This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2789179, IEEE Access

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Deep Regression Segmentation for Cardiac Bi-ventricle MR Images

XIUQUAN DU[1], WEIWEI ZHANG[1], HEYE ZHANG[2], JUN CHEN[1], YANPING ZHANG[1], JAMES CLAUDE WARRINGTON[3,4], GARY BRAHM[3,4] AND SHUO LI[3,4]

[1]School of Computer Science and Technology, Anhui University, Hefei 230039, China
[2]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[3]Department of Medical Imaging, Western University, London, ON N6A 3K7, Canada
[4]The digital imaging group of London, London, ON N6A 3K7, Canada

Corresponding author: Xiuquan Du (dxqllp@163.com) and Weiwei Zhang (mr.ww.zhang@gmail.com)

This work was supported in part by the National Science Foundation of China under Grant 61203290, 61673020 and 61771464, the Provincial Natural Science Research Program of Higher Education Institutions of Anhui province under Grant KJ2016A016, the Anhui Provincial Natural Science Foundation under Grand 1708085QF143, the National High Technology Research and Development Program (863 Program) under Grant SS2015AA020109 and the Shenzhen Innovation funding under Grant SGLH20150213143207911.

ABSTRACT Cardiac bi-ventricle segmentation can help physicians to obtain clinical indices such as mass and volume of left ventricle (LV) and right ventricle (RV). In this paper, we propose a regression segmentation framework to delineate boundaries of bi-ventricle from cardiac MR images by building a regression model automatically and accurately. First, we extract DAISY feature from images. Then, a point based representation method is employed to depict the boundaries. Finally, we use DAISY as input and boundary points as labels to train the regression model based on Deep Belief Network. Regression combined deep learning and DAISY feature can capture high level image information and accurately segment bi-ventricle with fewer assumptions and lower computational cost. In our experiment, the performance of the proposed framework is compared to manual segmentation on 145 clinical subjects (2900 images in total), which are collected from 3 hospitals affiliated with two health care centers (London Healthcare Center and St. Josephs HealthCare). The results of our method and manually segmented method are highly consistent. High Pearson's correlation coefficient (PCC) between automated boundaries and manual annotation is up to 0.995 (endocardium of LV), 0.997 (epicardium of LV), 0.985 (RV). Average Dice metric (DM) is up to 0.916 (endocardium of LV), 0.941 (epicardium of LV), 0.844 (RV). Altogether, experimental results are capable of demonstrating the efficacy of our regression segmentation framework for cardiac MR images.

INDEX TERMS Bi-ventricle segmentation, Regression, DAISY, Deep Belief Network

I. INTRODUCTION

Cardiac bi-ventricle segmentation (BVS) is conducive to quantitative analysis of cardiac MR images, which are typically necessary for diagnosis and treatment of cardiovascular pathologies [1]. However, clinical cardiac MR images with huge appearance variations and complex anatomical structures pose great challenges for manual processing and conventional computer processing methods. Manual segmentation by physicians is slow, tedious and inconsistent, and conventional automated BVS methods are leaky as follows: (1) image-based and pixel-classification based methods, such as thresholding, clustering, region growing and graph cuts, capture low level image information [2-5]. They are usually infeasible when the two ventricles and neighboring structures present analogous intensity profiles; (2) deformable models based methods, such as active contours and level sets, are sensitive to diverse boundary variations and local weak/no boundary/edge caused by different subjects and pathologies, and have high computational cost [6-8]; (3) statistical shape and appearance models based methods, such as active shape model and active appearance model, are greatly affected by the size of training data [9, 10].

In order to increase BVS accuracy, improved methods based on conventional model have been proposed. For example, Hautvast et al. have developed a new active contour model which obtained an invariant contour environment by comparing with gray values in the vertical

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2789179, IEEE Access

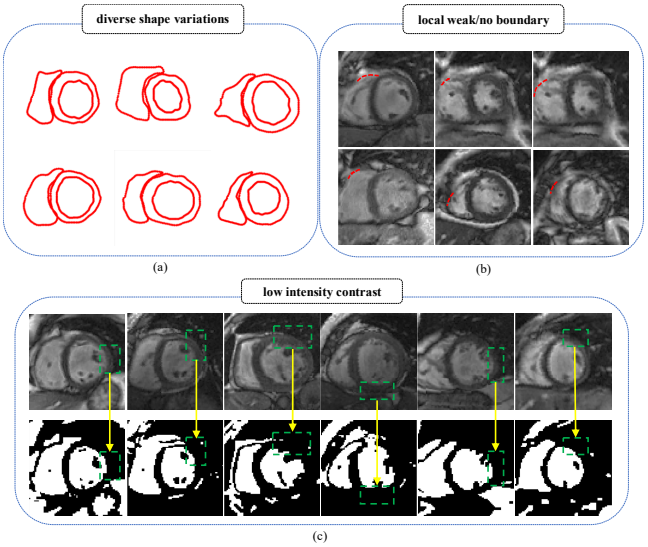

FIGURE 1. The infeasibility of conventional deformation model based methods for following scenes: (a) boundary diversity; (b) local weak/no boundary (indicated by the red dotted lines); (c) low intensity contrast (indicated by the green dashed boxes).

profile. The contours should be kept in the constant position relative to adjacent tissues, which leads to better performance [11]. Differently, Billet et al. have used an electromechanical model to conduct cardiac motion recovery to segment bi-ventricle. The coupling between model and data is realized by a constrained minimization of energy. To reach high segmentation accuracy, the method utilizes special image feature, image intensity and gradient information to perform a better cardiac motion recovery [12]. Among previously improved BVS methods, Bayesian-based methods also have achieved favorable results. For example, Senegas et al. have proposed a full Bayesian model that uses cardiac shape and motion shape as prior information. The method achieves BVS by means of probability distribution of the parameters and computing though Monte Carlo techniques. However, relying on strong prior assumptions can limit its practical application [13]. After that, segmentation-based methods such as active shape model (ASM) or active appearance model (AAM) are introduced. For example, Zhang et al. have proposed a new technique to combine ASM and AAM for 4D cardiac MR images. This method is divided into two stages: the first stage is that it uses the 4D model for initial segmentation in all cardiac phases and the second is applying a 3D model for improvement in accuracy of each phase. It possesses relatively high robustness and precision against ventricular variability caused by tetralogy of Fallot (TOF) [14].

However, many of previous model-based methods are subject to one or more limitations: (1) relying on first boundary initialization step; (2) sensitive to diverse shape variations, low intensity contrast and local weak/no boundary, as shown in Fig. 1; (3) susceptible to local noise and intensity inhomogeneity. Therefore, it is worthwhile to introduce a new general approach, which is more robust to cardiac shape variations and has less preprocessing, for more accurate segmentation of bi-ventricle.

