Dynamically constructed network with error correction for accurate ventricle volume estimation

Gongning Luo[a], Wei Wang[a,*], Clara Tam[d,e], Kuanquan Wang[a], Shaodong Cao[b], Henggui Zhang[a,c], Bo Chen[d,e], Shuo Li[d,e,*]

[a] *School of Computer Science and Technology, Harbin Institute of Technology, Harbin, 150001, China.*
[b] *The Department of Radiology, The Fourth Hospital of Harbin Medical University, Harbin 150001, China.*
[c] *School of Physics and Astronomy, University of Manchester, Manchester, UK.*
[d] *The Department of Medical Imaging, Western University, London, Canada.*
[e] *The Digital Imaging Group of London, London, ON N6A 3K7, Canada.*

Abstract

Automated ventricle volume estimation (AVVE) on cardiac magnetic resonance (CMR) images is very important for clinical cardiac disease diagnosis. However, current AVVE methods ignore the error correction for the estimated volume. This results in clinically intolerable ventricle volume estimation error and further leads to wrong ejection fraction (EF) assessment, which significantly limits the application potential of AVVE methods. The objective of this paper is to address this problem with AVVE and further make it more clinically applicable. We proposed a dynamically constructed network to achieve accurate AVVE. First, we introduced a novel dynamically constructed deep learning framework, that evolves a single model into a bi-model volume estimation network. In this way, the EF correlation can be built directly based on the bi-model network. Second, we proposed an error correction strategy using dynamically created residual nodes, which is based on stochastic configurations with an EF correlation constraint. Finally, we formulated the proposed method into an end-to-end joint optimization framework for accurate ventricle volume estimation with effective error correction. Experiments and comparisons on large-scale cardiac

*Corresponding author
Email addresses: wangwei2019@hit.edu.cn (Wei Wang), slishuo@gmail.com (Shuo Li)

magnetic resonance datasets were carried out. Results show that the proposed method outperforms state-of-the-art methods, and has good potential for clinical application. Besides, the proposed method is the first work to achieve error correction for AVVE and also has the potential to be extended to other medical index estimation tasks.

Keywords: Dynamically constructed network, Residual correction, Ventricle volume estimation, Ejection fraction correlation.

2010 MSC: 00-01, 99-00

1. Introduction

Cardiac diseases are the leading causes of death in the world with ventricle volume as a key clinical index for diagnosis (Zhen et al., 2016). Currently, cardiac magnetic resonance (CMR) images are widely used as a gold standard of ventricle volume assessment. However, clinical ventricle volume assessment relies on manual label and computation, which is time-consuming and subjective. Hence, automated ventricle volume estimation (AVVE) based on CMR images is very important for diagnosing cardiac diseases (Liao et al., 2017; Yu et al., 2020). It not only saves time and labor, but also avoids subjective variances.

However, current AVVE methods ignore the error correction for the estimated volume. This results in clinically intolerable ventricle volume estimation error and further leads to wrong ejection fraction (EF) assessment, which significantly limits the application potential of AVVE methods (Gu et al., 2018). The AVVE methods based on CMR can be classified into two types, i.e., the ventricle segmentation method and the direct volume estimation method.

a) **Ventricle segmentation**: Ventricle segmentation method was originally used to address the ventricle volume estimation problem (Petitjean and Dacher, 2011). Before 2011, most papers focused on addressing the ventricle segmentation problem. Then, the ventricle function indices, such as the left ventricle volume (LVV), right ventricle volume (RVV), and EF, were further evaluated based on the segmentation results. More details about ventricle segmentation

methods before 2011 can be found in the review papers (Petitjean and Dacher, 2011) and (Petitjean et al., 2015). After 2011, the main automated ventricle segmentation methods were divided into two types, i.e., the on-line learning methods and the off-line learning methods. 1) For the on-line learning methods, the snake (Kass et al., 1988), level set (Cremers et al., 2007) and atlas (Zhuang et al., 2010; Oguz et al., 2016; Wang et al., 2018; Dong et al., 2020) were frequently used for ventricle segmentation. In (Hajiaghayi et al., 2017), Hajiaghayi et al. extended the traditional 2D snake into a 3D active contour method to achieve automated left ventricle (LV) segmentation on CMR images. In (Khalifa et al., 2012), Khalifa et al. adopted the level set method to finish fast bi-ventricle segmentation and analysis. In (Yang et al., 2017), Yang et al. extended the conventional level set method into two-layer level sets for LV segmentation. (Zhuang et al., 2010) and (Bai et al., 2013) combined the cardiac atlas prior into a CMR image registration framework for cardiac segmentation. In (Oguz et al., 2016), Oguz et al. incorporated shape priors into a multi-atlas segmentation method to achieve a global optimal solution. In (Wang et al., 2018), Wang et al. adopted corrective learning to achieve a quicker run-time for multi-atlas segmentation. However, since most on-line segmentation methods relied heavily on large amounts of on-line iteration computations, on-line learning methods have low computational efficiency. 2) For off-line learning methods, some learning methods were designed mainly based on machine learning technology using large numbers of hand-drafted features and prior knowledge. In (Albà et al., 2018), Albà et al. proposed a method based on random forest to address ventricle segmentation and quality control. In (Eslami et al., 2013), Eslami et al. used the guided random walks method to finish LV segmentation. Futhermore, deep learning (DL) was the most widely used technique for ventricle segmentation in the past three years due to its strong feature representation ability. The fully convolutional network (FCN) was adopted to achieve ventricle segmentation on CMR images for the first time in (Tran, 2016). In (Tan et al., 2017), Tan et al. utilized DL to achieve boundary regression on converted polar coordinates. To achieve consistent ventricle segmentation along long axis slices,

Zheng et al. fused spatial propagation information into a DL network to address the consistent ventricle segmentation task (Zheng et al., 2018). In (Wong et al., 2018), Wong et al. proposed an exponential logarithmic loss for 3D deep learning network to handle highly unbalanced object sizes especially on 3D ventricle segmentation task. In (Tang et al., 2018), Tang et al. proposed a deep learning framework based on a dice similarity coefficient loss to achieve segmentation for cardiac anatomical structures. (Luo et al., 2016) and (Vigneault et al., 2018) proposed a spatial localization network that formulated the ventricle center localization and segmentation tasks into a unified optimization framework to achieve right ventricle (RV) and LV segmentation respectively. (Ngo et al., 2017) and (Avendi et al., 2016) used DL to localize ventricle and used active contour model to refine the LV segmentation results. In (Qin et al., 2018), Qin et al. combined the optical flow information to formulate the ventricle motion information into FCN to achieve ventricle segmentation. In (Mo et al., 2018), Mo et al. converted the traditional snake model into a DL problem based on gradient regression to address the LV segmentation problem. Although off-line learning methods can achieve relatively higher inference efficiency than on-line learning methods, current automated ventricle segmentation methods perform poorly on the complex geometry of the base slices and ambiguous anatomy of the apex slices. Futhermore, the ventricle segmentation method is a two-stage model for ventricle volume estimation. The geometry computation in the second stage relies on the accuracy of the segmentation in the first stage. A natural question is asked: Could a one-stage volume estimation method achieve a higher accuracy? Hence, the direct ventricle volume estimation task becomes a new research topic.

b) **Direct volume estimation**: Direct volume estimation is another kind of AVVE method based on CMR images without the segmentation procedure. The direct estimation is an effective data screening and analysis strategy for large numbers of CMR data, especially on the aspect of historical data analysis and population study of heart disease (Afshin et al., 2011, 2012). In the past six years, the direct volume estimation method, bypassing segmentation, has

achieved great success in cardiac medical image analysis (Xue et al., 2017a,b, 2018; Ge et al., 2020; Luo et al., 2018). The direct ventricle indices estimation methods are mainly based on two kinds of technologies, i.e., the traditional statistical learning technology and the DL technology. 1) Traditional statistical learning technology was first applied on bi-ventricle volume prediction (Wang et al., 2014; Zhen et al., 2014, 2015, 2016, 2017) based on hand-crafted features and multi-task learning theory (Zhen et al., 2018b, 2017, 2018a). In (Wang et al., 2014), Wang et al. proposed a bi-ventricle volume estimation method based on the improved Bayesian classifier. In (Zhen et al., 2015), Zhen et al. adopted random forest to achieve bi-ventricle volume regression. 2) DL was the only technology used for direct ventricle volume estimation task in the past four years. In (Zhen et al., 2016), Zhen et al. adopted multi-scale DL networks and random forests to address the bi-ventricular volume estimation problem. In (Ge et al., 2019), Ge et al. adopted paired apical views-based DL technology to achieve direct left ventricle multitype indices estimation. In (Luo et al., 2018), Luo et al. denoted the whole ventricle geometry using the multi-slice combination as the input of DL to achieve LV volume estimation. In (Luo et al., 2019), Luo et al. combined ventricle segmentation and index estimation into a unified framework to achieve direct ventricle volume estimation.

In summary, though direct ventricle volume estimation methods enable one-stage volume estimation, they do not provide an error correction for the estimated volume. This limits the accuracy of these methods. Hence, the accuracy of the ventricle volume estimation task still has large space to be improved for both the segmentation method and the direct ventricle volume estimation method. Besides, a higher volume estimation error leads to a larger EF estimation error, which significantly limits the clinical application potential of AVVE.

Specifically, two important problems are to be addressed urgently. First, the error correction should be addressed for more accurate AVVE. Second, EF correlation should be considered for AVVE because EF is very sensitive to volume changes and sometimes a small volume error will result in a large EF error, which is not tolerable for clinical application. Note that in this paper, we de-

note the volume estimation error as a volume residual in the view of numerical optimization.

The standard EF computation equation is:

$$EF_T = \frac{(EDV_T - ESV_T)}{EDV_T} = 1 - \frac{ESV_T}{EDV_T} \tag{1}$$

where EF_T denotes the ground truth EF, EDV_T is the ground truth volume from the end-diastole frame (EDV), ESV_T is the ground truth volume from the end-systole frame (ESV). The estimated EF based on AVVE is:

$$EF_E = 1 - \frac{ESV_E}{EDV_E} = 1 - \frac{ESV_T - \varepsilon_{ESV}^T}{EDV_T - \varepsilon_{EDV}^T} \tag{2}$$

where EF_E denotes the estimated EF, EDV_E is the estimated EDV, ESV_E is the estimated ESV, ε_{EDV}^T is the ground truth residual between EDV_E and EDV_T, ε_{ESV}^T is the ground truth residual between ESV_E and ESV_T. In some cases, the smaller residual between volume estimation value and ground truth value leads to a bigger EF computation error. For instance, we assume $EDV_T = 250ml$, $ESV_T = 100ml$, hence the $EF_T = 60\%$. We further assume that $\varepsilon_{EDV}^T = -15ml$, $\varepsilon_{ESV}^T = 15ml$, and then the $EF_E = 67.9\%$ (The difference with the ground truth is $|EF_T - EF_E| = 7.9\%$). This case indicates that some methods (Luo et al., 2018)(Wang et al., 2014)(Zhen et al., 2015) can achieve a relatively accurate volume estimation but larger EF quantification error in some cases. Hence, EF correlation and residual correction between the estimated volume value and ground truth volume value are very important factors, which should be considered in the volume estimation model.

