Direct Segmentation-based Full Quantification for Left Ventricle via Deep Multi-task Regression Learning Network

Xiuquan Du[1], Renjun Tang[1], Susu Yin[1], Yanping Zhang[1], and Shuo Li[2,3]

***Abstract*—Quantitative analysis of the heart is extremely necessary and significant for detecting and diagnosing heart disease, yet there are still some challenges. In this study, we propose a new end-to-end segmentation-based deep multi-task regression learning model (Indices-JSQ) to make a holonomic quantitative analysis of the left ventricle (LV), which contains a segmentation network (Img2Contour) and multi-task regression network (Contour2Indices). First, Img2Contour, which contains a deep convolutional encoder-decoder module, is designed to obtain the LV contour. Then, the predicted contour is feed as input to Contour2Indices for full quantification. On the whole, we take into account the relationship between different tasks, which can serve as a complementary advantage. Meanwhile, instead of using images directly from the original dataset, we creatively use the segmented contour of the original image to estimate the cardiac indices to achieve better and more accurate results. We make experiments on MR sequences of 145 subjects and gain the experimental results of 157 mm^2, 2.43 mm, 1.29 mm, and 0.87 on areas, dimensions, regional wall thicknesses (RWTs), and Dice Metric (DM), respectively. It intuitively shows that the proposed method outperforms the other state-of-the-art methods and demonstrates that our method has a great potential in cardiac MR images segmentation, comprehensive clinical assessment, and diagnosis.**

***Index Terms*—Left Ventricle, Full Quantification, Multi-task Regression Learning**

I. Introduction

CARDIOVASCULAR disease has always been the biggest threat to human health. The information obtained from MR images, such as clinical indices, can reveal the underlying causes to some extent. Therefore, accurate estimation of relevant clinical indices is of great significance for physicians to discover and diagnose cardiovascular diseases. As shown in Fig.1, in this work, clinical indices obtained from cardiac MR images include two areas (Fig. 1(b)), three dimensions of left ventricle (LV) (Fig. 1(c)), and six regional wall thicknesses (RWTs) (Fig. 1(d)) [1].

In early clinical work, LV quantification was achieved by manually delineating the boundaries of the heart, which was still time-consuming. Although later, researchers tried to solve the problem by an automatic way, there are still some challenges including 1) the intricate heart structure and the various symptoms, 2) the sophisticated dynamic deformation of LV, and 3) the difficulty of analyzing the relationship among multiple indices.

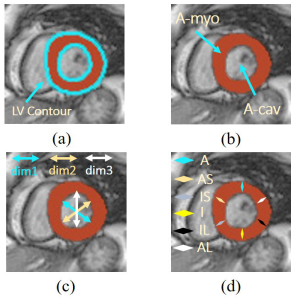

Fig. 1. Illustration of LV indices. (a) The contour of LV, which consists of endocardium and epicardium. (b) A-myo: myocardium area, A-cav: cavity area. (c) Three directional dimensions of cavity: dim1, dim2, dim3. (d) Six regional wall thickness. A: anterior; AS: anteroseptal; IS: inferoseptal; I: inferior; IL: inferolateral; AL: anterolateral.

To overcome the above challenges, we propose a new end-to-end segmentation-based deep multi-task regression framework (Indices-JSQ) to make a holonomic quantitative analysis for LV. In this framework, it contains a segmentation network (Img2Contour) and multi-task regression network (Contour2Indices). The proposed framework has the ability to express the image, extract the high-level features and capture the dynamic deformation of the heart structure. From this framework, multiple indices can be obtained and provide the quantitative diagnosis for clinicians by means of the final regression learning network of indices estimation.

A. Related Work for LV Segmentation and Quantification

1) Manual methods: In early clinical practice, manual methods are used to delineate the contour of the heart. Through the segmented contour, we can obtain the reliable quantification. However, it is still time-consuming and tiresome for clinicians to deal with a large number of cardiac MR images. Moreover,it is hard to avoid mistakes even for an expert with years of experience.

2) Segmentation-based methods: To overcome the drawback of manual methods, segmentation-based methods are designed to semi-automatically or automatically segment out the myocardium of LV. Taking it by and large, semi-automatic or fully automatic segmentation methods fall into two categories: i) image-driven methods and ii) model-driven methods.