In this paper, we propose a regression segmentation framework (Bi-DBN) to automatically segment bi-ventricle. This framework takes the BVS task as a boundary regression problem and aims at establishing a boundary regression model that implicates the nonlinear mapping relationship between cardiac MR images and desired object boundaries. Although regression-based methods have been previously made and already used for left ventricle (LV) segmentation [15], it is not yet applied on BVS with Deep Belief Network (DBN).

As illustrated in Fig. 2, the framework contains three building blocks: feature extraction, regression and boundary representation. Densely local DAISY feature extracted from cardiac MR image serves as input of the regression model. DAISY is similar to SIFT and can be efficiently computed at every pixel to make it more robust against photometric and geometric transformations [16, 17]. Outputs of the regression model are boundaries of left ventricle and right ventricle (RV) that are represented by a set of points and quantified as coordinates of these points. The point based boundary representation method has high flexibility for cardiac MR images with diverse shape variations. To build reliable regression model from input images to object boundaries, we employ multiple output DBN as regressor instead of multi-dimensional support vector regressor (MSVR) [18].

The regression model performs segmentation by a fashion of holistic regression output and holistic regression input. The holistic output fashion is that model regresses all the boundary points simultaneously rather than one after another. Based on input image and the learned global shape prior, the problem of local shape variations such as partial boundary deficiency is addressed effectively. The holistic input fashion is that model regresses each point using full image as input. It captures full context of each boundary point and reduces local noise disturbance.

In summary, our method makes the following contributions:

1) Application: It is an automated BVS method with fewer assumptions and substantially helps to assess cardiac function indices.
2) Approach: It considers the BVS task as a boundary regression problem and adopts a very flexible boundary representation strategy. It takes advantage of deep learning in a holistic fashion to obtain the optimal segmentation.
3) Methodology: It employs DAISY feature and boundary regression to model the highly nonlinear mapping relationship between diverse cardiac MR images and desired boundaries. DAISY feature serves as DBN's input to capture relevant context of boundary points. Boundary serves as DBN's output and is guided by the learned global shape prior.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2789179, IEEE Access

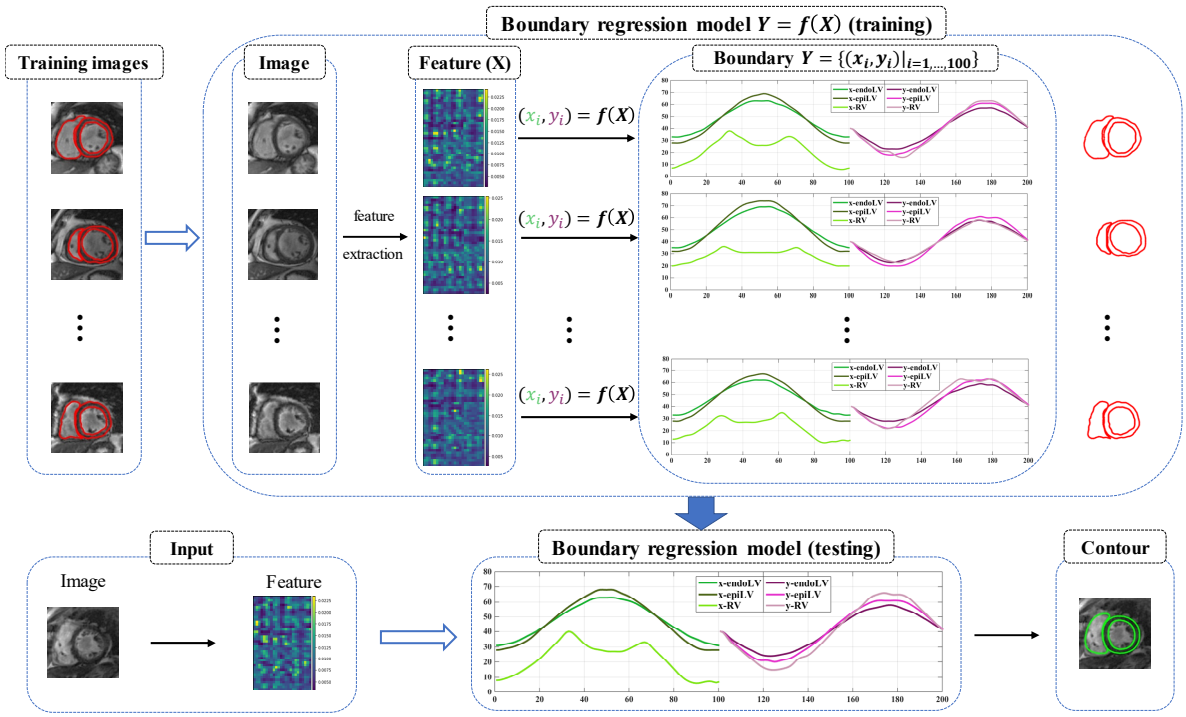

FIGURE 2. **The overview of the proposed Bi-DBN framework. In training, firstly extract DAISY feature of each image, and use 100 points to represent each contour of each image. And then the regression model is trained by the features and the coordinates of boundary points. In testing, desired boundaries are directly obtained from new input image based on the learned regression model. The detailed procedure is described in Section II.**

II. METHODDOLOGY

As shown in Fig. 2, the proposed regression framework mainly comprises of two phases: training and testing. In training phase, a nonlinear regression model $Y = f(X)$ is learned through DBN from training image set $\varsigma = \{(X_\kappa, Y_\kappa)|_{\kappa=1,...,N}\}$ with N images. In testing phase, the learned regression model is utilized to directly estimate endocardium of LV (endoLV), epicardium of LV (epiLV) and RV of the new input image.

For $Y = f(X)$, boundary Y, which serves as the output of the regression model, is represented by a set of points (as described in Section II-A). The flexible boundary representation method requires fewer assumptions and is able to represent any diverse boundaries and even local weak/no boundaries/edge. Image X served as the input of the regression model is represented by a discriminative DAISY feature (as described in Section II-B). The feature can capture relevant information about each boundary point for BVS. The nonlinear function $f(\bullet)$ is derived from DBN that eliminates the disadvantages of linear regression and simultaneously regress all the boundary points in a holistic style (as described in Section II-C). Based on full context information of each boundary point and shape prior learned from training images, the combination of the three foundation blocks (feature extraction, boundary representation and regression) builds the optimally nonlinear mapping relationship between input images and corresponding object boundaries.

A. BOUNDARY REPRESENTATION

In our method, boundary is represented by a set of discrete points and these points are described by their coordinates like [8]:

$$\{v_i = (x_i, y_i)|_{i=1,...,n}\} \quad (1)$$

where v_i is the i-th boundary point's coordinate and n is the number of discrete points of each boundary. Size of n depends on the interval size between every two adjacent points. The smaller interval size leads to the bigger n that provides more realistic representation of boundary and higher precision but with higher computation cost. To smoothly approximate boundary with complicated and diverse shapes, we empirically make n equal 100. The coordinate representation of boundaries is shown in Fig. 3 (a).