To the best of our knowledge, current methods ignored the EF correlation on volume estimation and no methods successfully model the EF correlation on volumes based on independent optimization for EDV and ESV. Furthermore, no methods modelled the residual between the ground truth volume value and the estimated volume value. In this paper, we proposed an accurate and direct AVVE method, using a dynamically constructed network through volume residual modeling with EF correlation, to achieve a more accurate ventricle volume estimation and higher generalization ability. The main contributions of this paper are four-fold:

1) For the first time, we proposed a dynamically constructed DL framework to achieve direct ventricle volume estimation on ED and ES frames independently and simultaneously, which enables the direct modelling of EF correlation.

2) For the first time, we modelled volume estimation residual in a DL framework using a stochastic configuration algorithm with EF correlation constraint to achieve error correction for more accurate ventricle volume estimation during dynamic model optimization.

3) We formulated the proposed method into a continuously differentiable framework to achieve end-to-end joint optimization for volume estimation and EF correlation constraint.

4) Experiments and comparisons on large-scale cardiac magnetic resonance datasets showed that the proposed method achieves best ventricle volume estimation and quantification performance compared with state-of-the-art methods.

The remainder of this paper is organized as follows. Section II introduces the proposed method for AVVE. The experiments and implementation details are given in section III, and we discuss and conclude the results and methods in section IV.

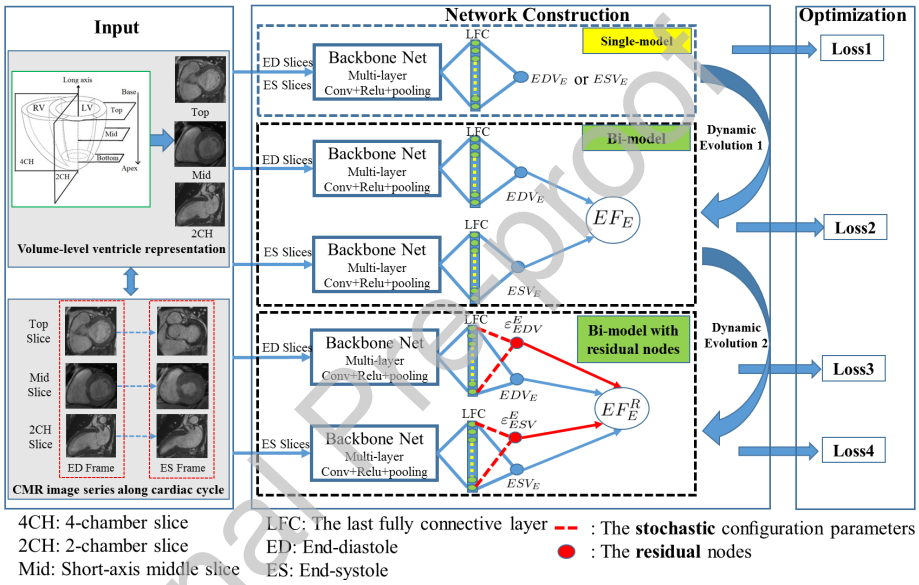


Figure 1: The framework of the proposed dynamic constructed network. The proposed framework can achieve network structure improvement with a dynamic loss function and an explicit residual estimation. This way not only enables independent and high-efficient optimization for direct EDV and ESV estimation, but also enables the modelling of the EF correlation, which is a critical ventricle function quantification index.

2. Method

Algorithm 1: The training strategy of DCN

1. Initializing the single-model network, setting the max iteration number (MIN).
2. **While** validation loss1> $Th1$ and $mu < MIN$:
3. Training single-model network using Back Propagation (BP), $mu = mu + 1$.
4. **End while**
5. Constructing the bi-model network, reset $mu = 0$.
6. Transfer parameter weight values from single-model network to bi-model network: $\omega^s_{base} \rightarrow \omega^b_{baseED}$, $\omega^s_{base} \rightarrow \omega^b_{baseES}$, $\omega^s_V \rightarrow \omega^b_{EDV}, \omega^s_V \rightarrow \omega^b_{ESV}$.
7. **While** validation loss2> $Th2$ and $mu < MIN$:
8. Training bi-model network using BP, $mu = mu + 1$.
9. **End while**
10. Stochastic configuration residual nodes using algorithm2, reset $mu = 0$.
11. **While** validation loss3> $Th3$ and $mu < MIN$:
12. Training parameters ω^b_{REDV} and ω^b_{RESV} connected to residual node using BP, **fixing other parameters**, $mu = mu + 1$.
13. **End while**
14. Reset $mu = 0$.
15. **While** validation loss4> $Th4$ and $mu < MIN$:
16. Training all the bi-network using BP, $mu = mu + 1$.
17. **End while**

2.1. Framework of the proposed method

The framework of the proposed Dynamically Constructed Network (DCN) is shown in Fig. 1. The DCN includes three modules, i.e., data input module,

network construction module, and optimization module. The data input module conducts the data preprocessing and selects the slices with high representation ability as an input of the DL network. The network construction module includes the dynamic DL networks, which changes from a single-model network to a bi-model network, and further into a bi-model network with stochastic configuration residual nodes. The optimization module includes multiple continuously differentiable dynamic loss functions, which completes the end-to-end optimization along with dynamic network construction.

2.2. Data preprocessing for ventricle representation

The data input module achieves good representation for the ventricle through CMR data preprocessing. The heart is a special organ with cyclic motion from end-systole (ES) to end-diastole (ED). As shown in the input module of Fig.1, CMR achieves the whole 3D spatial representation of the heart through multi-slice imaging, and the whole cardiac cycle representation through multi-frame imaging. Hence, we represented the CMR image series as $\chi = \{X_l^{f,n}\}$, where $f \in \{1....F\}$ denotes the frame, $l \in \{1....L\}$ denotes the slice set, and $n \in \{1...N\}$ denotes the subject.

In the aspect of data preprocessing, we followed the same preprocessing steps in our previous work (Luo et al., 2018) to normalize different datasets to the same parameter setting. Specifically, to achieve an effective cardiac representation and reduce the data input dimensionality, the basic data preprocessing operations, including the ROI localization, the physical space normalization, the pixel intensity level normalization, the image cropping, the slice selection, and the frame selection, were conducted:

1. We conducted the ROI localization using the localization method in (Luo et al., 2018) based on the intersections between the long and short-axis image planes.
2. We normalized all CMR images into the same physical space (1.4mm*1.4mm).
3. We conducted the pixel intensity level normalization for every CMR image using the zero-mean normalization method in (Luo et al., 2018).

4. We cropped the images into 100100 image patches along the ventricle center location.
5. In the aspect of slice selection, we had proven that the slice combination (including Top slice, 2 chamber (2CH) slice, and Mid slice) was optimal for representing the ventricle geometry in 3D space in our previous work (Luo et al., 2018). Hence, we selected the three slices (Top slice, 2CH slice, and Mid slice) as the input of the DL network, i.e., $l = \{Top, 2CH, Mid\}$.
6. In the aspect of frame selection, the volume labeling of the whole cardiac circle is impracticable. Thus far, the largest open accessible cardiac dataset in the world only includes the ventricle volume value label on ED and ES frames. Besides, in the clinical application, only ventricle volume values on ED and ES frames are used for EF calculation. Hence, in the model building and training stage, we only used the CMR images on ED and ES frames as the input of the DL network, i.e., $f \in \{ED, ES\}$.

2.3. Network with dynamic construction and stochastic configuration

The construction and optimization procedure of the proposed DCN is shown in Algorithm 1. To achieve a clear description in the following sections, we introduced the DCN according to the four stages in Algorithm 1.

*2.3.1. **Stage one: single-model network***

The single-model is the base for the dynamic construction of network. Until now, all the published direct AVVE methods based on DL technology adopted the single-model network to predict volume (Luo et al., 2018). The single-model network is a type of DL regression network with one node output, which denotes the ventricle volume value. In this stage (1-4 in Algorithm 1), to guarantee the generalization of the single-model network, we used slices in the ED frame or slices in the ES frame as the input of the single-model network (As shown in Fig. 1). The backbone net of the single-model network was configured in a flexible manner. The backbone net can be any popular DL network, such as VGG (Simonyan and Zisserman, 2014), ResNet (He et al., 2016), or DenseNet

(Huang et al., 2017). As shown in the Fig. 1, these popular networks have multiple convolutional layers, Relu activation functions, pooling layers, and skip connections. To fit the ventricle volume estimation application, we removed the last output layer of popular DL networks and added a fully connected layer with 1024 nodes (We denote this layer as the last fully connected (LFC) layer before the output node). Besides, the output node is always fully connected to the LFC layer. Hence, the single-model network can be represented as follows:

$$\begin{aligned} &V = F\left(\chi, \omega_{base}^{s}, \omega_{V}^{s}\right), \chi \in \{X_{l}^{ED,n}, X_{l}^{ES,n}\}, \\ &X_{l}^{ED,n} = [X_{Top}^{ED,n}, X_{2CH}^{ED,n}, X_{Mid}^{ED,n}], \\ &X_{l}^{ES,n} = [X_{Top}^{ES,n}, X_{2CH}^{ES,n}, X_{Mid}^{ES,n}] \end{aligned} \tag{3}$$

where V denotes the output volume value of the single-model network F, χ is the input data, ω_{V}^{s} denotes the network parameters of the output node, ω_{base}^{s} denotes the other network parameters of the single-model network except for ω_{V}^{s}, as well as $X_{l}^{ED,n}$ and $X_{l}^{ES,n}$ denote the input data with three channels from three slices (Top, 2CH, Mid) in ED and ES frames respectively. In this stage, the optimization objective is:

$$Loss1 = \frac{1}{2N} \sum_{i=1}^{2N} \frac{1}{2} \left(V^{i} - V_{T}^{i}\right)^{2} \tag{4}$$

where $Loss1$ is the loss in the stage of the single-model networks training, N is the number of training subjects (every subject includes two volumes. i.e., EDV and ESV), V^{i} is the V of case i, V_{T}^{i} is the ground truth volume value of case i.