[1]School of Computer Science and Technology, Anhui University, Hefei 230039, China.
[2]Department of Medical Imaging, Western University, London, ON N6A 3K7, Canada.
[3]The Digital Imaging Group of London, London, ON N6A 3K7, Canada.
This work was supported in part by the National Science Foundation of China under Grant 61673020, the Anhui Provincial Natural Science Foundation under Grand 1708085QF143.

i) Image-driven methods include thresholding, region-growing, clustering, pixel classification, active contour, deformable model, and level set. For instance, thresholding can be utilized to locate the region of interest (ROI) and sought the contour of LV. Pixel classification is used to classify every pixel of ROI and divide it into the corresponding class. Active contour-based methods find the chamber walls through different energy functions for a better result. However, the segmentation performance would be limited if they lack great user interaction [2]. ii) Model-driven methods include active shape model and active appearance model. They perform the segmentation by means of principal component analysis (PCA) for modeling and finding the optimal model in an image. Unfortunately, their segmentation performance depended on the quantity of training set and strong prior knowledge to a large extent.

3) Direct methods: In order to break through above limitations, direct estimation methods without segmentation are proposed for LV quantification. They generally can be divided into two classes: i) Two-stage estimation methods and ii) end-to-end estimation methods based on the deep neural network.

i) Two-stage estimation methods consist of two steps including cardiac image representation and indices estimation. For image representation, the image usually is expressed by feature obtained from manual extraction or other learning-based methods, such as the feature-based appearance model [3], supervised descriptor learning (SDL) algorithm [5], multi-scale convolutional deep belief network (MCDBN) [6]. By these features, final indices estimation can be effectively obtained through a regression model. ii) An end-to-end deep neural network-based methods. It directly estimates indices, such as direct volume estimation or ejection fraction estimation, multi-type indices estimation.

Although these methods have greatly improved in terms of time-saving and accuracy compared with previous methods, there are still some limitations: i) The mentioned direct methods did not adopt the strategy of combining segmentation to make more intuitive and accurate judgments for clinicians; ii) They also did not consider to use the predicted contour as input to estimate the multi-type clinical indices, yet it's worth noting that the obtained contour discards the background information of the image and plays an important role in data optimization. iii) Among the end-to-end deep neural network-based methods, the framework of fusing multiple independent CNN models was not deployed, which is capable of extracting more feature information from the input image.

The main contributions of our study are summarized below: (1) We propose a new end-to-end segmentation-based deep multi-task learning framework, which can be used to directly estimate different clinical indices. (2) We creatively use the segmented contour of the original image to estimate different clinical indices to achieve better and more accurate results. (3) The proposed framework takes into account the relationship between different tasks, which can serve as a complementary advantage, but others have largely failed to consider it. (4) We employ the structure of multiple CNN models fusion and combine LSTM units to achieve high accuracy of final clinical indices estimation.

II. Segmentation-based Full Quantification for Left Ventricle (Indices-JSQ)

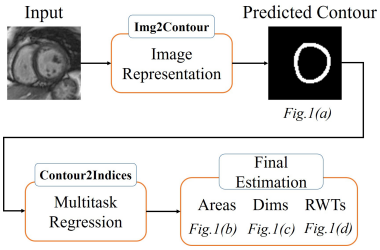

Fig. 2. Overview of Indices-JSQ, which contains Img2Contour network for cardiac MR image representation and Contour2Indices network for final indices estimation via multi-task regression learning.

The overview of the proposed framework is shown in Fig. 2. In our framework, in order to get the accurate results of the final estimation, we utilize the Img2Contour network to obtain the segmented contour of the original image. Then,we naturally adopt multi-task learning of Contour2Indices network to serve as a complementary advantage among these tasks. In the end, for the final estimation, we get the result of various indices via a regression learning model.

A. Deep Convolutional Encoder-Decoder for LV Segmentation (Img2Contour)

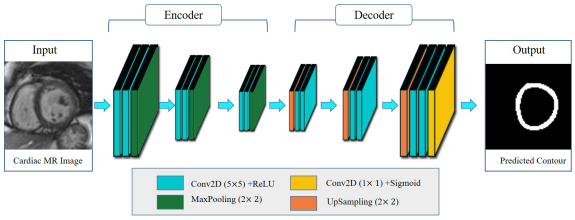

Fig. 3. The architecture of the Img2Contour network, which consists of an encoder module and decoder module.

In recent years, feature learning has become more and more popular in image processing [8] [9]. Especially, CNN has achieved great success in the medical image analysis. Some different CNN structures were proposed for cardiac MR image segmentation, such as 2D/3D U-net, ACNN framework, encoder-decoder architecture.

As shown in Fig. 3, a deep convolutional encoder-decoder (DCED) network is designed for LV segmentation to directly get the LV contour (including the boundary of endocardium and epicardium), which consists of an encoder module and decoder module.