We use spline method to obtain these discrete boundary points. The detail steps are as follows: firstly, fix the first point that is the intersection of boundary and horizontal center line, and then sample the remaining n-1 points along the clockwise boundary evenly and successively, as shown in Fig. 3 (a). We get discrete points of each boundary in a same way. Moreover, in both training and testing phases, to ensure the coordinates of boundary points of different images in a consistent space, a scale normalization step is

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2789179, IEEE Access

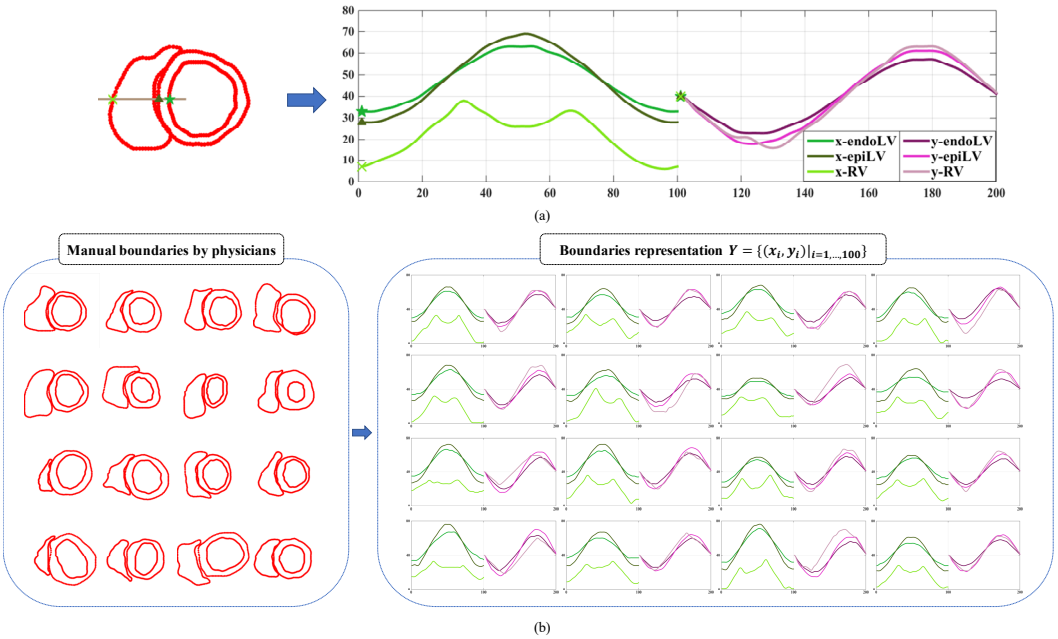

FIGURE 3. **Boundary representation: (a) The graphic illustration of boundary representation. X-mark, triangle and pentagram denote the first sampled boundary points of RV, epiLV and endoLV, respectively. In each boundary, the remaining n-1 points are evenly sampled in clockwise direction. The right-hand diagram illustrates the coordinate representation of 100 boundary points. (b) Flexibility of the point based boundary representation method exhibited on cardiac MR images with diverse appearances. Only coordinates of boundary points change when boundaries present huge deformation.**

performed for all the input images. During the testing process, we need to rescale coordinates of the estimated object boundary points back to the original size [18].

Comparing with other conventional and extensively used representation methods, such as PCA shape [19], our boundary representation method has fewer assumptions and more flexible structure. Therefore, it provides more effective boundary representation when handling cardiac MR images with wide shape variations caused by different subjects, pathologies, modalities and views.

B. REGRESSION INPUT BY DAISY FEATURE

To realize exact and credible boundary regression for cardiac MR images, DAISY feature [16] extracted from image is used as the input of regression model. DAISY is computed around sampled feature points on local regions and calculates histogram only once at each region. Gauss convolution is utilized to converge histogram and executed on gradients of particular direction via Gaussian kernel. Better partitioned strategy and fast computability of Gauss convolution make DAISY be extracted faster.

For a pixel point at location (x, y) of input image I, we firstly compute gradients on H directions (generally, $H = 8$) and obtain orientation maps:

$$G_o(x,y) = \left(\frac{\partial I(x,y)}{\partial o}\right)^+, \text{ subject to } (a)^+ = max(a, 0) \tag{2}$$

where $G_o(x, y)$ is the gradient of point (x, y) on direction o and $I(x, y)$ represents a point of the input image. Then, the same number of convolved orientation maps is obtained by Gaussian convolution:

$$G_o^{\Sigma}(x,y) = G_{\Sigma} * \left(\frac{\partial I(x,y)}{\partial o}\right)^+ \tag{3}$$

where $G_o^{\Sigma}(x, y)$ is the gradient after convolution operation and G_{Σ} is the corresponding Gaussian kernel with standard deviation Σ. By using different Gaussian kernels (Σ) for convolution, gradients of sampled points on concentric circulars and convolved orientation maps are achieved. Thirdly, the gradient histogram of the pixel point (x, y) is calculated as follows:

$$h_{\Sigma}(x,y) = [G_1^{\Sigma}(x,y), ..., G_H^{\Sigma}(x,y)]^{\tau} \tag{4}$$

where $G_1^{\Sigma}(x, y)$, $G_H^{\Sigma}(x, y)$ are gradients of specified directions after Σ -Gaussian convolution. Subsequently, normalization of $h_{\Sigma}(x, y)$ is performed to obtain normalized vector $\tilde{h}_{\Sigma}(x, y)$. Lastly, all the normalized histogram vectors of center pixel and other sampled points on circular layers are cascaded to form full descriptor $\mathrm{D}(x, y)$ as follows:

$$\begin{aligned}\mathrm{D}(x,y) = [&\tilde{h}_{\Sigma_1}^{\tau}(x,y),\\ &\tilde{h}_{\Sigma_1}^{\tau}(l_1(x,y,R_1)),...,\tilde{h}_{\Sigma_1}^{\tau}(l_T(x,y,R_1)),\\ &\tilde{h}_{\Sigma_2}^{\tau}(l_1(x,y,R_2)),...,\tilde{h}_{\Sigma_2}^{\tau}(l_T(x,y,R_2)),\\ &\cdots\\ &\tilde{h}_{\Sigma_Q}^{\tau}(l_1(x,y,R_Q)),...,\tilde{h}_{\Sigma_Q}^{\tau}(l_T(x,y,R_Q))]^{\tau}\end{aligned} \tag{5}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2789179, IEEE Access

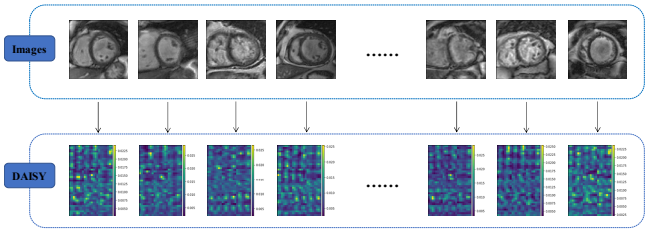

FIGURE 4. A set of cardiac MR images and corresponding DAISY features. The values represented by colors are given by the color scale on the right of each graphic DAISY descriptor.

where $l_d(x, y, R)$ is the location that has distance R and direction d away from (x, y), T and Q denote the number of directions and circular layers, respectively.