2.3.2. ***Stage two: dynamic construction from single-model network to bi-model network***

The bi-model network enables the modelling of the EF correlation based on the independent EDV and ESV output. Prior research in (Luo et al., 2018), (Wang et al., 2014), and (Zhen et al., 2015) has found that independent model optimization for EDV and ESV can achieve a more accurate volume estimation in theory. Besides, EF is an important clinical index for disease diagnosis. Hence, modelling the EF correlation between EDV and ESV as an effective

constraint enables more accurate volume estimation and higher clinical application value. In this stage (5-9 in Algorithm 1), to model the EF correlation based on simultaneous and independent EDV and ESV output, we constructed a bi-model network, which was dynamically constructed from a single-model network. The construction of network was three-fold, i.e., network structure construction, bi-model network parameters initialization, and optimization objective construction.

First: network structure construction, the network structure construction from the single-model network to bi-model network is shown in Fig.1. Specifically, the bi-model network has two independent paths corresponding to data from ED and ES frames respectively, and each path has the same feed-forward structure with a single-model network. Hence, the two paths of the bi-model network have an independent input (slices on ED and ES respectively) and an independent output (ventricle volume values on ED and ES respectively). The bi-model network can be represented as follows:

$$\begin{aligned} EDV_E &= F_{ED}\left(X_l^{ED,n}, \omega_{baseED}^b, \omega_{EDV}^b\right), \\ ESV_E &= F_{ES}\left(X_l^{ES,n}, \omega_{baseES}^b, \omega_{ESV}^b\right) \end{aligned} \tag{5}$$

where F_{ED} and F_{ES} denote the bi-model networks' two paths corresponding to the ED and ES frames respectively, EDV_E and ESV_E are ventricle volume values estimated from the bi-model networks' two paths respectively, ω_{baseED}^b and ω_{baseES}^b denote the parameters of the two backbone networks from the two paths of the bi-model network respectively except for the last output nodes, ω_{EDV}^b denotes the network parameters of the last output node for EDV, and ω_{ESV}^b denotes the network parameters of the last output node for ESV.

Second: bi-model network parameters' initialization, the bi-network parameters were initialized through directly transferring the parameters from the single-model network ($\omega_{base}^s \rightarrow \omega_{baseED}^b, \omega_{base}^s \rightarrow \omega_{baseES}^b, \omega_V^s \rightarrow \omega_{EDV}^b, \omega_V^s \rightarrow \omega_{ESV}^b$), as shown in the Algorithm 1, because each path of the bi-model network has the same feed-forward structure as the single-model network.

Third: optimization objective construction, the optimization objective was improved based on the EF regularization constraint. Compared to the

single-model network, the bi-model network has special superiority, benefiting from the EF regularization constraint, which can be modelled into the optimization objective starting from this stage. To the best of our knowledge, EF regularization constraint was proposed to improve the performance of AVVE for the first time, based on the simultaneous output of EDV and ESV in the bi-model network. In this stage, the constructive optimization objective was changed into:

$$\begin{aligned} Loss2 =& \frac{1}{N}\sum_{n=1}^{N}\frac{1}{2}\left(EDV_E^n - EDV_T^n\right)^2 + \frac{1}{2}\left(ESV_E^n - ESV_T^n\right)^2 \\ &+ \frac{\gamma_1}{2}\left(EF_E^n - EF_T^n\right)^2, EF_E^n = 1 - \frac{ESV_E^n}{EDV_E^n} \end{aligned} \tag{6}$$

where $Loss2$ is the loss in the stage of the bi-model networks training, γ_1 is the weight factor to control the EF regularization item, EDV_E^n and ESV_E^n are EDV_E and ESV_E of subject n in ED and ES frames respectively, EDV_T^n and ESV_T^n are the ground truth volumes of subject n in ED and ES frames respectively, EF_T^n is the ground truth EF value (EF_T) of subject n, and EF_E^n is the estimated EF value (EF_E) of subject n.

2.3.3. ***Stage three: dynamic construction with added residual nodes using stochastic configuration for error correction***

Residual exists in any regression model based on DL, and modelling the residual enables estimation error correction that improves the ventricle volume estimation accuracy. However, to the best of our knowledge, no methods can achieve residual modeling in an explicit manner (It is different from the latent residual modeling in ResNet (He et al., 2016), whose network structure is fixed). In this stage (10-13 in Algorithm 1), to model the volume estimation error in an explicit manner, we added residual nodes using a stochastic configuration method, which has been widely recognized for problem solving on incremental learning systems and dynamically constructive neural networks (Wang and Li, 2017; Kwok and Yeung, 1997; Barron, 1993). The stochastic configuration strategy is a good parameter initialization strategy for a dynamically added node. The stochastic configuration strategy was achieved by randomly out-

Algorithm 2: The stochastic parameters configuration process of the residual nodes

1.Input: The expected error tolerance ρ, the maximum times of stochastic configuration T_{max}.

2.Output: Network with residual nodes using stochastic parameters configuration.

3.Initialization: Computing the self-adjustment scope of stochastic parameters configuration on residual nodes' weights and biases, weights' range is $[min(W), max(W)]$, biases' scope is $[min(B), max(B)]$, and $EF_C^{min} = +\infty$.

4. **While** $EF_C > \rho$ and $t < T_{max}$:

5. **for** $k = 1....1024$ **do**:

6. Randomly assign $\omega_{REDV}^{b,k,weight}$ in range $[min(W), max(W)]$, where $\omega_{REDV}^{b,k,weight}$ denotes the weight of connection between residual node ε_{EDV}^{E} and kth nodes on LFC.

7. **End for**

8. Randomly assign $\omega_{REDV}^{b,bias}$ in range $[min(B), max(B)]$, where $\omega_{REDV}^{b,bias}$ is the bias of residual node ε_{EDV}^{E}.

9. **for** $k = 1....1024$ **do**:

10. Randomly assign $\omega_{RESV}^{b,k,weight}$ in range $[min(W), max(W)]$, where $\omega_{RESV}^{b,k,weight}$ denotes the weight of connection between residual node ε_{ESV}^{E} and kth nodes on LFC.

11. **End for**

12. Randomly assign $\omega_{RESV}^{b,bias}$ in range $[min(B), max(B)]$, where $\omega_{RESV}^{b,bias}$ is the bias of residual node ε_{ESV}^{E}

13. Computing the current EF constraint EF_C

14. **If** $EF_C < EF_C^{min}$

15. $EF_C^{min} = EF_C$

16. Saving randomly assigned parameters, i.e., ω_{REDV}^{b} and ω_{RESV}^{b}

17. **End If**

18. **End while**

putting weights until a predefined termination criterion was met (Kwok and Yeung, 1997).

In this way, a new node can be added with an explicit meaning. Besides, for a dynamically added node in a regression problem, the stochastic configuration is an excellent parameter initialization method, which has been proven in (Wang and Li, 2017). In the proposed framework, the residual node is exactly a dynamically added node for the volume residual regression. Hence, we used the stochastic configuration method to achieve good parameter initialization. The details on dynamically constructing stochastic configuration residual nodes are shown in Algorithm 2.

Specifically, we dynamically constructed the residual nodes $(\varepsilon_{EDV}^{E}, \varepsilon_{ESV}^{E})$ for EDV and ESV output nodes respectively in the bi-model network. Besides, inspired by the theory on randomized learning models in (Li and Wang, 2017), we proposed a range self-adjustment method to randomly select weights and biases from a self-adjustment scope. The self-adjustment scope contains the residual nodes weights' range ($[min(W), max(W)]$) and biases' scope ($[min(B), max(B)]$), where W denotes the weight parameter set of the network constructed after stage two and before adding the residual nodes, and B denotes the bias parameter set of the network constructed after stage two and before adding the residual nodes. Additionally, inspired by the regularization method on a stochastic parameter configuration model (Wang and Li, 2017), we improved the stochastic parameter configuration for residual nodes through the EF correlation constraint, which is different from the EF regularization constraint in equation (6) in stage two. The EF correlation constraint (EF_C) in this stage was computed based on the added residual node output and volume output (in detail, EF_C was computed according to equation (8)). In this stage, the bi-model network can be represented as follows:

$$\begin{aligned}\left[\varepsilon_{EDV}^{E}, EDV_E\right] &= F_{ED}\left(X_l^{ED,n}, \omega_{baseED}^{b}, \omega_{EDV}^{b}, \omega_{REDV}^{b}\right),\\ \left[\varepsilon_{ESV}^{E}, ESV_E\right] &= F_{ES}\left(X_l^{ES,n}, \omega_{baseES}^{b}, \omega_{ESV}^{b}, \omega_{RESV}^{b}\right)\end{aligned} \tag{7}$$

where ε_{EDV}^{E} and ε_{ESV}^{E} denote the estimated volume residuals on EDV and ESV respectively, and ω_{REDV}^{b} and ω_{RESV}^{b} denote the network parameters of the

last output layer for ε_{EDV}^{E} and ε_{ESV}^{E} respectively. The EF correlation (EF_C) constraint can be denoted as follows:

$$\begin{aligned} EF_C &= \frac{1}{N}\sum_{n=1}^{N}\frac{1}{2}\left(EF_E^{R,n} - EF_T^n\right)^2, \\ EF_E^{R,n} &= 1 - \frac{ESV_E^n + \varepsilon_{ESV}^{E,n}}{EDV_E^n + \varepsilon_{EDV}^{E,n}} \end{aligned} \tag{8}$$

where EDV_E^n and ESV_E^n denote the EDV_E and ESV_E of subject n, $\varepsilon_{EDV}^{E,n}$ and $\varepsilon_{ESV}^{E,n}$ denote the ε_{EDV}^{E} and ε_{ESV}^{E} of subject n, and $EF_E^{R,n}$ denotes the computed EF_E value of subject n in the bi-network with added residual nodes, and EF_T^n denotes EF_T of subject n. The optimization objective is:

$$\begin{aligned} Loss3 =& \frac{1}{N}\sum_{n=1}^{N}\left\{\frac{1}{2}\left(\varepsilon_{EDV}^{E,n} - \varepsilon_{EDV}^{T,n}\right)^2 + \frac{1}{2}\left(\varepsilon_{ESV}^{E,n} - \varepsilon_{ESV}^{T,n}\right)^2\right\} \\ &+ \gamma_2 EF_C \end{aligned} \tag{9}$$

where $\varepsilon_{EDV}^{T,n}$ and $\varepsilon_{ESV}^{T,n}$ denote the ε_{EDV}^{T} and ε_{ESV}^{T} of subject n ($\varepsilon_{EDV}^{T} = EDV_T - EDV_E, \varepsilon_{ESV}^{T} = ESV_T - ESV_E$), and γ_2 is the weight factor to control the EF_C constraint item. In this stage, only the parameters ω_{REDV}^{b} and ω_{RESV}^{b} connected to the residual nodes were updated using BP algorithm, to fix the other parameters of the DL network.