1) Image Representation by DCED: To direct segment the LV contour, we utilize a DCED for image representation, which belongs to a fully convolutional neural network architecture and consists of an encoder module and a corresponding decoder module. It regards the convolution layers as an encoder for efficient feature representation and the decoder network is used to map the low-resolution encoder feature maps to full input resolution feature maps. The decoder uses

pooling indices computed in the max-pooling step of the corresponding encoder to perform non-linear up-sampling.

i) *Encoder module of Img2Contour network.* As illustrated in Fig. 3, in the Img2Contour, the encoder of DCED consists of three convolution layers with ReLU activation function and max-pooling layers. The convolution layers with the kernel size of 5×5 and filters are used to generate a batch of feature maps. However, the information in the feature map and the number of parameters in the network are enormous. In order to filter the unnecessary redundant information in feature maps and reduce the number of parameters in the network, max-pooling layers are applied with a 2×2 window and stride 2. After that, efficient feature representation of cardiac MR images can be obtained by the encoder module.

ii) *Decoder module of Img2Contour network.* Compared with encoder module, the decoder module is formed by the corresponding convolution layers, activation function, and the up-sampling layers. Therefore, the kernel size, pool size, and stride of decoder module are the same as encoder module. Moreover, unlike the encoder module, the decoder module is used to clean input resolution feature maps by mapping the low-resolution feature maps for getting high dimensional feature representation.

In the Fig. 3, the last convolution layer, in order to get the LV contour, we add a convolutional layer with a convolution kernel of 1×1 to the output of the decoder module. The output of the decoder module is a feature map with the same resolution as the original image, we can classify each pixel in the feature map through the sigmoid activation function. We set the threshold at 0.5. In this way, we can get the binary image of left ventricular contour with the same resolution as the original image.

B. Deep Multi-task Regression Learning for LV Quantification (Contour2Indices)

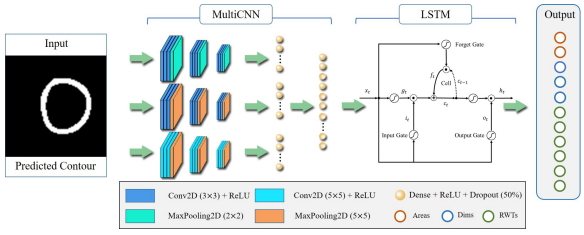

Fig. 4. The architecture of the Contour2Indices network, which consists of a MultiCNN module and LSTM module.

As shown in Fig. 4, a Contour2Indices is designed to make a holonomic quantitative analysis of LV. Considering the diversity and relevance of different tasks, we adopt deep multi-task learning to explore the relationship among multiple tasks. In the concrete implementation process, first, the fusion of multiple CNNs is used for feature extraction of the segmented contour image. Next, we employ RNN with LSTM units to capture the dynamic deformation of the heart structure. Finally, a multi-task regression model is designed to estimate the areas, dims and RWTs of LV. The learned shared representation between multiple tasks can reduce the number of data source, the scale of the whole model parameters, improve the generalization ability of the model, and make the estimation more efficiency.

1) Feature Representation by MultiCNN: Many studies have demonstrated improved performance by fusing multiple CNN. They combine multiple CNNs trained on different views or sources to extract more effective features. Therefore, we present a newly MultiCNN module, which including three independent CNN models as illustrated in Fig. 4. For each image, three different CNN models are used to extract the feature information. They have the same structure including three convolution layers with ReLU activation function, Dropout layers, Max-Pooling layers, and two fully connected layers. Their difference is kernel size and pool size. For the 1st CNN model, kernel size and pool size are 3×3 and 2×2. The 2nd model is 3×3 and 5×5. The 3rd model has the same size of kernel and pool with 5×5. Dropout layer is used to avoid overfitting problem. Finally, a fully connected layer is followed. It fuses the output of three different CNN models and feed into the next LSTM module as input. Compared with the simple CNN model, the fusion of multiple models will extract more effective information from the original image and improve the accuracy of the estimation network.

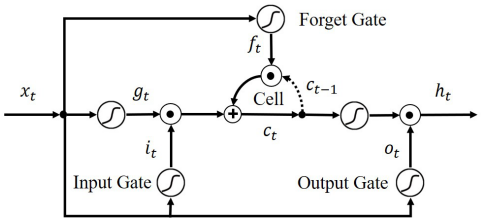

Fig. 5. Illustration of LSTM unit in RNN, which consists of the input gate, output gate, and forget gate for dynamic deformation modeling of cardiac MR image sequence.

2) Dynamic deformation by LSTM: Due to our cardiac MR images have strong temporal dependencies and RNNs are exactly designed for sequence learning. Therefore, RNNs can be used to address the sequence task. However, the gradient vanishing/exploding problem during model training process cause the low performance of RNNs. Hence, LSTM is proposed to solve above problems. Owing to the excellent performance of LSTM, it is widely used in image captioning [11]. Especially, in medical image analysis, LSTM is used to capture long-term dependencies of cardiac MR image sequence [12].