The shape and size of DAISY are determined by four parameters: radius (R, the distance between center point and the outermost sampled points), radius quantization (Q, the number of convolved orientations layers), angular quantization (T, the number of histograms of each layer), number of bins of each histogram (H) and spacing of adjacent sampled points (*step*). The dimension of descriptor for each sampled point is $(Q*T+1)*H$. Lager dimension brings higher accuracy, but requires higher computational cost. According to authors' analysis, different scenes correspond to different parameter sets. Multiple parameters settings have similarly high performance for the same application. Considering memory and speed requirements, we use the parameter set *step* = 20, $R=15$, $Q=2$, $T=6$, $H=8$ and the feature dimension is 936. Given an image set, the resulting DAISY descriptors are shown in Fig. 4.

C. REGRESSION BY DBN

Training image set $\varsigma = \{(X_\kappa, Y_\kappa)|_{\kappa=1,\ldots,N}\}$ comprises of N cardiac MR images. κ-th image provides feature vector X_κ and $2n$-dimensional coordinate vector Y_κ that are presented by (x, y) of n boundary points, as indicated in Fig. 2. The nonlinear function f is given by (6) based on DBN [20]:

$$Y = f(X) \tag{6}$$

DBN, which stacks by some Restricted Boltzmann Machines (RBMs) [21], is one of deep neural network techniques. The learning module, RBM, is easy to learn connection weights. It consists of only one hidden layer that served as feature detector, and one visible layer (i.e. input). A layer connects to another, which is not allowed in inner layer. As a result, each neuron of a layer has conditional independence, so we can parallel calculate the values of whole layer's neuron. Assumed that all nodes are random binary-state variable (0, 1), the joint configuration of visible (v) and hidden nodes (h) is described by an energy function based on the learned weight and bias [21]:

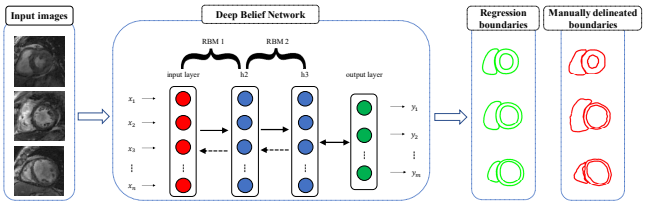

FIGURE 5. Diagrammatic sketch of our DBN architecture that consists of two RBMs and one output layer. Red neurons represent visible layer of the first RBM as well as input layer of DBN. Light green curves denote desired boundaries of the input cardiac MR images with myocardial hypertrophy, regional wall motion abnormalities and mildly enlarged LV, respectively, based on the graphic DBN architecture. Red curves denote corresponding manually segmented boundaries.

$$E(v,h) = -\sum_{i,j} v_i h_j w_{ij} - \sum_i a_i v_i - \sum_j b_j h_j \tag{7}$$

where v_i, a_i are the binary states and bias of visible node i, while h_j, b_j are the binary state and bias of hidden node j, and w_{ij} is the connection weight between i and j. For our model, because real-valued feature is in visible units, we calculate the $E(v,h)$ as followed [22]:

$$E(v,h) = \frac{1}{2}\sum_i v_i^2 - \sum_{i,j} v_i h_j w_{ij} - \sum_i a_i v_i - \sum_j b_j h_j \tag{8}$$

And then, the probability energy of an RBM is determined by function:

$$p(v,h) = \frac{1}{z}\exp(-E(v,h)) \tag{9}$$

where the partition function, Z, is the summation of all sets of hidden and visible vectors. Each RBM is trained via the contrastive divergence learning algorithm and all RBMs of DBN architecture are trained one by one [23]. Another layer with $2n$ units is connected to the last RBM and serves as the output of DBN. The fine-tune procedure is conducted using stochastic gradient ascent (SGD).

In our framework, we use discriminative feature of image as input of DBN and utilize multiple output of the net to realize regression. Our DBN network is comprised of two RBMs and one additional output layer by Theano and its detailed configuration is described in Table 1 [24]. We empirically set 'batch-size' to 64 and randomly select 20% samples from training set as validation set. In training, iteration parameter 'epoch' is an influential parameter. As

TABLE 1. DBN configuration in our Bi-DBN framework. Dense 2 is used as hidden layer of RBM 1 as well as visible (input) layer of RBM 2.

Structures	Layers	Size	Description
Input	—	936	—
RBM 1	Dense 1	768	ReLu + Dropout (50%)
RBM 2	Dense 2	256	ReLu + Dropout (50%)
	Dense 3	128	ReLu + Dropout (50%)
Output	Dense 4	200	Default

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2789179, IEEE Access

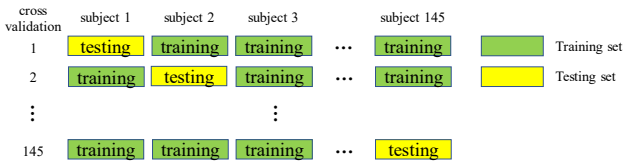

FIGURE 6. Illustration of leave-one-out cross validation. For each cross-validation process, use different subject for testing and the remainder for training.

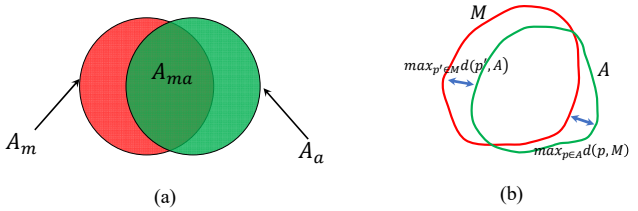

FIGURE 7. Graphic illustration of dice metric and Hausdorff distance. (a) Red (A_m) and green areas (A_a) are manually and automatically segmented, respectively. The intersection is represented by A_{ma} . (b) M and A denote manual and automated contour, respectively.

epoch increases, the segmentation performance is improved. To meet memory and speed requirements, we set epoch to 500. Fine-tune is conducted continuously until the error between estimated boundary by Bi-DBN and manually delineated boundary is small enough. From the learned regression model, the estimated boundaries of different pathologies such as myocardial hypertrophy, regional wall motion abnormalities and mildly enlarged LV are indicated in Fig. 5. It is worth mentioning that the network configuration is applicable for endoLV, epiLV and RV in our framework.

III. EXPERIMENTS

In this section, we describe the experimental setting from following aspects: data, cross validation, evaluation metric and comparison experiments.