2.3.4. ***Stage four: synchronous optimization***

To avoid the unbalanced optimization, in the last stage (14-17 in Algorithm 1), we trained the network in a synchronous manner, i.e., synchronously optimizing the predicted volume (EDV_E^n, ESV_E^n) and estimated residual ($\varepsilon_{EDV}^{E,n}$, $\varepsilon_{ESV}^{E,n}$). The optimization objective is:

$$\begin{aligned} Loss4 = \frac{1}{N}\sum_{n=1}^{N}\Bigg\{&\frac{1}{2}\left(EDV_E^n - EDV_T^n\right)^2 + \frac{1}{2}\left(ESV_E^n - ESV_T^n\right)^2 + \\ &\frac{1}{2}\left(\varepsilon_{EDV}^{E,n} - \varepsilon_{EDV}^{T,n}\right)^2 + \frac{1}{2}\left(\varepsilon_{ESV}^{E,n} - \varepsilon_{ESV}^{T,n}\right)^2\Bigg\} + \gamma_3 EF_C \end{aligned} \tag{10}$$

where γ_3 is the weight factor to control the EF_C constraint item, and EF_C is computed according to (8). In this way, the DCN can not only get an accurate volume estimation output, but also an accurate residual estimation and EF estimation output. Hence the final volume estimation value (sum of volume estimation output and residual estimation output) approaches the ground truth.

Besides, in the aspect of clinical application, the proposed method can achieve more accurate EF quantification benefiting from the EF correlation constraint.

2.4. Continuously differentiable property of dynamic loss functions

The loss functions in every stage of the proposed DCN are continuously differentiable, which is crucial for end-to-end model optimization. We can prove the continuously differentiable property as follows. In stage one, according to equation (4), $Loss1$'s derivative for any parameter ϖ is:

$$\frac{\partial Loss1}{\partial \varpi} = \frac{1}{2N}\sum_{i=1}^{2N}\left(V^i - V_T^i\right)\frac{\partial V^i}{\partial \varpi} \tag{11}$$

in which, the deep single-model network is continuously differentiable, hence $\frac{\partial V^i}{\partial \varpi}$ exists. So $Loss1$ has a derivative for any parameter ϖ. In stage two, according to equation (6), $Loss2$'s derivative for any parameter ϖ is:

$$\begin{aligned}\frac{\partial Loss2}{\partial \varpi} &= \frac{1}{N}\sum_{n=1}^{N}\left(EDV_E^n - EDV_T^n\right)\left(\frac{\partial EDV_E^n}{\partial \varpi}\right)\\ &+\left(ESV_E^n - ESV_T^n\right)\frac{\partial ESV_E^n}{\partial \varpi} + \gamma_1\left(EF_E^n - EF_T^n\right)\frac{\partial EF_E^n}{\partial \varpi},\\ \frac{\partial EF_E^n}{\partial \varpi} &= -\left(\frac{\partial ESV_E^n}{\partial \varpi}\bullet\frac{1}{EDV_E^n} - \frac{\partial EDV_E^n}{\partial \varpi}\bullet\frac{ESV_E^n}{(EDV_E^n)^2}\right)\end{aligned} \tag{12}$$

in which, each path in the deep bi-model network is continuously differentiable, hence $\frac{\partial ESV_E^n}{\partial \varpi}$ and $\frac{\partial EDV_E^n}{\partial \varpi}$ both exist. Besides, EDV_T^n, ESV_T^n, and EF_T^n are constants (the derivative of a constant is zero), so $Loss2$ has a derivative for any parameter ϖ.

Similarly, in stage three and stage four, according to equation (9) and (10), $Loss3$'s and $Loss4$'s derivatives for any parameter ϖ from the bi-model network are available, because $\frac{\partial \varepsilon_{EDV}^{E,n}}{\partial \varpi}$, $\frac{\partial \varepsilon_{ESV}^{E,n}}{\partial \varpi}$, $\frac{\partial ESV_E^n}{\partial \varpi}$, $\frac{\partial EDV_E^n}{\partial \varpi}$, and $\frac{\partial EF_E^{R,n}}{\partial \varpi}$ are available. Hence the proposed DCN is continuously differentiable in any time. Additionally, in stage four, the ground truths for volume are fixed. The ground truth of the residual is fixed within an iteration (within one forward and backward propagation). The ground truth of the residual is changed during the different iterations when training the network. Hence, the residual node can be optimized in an end-to-end way to fit the change of ground truth of the residual. The detailed proof of differentiable property of $Loss4$ for the optimization of residual nodes is shown in Appendix A.

3. Experiments

In this section, we designed the experiments to evaluate the performance of the proposed DCN and to evaluate the rationality and application potential of the proposed framework.

3.1. Datasets

We adopted large-scale open access datasets to train and validate the proposed DCN on ventricle volume estimation. These datasets include STACOM2011 (200 subjects) with LVV labels (Suinesiaputra et al., 2014), RV2012 (48 subjects) with RVV labels (Petitjean et al., 2015), ACDC2017 (150 subjects) with LVV and RVV labels (Bernard and A. Lalande, 2018), and Kaggle2016 (1140 subjects) with LVV labels (Kaggle, 2016). To achieve a fair comparison with state-of-the-art methods, we followed the same dataset split (training, validation, and testing) strategy with standard in (Kaggle, 2016) and (Bernard and A. Lalande, 2018) respectively. We trained the proposed model based on the training set of Kaggle2016 dataset for LVV estimation and based on the training set of ACDC2017 dataset for RVV estimation. Then, we tested the performance of the trained model on different datasets to evaluate the proposed method's generalization ability. Specifically, the STACOM2011 dataset, ACDC2017 dataset, and the testing set of the Kaggle2016 dataset are used for final performance testing for LVV estimation. The RV2012 dataset and the testing set of the ACDC2017 dataset are used for final performance testing for RVV estimation. We obtained the ground truth volume value based on the segmentation label of the datasets in (Suinesiaputra et al., 2014; Petitjean et al., 2015; Bernard and A. Lalande, 2018) according to the following equation (Radau et al., 2009):

$$volume = \sum_{h=1}^{H} area_h * HS \tag{13}$$

where H is the number of slices in short-axis CMR image series, $area_h$ denotes the cavity area of the hth slice from the manual labels, and HS is the space between the adjacent slices.

3.2. Experiments setting and metric

1) Though the proposed DCN is robust in most cases, in order to guarantee the repeatability of the experiments, we introduce the hyper-parameters that were used when conducting the experiments (This setting is the default for this paper, one could tune these hyper-parameters for better performance.). Firstly, to evaluate the robustness of the proposed DCN, the three representative backbone networks (including VGG16 in (Simonyan and Zisserman, 2014), ResNet50 in (He et al., 2016), and DenseNet121 in (Huang et al., 2017)) are used for configuring the DCN. Specifically, the reason of using VGG, ResNet, and DenseNet is that VGG, ResNet and DenseNet are the most widely used backbone network structures for a regression problem. The experimental results from the three network structures are representative. The weights were initialized using Gaussian distribution with a standard deviation of 0.001 for the single-model network. The BP algorithm with Stochastic Gradient Descent (SGD) using a fixed learning rate of 0.0001 was adopted for training DCN. There was no learning rate decay during the training for different stages. To achieve a fair comparison, we also adopted the same fixed learning rate 0.0001 for the following ablation models and baseline models. The batch size was 10, and data augmentation was conducted during training using random rotation and flipping. The hyper parameter setting of the proposed framework is shown in Table 1 for the following experiments. The γ_1, γ_2, and γ_3 were relatively bigger than other parameters for achieving the optimization balance between the volume loss value and EF constraint value (percentage). In the four datasets, the final model after four stage's training was used for performance evaluation on testing set in the following section.

2) We evaluated the ventricle volume estimation accuracy and EF quantification accuracy using widely recognized metrics, i.e., mean absolute error (MAE), regression analysis, and residual analysis. MAE is computed as follows:

$$MAE = \frac{1}{M}\sum_{i=1}^{M}\left|Y_E^i - Y_T^i\right| \tag{14}$$

where M is the number of testing cases, Y_E^i is the estimated value from the

Table 1: Hyper parameter setting for the proposed framework.

The kind of parameter	Value
Weight of loss function	$\gamma_1 = 10000$, $\gamma_2 = 10000$, $\gamma_3 = 10000$
The threshold of interation number	$MIN = 10000$, $Th_1 = 100$, $Th_2 = 150$, $Th_3 = 50$, $Th_4 = 30$
The parameters for stochastic configuration	$\rho = 0.00045$, $T_{max} = 1000$

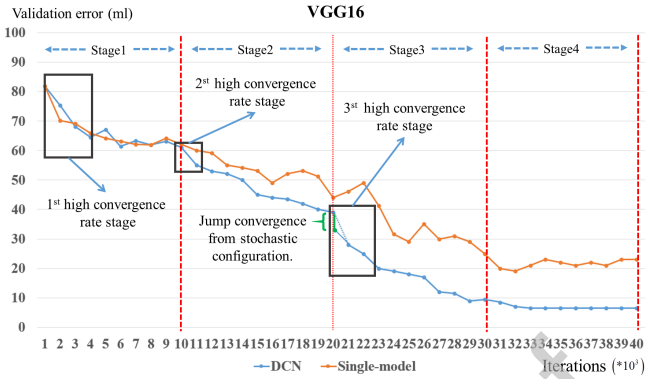
AVVE methods for case i, and Y_T^i is the ground truth value for case i. Note that, in stage one and two, the estimated volume is the volume output value, and in stage three and four, the estimated volume is the sum of the volume output value and residual value.

3.3. Optimization evaluation

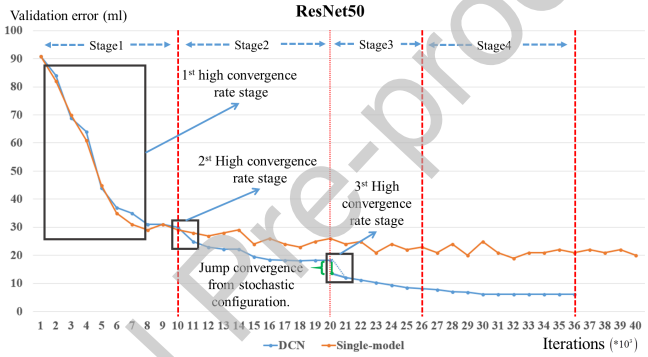
In this section, we conducted experiments to evaluate the optimization performance of DCN and the significance of the proposed multi-stage optimization strategy on Kaggle2016 dataset (Kaggle, 2016), which is the largest open access CMR dataset with LVV labels. Note that the experimental results for optimization performance evaluation on the other datasets were similar with the results on the Kaggle2016 dataset. Hence, to avoid redundancy, in some sections, we only showed the experimental results on the Kaggle2016 dataset.

