IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 12, 2019, accepted March 13, 2019, date of publication March 26, 2019, date of current version April 19, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2907564

Quantification of Full Left Ventricular Metrics via Deep Regression Learning With Contour-Guidance

WENJI WANG[1,2], YUANQUAN WANG[1], YUWEI WU[2], TAO LIN[1], SHUO LI[3], AND BO CHEN[3]

[1]School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China
[2]School of Computer Science, Beijing Institute of Technology, Beijing 100081, China
[3]Digital Imaging Group of London, Department of Medical Imaging, Western University, London N6A 4V2, Canada

Corresponding authors: Yuanquan Wang (wangyuanquan@scse.hebut.edu.cn) and Shuo Li (slishuo@gmail.com)

This work was supported in part by the National Science Foundation Program of China (NSFC) under Grant 61702037, in part by the Beijing Municipal Natural Science Foundation under Grant L172027, in part by the Key Program from NSF of Hebei Province under Grant F2016202144, in part by the NSF of Tianjin under Grant 16JCYBJC15600, and in part by the Youth Fund from the Department of Education of Hebei Province under Grant QN2016217.

ABSTRACT Quantifying full left ventricular (LV) metrics including cavity area, myocardium area, cavity dimensions and wall thicknesses from cardiac magnetic resonance (MR) images, and then assessing regional and global cardiac function plays a crucial role in clinical practice. However, due to highly variable cardiac structures across different subjects, it is challenging to obtain an accurate estimation of LV metrics. In this paper, we propose a novel deep learning framework, called cascaded segmentation and regression network (CSRNet), to improve the quantification results. The CSRNet consists of two components: a segmentation component and a regression component. The segmentation component yields myocardial contours of the left ventricle from the input cardiac MR images, and then the regression component learns hierarchical representations from the segmented images and estimates the desired LV metrics. By introducing the myocardial contours, the regression component can pay more attention to the left ventricle, which contributes to more accurate quantification results, although the cardiac structures are variable. The extensive experiments on a dataset of 145 subjects demonstrate that our framework outperforms the state-of-the-art methods.

INDEX TERMS Cardiac magnetic resonance images, quantification, left ventricular metrics, deep learning model, regression.

I. INTRODUCTION

Magnetic resonance imaging (MRI) is one of the most popular medical imaging modalities to detect cardiovascular disease due to its noninvasive and versatile nature. Quantification of full left ventricular (LV) metrics (as shown in Fig. 1) including cavity area, myocardium area, cavity dimensions and wall thicknesses from magnetic resonance (MR) images is of great significance in clinical practice. On the one hand, the heart volume can be obtained by accumulating cavity and myocardium areas spatially from the base to the apex and multiplying the slice thickness, and then the cardiac function parameters such as stroke volume, ejection fraction can be computed for assessing global cardiac function. On the other hand, cavity dimensions and wall thicknesses can provide regional cardiac function assessment and conduce to early disease diagnosis such as cardiac hypertrophy and myocardial infarction.

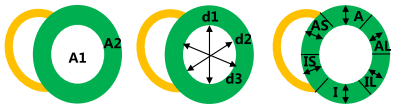


FIGURE 1. Full LV metrics to be quantified. (a) A1: cavity area, A2: myocardium area (b) d1-d3: cavity dimensions (c) A-AS: wall thicknesses.

There are many researches have been conducted to quantify the LV metrics for years. We roughly divide the existing methods into two-step methods [1]–[9] and end-to-end methods [10]–[15] according to how they are carried out. Two-step methods are usually carried out in two manners. One is based on segmentation [1], [3]–[5], where the endo- and epi-cardium of the left ventricle are extracted first, and then the desired LV metrics are estimated based on segmentation results. The other manner takes advantage of machine learning algorithms [7]–[9], where features are extracted from

The associate editor coordinating the review of this manuscript and approving it for publication was Mufti Mahmud.

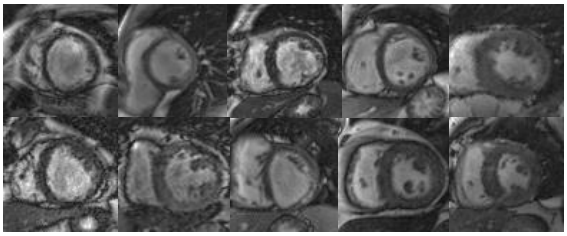


FIGURE 2. There are highly variable structures across different subjects, containing the variations of shape, size and papillary muscle in the blood pool, which is a big challenge when extracting robust representation of the left ventricle.

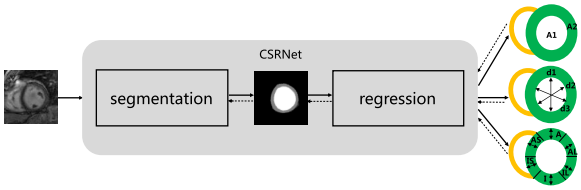


FIGURE 3. The overview of the proposed CSRNet, which is composed of segmentation component and regression component to estimate full LV metrics. The solid arrows denote forward propagation while the dashed arrows indicate backward.

the cardiac MR images and a regression learning model is followed to estimate the LV metrics. In both two-step manners, the two steps are independent of each other and the estimated LV metrics are heavily dependent on the results of first step, i.e., segmented contours or extracted features.

In contrast to the two-step methods, end-to-end methods [10]–[15] combing feature extraction and metrics regression together via using deep neural networks, have attracted considerable attention in very recent years. In this kind of methods, the Convolutional Neural Network (CNN) is usually utilized to learn common representations from cardiac MR images of different subjects, and the Recurrent Neural Network (RNN) is used to capture temporal dynamics over a cardiac cycle. However, due to the variations of the left ventricle in shape, size and papillary muscle in the blood pool across different subjects which are shown in FIG. 2, it is challenging to obtain accurate estimation of the LV metrics.

In this paper, we propose to introduce the myocardial contours of the left ventricle into an end-to-end deep learning framework to improve quantification results of LV metrics. The proposed end-to-end framework, called Cascaded Segmentation and Regression Network (CSRNet), which is shown in FIG.3. The CSRNet consists of two components, i.e., segmentation component and regression component. The segmentation component classifies each pixel in the cardiac MR images into one of the three distinct categories, i.e., background, myocardium and cavity, aiming to extract the myocardial contours of the left ventricle and remove those task-unrelated structures. The regression component extracts features from the segmented images and estimates the desired LV metrics. Since the proposed CSRNet takes into account the myocardial contours of the left ventricle, it has great superiority in learning more robust representation from images and thus getting more accurate quantification results. For clarity, the main contributions of this work are summarized as follows:

1) We propose an effective end-to-end framework to improve quantification results of full LV metrics, which includes two areas, three cavity dimensions and six regional wall thicknesses, aiming at precise regional and global cardiac function assessment.
2) The proposed framework is called Cascaded Segmentation and Regression Network, CSRNet in short, which integrates a segmentation module into a regression learning framework.
3) Experiments on a dataset of 145 subjects demonstrate the effectiveness of our framework. When compared to the state-of-the-art methods, the CSRNet achieves the lowest average Mean Absolute Error (MAE) of 134 mm^2, 1.54 mm, 1.16 mm and the largest average Correlation Coefficient ρ of 0.965, 0.978, 0.868 for areas, dimensions, and thicknesses, respectively.