A. DATA

A dataset that includes 2900 cardiac MR images from 145 subjects is used to evaluate the segmentation performance of our regression framework [25]. These subjects are between 16 and 97 years old, with an average age of 58.7. They derive from 3 hospitals affiliated with two health care centers (London Healthcare Center and St. Josephs HealthCare). The images are 2D short-axis cine MRI with pixel spacing from 0.6836 mm to 2.0833 mm and mode of 1.5625 mm. They present a mix of pathologies: regional wall motion abnormalities, myocardial hypertrophy, mildly enlarged LV, LV dysfunction, etc. 20 frames are collected for each subject throughout a full cardiac cycle. In each frame, LV and RV are divided into three slices (basal, mid-cavity, and apical), respectively, along to the short axis of heart. In this study, we chose the representative mid-cavity slice to validate our regression framework.

Before manually annotating ground truth contours, several preprocessing steps are executed to obtain ROI: (1) manually label junctions of RV wall with LV; (2) constantly rotate image to make the line between the two signs become vertical; (3) image is cropped to a squared area centered with the mid-perpendicular of this line as ROI, and the length of its sides are two times of the distance between the two signs; (4) resize ROI image to 80 × 80 pixels. After the preprocessing procedure, all ROI images are manually segmented to obtain endoLV, epiLV and RV, which are checked by two experts within this field.

B. LEAVE-ONE-OUT CROSS VALIDATION

We evaluate our framework using leave-one-out cross validation in an ordinary PC (3.4 GHz C2D PC with 4 GB RAM and Win 10). As depicted in Fig. 6, the detailed process is as follows: chose one subject (20 images) for testing, and the remaining 144 subjects (total 2880 images) as training set. This process is repeatedly executed 145 times and different subject is used as testing set each time. To improve robustness, we repeat the above leave-one-out cross validation procedure 5 times, the average of 5 times is used as the final evaluation value.

C. EVALUATION METRIC

In this paper, we apply several measures to evaluate performance of the proposed Bi-DBN framework: Pearson's correlation coefficient (PCC), Dice Metric (DM), Hausdorff Distance (HD). We explain DM and HD detailly in the following.

Dice Metric: Average dice metric (DM) [26] measures the overlap (A_{ma}) between manually segmented (A_m) and automatically segmented contour area (A_a) by our method. DM always ranges between [0, 1]. The larger DM value is, the higher consistency between manual and automated segmentation we get. As indicated in formula (10) and Fig. 7 (a):

$$\mathrm{DM}(A_m, A_a) = \frac{2A_{ma}}{A_m + A_a} \tag{10}$$

Hausdorff Distance: Another evaluation metric is the Hausdorff distance (HD) [27], as shown in Fig. 7 (b). First, we define the minimum distance (MD) from a point (p) on automated contour (A) to its nearest point (p') of manual contour (M), as followed:

$$d(p, M) = \min_{p' \in M} \| p - p' \| \tag{11}$$

then, for all points in A and M , we find the maximal MD $d(p, M)$ and $d(p', A)$. HD is the maximum of the two values and always during $[0, \infty]$. HD increases, the performance degrades.

$$\mathrm{HD}(A_m, A_a) = \max(\max_{p \in A} d(p, M),\ \max_{p' \in M} d(p', A)) \tag{12}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2789179, IEEE Access

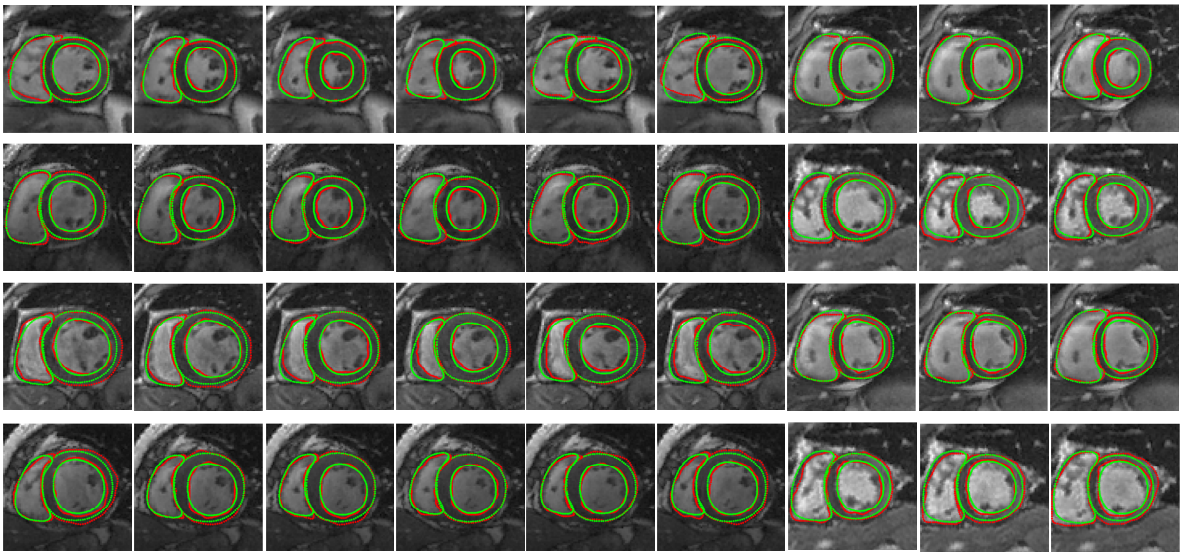

FIGURE 8. Visualization of segmentation results of representative cardiac MR images. These samples embody following challenges: diverse shape variations, having similar intensity with adjacent tissues, noise disturbance and local weak/no edge. The shown contours are defined for RV, epiLV and endoLV from left to right, respectively. The automated segmentation results of our method are depicted by green curves, and the corresponding ground truth manually delimited by experts is depicted by red curves.

D. EXPERIMENTS

A large number of experiments have been constructed to demonstrate the effectiveness and feasibility of our Bi-DBN from several aspects.

Firstly, we evaluate performance of Bi-DBN on our dataset of 2900 cardiac MR images by leave-one-out cross validation protocol (Section III-B).

Secondly, the advantage of regression by DBN is embodied by comparing with MSVR, a commonly used regression method [8, 18, 28, 29]. We have implemented MSVR using HOG and DAISY feature and the same protocol discussed in Section III-B for fair comparison.

Thirdly, advantage of DAISY is embodied by comparing with several representative features: GIST, HOG and PHOG [30-32]. Dimensions of the three feature descriptors are 1152, 864 and 680.

Finally, several popular conventional segmentation methods are conducted with our dataset: 1) graph cuts, which is proposed by Boykov et al. aimed to search the min cut between object and background based on max-flow algorithm [33]; 2) active contours, which performs segmentation by minimizing energy function via continuously approximating ground truth based on prior contour [6]; 3) level set, that desired boundary is evolved from original curve by means of a 3D surface [7].

We have realized these comparable experiments by our own codes following authors' description. In order to make a fair comparison, the input parameter set of each method is experimentally obtained to achieve their best performance and corresponding interior parameters remain fixed. Each method is repeatedly performed 5 times and average results are used for comparison.