3.3.1. Optimization performance evaluation of DCN

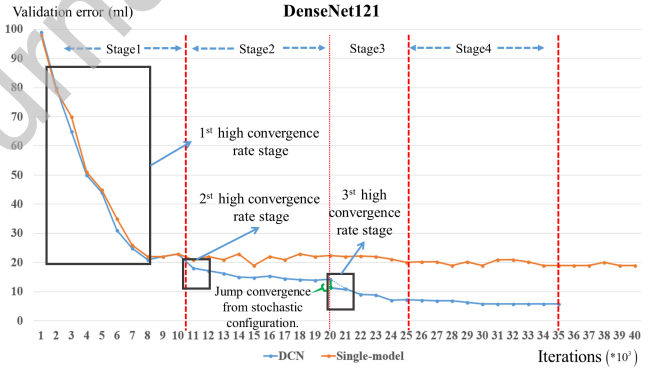
The proposed DCN has superior optimization performance with various backbone network structures. The superior optimization performance is shown in two aspects, i.e., fast convergence rate and high convergence stability, as well as robustness for various backbone network configurations. To show the



(a)



(b)



(c)

Figure 2: Optimization performance comparison between the proposed DCN and traditional single-model network on kaggle2016 dataset (Kaggle, 2016), and the mean MAE of EDV and ESV on validation set are computed as validation error during model training. a) The backbone network is VGG16. b) The backbone network is ResNet50. c) The backbone network is DenseNet121.

superior optimization performance, we conducted comparison experiments between the proposed DCN and the traditional single-model network using various backbone networks (VGG16, ResNet50, and DenseNet121) on the kaggle2016 dataset (Kaggle, 2016). Finally, the results were analyzed in the following two aspects:

Convergence rate and stability evaluation: As shown in Fig. 2, compared with the traditional single-model network training, the proposed DCN has a faster convergence rate and higher stability in every bi-model network training stage. In stage two and three, the DCN can significantly change the single-model network's convergence tendency to make it converge to a lower validation error with a faster rate. Specifically, the traditional single-model only has one high convergence rate (HCR) stage, which denotes the stage with stable and high-rate convergence tendency (HCR is shown in the black rectangular region in Fig. 2). However, the DCN has three HCR stages, and every HCR stage is the model's convergence momentum, which can make the model converge to a lower validation error with a high rate. Besides, the stochastic configuration algorithm in stage three enables the jump convergence (It is the green lines in Fig. 2, and in most cases jump convergence results in more than 5 ml direct decrease on MAE), which significantly changes the model's convergence tendency. Besides, in every bi-model network training stage, the DCN has higher training stability than the continuously trained single-model network, which has unstable convergence tendency. These phenomenons show that the proposed DCN enabled a faster convergence rate and higher convergence stability.

Robustness evaluation on backbone network configuration: The proposed DCN is robust to various backbone network configurations. As shown in Fig.2 (a-c), though the DCN has various convergence rates with different backbone network configurations (The ResNet50 and DenseNet121 have faster convergence rate than VGG16 in the single-model network training), DCN always improves the convergence tendency in stage two, three and four on three different backbone network configurations. Besides, we also can see that the ventricle volume estimation accuracy of DCN with various backbone networks is

always higher than state-of-the-art methods in the following Tables 2-4. Hence, the proposed DCN has robust optimization ability on flexible backbone network configuration.

3.3.2. Significance of multi-stage optimization of DCN

To show the significance of multi-stage optimization of DCN, we conducted the ablation experiments (i.e., S1, S12, S123, and S1234) for multi-stage optimization of DCN with residual analysis on Kaggle 2016 dataset. Specifically, S1 denotes that we removed the DCN's stage 2, 3, 4 and continued the training of stage 1 until 40000 iterations (to guarantee unified iteration numbers for fair comparison), S12 denotes that we removed the DCN's stage 3, 4 and continued the training of stage 2 until 30000 iterations (to guarantee unified iteration numbers for fair comparison), S123 denotes that we removed the DCN's stage 4 and continued the training of stage 3 until 20000 iterations (to guarantee unified iteration numbers for fair comparison), and S1234 denotes normal DCN training including four stages. Additionally, we removed the termination thresholds $Th1$, $Th2$, and $Th3$ for S1, S12, and S123 respectively. The models with the lowest MAE on validation set (200 subjects with 400 labeled volumes) were used for residual analysis. The residual analysis results are shown in Fig. 3, Fig. 4 and Fig. 5. We can see that every incremental stage can improve the validation accuracy. Besides, the three backbone networks (VGG16, ResNet50, and DenseNet121) have consistent tendencies along with incremental stages. This phenomenon indicates that every training stage of DCN is important for model optimization, and multi-stage training with a dynamic network structure can improve the ventricle volume estimation accuracy.

Finally, we also evaluated DCN(S1234)'s performance on the testing set of Kaggle2016 dataset (440 subjects with 880 labeled volumes) by the correlation analysis between the ground truth volume and the estimated volume. As shown in the Fig. 6, DCN achieved higher correlation coefficients with different backbone networks (0.981, 0.986, and 0.988 for VGG16, ResNet50, and DenseNet121 respectively). These results show that the proposed method has high estimation

performance in the aspect of the correlation between the ground truth volume and the estimated volume.

3.3.3. Optimization performance of the added residual nodes

In the proposed method, the accurate residual estimation plays an important role for reducing the error of the volume estimation. To evaluate the performance of residual estimation, we analyzed the correlation between the ground truth residuals and the estimated residuals. Specifically, the estimated residuals after the last iteration in stage 3 and in stage 4 were analyzed respectively based on the validation set of Kaggle2016 dataset.

As shown in Fig. 7, the proposed method achieved higher correlation between the ground truth residuals and the estimated residuals. The correlation coefficient achieves 0.967, 0.9656, and 0.954 for VGG16, ResNet50, and DenseNet121 respectively after stage 3. The correlation coefficient achieves 0.9961, 0.9937, and 0.9897 for VGG16, ResNet50, and DenseNet121 respectively after stage 4. These results indicate that the proposed method can achieve relatively accurate residual estimation based on the added residual nodes after stage 3. Additionally, the correlation coefficient was improved after optimization in stage 4. This phenomenon indicates that the performance of residual estimation can be further improved based on the synchronous optimization of stage 4.

3.4. Significance of bi-model network for independent optimization of EDV and ESV

The bi-model network inherently is an improvement for the proposed DCN. To show the significance of the bi-model network for independent optimization of EDV and ESV, we conducted the ablation experiments (i.e., D_V_S1, D_V_B, and D_V_S12) for DCN with VGG16 as the backbone network (The experimental results of the ablation models with other backbone networks were similar with the results of the ablation models with VGG16. Hence, to avoid redundancy, in this section, we only showed the experimental results of the ablation model with VGG16 as backbone network).

Specifically, D_V_S1 denotes S1 with VGG16 as backbone, D_V_B denotes S12 with VGG16 as backbone and without the EF regularization constraint in the $loss2$, and D_V_S12 denotes S12 with VGG16 as backbone and with the EF regularization constraint. Additionally, we removed the termination thresholds $Th1$ for D_V_S1, and removed the termination thresholds $Th2$ for D_V_B and D_V_S12. The models with the lowest MAE on the validation set (200 subjects with 400 labeled volumes) were used for experiment analysis.

As shown in Tables 5 and 6, for LVV estimation and RVV estimation, D_V_B always achieved lower estimation MAE than D_V_S. Besides, D_V_S12 always achieved lower estimation MAE than D_V_B. This phenomenon indicates that the bi-model network inherently improved the performance of both EDV estimation and ESV estimation. At the same time, this phenomenon also shows that the EF regularization constraint is a valuable constraint, which further improves the volume estimation accuracy.

3.5. Application performance evaluation

To show the generalization ability and clinical application potential, we conducted comparison experiments with state-of-the-art methods on multi-scale CMR datasets for LVV and RVV estimation performance evaluation. For LVV and LV EF estimation performance evaluation, we conducted comparison experiments with methods (Zhen et al., 2016), (Liao et al., 2017), (Ngo et al., 2017), and (Avendi et al., 2016) (Up to now, in all the published papers, (Zhen et al., 2016) achieved best direct bi-ventricle volume prediction accuracy on a private dataset, (Liao et al., 2017) achieved the best direct LVV estimation accuracy on Kaggle2016 dataset, (Ngo et al., 2017) and (Avendi et al., 2016) achieved the best results for LV segmentation.), and results were shown in Tables 2 and 4. For RVV and RV EF estimation performance evaluation, we conducted comparison experiments with methods (Zhen et al., 2016), (Luo et al., 2016), and (Avendi et al., 2017) (Up to now, in all the published papers, (Zhen et al., 2016) achieved the best direct bi-ventricle volume prediction accuracy on a private dataset, (Luo et al., 2016) and (Avendi et al., 2017) achieved the best results for RV

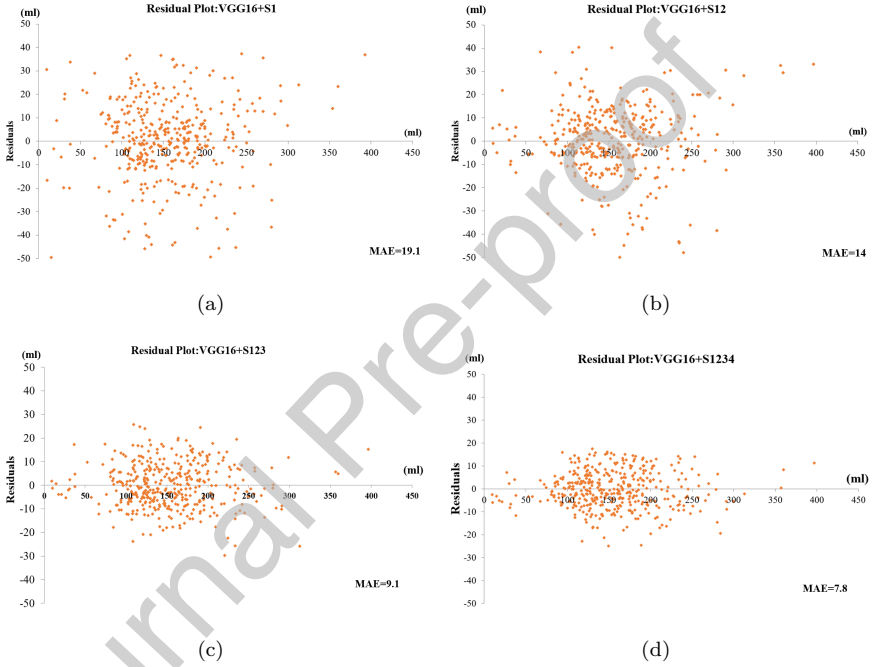


Figure 3: The volume estimation residual analysis plot with stage ablation experiments on Kaggle2016 dataset.The residual plot with stage ablation using VGG16 as the backbone network.