The remainder of this paper is organized as follows: some works related to quantification of LV metrics are introduced briefly in Section II, and the details of the proposed CSRNet are presented in Section III. Following that, experiment settings are introduced in Section IV, and experimental results and discussions are reported in Section V. Finally, the conclusion is drawn in Section VI.

II. RELATED WORKS

A. TWO-STEP METHODS FOR QUANTIFICATION OF LV METRICS

The two-step methods for quantification of LV metrics are carried out in two manners. The first manner is based on segmentation, in which the myocardial contours of the LV are first extracted manually or automatically, and then the desired LV metrics are measured based on the segmentation results. Suinesiaputra *et al.* [16] invited seven experienced experts to depict the myocardial contours and further calculated the LV metrics including cavity volumes at end-systole (ES) and end-diastole (ED) and myocardial mass at ED to assess cardiac function. However, manual segmentation is time-consuming, experience-driven and irreproducible. To circumvent these limitations, some automatic segmentation algorithms, such as deformable models [1], [3], [17], [18], statistical models [4], [19], and deep neural networks [20]–[22] were proposed. Among them, deformable models and statistical models generally require prior knowledge or/and user interaction, which leads to an inefficient and inaccurate procedure. The deep neural networks have been proved to be promising and useful in segmenting cardiac images, for example, Fully Convolutional Neural Network (FCNN) [6], [20], U-Net [21], some deep neural networks combined with traditional deformable models [5], [23], [24], and recently proposed deep regression segmentation method [25]. However, the quantification results are completely dependent on the accuracy of segmentation results in this kind of segmentation-based method.

Li *et al.* coined the direct manner of two-step methods [7]–[9] to quantify LV metrics. In this manner,

segmentation is no longer necessary, and machine learning algorithms are employed, in which features are extracted first and an independent regression learning model follows to estimate LV metrics. For example, Wang *et al.* [7] proposed an adaptive Bayesian formulation based on blob, homogeneity, and edge features to calculate bi-ventricular volumes. Zhen *et al.* [8] first extracted hierarchical representations by using multi-scale deep neural network, and then put them into a random regression forest to estimate the volume of left ventricle. Furthermore, four chamber volumes are estimated directly by using supervised descriptor learning (SDL) [9]. In this kind of two-step methods, there is only forward connection and no feedback from the second step. And thus features extracted in the first step may not be closely relevant to the target tasks, as a result, the estimation results in the second step may not be so accurate.

B. END-TO-END METHODS FOR THE LV QUANTIFICATION

Very recently, deep learning is employed to quantify the LV metrics in an end-to-end manner, where features extraction and metrics regression are integrated together. Luo *et al.* [10], [26] and Kabani and El-Sakka [11] took advantage of CNN to calculate the volume at ED and ES. In [26], multi-views fusion strategy was employed to improve the estimated volumes. Xue *et al.* [12] proposed an integrated model to yield multiple LV metrics, including two areas and six regional wall thicknesses on each frame over a cardiac cycle. In order to capture temporal dynamics of cardiac sequences, Xue et. al [13] employed the RNN following the CNN module to estimate six regional wall thicknesses. Furthermore, Xue *et al.* [14], [15] focused on the quantification of full LV metrics, which requires to estimate areas, directional dimensions and regional wall thicknesses at one time for each MR image. In order to attain more accurate results, they not only employed the CNN and RNN modules, but also modeled the correlations among different LV metrics. However, there are still difficulties for the end-to-end methods to extract discriminative features because of highly variable cardiac structures across different subjects. Following the work of Xue *et al.*, we build the Cascaded Segmentation and Regression Network (CSRNet), in which the segmentation component extracts myocardial contours of the left ventricle and the regression component estimates the desired LV metrics from those segmented images. By integrating a segmentation module into a regression learning framework, the network can learn more robust image representation and get more accurate estimation results.

III. CSRNET MODEL

The overview of the proposed CSRNet is shown in FIG.3, where the segmentation component takes cardiac MR images as input so that myocardial contours of the left ventricle are extracted and task-unrelated structures are removed, and then, the regression component learns hierarchical representations from the segmented images and yields the desired LV metrics. The objective function of proposed CSRNet can be written as follows,

$$\hat{y}^t_{s,f} = \mathbf{f}_{CSRNet}(x_{s,f}|w_{CSRNet}), \tag{1}$$

where $\mathcal{X} = x_{s,f}$ are the input cardiac MR images, and $\mathcal{Y} = \hat{y}^t_{s,f}$ are the output LV metrics. $s = 1 \cdots S$ denotes subjects, and $f = 1 \cdots F$ represents frame sequence, $t \in \{areas, dimensions, wall\ thicknesses\}$. w_{CSRNet} is the parameters set of the CSRNet. The segmentation and regression component will be detailed in the following subsections.

A. SEGMENTATION COMPONENT

We employ a Densely Connected Convolutional Network (DenseNet) [27] architecture for segmentation, which exploits dense connectivity between layers. This architecture can reduce the number of parameters to be learned and encourage feature reuse throughout the network. As shown in FIG.4, the DenseNet architecture extracts the features of cardiac MR images mainly through three dense blocks and three transition blocks. FIG.4(b) shows the details of the dense block. There are three identical "dense layer" including BN-ReLU-conv (1×1) -BN-ReLU-conv (3×3) operations. As reported in [27], although each layer in the dense block only produces k (growth rate) output feature-maps, it typically has many more inputs through concatenation. Therefore, 1×1 convolution before 3×3 is employed to reduce the number of feature-maps. As for the transition block, except for a 1×1 convolution with the same effect, average pooling operation is also employed for reducing image spatial resolution. After a series of convolution and pooling operations, feature-maps with different scales are upsampled to the original resolution by utilizing transposed convolutions. To combine information from the coarse high layer with that from the fine low layer, all transposed feature-maps are concatenated. Finally, three probability maps corresponding to three different categories, i.e., background, myocardium and cavity are generated by an extra transposed convolution with the softmax function. Different from traditional segmentation tasks, a soft classification result instead of a true category label map is needed here, so that the CSRNet can work in an end-to-end manner.