IV. RESULTS AND ANALYSIS

In this section, we demonstrate effectiveness and feasibility of Bi-DBN by the above-mentioned experiments and also analyses the advantages of Bi-DBN over several popular segmentation methods.

A. SEGMENTATION PERFORMANCE ANALYSIS OF BI-DBN

The accuracy and robustness of the proposed Bi-DBN framework are explained by following qualitative and quantitative analysis. Qualitative analysis intuitively shows the effectiveness and robustness of our framework for BVS of challenging cardiac MR images with diverse boundaries and local weak/no boundary. Quantitative analysis numerically depicts the efficiency and robustness of our framework using three evaluation metrics.

Qualitative Analysis: Fig. 8 shows some representative cardiac MR images possessed some pathologies, such as regional wall motion abnormalities, myocardial hypertrophy, which are challenging for published methods. The red dashed represents ground-true contours, and the green dashed is automated segmentation results by our framework. Three contours indicate RV, epiLV and endoLV from left to right, respectively. It shows that the intersection of estimated and manual area is large, which indicates they have high consistency. The visualization results clearly illustrate that Bi-DBN is capable to accurately segment bi-ventricle for cardiac MR images, though diverse shape variations happen owing to heart motion or pathologies.

As followed, we discuss the effectiveness and robustness of our framework from two aspects. Firstly, in our

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2789179, IEEE Access

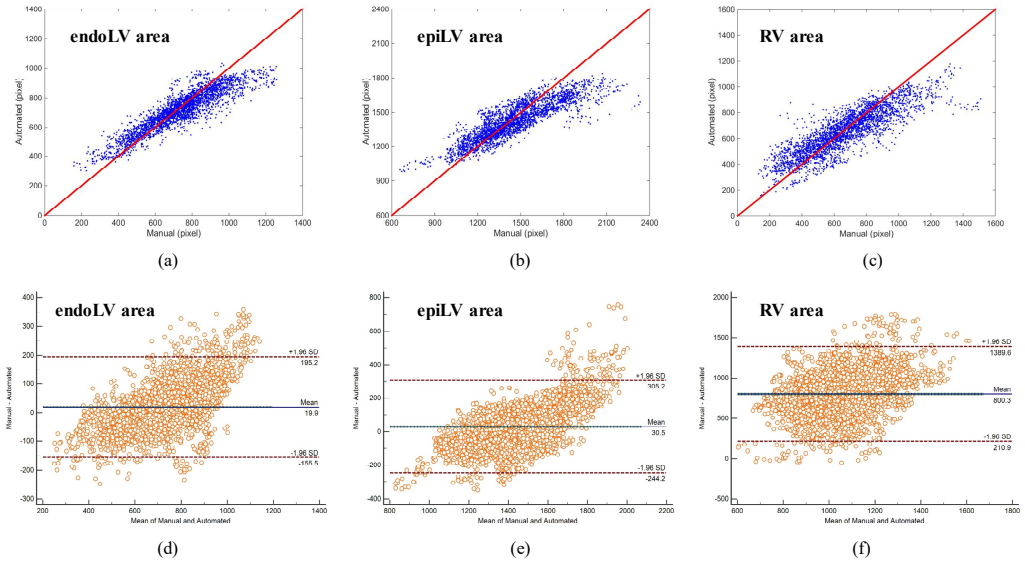

FIGURE 9. Comparison between automated areas and manual areas. Linear regression analysis: (a) endoLV, (b) epiLV and (c) RV. Bland-Altman analysis: (d) endoLV, (e) epiLV and (f) RV.

framework, we use a point based method to represent each boundary (in Section II-A and Fig. 3). Based on this strategy, each boundary is quantified as a coordinate vector. As shown in Fig. 3 (b), if boundaries present huge shape variations, right-hand coordinate curves only show slight fluctuations, which make our framework to have high robustness. Secondly, during testing procedure, all boundary points are regressed simultaneously and each point is regressed using full image as input of the learned regression model. Abundant information of boundary points is captured. The impacts of local noise disturbance and intensity inhomogeneity are greatly weakened. The local missing parts of boundary, always causes serious segmentation problem, can be padded based on learned global shape prior. Consequently, the proposed regression framework realizes accurate and robust BVS for cardiac MR images.

Quantitative Analysis: Table 2 shows the average results of endoLV, epiLV and RV for 2900 images, respectively. The average PCC, DM, HD of endoLV and epiLV are 0.995, 0.916, 2.902 mm and 0.997, 0.941, 3.000 mm, respectively, owing to their more obvious boundaries and fewer shape variations. Relatively speaking, RV achieves lower PCC (0.985), DM (0.844) and higher HD (5.468 mm) for its crescent shape and resulting complex diversity. Small standard deviations indicate the robustness of our framework against cardiac MR images with diverse deformed boundaries. For each image, higher PCC between manual and automated boundary points indicates better segmentation performance. Fig. 9 gives the linear regression analysis and Bland-Altman analysis about automated area (A_a) by Bi-DBN and manual obtained area (A_m). Each blue point is closer to the $y = x$ line, the overlap between automated and manual area is larger. Most of the blue points around the $y = x$ line demonstrate the efficiency of our framework for BVS task.

TABLE 2. The total performance of the Bi-DBN in terms of PCC, DM and HD (mean value ± standard deviation).

Structures	PCC	DM	HD (mm)
endoLV	0.995	0.916 ± 0.040	2.902 ± 1.142
epiLV	0.997	0.941 ± 0.026	3.000 ± 1.128
RV	0.985	0.844 ± 0.074	5.468 ± 2.552

B. PERFORMANCE COMPARISON WITH MSVR

Fig. 10 shows average DM values of 30 subjects by Bi-DBN and MSVR. Average DM by Bi-DBN are 0.916 (endoLV), 0.941 (epiLV) and 0.844 (RV) (green lines). And MSVR achieves average DM of 0.847 (endoLV), 0.900 (epiLV) and 0.749 (RV) using HOG feature (black lines), average DM of 0.892 (endoLV), 0.924 (epiLV) and 0.808 (RV) using DAISY feature (purple lines). For each group bins, we can see that majority bins of Bi-DBN are higher than others, which demonstrates that our method performs better than MSVR with HOG or DAISY for our BVS task.

