras deep learning framework [14], which is widely used for building a deep learning application. The workstation's configurations are 3.3GHz Core i5 CPU, 128GB RAM, Nvidia 1080 Ti (11GB memories). The *MaxIteration* is set as 10000, and the number of training subjects is 100, and the Gaussian weight initializations are used to achieve the good initialization of the proposed framework. We divided the MICCAI 2012 LVSC data [12] into 100 subjects for training, 50 subjects for validation, and 50 subjects for testing.

3.3. Prediction Accuracy

The LVV estimation performance and the clinical correlation coefficient index has been shown in the Table 1. Compared with the state-of-the-art LVV estimation methods, the proposed method achieves the relatively better estimation accuracy in the aspect of End-diastole (ED) volume and End-systole (ES) volume. Additionally, the models trained by two training patterns have different performance. IOP has better LVV estimation accuracy on ES phase, and JOP has better LVV estimation accuracy on ED phase. This result also proves that the proposed method can utilize the inherent ranking relationship between adjacent CMR frames on spatial dimension and temporal dimension to generate more robust and accurate LVV estimation results.

Table1. The comparison with the state-of-the-art methods among different training pattern on volume index quantification performance.

Method	LV_ED		LV_ES	
	R	MAE (ml)	R	MAE (ml)
IOP	0.83	10.1±5.2	**0.93**	**8.5±2.5**
JOP	**0.92**	**9.5±3.2**	0.92	9.1±3.2
[2]	**0.92**	14.4±5.5	**0.93**	10.2±7.2
[7]	0.85	15.6±5.2	0.9	13.3±4.5
[8]	0.9	12.3±7.7	0.88	14.1±3.9

4. Conclusion

In the paper, we proposed a novel spatio-temporal self-supervised deep learning framework to address the LV volume estimation problem. It introduces the spatio-temporal self-supervised relationship into end-to-end LV volume estimation framework though a novel self-supervised loss for the first time. This framework can use the spatio-temporal self-supervised ranking relationship between adjacent CMR slices to improve good accuracy and robustness on volume estimation results. The experiment results on the open accessible CMR datasets prove the proposed method has the big potential to be extended into wider application field in the clinical cardiac disease diagnosis.

Acknowledgements

This work was supported by the National Key R&D Program of China under Grant 2017YFC0113000

References

[1] G. N. Luo, G. X. Sun, K. Q. Wang, S. Y. Dong, and H. G. Zhang, "A Novel Left Ventricular Volumes Prediction method based on deep learning network in cardiac MRI," presented at the Computing in Cardiology, 2016.

[2] G. Luo, S. Dong, K. Wang, W. Zuo, S. Cao, and H. Zhang, "Multi-views Fusion CNN for Left Ventricular Volumes Estimation on Cardiac MR Images," IEEE Transactions on Biomedical Engineering, vol. PP, no. 99, pp. 1-1, 2018.

[3] G. N. Luo, R. An, K. Q. Wang, S. Y. Dong, and H. G. Zhang, "A Deep Learning Network for Right Ventricle Segmentation in Short-axis MRI," presented at the Computing in Cardiology, 2016.

[4] W. Xue, A. Islam, M. Bhaduri, and S. Li, "Direct Multitype Cardiac Indices Estimation via Joint Representation and Regression Learning," IEEE Transactions on Medical Imaging, vol. 36, no. 10, pp. 2057-2067, 2017.

[5] C. Petitjean and J. N. Dacher, "A Review of Segmentation Methods in Short Axis Cardiac MR Images,", Medical Image Analysis, vol. 15, no. 2, pp. 169-184, Apr 2011.

[6] X. Zhen, Z. Wang, A. Islam, M. Bhaduri, I. Chan, and S. Li, "Direct Estimation of Cardiac Bi-ventricular Volumes with Regression Forests," in International Conference on Medical Image Computing and Computer-Assisted Intervention, 2014, pp. 586-593: Springer.

[7] X. Zhen, Z. Wang, A. Islam, M. Bhaduri, I. Chan, and S. Li, "Multi-scale Deep Networks and Regression Forests for Direct Bi-ventricular Volume Estimation," Med Image Anal, vol. 30, pp. 120-9, May 2016.

[8] Z. Wang, M. B. Salah, B. Gu, A. Islam, A. Goela, and S. Li, "Direct Estimation of Cardiac Biventricular Volumes with An Adapted Bayesian Formulation," IEEE Trans Biomed Eng, vol. 61, no. 4, pp. 1251-60, Apr 2014.

[9] X. Zhen, A. Islam, M. Bhaduri, I. Chan, and S. Li, "Direct and Simultaneous Four-Chamber Volume Estimation by Multi-Output Regression," in Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015, Cham, 2015, pp. 669-676: Springer International Publishing.

[10] X. Zhen, H. Zhang, A. Islam, M. Bhaduri, I. Chan, and S. Li, "Direct and Simultaneous Estimation of Cardiac Four Chamber Volumes by Multioutput Sparse Regression," Medical Image Analysis, vol. 36, pp. 184-196.

[11] G. N. Luo, S. Y. Dong, K. Q. Wang, and H. G. Zhang, "Cardiac Left Ventricular Volumes Prediction Method Based on Atlas Location and Deep Learning,", 2016 IEEE International Conference on Bioinformatics And Biomedicine (BIBM), pp. 1604-1610, 2016.

[12] A. Suinesiaputra et al., "A Collaborative Resource to Build Consensus for Automated Left Ventricular Segmentation of Cardiac MR Images," (in English), Medical Image Analysis, vol. 18, no. 1, pp. 50-62, Jan 2014.

[13] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-scale Image Recognition," presented at the International Conference on Learning Representations, 2015.

[14] F. Chollet. (2015). Keras. Available: https://github.com/fchollet/keras

Address for correspondence.

Kuanquan Wang
Mailbox 332, Harbin Institute of Technology
Harbin 150001, China
wangkq@hit.edu.cn