As shown in Fig. 5, to capture the long-term dependencies and avoid the gradient vanishing/exploding problem of cardiac MR image sequence, LSTM unit is deployed. It contains several purpose-designed memory cells to save information and a few gates are used to manage the passing of information along the sequences at each time step. For each cardiac MR image during a cardiac cycle (20 frames), after the feature representation of MultiCNN module, it has a vector as the input of LSTM module and its information flows are computed by the following equations:

$$
\begin{aligned}
i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \\
f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \\
o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \\
g_t &= \mu(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \\
c_t &= f_t \odot c_{t-1} + i_t \odot g_t \\
h_t &= o_t \odot \mu(c_t)
\end{aligned} \tag{1}
$$

where $\sigma(\cdot)$ and $\mu(\cdot)$ denote element-wise logistic sigmoid and hyperbolic tangent non-linearity functions respectively. $\odot$ denotes element-wise products. x_t is feed into the input gate i_t, forget gate f_t, and output gate o_t as the current input to control information flow. g_t denotes the previous memory states of the cell, c_t denotes the current memory states of the cell, h_t denotes the current output. W is the weight matrices that lead to different control gates and connections and b denotes the corresponding bias.

3) Multi-task learning by regression learning: For LV quantification, three multi-type clinical indices are need to be estimated. There are both connections and differences between these indices. However, it is time-consuming to estimate these indices independently, and the relationship between them is neglected. Therefore, it seems to be the best choice to simultaneously analyze the relationship among multiple indices using the multi-task learning framework. And some solutions have been proposed [13] [14]. They take the estimation task as a multi-task regression problem and formalize it, transfer knowledge among the tasks to improve the generalization performance.

The task of LV segmentation and various clinical indices estimation can be expressed as solving a single task regression model and a multi-task regression model. The single task regression model of Img2Contour is designed only for LV segmentation. The multi-task regression model of Contour2Indices contains three tasks $y_t^{s,f}$ ($t \in \{area, dim, rwt\}$), which are contemporaneously learned to predict frame-wise values of the mentioned LV indices. Their objective function as shown below:

$$
\begin{aligned}
&\min_{\theta} -\frac{1}{2S \times F} \sum_{f} \sum_{s} Q\left(X_f^s|\theta\right) log\ y_f^s \\
&+(1 - Q\left(X_f^s|\theta\right))log\ \left(1 - Q\left(X_f^s|\theta\right)\right) + \lambda R(\theta)
\end{aligned} \tag{2}
$$

$$
\min_{\alpha} \frac{1}{S \times F} \sum_{s,f} \sum_{t} C_t\left(y_t^{s,f}, J\left(X_t^{s,f}|\alpha\right)\right) + \beta R\left(\alpha\right) \tag{3}
$$

where $s = 1 \cdots S$ denotes the subject and $f = 1 \cdots F$ denotes the frame for a set of cardiac MR sequences $x = \left\{X_f^s\right\}$. Q, J denote the Img2Contour network and Contour2Indices network, respectively. C_t is the cost function of task t and $R\left(\theta\right)/R\left(\alpha\right)$ regularizes the parameter vector.

III. Experimental

A. Dataset

Our dataset consists of 2900 short-axis cardiac MR images of 145 subjects, which is collected from 3 hospitals affiliated with two health care centers (London Healthcare Center and St. Josephs HealthCare). The subjects age from 16 yrs to 97 yrs, with average of 58.9 yrs. The pixel spacing of the MR images ranges from 0.6836 mm/pixel to 2.0833 mm/pixel, with the mode of 1.5625 mm/pixel. Diverse pathologies are in presence including regional wall motion abnormalities, myocardial hypertrophy, mildly enlarged LV, atrial septal defect, LV dysfunction, etc. Each subject contains 20 frames during a cardiac cycle. In each frame, LV is divided into equal thirds (basal, mid-cavity, and apical) perpendicular to the long axis of the heart following the standard AHA prescription and a representative mid-cavity slice is selected for validation of our Indices-JSQ. We obtained the clinical indices including two areas (cavity and myocardium), three LV cavity dimensions, and six regional wall thicknesses (RWTs) by calculating the manual contours of LV myocardium. In our experiments, before inputting the images to the proposed framework, it is cropped and resized to 80×80. Furthermore, the values of RWTs and dimensions are normalized by the image dimension 80, while areas of LV cavity and myocardium are normalized by image area 80×80.

B. Evaluation criteria

1) Dice Metric (DM): Metric learning [15] is used to measure the differences between data. In order to evaluate the performance of the LV segmentation, we use DM to calculate the overlapping region between the predicted contour and manual contour (ground truth). The formulas as shown below:

$$
DM\left(C_m, C_p\right) = \frac{2 \times C_{mp}}{C_m + C_p} \tag{4}
$$

where C_m and C_p denote the region of manual and predicted contour respectively, and C_{mp} denotes their intersection.