We discuss why DBN serves as the regressor rather than MSVR in our framework. The main component of DBN is RBM. RBM hides some variables and expresses complex distributions of input data powerfully [21]. The characteristic just indicates diverse shape variations of cardiac MR images. The hidden layer of DBN is also called feature detector, adjusting the number and size of hidden layers can help to simulate complex nonlinear mapping function about cardiac MR images and object boundaries.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2789179, IEEE Access

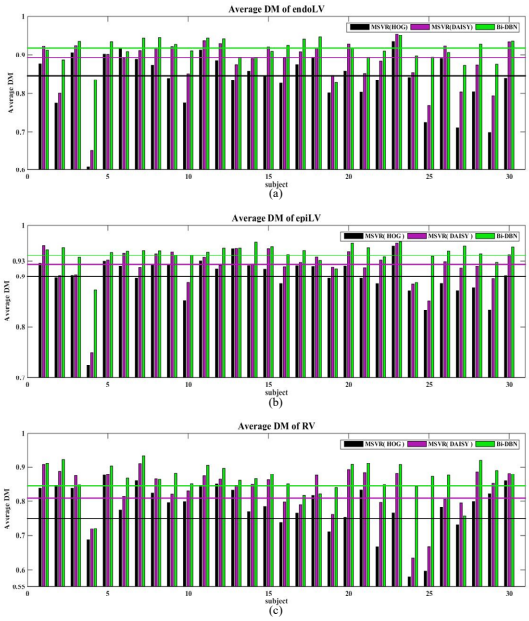

FIGURE 10. DM histograms of top 30 subjects by Bi-DBN and MSVR: (a) endoLV, (b) epiLV and (c) RV. A group bins, i.e. three adjacent bins with different colors, depict comparative results of the same subject based on different methods. Three lines denote average DMs of 145 subjects by MSVR (HOG), MSVR (DAISY) and Bi-DBN, respectively.

As a deep learning technology, DBN is capable to deal with large dataset compared to MSVR.

C. PERFORMANCE COMPARISON WITH VARIOUS FEATURES

Fig. 11 shows the box plot analysis of DAISY, GIST, HOG and PHOG. Bi-DBN achieves highest DM values, 0.916 (endoLV), 0.941 (epiLV) and 0.844 (RV) using DAISY feature. The next best is HOG, and it is M=0.905, due to applying the information of gradient and preserving good invariance of image geometry, but it is still inferior to DAISY. The shortest whisker of DAISY demonstrates that it is effective and robust against challenging cardiac MR images. The results indicate that DAISY is more suitable for BVS task in our Bi-DBN framework.

D. PERFORMANCE COMPARISON WITH CONVENTIONAL METHODS

Fig. 12 intuitively illustrates the segmentation results of 10 representative cardiac MR images for different methods. These images are selected based on some challenges for BVS: (1) diverse shape variations; (2) local weak/no boundary; (3) low intensity contrast (as shown in Fig. 12(a)). Results of other three conventional methods are obviously worse than our Bi-DBN. The comparative qualitative metrics are listed in Table 3, which shows that Bi-DBN obtains the highest DM values.

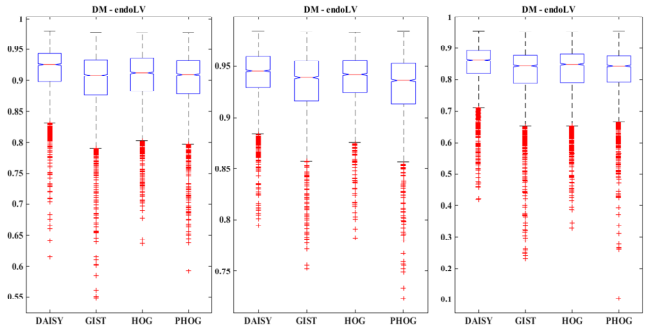

FIGURE 11. Box plot of DM values used different features. They denote endoLV, epiLV and RV from left to right, respectively.

In the following, we discuss advantages of Bi-DBN over the above-mentioned methods from several aspects. Firstly, our Bi-DBN framework formulates bi-ventricle segmentation task as a boundary regression problem. It establishes the nonlinear relationship between cardiac MR image and boundaries of LV and RV through a regression model, rather than performs segmentation from region level and pixel classification. This idea greatly reduces impacts of noise disturbance and intensity inhomogeneity, which often obstruct the optimal result of graph cuts (as shown in Fig. 12 (d)). Secondly, Bi-DBN utilizes a point based boundary representation strategy to flexibly handle boundary diversity. All points of each boundary are obtained simultaneously in a regression process based on the corresponding MR image and learned global shape prior. Consequently, diverse shape variations and missing parts of boundary that caused poor performances for active contours (Fig. 12 (e)) and level set (Fig. 12 (f)) are effectively addressed. Thirdly, active contours and level set methods are semi-automatic and their segmentation results are affected by the quality of initial contour. Our Bi-DBN hardly needs assumption and conducts regression using DBN, an end-to-end network, which largely improves its practical application ability. Fig. 12 (c) and Table 3 intuitively and quantitatively indicate the efficacy of our segmentation framework, respectively.

TABLE 3. Comparison between our method and several popular segmentation methods in terms of DM, HD (mean value ± standard deviation).

Methods	DM			HD (mm)		
	endoLV	epiLV	RV	endoLV	epiLV	RV
Graph cuts	0.801 ± 0.159	0.842 ± 0.012	0.795 ± 0.165	9.290 ± 5.800	4.200 ± 0.900	13.686 ± 12.911
Active contours	0.800 ± 0.102	0.921 ± 0.029	0.788 ± 0.066	4.700 ± 0.982	3.483 ± 0.673	7.708 ± 1.021
Level set	0.846 ± 0.106	0.909 ± 0.023	0.831 ± 0.064	5.739 ± 1.819	5.484 ± 1.216	5.900 ± 1.409
Bi-DBN	**0.916 ± 0.040**	**0.941 ± 0.026**	**0.844 ± 0.074**	**2.902 ± 1.142**	**3.000 ± 1.128**	**5.468 ± 2.552**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2789179, IEEE Access

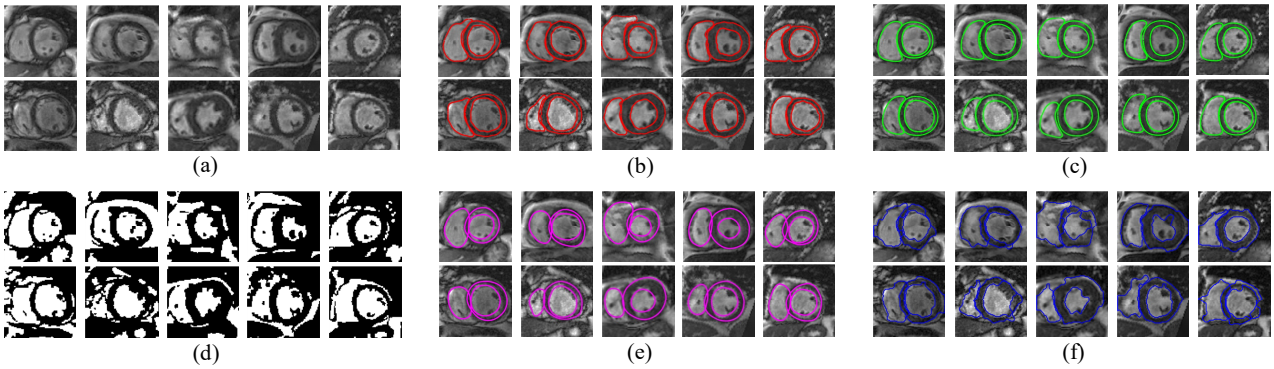

FIGURE 12. Visualization of BVS performance of 10 images based on different methods. (a) Original images with great diversity in boundary shape, local weak/no boundaries and analogous intensity profile with adjacent tissues; (b) manual contours signed by experts (red dashed); (c) results of Bi-DBN; (d) results of graph cuts; (e) results of active contours (mauve dashed); (f) results of level set (blue dashed).