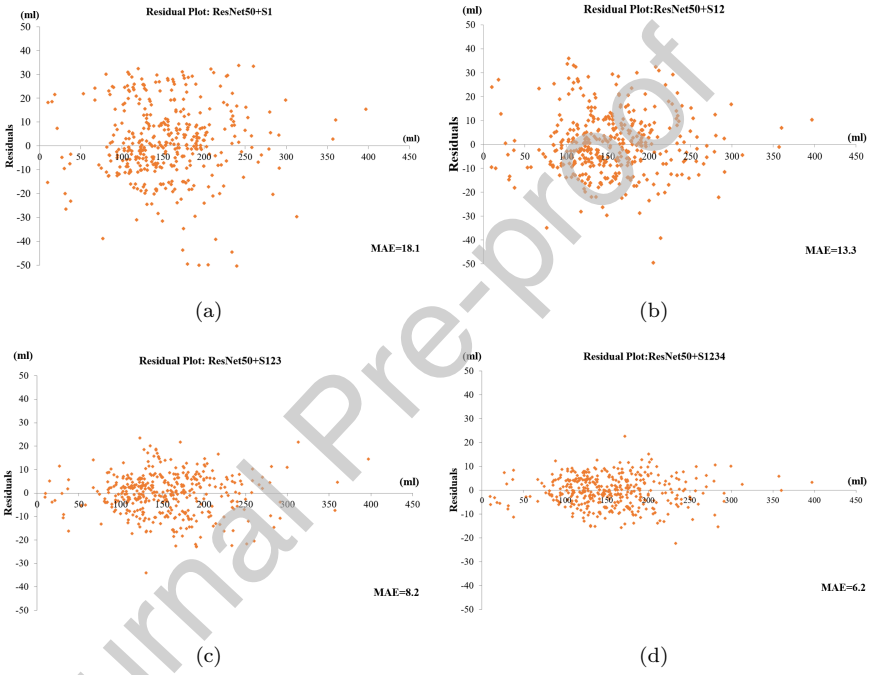


Figure 4: The volume estimation residual analysis plot with stage ablation experiments on Kaggle2016 dataset.The residual plot with stage ablation using ResNet50 as the backbone network.

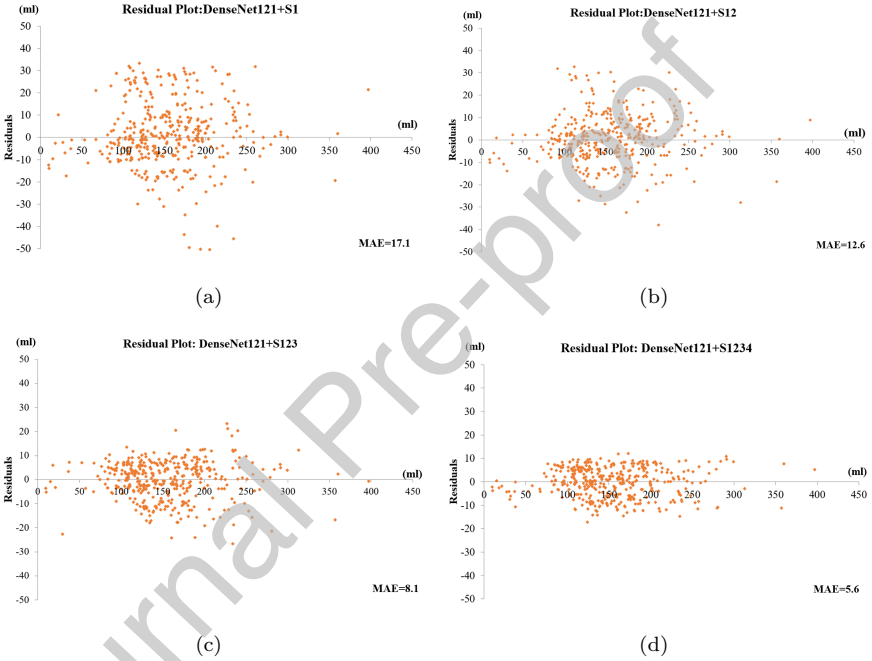
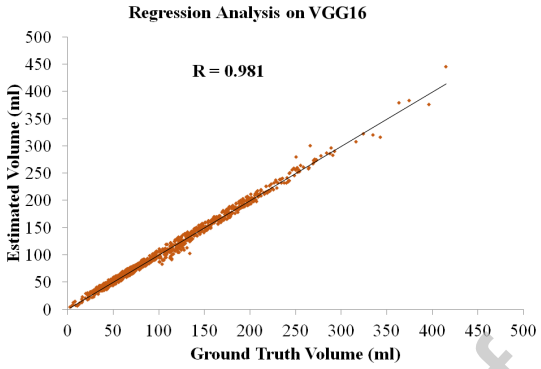
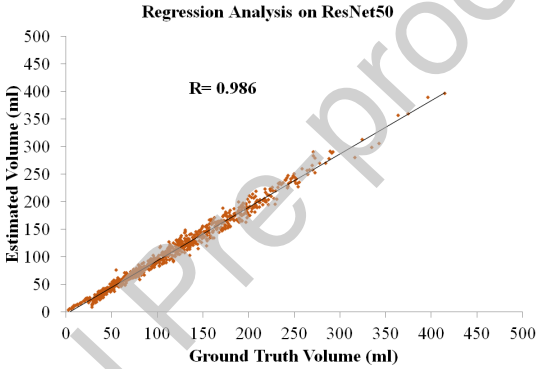


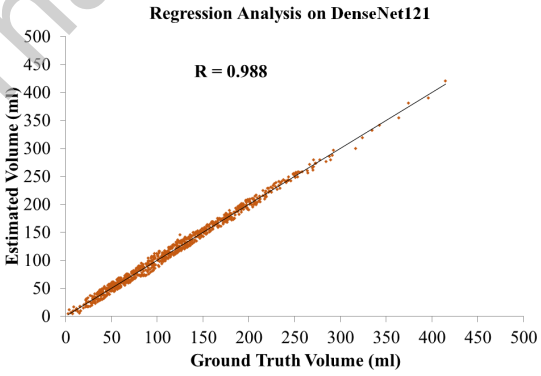
Figure 5: The volume estimation residual analysis plot with stage ablation experiments on Kaggle2016 dataset. The residual plot with stage ablation using DenseNet121 as the backbone network.



(a)



(b)



(c)

Figure 6: The correlation analysis between the ground truth volume and the estimated volume. a) The backbone network is VGG16. b) The backbone network is ResNet50. c) The backbone network is DenseNet121.

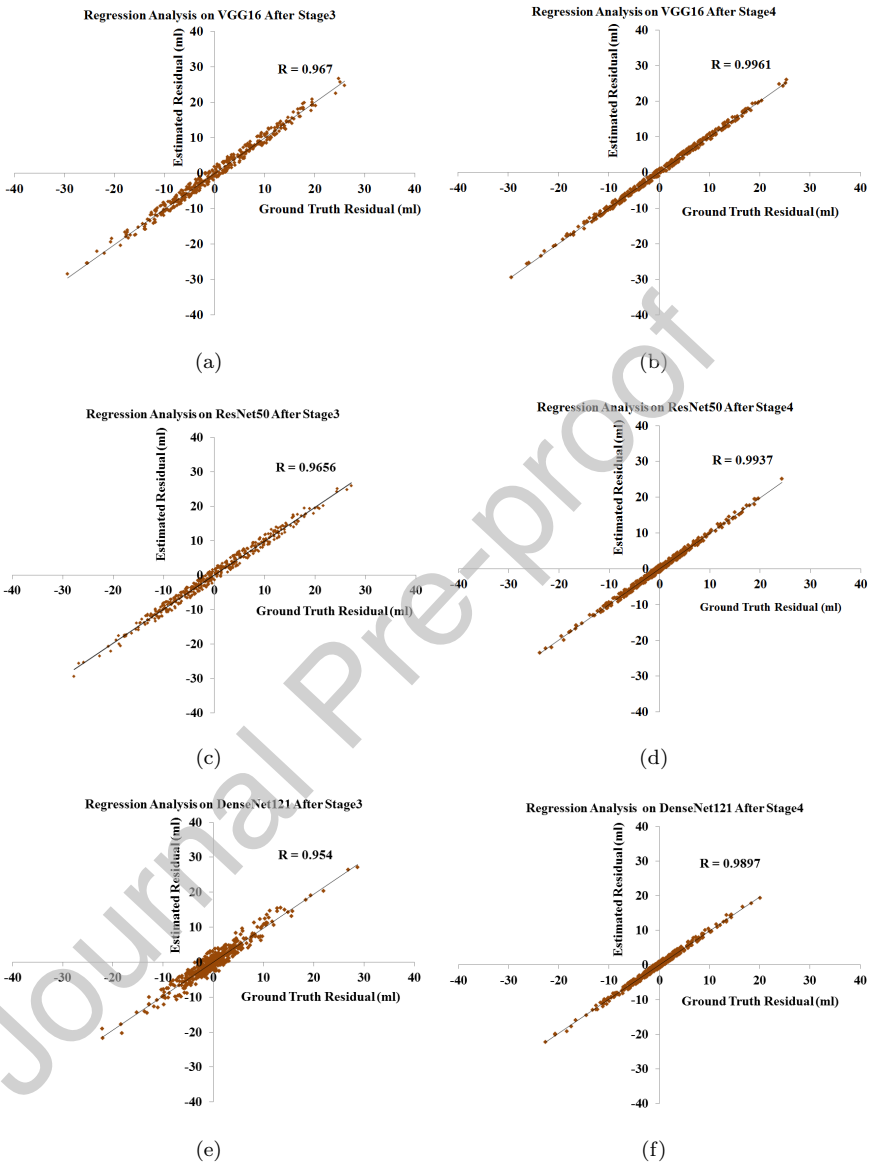


Figure 7: The correlation analysis between the ground truth residual and the estimated residual. a) The result with VGG16 after stage 3. b) The result with VGG16 after stage 4. c) The result with ResNet50 after stage 3. d) The result with ResNet50 after stage 4 e) The result with DenseNet121 after stage 3. f) The result with DenseNet121 after stage 4.

Table 2: Performance of DCN under different backbone network configurations and its competitors for LVV estimation on ED and ES frames. MAE± standard deviation (std) is used as metric of accuracy evaluation. D_V denotes DCN with VGG16 as backbone network, D_R denotes DCN with ResNet50 as backbone network, and D_D denotes DCN with DenseNet121 as backbone network.