2) Mean Absolute Error (MAE): We use the mean absolute error (MAE) to evaluate the performance of Contour2Indices network. The formulas as displayed below:

$$
MAE\left(y_t, y_p\right) = \frac{1}{N} \sum_{k=1}^{N} \left|y_t^k - y_p^k\right| \tag{5}
$$

where y_t denotes ground truth and y_p denotes the predicted ones.

C. Results and analysis

We evaluate the performance of the proposed method by comparing the results of the three experiments below including 1) Performance of Img2Contour network, 2) Performance of Contour2Indices network, and 3) Performance comparison with state-of-the-art methods.

1) Performance of Img2Contour Network: In order to evaluate the performance of Img2Contour network, we use DM to calculate the overlapping region between the predicted contour and ground truth. Fig. 6 shows the DM values of the MR images among 145 subjects and each subject has 20 MR images. For convenience, each column in Fig. 6 shows the average DM values of the 20 MR images that belong to each subject. It can be seen that most DM values fluctuate between

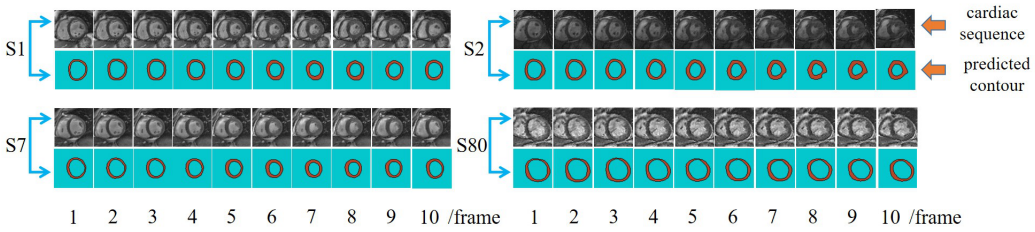

Fig. 7. Results for four MR sequences and the corresponding predicted contour. Every two rows belong to one subject. Qualitative diagnosis and visual assistance can be obtained through the result of Img2Contour network for clinicians.

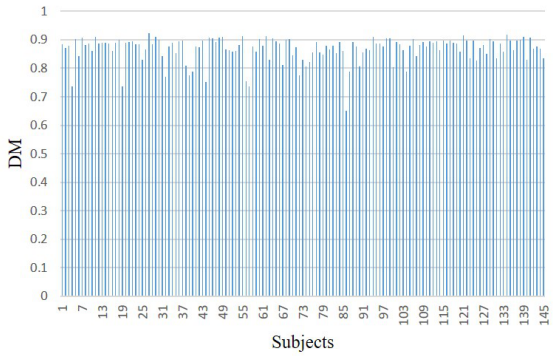

Fig. 6. DM histogram of 145 subjects short-axis cardiac MR images. Each DM value denotes the mean DM of 20 frames during one cardiac cycle, which belongs to one subject.

0.85 and 0.9 (1 denotes the best result) on 145 subjects. Although there have several poor results, it still intuitively shows that our Img2Contour framework has a great potential in cardiac MR images segmentation. There may be several reasons for the poor results: i) some subjects have blurred LV contour or some regions are lost in the MR image, and ii) some subjects have a very dark MR image. All of these perhaps lead to the result is not good enough.

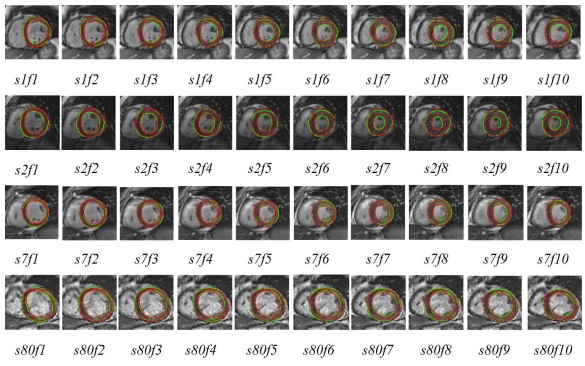

Fig. 8. Illustration of ground truth (green circle) and predicted contour (red circle). (s denotes the subject, f denotes the frame). The results of the predicted contour are very precise and it can provide the clinicians with a reliable visual assistance and qualitative diagnosis.

Thanks to the excellent performance of the Img2Contour, we gain the predicted contour as shown in Fig. 7. The Fig.7 shows 4 patients and 10 consecutive cardiac MR images from one cardiac cycle (20 frames) of each patient. The corresponding contour images obtained through the Img2Contour network is displayed in the Fig. 7. They clearly show the excellent performance of the network. In order to get a more clearly segmentation effect, we plot the obtained contour (red circle) and the ground truth (green circle) on the original image in the Fig. 8. It clearly shows that our model has high accuracy for the segmentation task and the obtained contour clearly provides clinicians visual assistance for qualitative diagnosis.