V. CONCLUSION

Cardiac bi-ventricle segmentation helps to obtain some clinical indices such as mass and volume of LV and RV and plays an important role in the diagnostic procedure of cardiovascular pathologies. In this paper, we have proposed a regression segmentation framework that considers the bi-ventricle segmentation task as a mathematic regression problem. The framework aims to build a regression model to simulate highly nonlinear relationship of cardiac MR images and object boundaries of LV and RV. The regression model is trained by DIASY extracted from images instead of previously frequently-used feature. To flexibly depict diverse boundaries, a point based representation method is applied. Moreover, Deep Belief Network is used to formulate nonlinear relationship of images and corresponding boundaries. Performance evaluation of our framework is performed on a dataset of cardiac MR images from 145 patients. The experimental results show that our framework has extremely approximate accuracy with manual delineated method and can be considered effective for bi-ventricle segmentation task.

REFERENCES

[1] X. Zhen, Y. Yin, M. Bhaduri, I. B. Nachum, D. Laidley, and S. Li, "Multi-task shape regression for medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016, pp. 210-218: Springer.

[2] A. Katouzian, A. Prakash, and E. Konofagou, "A New Automated Technique for Left-and Right-Ventricular Segmentation in Magnetic Resonance Imaging," in *Engineering in Medicine and Biology Society, 2006. Embs '06. International Conference of the IEEE*, 2006, pp. 3074-3077.

[3] M. Lynch, O. Ghita, and P. F. Whelan, "Automatic segmentation of the left ventricle cavity and myocardium in MRI data," *Computers in biology and medicine,* vol. 36, no. 4, pp. 389-407, 2006.

[4] G. Mühlenbruch *et al.*, "Global left ventricular function in cardiac CT. Evaluation of an automated 3D region-growing segmentation algorithm," *European radiology,* vol. 16, no. 5, pp. 1117-1123, 2006.

[5] A. Suga, K. Fukuda, T. Takiguchi, and Y. Ariki, "Object recognition and segmentation using SIFT and Graph Cuts," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, 2008, pp. 1-4: IEEE.

[6] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International journal of computer vision,* vol. 1, no. 4, pp. 321-331, 1988.

[7] C. Li, C. Xu, C. Gui, and M. D. Fox, "Distance regularized level set evolution and its application to image segmentation," *IEEE transactions on image processing,* vol. 19, no. 12, pp. 3243-3254, 2010.

[8] X. He, A. Lum, M. Sharma, G. Brahm, A. Mercado, and S. Li, "Automated segmentation and area estimation of neural foramina with boundary regression model," *Pattern Recognition,* vol. 63, pp. 625-641, 2017.

[9] J. Abi-Nahed, M.-P. Jolly, and G.-Z. Yang, "Robust active shape models: A robust, generic and simple automatic segmentation tool," *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006,* pp. 1-8, 2006.

[10] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 23, no. 6, pp. 681-685, 2001.

[11] G. Hautvast, S. Lobregt, M. Breeuwer, and F. Gerritsen, "Automatic contour propagation in cine cardiac magnetic resonance images," *IEEE transactions on medical imaging,* vol. 25, no. 11, pp. 1472-1482, 2006.

[12] F. Billet, M. Sermesant, H. Delingette, and N. Ayache, "Cardiac motion recovery and boundary conditions estimation by coupling an electromechanical model and cine-MRI data," *Functional Imaging and Modeling of the Heart,* pp. 376-385, 2009.

[13] J. Senegas, C. A. Cocosco, and T. Netsch, "Model-based segmentation of cardiac MRI cine sequences: a Bayesian formulation," in *Proc. of SPIE Vol*, 2004, vol. 5370, p. 433.

[14] H. Zhang, A. Wahle, R. K. Johnson, T. D. Scholz, and M. Sonka, "4-D cardiac MR image analysis: left and right ventricular morphology and function," *IEEE transactions on medical imaging,* vol. 29, no. 2, pp. 350-364, 2010.

[15] L. K. Tan, Y. M. Liew, E. Lim, and R. A. McLaughlin, "Convolutional neural network regression for short-axis left ventricle segmentation in cardiac cine MR sequences," *Medical Image Analysis,* vol. 39, pp. 78-86, 2017.

[16] E. Tola, V. Lepetit, and P. Fua, "DAISY: An Efficient Dense Descriptor Applied to Wide-Baseline Stereo," *IEEE Transactions on Software Engineering,* vol. 32, no. 5, pp. 815-830, 2010.

[17] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision,* vol. 60, no. 2, pp. 91-110, 2004.

[18] Z. Wang, X. Zhen, K. Tay, and S. Osman, "Regression Segmentation for M3 Spinal Images," *IEEE Transactions on Medical Imaging,* vol. 34, no. 9, pp. 1989-1989, 2014.

[19] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer vision and image understanding,* vol. 61, no. 1, pp. 38-59, 1995.

[20] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation,* vol. 18, no. 7, p. 1527, 2006.

[21] G. Hinton, "A practical guide to training restricted Boltzmann machines," *Momentum,* vol. 9, no. 1, p. 926, 2010.

[22] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *International Conference on Machine Learning, ICML 2009, Montreal, Quebec, Canada, June*, 2009, pp. 609-616.

[23] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science,* vol. 313, no. 5786, p. 504, 2006.

[24] J. Bergstra *et al.*, "Theano: Deep learning on gpus with python," 2011.

[25] 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,* 2017.

[26] M. Lynch, O. Ghita, and P. F. Whelan, "Segmentation of the Left Ventricle of the Heart in 3-D+t MRI Data Using an Optimized Nonrigid Temporal Model," *IEEE Transactions on Medical Imaging,* vol. 27, no. 2, pp. 195-203, 2008.

[27] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 15, no. 9, pp. 850-863, 1993.

[28] 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," *IEEE Transactions on Signal Processing,* vol. 52, no. 8, pp. 2298-2307, 2004.

[29] G. Camps-Valls and E. Soria-Olivas, "Multi-dimensional Function Approximation and Regression Estimation," in *International Conference on Artificial Neural Networks*, 2002, pp. 757-762.

[30] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International journal of computer vision,* vol. 42, no. 3, pp. 145-175, 2001.

[31] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, vol. 1, pp. 886-893: IEEE.

[32] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," in *Proceedings of the 6th ACM international conference on Image and video retrieval*, 2007, pp. 401-408: ACM.