Datasets	STACOM2011(ml)		ACDC2017(ml)		Kaggle2016(ml)	
Methods	EDV	ESV	EDV	ESV	EDV	ESV
D_V	**9.6±3.1**	**6.9±3.5**	**9.2±4.5**	**5.9±4.5**	**9.4±5.1**	**5.1±3.1**
D_R	**8.9±3.2**	**5.8±3.9**	**9.5±5.5**	**6.2±2.3**	**9.2±2.9**	**4.1±2.4**
D_D	**8.7±2.4**	**4.7±2.3**	**8.9±1.6**	**5.5±3.3**	**8.1±1.5**	**3.6±1.9**
(Zhen et al., 2016)	13.9±9.2	9.8±8.9	14.8±8.9	10.2±7.9	12.5±6.1	9.1±7.2
(Liao et al., 2017)	12.1±8.3	9.6±9.5	15.8±9.6	9.9±9.5	11.7±8	8.9±5.1
(Ngo et al., 2017)	15.1±15.2	16.6±10.4	17.1±11.5	16.8±12.5	17±9.1	15.5±10.1
(Avendi et al., 2016)	14.9±12.1	15.8±12.4	17.5±16.9	19.2±20.3	16±9.5	14.5±12.3

Table 3: Performance of DCN under different backbone network configurations and its competitors for RVV estimation on ED and ES frames. MAE±std is used as metric of accuracy evaluation. D_V denotes DCN with VGG16 as backbone network, D_R denotes DCN with ResNet50 as backbone network, and D_D denotes DCN with DenseNet121 as backbone network.

Datasets	ACDC2017		RV2012	
Methods	EDV(ml)	ESV(ml)	EDV(ml)	ESV(ml)
D_V	**8.1±3.5**	**4.9±3.1**	**9.3±5.5**	**9.6±5.2**
D_R	**7.9±2.1**	**4.5±2.6**	**9.5±3.5**	**7.3±2.2**
D_D	**7.7±1.1**	**3.9±2.3**	**9.2±2.1**	**6.9±1.5**
(Zhen et al., 2016)	12.4±5.2	10.9±9.7	14.1±7.6	10.9±8.9
(Luo et al., 2016)	14.5±8.4	13.6±15.5	15.8±8.6	12.9±10.9
(Avendi et al., 2017)	16.1±13.1	14.1±16.6	16.2±10.5	15.9±13.5

Table 4: Performance of DCN under different backbone network configurations and its competitors for EF estimation. MAE±std is used as metric of accuracy evaluation. N/A denotes 'not suitable'. D_V denotes DCN with VGG16 as backbone network, D_R denotes DCN with ResNet50 as backbone network, and D_D denotes DCN with DenseNet121 as backbone network.

indices	RVEF(%)		LVEF(%)		
Methods	ACDC2017	RV2012	STACOM2011	ACDC2017	Kaggle2016
D_V	**4.1±3.5**	**3.7±2.1**	**3.6±2.3**	**4.6±3.3**	**3.5±2.4**
D_R	**3.9±2.5**	**3.5±2.5**	**2.9±2.5**	**4.1±2.1**	**2.9±2.1**
D_D	**3.7±1.1**	**3.9±1.3**	**2.1±1.5**	**3.9±1.4**	**2.6±1.5**
(Zhen et al., 2016)	7.4±6.1	8.2±4.7	8.2±6.3	8.9±6.1	7.1±5.3
(Liao et al., 2017)	N/A	N/A	8.1±6.6	7.9±6.4	7.7±5.1
(Ngo et al., 2017)	N/A	N/A	9.1±8.9	9.2±6.5	9.8±6.1
(Avendi et al., 2016)	N/A	N/A	9.9±8.3	9.7±8.3	8.9±7.5
(Luo et al., 2016)	8.5±6.4	9.6±5.5	N/A	N/A	N/A
(Avendi et al., 2017)	8.1±7.9	9.2±7.6	N/A	N/A	N/A

Table 5: Performance comparison between bi-model network and single-model network for LVV estimation on ED and ES frames. MAE±std is used as metric of accuracy evaluation.

Datasets	STACOM2011(ml)		ACDC2017(ml)		Kaggle2016(ml)	
Methods	EDV	ESV	EDV	ESV	EDV	ESV
D_V_S1	12.7±4.2	10.2±5.5	12.2±6.3	10.4±4.9	11.5±5.9	9.4±4.3
D_V_B	**10.9±3.8**	**9.3±5.1**	**11.3±5.5**	**9.6±4.8**	**10.2±5.4**	**8.8±3.5**
D_V_S12	**10.1±3.5**	**8.7±4.3**	**10.8±5.2**	**7.9±4.7**	**9.9±5.5**	**7.6±3.9**

Table 6: Performance comparison between bi-model network and single-model network for RVV estimation on ED and ES frames. MAE±std is used as metric of accuracy evaluation.

Datasets	ACDC2017		RV2012	
Methods	EDV(ml)	ESV(ml)	EDV(ml)	ESV(ml)
D_V_S1	12.5±5.9	11.9±6.9	13.1±5.2	11.6±6.2
D_V_B	**11.6±5.5**	**10.8±6.1**	**12.4±4.7**	**10.9±5.5**
D_V_S12	**10.3±5.3**	**9.6±4.5**	**11.9±4.1**	**10.2±5.2**

segmentation.), and results were shown in Tables 3 and 4. We can see that the proposed DCN achieved the lowest MAE on RVV, LVV and EF estimation, compared to the state-of-the-art ventricle volume estimation methods, though results were different when DCN was configured with different backbone networks. These results show that the proposed DCN had a strong generalization ability on multi-scale datasets, and also show that the proposed method had high robustness for different backbone network configurations. Additionally, we also found some phenomenons, which were useful for further research. First, the DCN using DenseNet121 as backbone network achieved relatively more accurate ventricle estimation than the other two popular backbone networks (The reason may be that DenseNet has more skip connections between different layers). Second, the direct ventricle estimation methods (DCN, (Zhen et al., 2016), and (Liao et al., 2017)) were more accurate than the indirect segmentation methods ((Ngo et al., 2017), (Avendi et al., 2016), (Luo et al., 2016), and (Avendi et al., 2017)). These two phenomenons showed that the proposed DCN (Direct estimation method) had bigger clinical application potential.

4. Discussion

In this paper, we proposed a novel dynamic construction DL network to address the ventricle volume estimation problem. Compared to the state-of-the-art methods, the proposed DCN achieved better optimization performance and volume estimation accuracy on LVV and RVV across different datasets, and had good generalization ability. In this section, we discussed the proposed method on three aspects, i.e., the insight of the direct volume estimation method, the advantage of the proposed method, and the limitations and challenges of the proposed method.

Potential of direct volume estimation method: Compared with indirect volume estimation method, the direct volume estimation method has greater potential in the following aspects:

1) Analyzing large scale historical data with high computational efficiency:

the historical data analysis is necessary for clinical diagnosis and treatment, along with the increase of clinical data. During analyzing such large numbers of historical data (about more than 240 CMR images for every patient), a clinician generally focuses on some important indexes (for instance, volume and ejection fraction). In this case, the direct index estimation strategy is a good choice for analyzing historical data efficiently.

2) Data screening strategy for a population study of cardiac disease: The population study of heart disease is important for clinical scientists. The direct estimation is an effective data screening strategy for a population study. For large numbers of CMR data, checking segmentation result slice-by-slice is very time-consuming and impossible for clinicians in most of time. After the effective data screening, clinician can further conduct analysis, research for the selected or filtered data. In this way, the direct estimation method not only saves time but also makes analysis and research more targeted.

3) Utilizing large numbers of historical CMR data without segmentation records to build a more accurate direct estimation model: The direct volume estimation strategy is the only way to utilize large numbers of historical CMR data without segmentation results to achieve good volume estimation based on deep learning. Specifically, in a hospital, we find that there are a large number of historical CMR data only with ventricle quantification index records and without segmentation records. How to utilize such large-scale data to train accurate ventricle volume estimation model is a significant research topic, especially for deep learning methods, which relies heavily on data.

Whats more, the direct volume estimation is a representative example to show the value of the proposed dynamic construction network and residual correction strategy. The potential applications of the proposed method may be not limited to the above three aspects. It also has great potential to be extended to other applications or research fields.

Advantages of the proposed method: In the proposed framework, the good performance and application potential benefit from the following three aspects:

1) The dynamic construction from a single-model network to a bi-model network enables the modelling of the EF constraint, which has significant clinical meaning and a continuously differentiable property. Hence, finally the dynamic network with EF constraint enables high-efficiency optimization and good ventricle estimation performance.

2) Dynamically constructed DL framework enables volume residual modeling through an explicit manner. It is different from the latent residual modeling in most of traditional DL technology through identity mapping (He et al., 2016) (Huang et al., 2017), i.e., skip structure. Experimental results show that the residual modeling strategy in this paper can improve the ventricle estimation accuracy, and may be useful for other similar regression tasks.

3) Compared with the traditional direct volume estimation methods based on DL technology, the proposed method has an interpretation-ability to some extent. Every stage of the network structure construction has a clear meaning, because the dynamic and incremental network structure can provide interpretation for model construction. In this aspect, the proposed DCN improves the interpretation ability of traditional direct ventricle volume estimation methods based on DL technology.

Limitation and challenge: The main limitation of the proposed framework is that it cannot provide visual interpretation (segmentation results) for the estimated volume to a physician, though the interpretation-ability of the proposed framework has been improved compared with the state-of-the-art direct volume estimation methods. Hence, this limitation is also a challenge for the proposed framework. In the future, we will further improve the interpretation-ability of the proposed framework and collect more CMR data to boost the proposed method's performance and achieve a more accurate volume estimation for clinical application.

5. Conclusion

In this paper, we proposed a dynamically constructed DL framework to achieve error correction for accurate ventricle volume estimation. The proposed DCN enables the modelling of the EF correlation and the modelling of volume residual by dynamic network construction, dynamic optimization, and stochastic configuration theory. The experimental results on the large-scale CMR datasets show that the proposed method can achieve a more accurate volume estimation than state-of-the-art methods, and has high robustness as well as strong generalization ability. Moreover, the proposed DCN not only improves the volume estimation task, but also has potential to be extended to other medical index estimation tasks.

6. Acknowledgement

This work was supported by the National Key R&D Program of China under Grant 2017YFC0113000 and by the National Natural Science Foundation of China under Grant 61701135. The authors thank Dr. Olga Shuilovich, Dr. Yue Zhang, and Dr. Ashley Mercado for providing the large amount of labelled data, as well as Nvidia for donating the Titan X.