2) Performance of Contour2Indices Network: Fig. 9 shows the results of the actual clinical indices (Ground Truth) of a randomly selected sample of 145 subjects compared with prediction indices by our method. The solid line in the figure is the estimation result of our method, and the dotted line shows the ground truth. From the estimation results of three multi-type clinical indices, we can see that they are very close to their ground truth. The predicted value fluctuates around the ground truth value and basically maintains the same development trend. The three figures of experimental result illustrate that the Contour2Indices has a significant effect on the indices estimation.

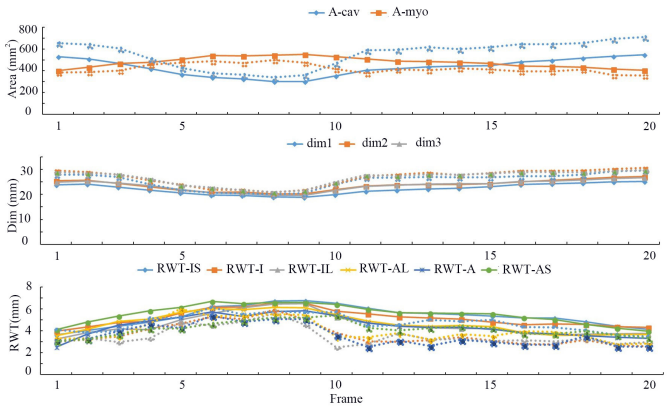

Fig. 9. Example of LV indices estimation by Indices-JSQ for a random patient during a cardiac cycle (20 frames). The predicted results are shown with the solid line and the corresponding ground truth values are displayed with the dotted line in the same color.

To know the estimation performance of Contour2Indices for all subjects, we display the indices through scatter diagram. As shown in Fig.10, we compare the results obtained from our method with the ground truth of cardiac LV indices. It can be clearly seen that the coordinate points formed by the predicted and true results are basically fluctuating around the standard line (red line). It shows that the estimation results are

TABLE II
PERFORMANCE (MAE) COMPARISON BETWEEN FULL LV QUANTIFICATION FOR EXISTING ADVANCED METHODS AND THE PROPOSED INDICES-JSQ.

Method		Max Flow	Multi-Features	SDL	MCDBN	Indices-Net	FullLVNet	DMTRL	Indices-JSQ
$Areas$	A-cav	**156±193**	231±193	198±169	208±166	185±162	181±155	172±148	157±145
(mm^2)	A-myo	339±272	291±246	286±242	269±217	223±193	199±174	189±159	**157±161**
Average		247±201	261±165	242±158	239±135	204±133	190±128	180±118	**157±120**
$Dims$	dim1	2.81±2.76	3.53±2.77	2.99±2.43	2.88±2.48	\	**2.26±2.09**	2.47±1.95	2.43±1.91
(mm)	dim2	2.60±2.62	3.49±2.87	2.55±2.30	2.45±2.01	\	2.64±2.12	2.59±2.07	**2.32±1.77**
	dim3	2.49±2.88	3.91±3.23	3.10±2.54	2.93±2.49	\	2.77±2.22	**2.48±2.34**	2.54±1.97
Average		2.65±2.33	3.64±2.61	2.88±2.03	2.75±1.90	\	2.68±1.64	2.51±1.58	**2.43±1.62**
$RWTs$ (mm)	IS	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	**1.16±1.03**
	I	3.23±2.83	1.71±1.34	1.67±1.40	1.68±1.41	1.51±1.21	1.38±1.10	1.40±1.10	**1.33±1.07**
	IL	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	**1.42±1.20**
	AL	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	**1.37±1.18**
	A	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	1.21±1.07
	AS	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	**1.24±1.08**
Average		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	**1.29±0.70**

† A-cav, A-myo: the area of the left ventricular cavity and myocardium.

† dim1, dim2, dim3: dimensions in three different directions of the left ventricular cavity.

† IS, I, IL, AL, A, AS: regional wall thickness in six different directions of the left ventricular myocardium.

basically the same as the real ones and reveals the excellent performance of Contour2Indices.

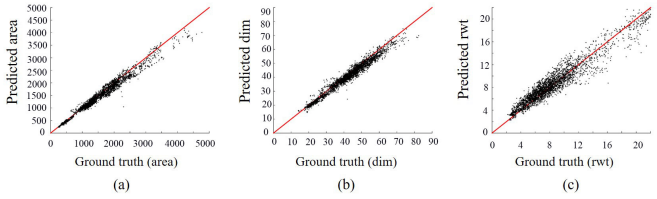

Fig. 10. The comparison between predicted indices and ground truth obtained manually. (a) Two areas. (b) Three LV dimensions. (c) Six regional wall thicknesses (RWTs).