B. REGRESSION COMPONENT

A Convolutional Neural Network (CNN) is responsible for the regression learning, which is shown in FIG.5. The CNN takes soft segmentation results from the DenseNet as input, and extracts hierarchical features through a stack of convolutional layers and max-pooling layers. All feature-maps of size 6×6 obtained from the last max-pooling layer are converted into a feature vector. Following that, the first fully connected layer selects common representations from this vector for all desired LV metrics and the second fully connected layer matches some more relevant features to the specific LV metrics. Here, we employ a larger convolution kernel of size 5×5, instead of 3×3 used in the DenseNet, since there is no necessity to care about finer information when those task-unrelated structures have been removed. This simple network

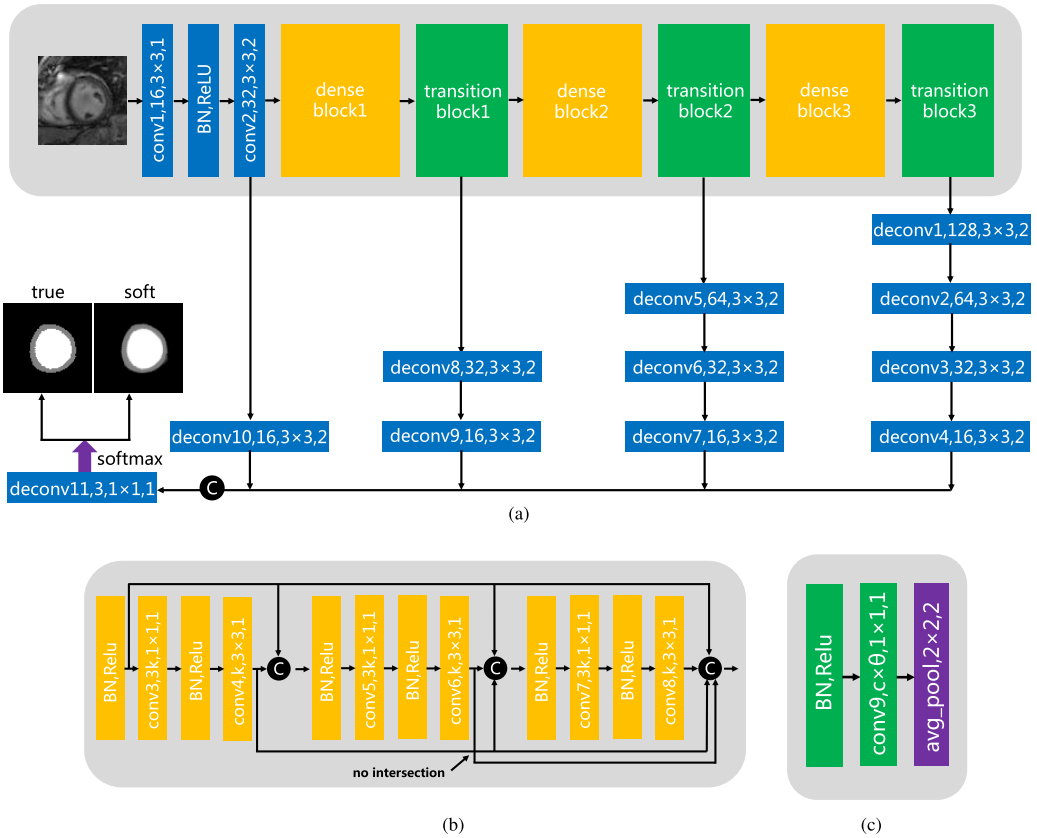


FIGURE 4. The overview of the segmentation component. (a) The DenseNet architecture for segmentation. Features of cardiac MR images are extracted mainly through three dense blocks and three transition blocks. Each "deconv" here corresponds to the sequence "deconv-BN-ReLU" and there are two different segmentation results from the DenseNet: the left "true" segmentation result is generated from the predicted category labels while the right "soft" one is produced by weighting three probability maps. Details of the dense block and the transition block are illustrated in (b) and (c), where k (= 16 in our network) represents the growth rate of feature channels, and θ (= 0.5 in our network) determines the output number of channels. (a) The DenseNet architecture for segmentation. (b) Dense block. (c) Transition block.

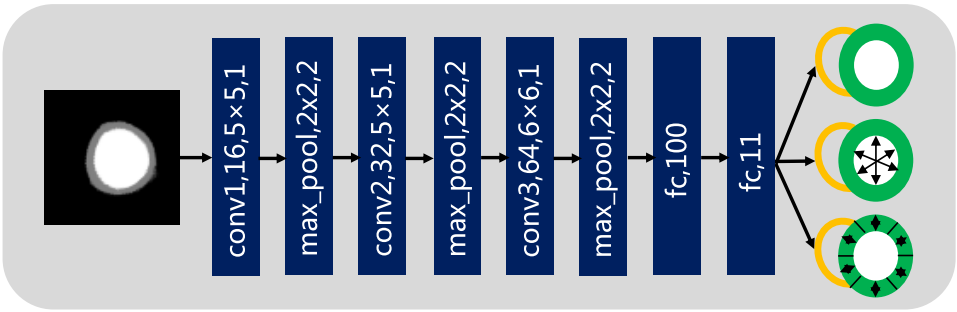


FIGURE 5. The CNN for regression component. The CNN consists of three convolution layers and two fully connected layers, each convolution contains convolution and ReLU operations. The soft segmentation results of cardiac MR image from the DenseNet are as input for the CNN.

is effective enough for the regression task, since it can easily learn more task-relevant representations from the segmented images.

C. TRAINING STRATEGY

1) PRE-TRAINING THE DENSENET

We first pre-train the DenseNet to provide good initial parameters for the following end-to-end training. Objective function of the DenseNet can be expressed as follows:

$$\hat{y}^c_{s,f,(m,n)} = \mathbf{f}_{Dense}(x_{s,f}|w_{Dense}), \tag{2}$$

where $\mathcal{X} = x_{s,f}$ are the input cardiac images, and $\mathcal{Y} = \hat{y}^c_{s,f,(m,n)}$ are the category labels for each pixel. $s = 1 \cdots S$ denotes diverse subjects, and $f = 1 \cdots F$ represents frame sequence, $c \in \{background, myocardium, cavity\}$. (m, n) denotes the pixel index, and w_{Dense} is the parameters set of DenseNet. The DenseNet is trained by minimizing the mean log-likelihood cost, and the loss for category prediction is

$$\mathcal{L}_{log} = -\frac{\sum_{s,f,(m,n)} \sum_c y^c_{s,f,(m,n)} log\hat{y}^c_{s,f,(m,n)}}{S \times F \times M \times N}, \tag{3}$$

where $y^c_{s,f,(m,n)}$ is the annotated category for each pixel.

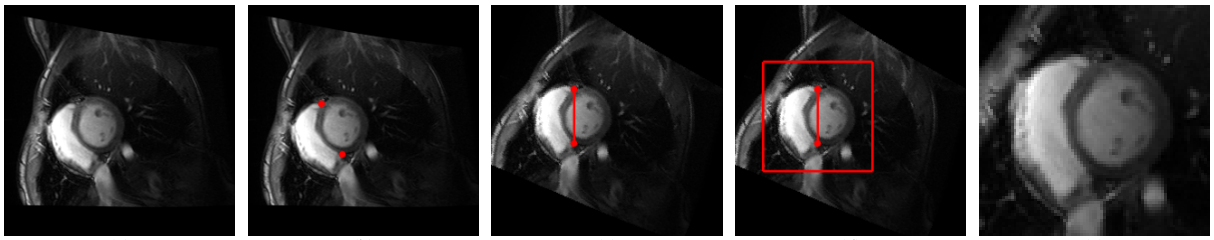


FIGURE 6. Pre-processing of original images. (a) Original image. (b) Landmarks labeling. (c) Rotation. (d) ROI cropping. (e) Resizing (80 × 80).

