Research Article

A Combined Fully Convolutional Networks and Deformable Model for Automatic Left Ventricle Segmentation Based on 3D Echocardiography

Suyu Dong,[1] Gongning Luo,[1] Kuanquan Wang [1], Shaodong Cao,[2] Qince Li,[1] and Henggui Zhang [1,3,4]

[1]*School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China*
[2]*Department of Radiology, The Fourth Hospital of Harbin Medical University, Harbin 150001, China*
[3]*School of Physics and Astronomy, University of Manchester, Manchester, UK*
[4]*Space Institute of Southern China, Shenzhen, Guangdong, China*

Correspondence should be addressed to Henggui Zhang; henggui.zhang@manchester.ac.uk

Segmentation of the left ventricle (LV) from three-dimensional echocardiography (3DE) plays a key role in the clinical diagnosis of the LV function. In this work, we proposed a new automatic method for the segmentation of LV, based on the fully convolutional networks (FCN) and deformable model. This method implemented a coarse-to-fine framework. Firstly, a new deep fusion network based on feature fusion and transfer learning, combining the residual modules, was proposed to achieve coarse segmentation of LV on 3DE. Secondly, we proposed a method of geometrical model initialization for a deformable model based on the results of coarse segmentation. Thirdly, the deformable model was implemented to further optimize the segmentation results with a regularization item to avoid the leakage between left atria and left ventricle to achieve the goal of fine segmentation of LV. Numerical experiments have demonstrated that the proposed method outperforms the state-of-the-art methods on the challenging CETUS benchmark in the segmentation accuracy and has a potential for practical applications.

1. Introduction

The assessment of left ventricle (LV) function on the echocardiography plays a key role in the diagnosis of heart disease. The LV segmentation based on echocardiographic images is an essential step for LV function assessment in terms of characterizing the ventricular volume, ejection fraction, wall motion abnormalities, and myocardial contractility [1, 2]. Compared to the traditional 2D echocardiography, 3D echocardiography (3DE) allows a real-time 3D visualization of the heart [3]. However, due to some intrinsic limits such as low signal-noise ratio, low spatial and temporal resolutions, and presence of motion artifacts, the fully automatic segmentation of the LV in 3DE is still an open and challenging task [4]. Moreover, manual segmentation is time-consuming and prone to subjective variability [5]. Therefore, an automatic

and accurate LV segmentation method is desirable for accessing the LV function. Current methods for LV segmentation on echocardiography can be classified as deformable models, statistical models, and machine learning methods [6].

Deformable models are widely used for the LV segmentation in echocardiography [7–13]. With this method, an energy function is defined and minimized for accurate detection of the LV boundary. The widely used deformable models can be divided into snake model and level-set model, which are based on the boundary and region information, respectively [14, 15]. Usually, some new constraints are integrated into the common energy function to achieve a more accurate segmentation [10, 12, 16]. Although they are useful in LV segmentation on multimodality medical image data [10, 17, 18], these models have intrinsic limitations as they depend much on the initialization and image quality. In general, good
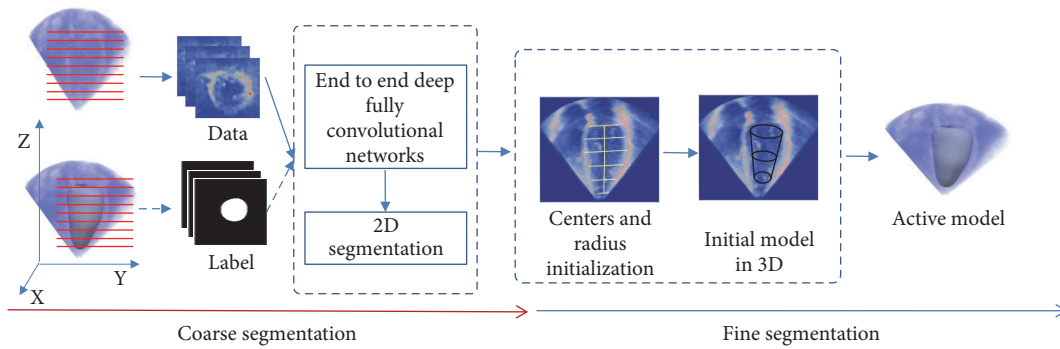
FIGURE 1: The overview of the proposed framework from coarse segmentation to fine segmentation.

initialization is necessary, especially for ultrasound images, which are vague and noisy. Currently the common initialization method of LV segmentation is manual or semiautomatic. Therefore, an accurate and automatic initialization method is the key for fully automatic LV segmentation.

Statistical models, such as active shape model (ASM) [19, 20] and active appearance model (AAM) [21], are based on the statistical information from large labeled data from experts [22]. The statistical information from the labeled data is modeled using some parameters, which are mainly based on contour borders and image textures information in image. Recently, the AAM and ASM have been widely applied to the LV segmentation problem of echocardiography [23–28]. However, due to the dependence on the large number of annotated images, initialization, and assumption of shape and appearance [29–32], statistical models exhibit clear limitations on the automatic segmentation of LV.