TABLE I
PERFORMANCE COMPARISON (DM) OF LV SEGMENTATION BETWEEN OUR PROPOSED METHOD AND EXISTING ADVANCED FULLY AUTOMATED/SEMI-AUTOMATED SEGMENTATION METHODS.

Method	† FA / SA	Dice Metric
MSVR	SA	0.68(0.13)
Bi-DBN	SA	0.81(0.08)
FCN	FA	**0.88(0.06)**
Stacked AE	FA	0.76(0.11)
Seg-Net	FA	0.83(0.01)
Conv-Deconv	FA	0.85(0.02)
Indices-JSQ	FA	**0.87(0.06)**

† FA / SA: Fully Automated / Semi-Automated

3) Performance comparison: Table 1 shows the results between our method and the existing advanced fully automatic (FCN [16], Stacked AE [17], Seg-Net [10], Conv-Deconv [18]) /semi-automatic (MSVR [19], Bi-DBN [20]) segmentation methods for LV segmentation. It can be seen that the DM value of our method exceeds the most methods except FCN. The DM value of FCN is 0.88, while the DM value of our method is 0.87. Although our method is not the best, it is almost at the same level as the best method. It shows that our method is capable of achieving high performance for LV segmentation. Note that although our method is 0.01 lower than the best result, the proposed framework not only achieves LV segmentation but also is capable of doing the task of LV quantification, which is not achieved by FCN.

Full LV quantification is the final purpose of our work. Therefore, we compare the proposed method (Indices-JSQ) with the existing advanced methods (Max Flow [21], Multi-Features [4], SDL [5], MCDBN [6], Indices-Net [22], FullLVNet [23], and DMTRL [7]) to evaluate the performance of our method. The results are showed in Table 2, for the last column, we can clearly see that the proposed method completely outperforms the other methods. The average MAE is 157mm^2, 2.43mm and 1.29mm for areas, dims and RWTs, which is better than any other method and lower than the previous best method (DMTRL) of 23mm^2, 0.08mm, and 0.1mm. As we all know, among these clinical indices, the RWTs and the area of the myocardium are more difficult than the area of cavity and dimensions. However, Indices-JSQ still shows exceptional performance in this respect. The reason is obvious that the predicted contour is the border of endocardium and epicardium, which forms the myocardium. Therefore, the clinical indices of myocardium such as areas and RWTs can be predicted well through our method. Meanwhile, the result of dim is also surprisingly good. It distinctly reveals that the strategy of jointly segmentation and quantification and multi-task learning can achieve better and more accurate results for LV segmentation and clinical indices estimation.

As shown in Fig. 11, To more intuitively reflect the performance differences among each algorithm, we utilize bar charts for each specific clinical index. i) *Indices comparison of areas.* For each bar chart, the first approach Indices-JSQ is our proposed method. From Fig. 11(a), it is obvious that our method outperforms the other existing advanced methods. Although the MAE value of cavity area is almost equal to the method of Max Flow, the Max Flows estimation performance of myocardium area is worst among other methods and our method has better performance on areas estimation generally. ii) *Indices comparison of dims.* From Fig. 11(b), the best MAE value of dims estimation comes from our methods (Indices-JSQ) and the previous best method DMTRL. It reveals that our method has the same great performance on dims estimation as DMTRL. In fact, our method has a little better on average MAE. iii) *Indices comparison of RWTs.* The MAE value of

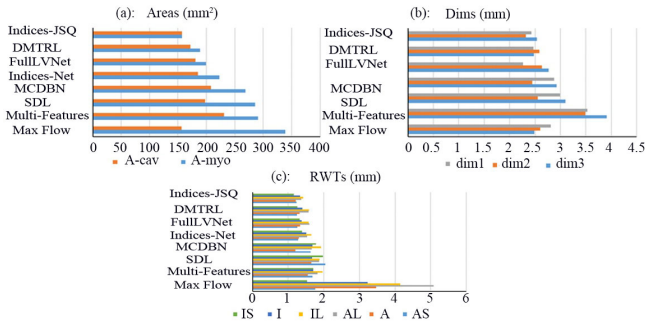

Fig. 11. The MAE comparison of the proposed method Indices-JSQ and existing advanced methods on Areas (a), Dims (b), and RWTs (c) of clinical indices.

RWTs estimation among our method and the other advanced methods is shown in Fig. 11(c). We can see that except Max Flow, the differences between the other methods are not obvious, and the MAE values of multiple methods are very close due to the ground truth value of RWTs is very small. However, if we look at it carefully, our method is still superior to the other methods for every index of RWTs.