2) TRAINING THE CSRNET

The CSRNet is trained in an end-to-end manner, in which the segmentation component takes the pre-trained parameters from the DenseNet as initialization, and the regression component employs the parameters from scratch. We minimize the mean squared errors (MSE) to optimize the parameters, and the loss for estimated LV metrics is

$$\mathcal{L}_{mse} = \frac{\sum_{s,f} \sum_{t} \left(\hat{y}^{t}_{s,f} - y^{t}_{s,f} \right)^2}{2 \times S \times F}, \tag{4}$$

where $y^{t}_{s,f}$ are the ground truth of metrics. $s = 1 \cdots S$ denotes diverse subjects and $f = 1 \cdots F$ represents frame sequence, $t \in \{areas, dimensions, wall\ thicknesses\}$.

IV. EXPERIMENT SETTINGS

A. DATASET

The dataset employed in this work has been utilized in our previous works [12]–[15], which includes 2900 2D short-axis cine MR images of 145 subjects. For each subject, the mid-cavity slice of 20 frames over a cardiac cycle is selected. The subjects age from 16 years to 97 years and the pixel spacings of MR images range from 0.6836 mm/pixel to 2.0833 mm/pixel, with mode of 1.5625 mm/pixel. Diverse pathologies are in presence including regional wall motion abnormalities, myocardial hypertrophy, atrial septal defect, etc. More details of the dataset can be also found in [12]–[15]. Before the experiments, several pre-processing steps (shown in FIG.6) are carried out on the original MR images, which include the following operations,

1) LANDMARKS LABELLING

The first frame of each subject is required to manually label two landmarks, and remaining frames use the same coordinates generated from these two landmarks to keep consistent. The landmarks locate in the intersection of left and right ventricular wall, which are shown in FIG.6(b).

2) ROTATION

Connect the two landmarks into a line, and rotate the image until the line is perpendicular to the horizon (shown in FIG.6(c)). Notably, the resulting image after rotation should be taken to ensure that the right ventricle is on the left of the left ventricle.

3) ROI CROPPING

Taking the midpoint of the line between two labeled landmarks as the center, twice the length of the line as the size, we crop the ROI into a square as shown in FIG.6(d).

4) RESIZING

In this step, all cropped images are resized to 80 × 80. In order to display easily, the resized image is zoomed in FIG.6(e).

After all above pre-processing steps, the myocardial borders are manually contoured by two experienced radiologists and then the ground truth of LV metrics can be measured according to the contoured borders, more details shown in [12].

B. CONFIGURATIONS

To demonstrate the performance of our model with limited cardiac images, we divide the dataset equally into five groups, performing five-fold cross validation. During this procedure, four groups with a total of 2320 images are employed for training and the left group of 580 images for testing. The training operation will be performed five times so as to generate five different models to test all 2900 images. Each time the DenseNet is pre-trained first, and the CSRNet is further trained in an end-to-end manner, and they both use the same training data with 2320 images. In our experiment, all networks are implemented by Tensorflow [28] with AdamOptimizer [29], 100 training steps for pre-trained DenseNet and 200 training steps for the end-to-end trained CSRNet. Both training and testing procedure are carried out on a geforce gtx 1080 ti GPU.

C. EVALUATION CRITERIA

1) EVALUATION OF SEGMENTATION RESULTS

We evaluate the segmentation results quantitatively with Dice Coefficient (DC) and Hausdorff Distance (HD). They are defined as follows:

$$DC = \frac{2|A \cap B|}{|A| + |B|}, \tag{5}$$

$$\begin{aligned} HD &= \max(h(A,B), h(B,A)), \\ h(A,B) &= \max_{(a \in A)} \min_{(b \in B)} \|a - b\|, \\ h(B,A) &= \max_{(b \in B)} \min_{(a \in A)} \|b - a\|, \end{aligned} \tag{6}$$

where A and B are the coordinate set of pixels for segmentation results and annotation respectively, and $\|\cdot\|$ represents the distance paradigm.

TABLE 1. Comparison of segmentation results between U-Net [21] and DenseNet. dice coefficient and hausdorff distance (mm) between the segmentation results and annotation are reported.

Method	Dice Coefficient			Hausdorff Distance (mm)		
	background	myocardium	cavity	background	myocardium	cavity
U-Net [21]	0.988	0.873	0.950	6.098	6.685	5.220
DenseNet	0.989	0.886	0.959	4.881	5.433	3.557

2) EVALUATION OF QUANTIFICATION RESULTS

Quantification results are evaluated according to two criteria, i.e, Mean Absolute Error (MAE) and Correlation Coefficient ρ, which are defined as follows:

$$MAE^t = \frac{1}{S \times F} \sum_{s=1}^{S} \sum_{f=1}^{F} \left| y_{s,f}^t - \hat{y}_{s,f}^t \right|, \tag{7}$$

$$\rho^t = \frac{2 \sum_{s=1}^{S} \sum_{f=1}^{F} \left(y_{s,f}^t - y_m^t \right) \left(\hat{y}_{s,f}^t - \hat{y}_m^t \right)}{\sum_{s=1}^{S} \sum_{f=1}^{F} \left((y_{s,f}^t - y_m^t)^2 + (\hat{y}_{s,f}^t - \hat{y}_m^t)^2 \right)}, \tag{8}$$

where y_m^t and $\hat{y}_m^t$ are the mean values of ground truth and estimated results, respectively. As can be seen from the above equations, MAE reflects mean absolute difference between the values of ground truth and estimated results, and the ρ reflects linear correlation between them. The lower MAE is, the higher estimated accuracy is, and the larger ρ is, the closer the distribution between values of ground truth and estimated results is.

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. PERFORMANCE OF DENSENET FOR SEGMENTATION

To better evaluate the performance of DenseNet, we used a slightly modified version of U-Net [21] for comparison, which employs padding operation in all 3×3 convolutions to prevent information loss of border pixels and thus no longer requires the cropping operation when a short cut is connected between the encoder and decoder. All training details of U-Net are the same as that of DenseNet. Dice Coefficient and Hausdorff Distance between the results of segmentation and annotation are reported in the TABLE 1. We can see that both methods achieve accurate segmentation results, and the DenseNet performs slightly better than the U-Net. Among three categories, the segmentation results of background are the best, followed by the cavity and the worst for myocardium. The reason behind this observation is that the cavity and myocardium change a lot across different subjects and during a cardiac cycle, which makes them more difficult to be recognized. Furthermore, since the myocardium locates between background and cavity, it is easier for the algorithm to make mistakes. Other details of network such as parameters are shown in TABLE 2. We can see that the number of parameters of U-Net is one hundred times that of the DenseNet, and correspondingly, training time and testing time are significantly longer than that of the DenseNet. It is well-known that the annotated data are always limited in the community of medical image analysis, and as a result, fewer parameters in the deep learning model contribute to avoid overfitting during the training procedure.