Appendix A. The detailed deduction for optimization of *Loss*4

The detailed deduction of back propagation for $Loss4$ in the stage four is shown as follows:

$$\begin{aligned}
\frac{\partial Loss4}{\partial \varpi} &= \frac{1}{N}\sum_{n=1}^{N}\Bigg\{(EDV_E^n - EDV_T^n)\frac{\partial EDV_E^n}{\partial \varpi} + (ESV_E^n - ESV_T^n)\frac{\partial ESV_E^n}{\partial \varpi} \\
&+ \left(\varepsilon_{EDV}^{E,n} - \varepsilon_{EDV}^{T,n}\right)\left(\frac{\partial \varepsilon_{EDV}^{E,n}}{\partial \varpi} - \frac{\partial \varepsilon_{EDV}^{T,n}}{\partial \varpi}\right) + \left(\varepsilon_{ESV}^{E,n} - \varepsilon_{ESV}^{T,n}\right)\left(\frac{\partial \varepsilon_{ESV}^{E,n}}{\partial \varpi} - \frac{\partial \varepsilon_{ESV}^{T,n}}{\partial \varpi}\right)\Bigg\} \\
&+ \gamma_3 \frac{\partial EF_C}{\partial \varpi}, \\
&= \frac{1}{N}\sum_{n=1}^{N}\Bigg\{(EDV_E^n - EDV_T^n)\frac{\partial EDV_E^n}{\partial \varpi} + (ESV_E^n - ESV_T^n)\frac{\partial ESV_E^n}{\partial \varpi} \\
&+ \left(\varepsilon_{EDV}^{E,n} - (EDV_T^n - EDV_E^n)\right)\left(\frac{\partial \varepsilon_{EDV}^{E,n}}{\partial \varpi} - \frac{\partial (EDV_T^n - EDV_E^n)}{\partial \varpi}\right) \\
&+ \left(\varepsilon_{ESV}^{E,n} - (ESV_T^n - ESV_E^n)\right)\left(\frac{\partial \varepsilon_{ESV}^{E,n}}{\partial \varpi} - \frac{\partial (ESV_T^n - ESV_E^n)}{\partial \varpi}\right)\Bigg\} \\
&+ \gamma_3 \frac{\partial EF_C}{\partial \varpi}, \\
&= \frac{1}{N}\sum_{n=1}^{N}\Bigg\{(EDV_E^n - EDV_T^n)\frac{\partial EDV_E^n}{\partial \varpi} + (ESV_E^n - ESV_T^n)\frac{\partial ESV_E^n}{\partial \varpi} \\
&+ \left(\varepsilon_{EDV}^{E,n} - (EDV_T^n - EDV_E^n)\right)\left(\frac{\partial \varepsilon_{EDV}^{E,n}}{\partial \varpi} + \frac{\partial EDV_E^n}{\partial \varpi}\right) \\
&+ \left(\varepsilon_{ESV}^{E,n} - (ESV_T^n - ESV_E^n)\right)\left(\frac{\partial \varepsilon_{ESV}^{E,n}}{\partial \varpi} + \frac{\partial ESV_E^n}{\partial \varpi}\right)\Bigg\} \\
&+ \gamma_3 \frac{\partial EF_C}{\partial \varpi},
\end{aligned} \tag{A.1}$$

where the parameters ϖ, $\varepsilon_{EDV}^{E,n}$, $\varepsilon_{ESV}^{E,n}$, ESV_E^n, EDV_E^n, $\varepsilon_{EDV}^{T,n}$, $\varepsilon_{ESV}^{T,n}$, ESV_T^n, and EDV_T^n have same meaning with equation (10) and (12).

References

Afshin, M., Ayed, I. B., Islam, A., Goela, A., Peters, T. M., Li, S., 2012. Global assessment of cardiac function using image statistics in mri. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 535–543.

Afshin, M., Ayed, I. B., Punithakumar, K., Law, M. W., Islam, A., Goela, A., Ross, I., Peters, T., Li, S., 2011. Assessment of regional myocardial function

via statistical features in mr images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 107–114.

Albà, X., Lekadir, K., Pereañez, M., Medrano-Gracia, P., Young, A. A., Frangi, A. F., 2018. Automatic initialization and quality control of large-scale cardiac mri segmentations. Medical image analysis 43, 129–141.

Avendi, M., Kheradvar, A., Jafarkhani, H., 2016. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac mri. Medical image analysis 30, 108–119.

Avendi, M. R., Kheradvar, A., Jafarkhani, H., 2017. Automatic segmentation of the right ventricle from cardiac mri using a learning-based approach. Magnetic resonance in medicine 78 (6), 2439–2448.

Bai, W., Shi, W., O'Regan, D. P., Tong, T., Wang, H., Jamil-Copley, S., Peters, N. S., Rueckert, D., 2013. A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: application to cardiac mr images. IEEE transactions on medical imaging 32 (7), 1302–1315.

Barron, A. R., 1993. Universal approximation bounds for superpositions of a sigmoidal function. IEEE Transactions on Information theory 39 (3), 930–945.

Bernard, O., A. Lalande, e., 2018. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: Is the problem solved? IEEE Transactions on Medical Imaging, 1–1.

Cremers, D., Rousson, M., Deriche, R., 2007. A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. International journal of computer vision 72 (2), 195–215.

Dong, S., Luo, G., Tam, C., Wang, W., Wang, K., Cao, S., Chen, B., Zhang, H., Li, S., 2020. Deep atlas network for efficient 3d left ventricle segmentation on echocardiography. Medical image analysis 61, 101638.

Eslami, A., Karamalis, A., Katouzian, A., Navab, N., 2013. Segmentation by retrieval with guided random walks: application to left ventricle segmentation in mri. Medical image analysis 17 (2), 236–253.

Ge, R., Yang, G., Chen, Y., Luo, L., Feng, C., Ma, H., Ren, J., Li, S., 2020. K-net: Integrate left ventricle segmentation and direct quantification of paired echo sequence. IEEE Transactions on Medical Imaging 39 (5), 1690–1702.

Ge, R., Yang, G., Chen, Y., Luo, L., Feng, C., Zhang, H., Li, S., 2019. Pv-lvnet: Direct left ventricle multitype indices estimation from 2d echocardiograms of paired apical views with deep neural networks. Medical image analysis 58, 101554.

Gu, B., Shan, Y., Sheng, V. S., Zheng, Y., Li, S., 2018. Sparse regression with output correlation for cardiac ejection fraction estimation. Information Sciences 423, 303–312.

Hajiaghayi, M., Groves, E. M., Jafarkhani, H., Kheradvar, A., 2017. A 3-d active contour method for automated segmentation of the left ventricle from magnetic resonance images. IEEE Transactions on Biomedical Engineering 64 (1), 134–144.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778.

Huang, G., Liu, Z., Weinberger, K. Q., van der Maaten, L., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. Vol. 1. p. 3.

Kaggle, 2016. the 2016 data science bowl cardiac challenge data in kaggle (ds-bccd). https://www.kaggle.com/c/second-annual-data-science-bowl/data.

Kass, M., Witkin, A., Terzopoulos, D., 1988. Snakes: Active contour models. International journal of computer vision 1 (4), 321–331.

Khalifa, F., Beache, G. M., Gimelrfarb, G., Giridharan, G. A., El-Baz, A., 2012. Accurate automatic analysis of cardiac cine images. IEEE Transactions on Biomedical Engineering 59 (2), 445–455.

Kwok, T.-Y., Yeung, D.-Y., 1997. Objective functions for training new hidden units in constructive neural networks. IEEE Transactions on neural networks 8 (5), 1131–1148.

Li, M., Wang, D., 2017. Insights into randomized algorithms for neural networks: Practical issues and common pitfalls. Information Sciences 382, 170–178.

Liao, F., Chen, X., Hu, X., Song, S., 2017. Estimation of the volume of the left ventricle from mri images using deep neural networks. IEEE Transactions on Cybernetics 1 (1), 1–1.

Luo, G., An, R., Wang, K., Dong, S., Zhang, H., 2016. A deep learning network for right ventricle segmentation in short-axis mri. In: Computing in Cardiology Conference (CinC), 2016. IEEE, pp. 485–488.

Luo, G., Dong, S., Wang, K., Zuo, W., Cao, S., Zhang, H., 2018. Multi-views fusion cnn for left ventricular volumes estimation on cardiac mr images. IEEE Transactions on Biomedical Engineering 65 (9), 1924–1934.

Luo, G., Dong, S., Wang, W., Wang, K., Cao, S., Tam, C., Zhang, H., Howey, J., Ohorodnyk, P., Li, S., 2019. Commensal correlation network between segmentation and direct area estimation for bi-ventricle quantifcation. Medical image analysis 59, 11253.

Mo, Y., Liu, F., McIlwraith, D., Yang, G., Zhang, J., He, T., Guo, Y., 2018. The deep poincaré map: A novel approach for left ventricle segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 561–568.

Ngo, T. A., Lu, Z., Carneiro, G., 2017. Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance. Medical image analysis 35, 159–171.

Oguz, I., Kashyap, S., Wang, H., Yushkevich, P., Sonka, M., 2016. Globally optimal label fusion with shape priors. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 538–546.

Petitjean, C., Dacher, J. N., 2011. A review of segmentation methods in short axis cardiac mr images. Medical Image Analysis 15(2), 169–184.

Petitjean, C., Zuluaga, M. A., Bai, W., Dacher, J.-N., Grosgeorge, D., Caudron, J., Ruan, S., Ayed, I. B., Cardoso, M. J., Chen, H.-C., et al., 2015. Right ventricle segmentation from cardiac mri: a collation study. Medical image analysis 19 (1), 187–202.

Qin, C., Bai, W., Schlemper, J., Petersen, S. E., Piechnik, S. K., Neubauer, S., Rueckert, D., 2018. Joint learning of motion estimation and segmentation for cardiac mr image sequences. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. Springer International Publishing, Cham, pp. 472–480.

Radau, P., Lu, Y., Connelly, K., Paul, G., Dick, A., Wright, G., 2009. Evaluation framework for algorithms segmenting short axis cardiac mri. The MIDAS Journal-Cardiac MR Left Ventricle Segmentation Challenge 49.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Suinesiaputra, A., Cowan, B. R., Al-Agamy, A. O., Elattar, M. A., Ayache, N., Fahmy, A. S., Khalifa, A. M., Medrano-Gracia, P., Jolly, M.-P., Kadish, A. H., et al., 2014. A collaborative resource to build consensus for automated left ventricular segmentation of cardiac mr images. Medical image analysis 18 (1), 50–62.

Tan, L. K., Liew, Y. M., Lim, E., McLaughlin, R. A., 2017. Convolutional neural network regression for short-axis left ventricle segmentation in cardiac cine mr sequences. Medical image analysis 39, 78–86.