IV. Conclusion

In this study, we propose a new end-to-end segmentation-based deep multi-task regression learning neural network model (Indices-JSQ) to make a holonomic quantitative analysis of the LV. The experimental results show that Indices-JSQ with segmented contours instead of original images, which has better performance and more accurate result for LV quantification.

References

[1] M. D. Cerqueira *et al.*, " Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: A statement for healthcare professionals from the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association," *Journal of the American Society of Echocardiography*, vol. 15, no. 5, pp. 463-467, May. 2002.

[2] C. Petitjean and J.-N. Dacher, "A review of segmentation methods in short axis cardiac MR images," *Medical Image Analysis*, vol. 15, no. 2, pp. 169-184, Apr. 2011.

[3] 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 Transactions on Biomedical Engineering*, vol. 61, no. 4, pp. 1251-1260, Apr. 2014.

[4] X. Zhen, Z. Wang, A. Islam, M. Bhaduri, I. Chan, and S. Li, "Direct estimation of cardiac bi-ventricular volumes with regression forests," in *International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI*, Boston, MA, 2014, pp. 586-593.

[5] X. Zhen, A. Islam, M. Bhaduri, I. Chan, and S. Li, "Direct and Simultaneous Four-Chamber Volume Estimation by Multi-Output Regression," *Prediction of Trabecular Bone Anisotropy from Quantitative Computed Tomography Using Supervised Learning and a Novel Morphometric Feature Descriptor*, Berlin, Germany: Springer, 2015, pp. 669-676.

[6] 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," *Medical Image Analysis*, vol. 30, pp. 120-129, Jul. 2015.

[7] W. Xue, G. Brahm, S. Pandey, S. Leung, and S. Li, "Full left ventricle quantification via deep multitask relationships learning," *Medical Image Analysis*, vol. 43, pp. 54-65, Dec. 2017.

[8] J. Shi, J. Wu, Y. Li, Q. Zhang and S. Ying, "Histopathological Image Classification With Color Pattern Random Binary Hashing-Based PCANet and Matrix-Form Classifier," in *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 5, pp. 1327-1337, Sept. 2017.

[9] J. Shi, X. Zheng, Y. Li, Q. Zhang and S. Ying, "Multimodal Neuroimaging Feature Learning With Multimodal Stacked Deep Polynomial Networks for Diagnosis of Alzheimer's Disease," in *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 1, pp. 173-183, Jan. 2018.

[10] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Scene Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481-2495, Dec. 2017.

[11] J. Johnson, A. Karpathy and L. Fei-Fei, "DenseCap: Fully Convolutional Localization Networks for Dense Captioning," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 2016, pp. 4565-4574.

[12] 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," *Information Processing in Medical Imaging*, Berlin, Germany: Springer, 2017, pp. 505-516.

[13] J. Zhou, L. Yuan, J. Liu, and J. Ye, "A multi-task learning formulation for predicting disease progression," in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Diego, California, 2011, pp. 814-822.

[14] K. Lin and J. Zhou, "Interactive Multi-task Relationship Learning," *2016 IEEE 16th International Conference on Data Mining (ICDM)*, Barcelona, Spain, 2016, pp. 241-250.

[15] S. Ying, Z. Wen, J. Shi, Y. Peng, J. Peng and H. Qiao, "Manifold Preserving: An Intrinsic Approach for Semisupervised Distance Metric Learning," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 7, pp. 2731-2742, July 2018.

[16] P. V. Tran, " A Fully Convolutional Neural Network for Cardiac Segmentation in Short-Axis MRI," https://arxiv.org/abs/1604.00494.

[17] M. Avendi et al., "A Combined Deep-Learning and Deformable-Model Approach to Fully Automatic Segmentation of the Left Ventricle in Cardiac MRI," *Medical Image Analysis*, vol. 30, pp. 108-119, May 2016.

[18] H. Noh, S. Hong and B. Han, "Learning Deconvolution Network for Semantic Segmentation," *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, 2015, pp. 1520-1528.

[19] M. Sanchez-Fernandez, M. de-Prado-Cumplido, J. Arenas-Garcia and F. Perez-Cruz, "SVM multiregression for nonlinear channel estimation in multiple-input multiple-output systems," in *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2298-2307, Aug. 2004.

[20] X. Du et al., "Deep Regression Segmentation for Cardiac Bi-Ventricle MR Images," in *IEEE Access*, vol. 6, pp. 3828-3838, Jan. 2018.

[21] A. I. Ben, 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," *Medical Image Analysis*, vol. 16, no. 1, pp. 87-100, Jan.2012.

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

[23] 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 *International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI*, Quebec City, Canada, 2017, pp. 276-284.