TABLE 2. U-Net [21], DenseNet and CSRNet in terms of parameters, epochs and times are reported.

	parameters	epochs	training time	testing time
U-Net [21]	$\sim$ 33 M	100	25.53 mins	1.72 s
DenseNet	$\sim$ 0.3 M	100	11.56 mins	0.98 s
CSRNet	$\sim$ 0.6 M	200	17.62 mins	1.20 s

We also show some qualitative segmentation results of DenseNet in FIG.7. The first four rows show the cardiac images, annotated masks, estimated masks and misclassified pixels of frame 1-10, respectively, while the last four rows display that of frame 11-20. The estimated masks obtained from the DenseNet are in good agreement with the masks annotated by the experts. Temporal dynamics during cardiac cycle accounts for the slight difference between the results of algorithm and by experts.

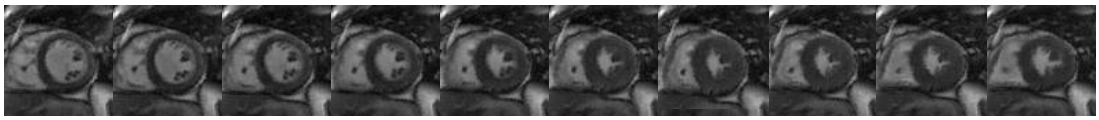
B. PERFORMANCE OF CSRNET FOR QUANTIFICATION

We compute MAE and ρ of CSRNet on 145 subjects and make comparison with the state-of-the-art methods, including Max Flow model [3], two-step methods in a direct way [8], [9], [30] and end-to-end methods [13]–[15].

1) COMPARISON WITH THE STATE-OF-THE-ART METHODS

Max Flow model [3] is a two-step method based on segmentation, where myocardial contours are extracted first and then LV metrics are measured. In Max Flow model, the first frame for each subject is required to be delineated manually, and all subsequent frames are automatically segmented by using the first frame as guidance. As shown in TABLE 3, the Max Flow model has high MAEs of LV metrics, especially for regional wall thicknesses. However, an interesting finding is that all ρ values are larger than that of two-step methods in a direct way, and even some ones are larger than that of end-to-end methods. This is because that the measured LV metrics based on extracted contours change regularly over a whole cardiac cycle, so that the results from Max-flow model have a good correlation with the metrics of ground-truth.

Two-step methods without segmentation employed for comparison are Multi-features + RF [30], SDL + AKRF [9], and MCDBN + RF [8], and the corresponding MAEs and ρ values are recorded in the second column to the fourth

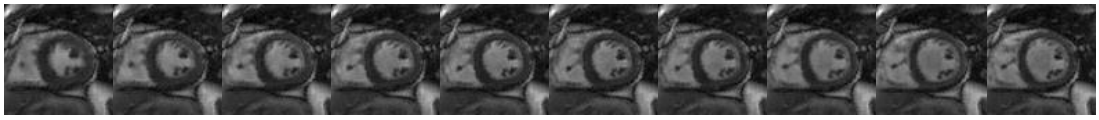


(a) Input cardiac MR images (frame 1-10).

(b) Manual annotation for ground truth (frame 1-10).

(c) Estimated results from DenseNet (frame 1-10).

(d) Misclassified pixels (frame 1-10).



(e) Input cardiac MR images (frame 11-20).

(f) Manual annotation for ground truth (frame 11-20).

(g) Estimated results from DenseNet (frame 11-20).

(h) Misclassified pixels (frame 11-20).

FIGURE 7. Qualitative segmentation results of DenseNet. (a)-(d) are the original input cardiac MR images, manual annotation masks, estimated results from the DenseNet, and misclassified pixels of frame 1-10, respectively. (e)-(h) are that for frame 11-20.

column of TABLE 3, respectively. Two-step methods without segmentation first extract cardiac features, and then put them into a regression model to calculate desired LV metrics. As shown in TABLE 3, the two-step methods without segmentation not only have a high mean absolute difference but also a poor correlation with the ground-truth metrics. Extracted features from the two-step methods are not closely related to target tasks can account for high MAEs and low ρ values.

The fifth to the seventh columns of TABLE 3 show performance of end-to-end methods including Indices-Net [12], FullLVNet [14], and DMTRL [15]. Indices-Net [12] is the

TABLE 3. Quantification results of state-of-the-art methods and CSRNet. There are two criteria, i.e., MAE and ρ employed for each LV metrics in this table. The best results are highlighted in bold.

Method		Max Flow features+RF	Multi- +AKRF	SDL +RF	MCDBN	Indices-Net	FullLVNet	DMTRL	CSRNet
Cavity-area (mm^2)	MAE	156±193	231±193	198±169	208±166	185±162	181±155	172±148	**107**±98
	ρ	0.958	0.924	0.942	0.926	0.953	0.940	0.943	**0.982**
Myocardium-area (mm^2)	MAE	339±272	291±246	286±242	269±217	223±193	199±174	189±159	**162**±127
	ρ	0.851	0.729	0.742	0.723	0.853	0.935	**0.947**	0.928
Average (mm^2)	MAE	247±201	261±165	242±158	239±135	204±133	190±128	180±118	**134**±117
	ρ	0.904	0.827	0.842	0.824	0.903	0.937	0.945	**0.965**
Dimension1 (mm)	MAE	2.81±2.76	3.53±2.77	2.99±2.43	2.88±2.48	–	2.62±2.09	2.47±1.95	**1.57**±1.42
	ρ	0.937	0.885	0.914	0.895	–	0.952	0.957	**0.974**
Dimension2 (mm)	MAE	2.60±2.62	3.49±2.87	2.55±2.30	2.45±2.01	–	2.64±2.12	2.59±2.07	**1.48**±1.36
	ρ	0.946	0.897	0.938	0.932	–	0.881	0.894	**0.979**
Dimension3 (mm)	MAE	2.49±2.88	3.91±3.23	3.10±2.54	2.93±2.49	–	2.77±2.22	2.48±2.34	**1.56**±1.33
	ρ	0.945	0.865	0.916	0.903	–	0.935	0.943	**0.979**
Average (mm)	MAE	2.65±2.33	3.64±2.61	2.88±2.03	2.75±1.90	–	2.68±1.64	2.51±1.58	**1.54**±1.37
	ρ	0.943	0.882	0.923	0.910	–	0.917	0.925	**0.978**
IS (mm)	MAE	1.53±1.73	1.70±1.47	1.98±1.58	1.78±1.40	1.39±1.13	1.32±1.09	1.26±1.04	**1.06**±0.87
	ρ	0.796	0.729	0.599	0.611	0.824	0.840	0.856	**0.895**
I (mm)	MAE	3.23±2.83	1.71±1.34	1.67±1.40	1.68±1.41	1.51±1.21	1.38±1.10	1.18±0.93	**1.33**±1.14
	ρ	0.720	0.603	0.582	0.462	0.701	0.751	0.747	**0.812**
IL (mm)	MAE	4.15±3.17	1.97±1.54	1.88±1.63	1.92±1.45	1.65±1.36	1.57±1.35	1.59±1.29	**1.33**±1.09
	ρ	0.743	0.483	0.515	0.435	0.671	0.691	0.693	**0.788**
AL (mm)	MAE	5.08±3.95	1.82±1.41	1.87±1.55	1.66±1.20	1.53±1.25	1.60±1.36	1.57±1.34	**1.32**±1.09
	ρ	0.706	0.533	0.493	0.547	0.698	0.651	0.659	**0.770**
A (mm)	MAE	3.47±3.25	1.55±1.33	1.65±1.45	1.20±1.01	1.30±1.12	1.34±1.11	1.32±1.10	**1.08**±0.92
	ρ	0.724	0.685	0.599	0.661	0.781	0.768	0.777	**0.840**
AS (mm)	MAE	1.76±1.80	1.68±1.43	2.04±1.59	1.63±1.23	1.28±1.00	1.26±1.10	1.25±1.01	**0.97**±0.80
	ρ	0.785	0.777	0.626	0.726	0.871	0.864	0.877	**0.919**
Average (mm)	MAE	3.21±1.98	1.73±0.97	1.85±1.03	1.65±0.77	1.44±0.71	1.41±0.72	1.39±0.68	**1.16**±0.97
	ρ	0.746	0.635	0.569	0.573	0.758	0.761	0.768	**0.868**

first method to estimate multiple LV metrics in an end-to-end manner. We can see that the Indices-Net has a reduction of 17.4% for average areas and 55.1% for average wall thicknesses in terms of MAE when compared to Max Flow model. FullLVNet [14] and DMTRL [15] focus on modeling relationships between different LV metrics, and capture temporal dynamics by utilizing RNN module, which further improve the estimation results as shown in the sixth and the seventh columns of TABLE 3. Experimental results show that end-to-end methods outperform all above two-step methods and have great potential to achieve more accurate quantification results of LV metrics.

MAEs and ρ values between the estimated results from proposed CSRNet and ground truth are reported in the last column of TABLE 3. We can see that CSRNet yields the lowest average MAEs of 134 mm^2, 1.54 mm, 1.16 mm and the largest average ρ values of 0.965, 0.978, 0.868 for area, cavity dimension, and regional wall thickness, respectively. There is a reduction of 37.8% for cavity area and 38.6% for average cavity dimension in terms of MAE when compared to the DMTRL method. However, fewer improvements on myocardium area and regional wall thicknesses are made in terms of MAE, because that the segmentation results from DenseNet for cavity are better than that for myocardium. But unlike two-step methods based on segmentation, on the one hand, the following regression network only requires the extracted contours as guidance, not completely depend on the segmentation results. On the other hand, the CSRNet is trained in an end-to-end manner, which makes it possible to further improve segmentation results under the supervision of ground truth metrics and conduce to more accurate quantification results in turn. In addition, the higher average ρ value means that there is a better linear relationship between the estimation results from CSRNet and the ground truth, which is illustrated in FIG.8.

2) TEMPORAL DYNAMICS

Besides high variability of LV structures, temporal dynamics also bring challenges to achieve accurate quantification results of LV metrics. Twenty frames of one subject are illustrated in FIG.9 to show complicated temporal dynamics over a cardiac cycle. One can see that the LV structure

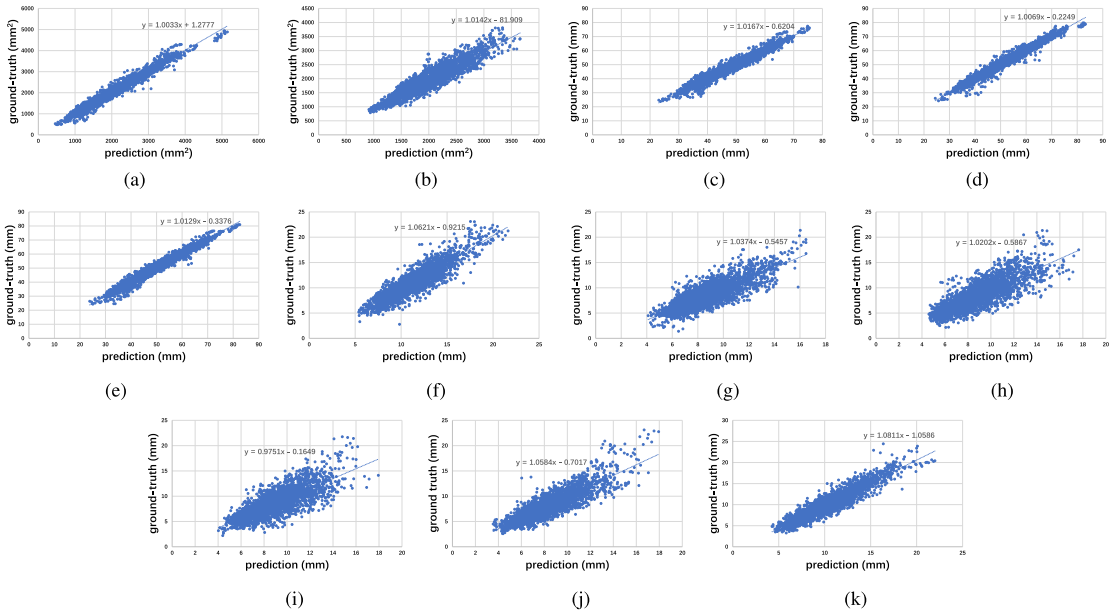


FIGURE 8. A plot for prediction from the CSRNet vs ground-truth of LV metrics. (a) cavity area. (b) myocardium area. (c) dimension 1. (d) dimension 2. (e) dimension 3. (f) IS. (g) I. (h) IL. (i) AL. (j) A. (k) AS.

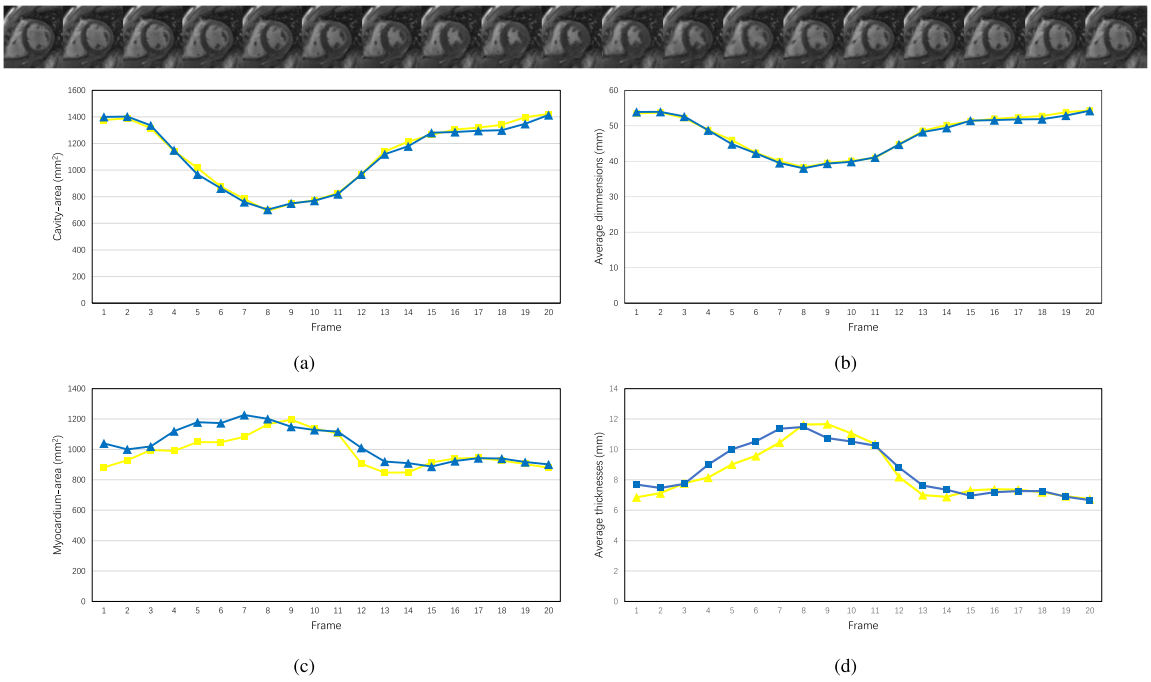


FIGURE 9. The ability of CSRNet to handle cardiac temporal dynamics. Images in the top row illustrate a whole cardiac cycle of a randomly selected subject. (a) and (c) are areas of the cavity and myocardium, respectively, while (b) and (d) are the average values of three cavity dimensions and six regional wall thicknesses. Yellow: estimated values. Blue: ground truth.

changes regularly from end-systole to end-diastole, and it is believed to get more accurate quantification results if the temporal information can be taken into consideration. In the CSRNet, myocardial contours extracted by DenseNet can provide guidance on capturing temporal information to some extent. The LV metrics including cavity area, average cavity dimension, myocardium area and average regional wall thickness are calculated and plotted in FIG.9(a) to FIG.9(d), respectively. The estimated cavity area and average cavity dimension have almost the same trend as those of ground truth. On the other hand, high ρ values of 0.982 for cavity area and 0.978 for average cavity dimension also manifests this agreement. As for the myocardium area and average regional wall thickness, the estimated results have a similar tendency as ground truth. FIG.9 shows that, with the contour guidance, CSRNet has ability to capture temporal dynamic, i.e., the cavity area and dimensions will become smaller during the systolic period, and get bigger over the diastolic phase, while the myocardium area and wall thicknesses just change on the contrary.

TABLE 4. LV metrics estimated by the method of segment-based and CSRNet are compared under the MAE.

Metrics	Segment-based	CSRNet	Metrics	Segment-based	CSRNet
Cavity-area (mm^2)	106±93	107±98	IS (mm)	1.81±1.13	1.06±0.87
Myocardium-area (mm^2)	203±147	162±127	I (mm)	2.25±1.38	1.33±1.14
Average (mm^2)	154±132	134±117	IL (mm)	1.95±1.31	1.33±1.09
Dimension1 (mm)	1.84±1.48	1.57±1.42	AL (mm)	1.95±1.36	1.32±1.09
Dimension2 (mm)	2.01±1.54	1.48±1.36	A (mm)	1.65±1.11	1.08±0.92
Dimension3 (mm)	1.72±1.38	1.56±1.33	AS (mm)	1.52±0.99	0.97±0.80
Average (mm)	1.86±1.47	1.54±1.37	Average (mm)	1.85±1.25	1.16±0.97

C. DISCUSSION

1) EFFECTIVENESS

Extensive experiments on the dataset have demonstrated the effectiveness of CSRNet. By integrating the myocardial contour into a regression learning framework, the CSRNet combines the advantages of two-step methods based on segmentation with end-to-end methods. On the one hand, the segmentation module of CSRNet can remove task-unrelated structures so that the following regression network can extract discriminative features from the segmented images. On the other hand, the end-to-end framework makes it possible to further improve segmentation results under the supervision of ground truth metrics and conduce to more accurate estimation of LV metrics in turn. The introduced contours only guide the following regression task, but do not completely determine the accuracy of quantification results like the two-step methods based on segmentation. As a comparison, we also calculate the LV metrics based on segmentation results from DenseNet trained in 100 epochs. The planar LV metrics such as cavity area, myocardium area are computed by counting the segmented pixels, and the linear ones including cavity dimension and wall thickness are measured according to some certain rules like using 2D centerline method [31] mentioned in [12]. The MAEs for both segment-based method and CSRNet are reported in TABLE 4. We can see that the CSRNet outperforms the segment-based method. Furthermore, from TABLE 2, we can see that the CSRNet has about 0.6 million parameters, and takes less than 18 minutes to train 200 epochs. In the testing phase, LV metrics of 29 subjects with 580 images are estimated in only 1.2 seconds, which demonstrates the real-time nature of CSRNet.

2) LIMITATIONS

However, because of the limitation of the dataset employed, it is difficult to transform estimated LV metrics into cardiac function parameters such as stroke volume, ejection fraction, which are related to all slices from the base to the apex. In the future, we will perform our framework on more available datasets which contain both temporal and spatial information.

VI. CONCLUSION

In this paper, we proposed an effective end-to-end framework to improve quantification results of full LV metrics. The proposed framework is called Cascaded Segmentation and Regression Network (CSRNet), which consists of two components, i.e., DenseNet for segmentation and CNN for regression. Although the LV structures across different subjects are highly variable, the CSRNet can learn more robust representation from the images and get more accurate estimation of the LV metrics. We evaluated the CSRNet on 145 different subjects and accurate quantification results showed its effectiveness. In the future, we will perform our framework on more available datasets which contain both temporal and spatial information, so that the cardiac function parameters can be acquired and precise assessments of the heart can be provided.

REFERENCES

[1] Y. Wang and Y. Jia, "Segmentation of the left ventricle from cardiac MR images based on degenerated minimal surface diffusion and shape priors," in *Proc. 18th Int. Conf. Pattern Recognit.*, 2006, pp. 671–674.

[2] C. Santiago, J. C. Nascimento, and J. S. Marques, "Fast segmentation of the left ventricle in cardiac MRI using dynamic programming," *Comput. Methods Programs Biomed.*, vol. 154, pp. 9–23, Feb. 2018.

[3] I. B. Ayed, H.-M. Chen, K. Punithakumar, I. Ross, and S. Li, "Max-flow segmentation of the left ventricle by recovering subject-specific distributions via a bound of the Bhattacharyya measure," *Med. Image Anal.*, vol. 16, no. 1, pp. 87–100, 2012.

[4] J. Senegas, C. A. Cocosco, and T. Netsch, "Model-based segmentation of cardiac MRI cine sequences: A Bayesian formulation," *Proc. SPIE*, vol. 5370, pp. 432–444, May 2004.

[5] T. A. Ngo, Z. Lu, and G. Carneiro, "Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance," *Med. Image Anal.*, vol. 35, pp. 159–171, Jan. 2017.

[6] W. Bai *et al.*, "Automated cardiovascular magnetic resonance image analysis with fully convolutional networks," *J. Cardiovascular Magn. Reson.*, vol. 20, no. 1, p. 65, 2018.

[7] 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. 1251–1260, Apr. 2014.

[8] 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. 120–129, May 2016.

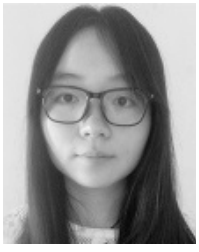
[9] X. Zhen, H. Zhang, A. Islam, M. Bhaduri, I. Chan, and S. Li, "Direct and simultaneous estimation of cardiac four chamber volumes by multioutput sparse regression," *Med. Image Anal.*, vol. 36, pp. 184–196, Feb. 2017.

[10] G. Luo, S. Dong, K. Wang, and H. Zhang, "Cardiac left ventricular volumes prediction method based on atlas location and deep learningg," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2016, pp. 1604–1610.

[11] A. Kabani and M. R. El-Sakka, "Estimating ejection fraction and left ventricle volume using deep convolutional networks," in *Proc. Int. Conf. Image Anal. Recognit.*, 2016, pp. 678–686.

[12] W. Xue, A. Islam, M. Bhaduri, and S. Li, "Direct multitype cardiac indices estimation via joint representation and regression learning," *IEEE Trans. Med. Imag.*, vol. 36, no. 10, pp. 2057–2067, Oct. 2017.

[13] W. Xue, I. B. Nachum, S. Pandey, J. Warrington, S. Leung, and S. Li, "Direct estimation of regional wall thicknesses via residual recurrent neural network," in *Proc. Int. Conf. Inf. Process. Med. Imag.*, 2017, pp. 505–516.
[14] W. Xue, A. Lum, A. Mercado, M. Landis, J. Warrington, and S. Li, "Full quantification of left ventricle via deep multitask learning network respecting intra- and inter-task relatedness," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2017, pp. 276–284.
[15] W. Xue, G. Brahm, S. Pandey, S. Leung, and S. Li, "Full left ventricle quantification via deep multitask relationships learning," *Med. Image Anal.*, vol. 43, pp. 54–65, Jan. 2017.
[16] A. Suinesiaputra *et al.*, "Quantification of LV function and mass by cardiovascular magnetic resonance: multi-center variability and consensus contours," *J. Cardiovascular Magn. Reson.*, vol. 17, no. 1, p. 63, 2015.
[17] C. Xu and J. L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 359–369, Mar. 1998.
[18] Y. Wu, Y. Wang, and Y. Jia, "Segmentation of the left ventricle in cardiac cine MRI using a shape-constrained snake model," *Comput. Vis. Image Understand.*, vol. 117, no. 9, pp. 990–1003, 2013.
[19] A. Andreopoulos and J. K. Tsotsos, "Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI," *Med. Image Anal.*, vol. 12, no. 3, pp. 335–357, 2008.
[20] P. V. Tran. (2016). "A fully convolutional neural network for cardiac segmentation in short-axis MRI." [Online]. Available: https://arxiv.org/abs/1604.00494
[21] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.
[22] H. B. Winther *et al.* (2017). "*v*-Net: Deep learning for generalized biventricular cardiac mass and function parameters." [Online]. Available: https://arxiv.org/abs/1706.04397
[23] T. A. Ngo and G. Carneiro, "Left ventricle segmentation from cardiac MRI combining level set methods with deep belief networks," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 695–699.
[24] M. R. Avendi, A. Kheradvar, and H. Jafarkhani, "A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI," *Med. Image Anal.*, vol. 30, pp. 108–119, May 2016.
[25] X. Du *et al.*, "Deep regression segmentation for cardiac bi-ventricle MR images," *IEEE Access*, vol. 6, pp. 3828–3838, 2018.
[26] 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 Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1924–1934, Sep. 2018.
[27] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
[28] M. Abadi *et al.* (2016). "TensorFlow: Large-scale machine learning on heterogeneous distributed systems." [Online]. Available: https://arxiv.org/abs/1603.04467
[29] D. P. Kingma and J. Ba. (2014). "Adam: A method for stochastic optimization." [Online]. Available: https://arxiv.org/abs/1412.6980
[30] X. Zhen, Z. Wang, A. Islam, M. Bhaduri, I. Chan, and S. Li, "Direct estimation of cardiac bi-ventricular volumes with regression forests," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2014, pp. 586–593.
[31] V. G. M. Buller, R. J. van der Geest, M. D. Kool, E. E. van der Wall, A. de Roos, and J. H. Reiber, "Assessment of regional left ventricular wall parameters from short axis magnetic resonance imaging using a three-dimensional extension to the improved centerline method," *Invest. Radiol.*, vol. 32, no. 9, pp. 529–539, 1997.



WENJI WANG received the B.S. degree from the School of Artificial Intelligence, Hebei University of Technology (HEBUT), in 2017. She is currently pursuing the master's degree with the School of Computer Science, Beijing Institute of Technology (BIT). She works closely with Prof. Y. Wang, and focuses on medical image analysis using deep learning methods.

YUANQUAN WANG received the master's and Ph.D. degrees with the Nanjing University of Science and Technology (NJUST), in 1998 and 2004, respectively. He was a Postdoc with the Beijing Institute of Technology (BIT), in 2004. From 1998 to 2001, he was a Software Engineer in industry. He is currently a Professor with the School of Artificial Intelligence, Hebei University of Technology (HEBUT), China.

He has published 12 papers in reputable journals and conference proceedings, covering the areas of computer vision, medical image analysis, and deep learning. His papers have got over 400 citations, including the citation by *Nature Methods journal*. His research interest includes intelligent analysis of cardiac MR images and MRA, development of technologies to transfer research concepts into tools for clinical study, and semantic and instance segmentation based on deep learning.

Dr Wang received the Most Outstanding Postgraduate Award from NJUST, in 1998.

YUWEI WU received the Ph.D. degree in computer science from the Beijing Institute of Technology (BIT), Beijing, China, in 2014, where he is currently an Assistant Professor with the School of Computer Science. From 2014 to 2016, he was a Postdoctoral Research Fellow with the Rapid-Rich Object Search (ROSE) Laboratory, School of Electrical and Electronic Engineering (EEE), Nanyang Technological University (NTU), Singapore. He received the Outstanding Ph.D. Thesis Award from BIT, and Distinguished Dissertation Award Nominee from China Association for Artificial Intelligence (CAAI).

TAO LIN received the master's degree from Tianjin University, in 1999, and the Ph.D. degree from the Hebei University of Technology (HEBUT), China, in 2007. From 2010 to 2013, he was a Postdoc with the Tianjin Development Zone Aojin High-tech Co., Ltd. Workstation, Hebei University of Technology Mobile Station. He is currently a Professor with the School of Artificial Intelligence, HEBUT.

SHUO LI received the Ph.D. degree in computer science from Concordia University, Montreal, QC, Canada, in 2006. He is currently an Adjunct Research Professor with Western University and an Adjunct Scientist with the Lawson Health Research Institute. He is also leading the Digital Imaging Group of London as the Scientific Director. His current research interests include automated medial image analysis and visualization.

BO CHEN received the Ph.D. degree from McMaster University, Canada, in 2006. In the past ten years, she was a Research Scientist with the Canada Headquartered International Corporation (Alcohol Countermeasure Systems). She joined the Digital Imaging Group of London, in 2018. Her current interest is the advanced deep learning methods for medical imaging.