Different from statistical models, machine learning methods do not depend on the assumption of shape and appearance. Machine learning methods have demonstrated excellent performance in natural image segmentation. Compared to the traditional machine learning methods, which are based on the hand-crafted features, deep convolutional networks achieved milestone segmentation results in natural images [33–36]. Inspired by the remarkable success in natural image segmentation, recently some studies have successfully implemented the deep convolutional networks for LV segmentation [14, 17, 29, 37, 38]. However, compared to the natural image segmentation, the LV segmentation of 3DE is limited by a lack of large training datasets and low signal-to-noise ratio. So far, few researchers try to formulate the LV segmentation task on 3DE into a deep learning task. Some research combined deep learning method and deformable model to segment LV on cardiac MR images [17, 38], in which deep learning methods were employed to produce a rectangle to detect the region of interest (ROI) of LV, and then other postprocessing method was used to make a final segmentation of LV.

Different from these researches, we proposed a new fully automatic method which employed fully convolutional networks and deformable model for LV segmentation on the 3DE. In this work, we advance our preliminary attempt on

LV segmentation of 3DE [39]. The main contributions are following points:

(1) We formulated the fully automatic LV segmentation task into a coarse-to-fine framework, which includes the coarse segmentation based on deep learning technology and fine segmentation based on the 3D snake respectively.

(2) We employed data augmentation method and the transfer ability of deep convolutional networks between the natural image and 3DE and proposed a new fusion network structure to further improve the coarse segmentation results.

(3) Based on the coarse segmentation results, we proposed a new initialization method utilizing the relation between the spatial position and region sizes of LV in the coarse results and further improved the traditional 3D snake model based on the spatial regularization to generate the fine segmentation results. Besides, compared to the purely end-to-end deep learning method, the proposed method has high interpretability.

(4) Numerical experiments have been carried out to demonstrate that the proposed method outperforms the state-of-the-art methods on the challenging CETUS benchmark in the segmentation accuracy and has potential to be applied into clinical application.

The manuscript is structured as follows. In Section 2, our method is described in detail. In Section 3, some experimental results are presented, including a comparison with state-of-the-art methods. In Section 4, we discuss the results and characteristics of the proposed framework. Finally, we conclude this work and discuss its future applications in Section 5.

2. Method

As shown in Figure 1, we formulate the LV segmentation problem into a coarse-to-fine framework. First, to get enough training data for deep learning, we employed an appropriate data augmentation method based on the conversion from 3D volume to 2D slices. Second, an improved deep fully convolutional network (FCN) based on feature fusion and transfer learning was applied to make initial segmentation of all original 2D slices from 3DE. And then, based on the initial segmentation results, the initialization models for the deformable model were constructed. Finally, an improved 3D
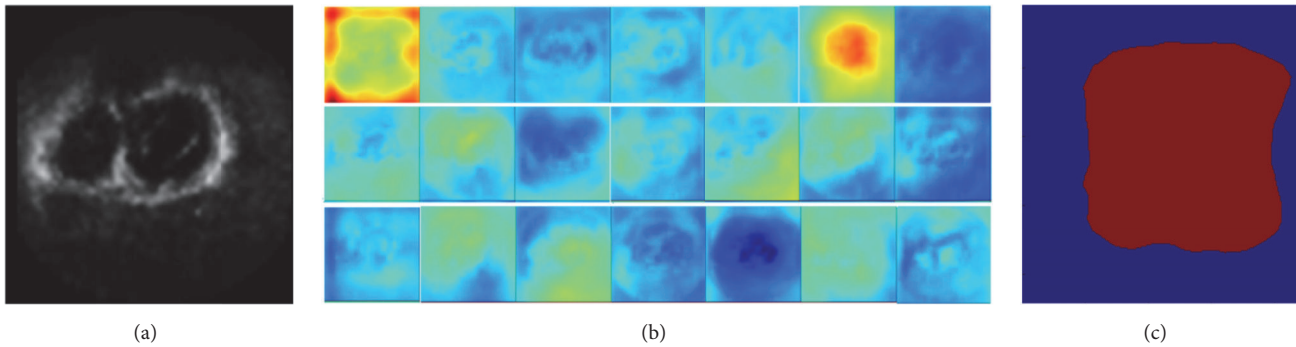
FIGURE 2: The feature maps of the last layer of FCN model with pretrained weights on VOC datasets, adopting the 2D echocardiography as input. (a) The original input 2D slice. (b) 21 feature mappings on the last layer. (c) The final output result of FCN.

snake model was used to segment 3DE. The details of the proposed method are as follows.

2.1. Coarse Segmentation.

We used the deep FCN in [33], which does not contain fully connected layer and uses skip structure for the first time, to achieve the initial segmentation. FCN can provide an accurate localization and initial shape of LV for the further fine segmentation. However, due to limited training datasets, few researchers try to adopt the deep learning technology to address the LV segmentation problem on 3DE. To overcome the problem of insufficient training datasets, we adopted the transfer learning method to initialize the deep FCN model. Additionally, to improve the segmentation accuracy of deep FCN, we proposed a new FCN structure based on the feature fusion across different layers and residual module. Subsequently, we introduce the coarse segmentation method in the aspect of data augmentation and transfer learning and improved the structure of the FCN, respectively.

2.1.1. Date Augmentation.

The CETUS benchmark [3] is believed to have the largest open accessible 3DE datasets. Hence, we adopted the CETUS benchmark datasets to train and validate the FCN model. The CETUS dataset includes 45 3DE subjects, which were acquired from three ultrasound machines of three different vendors. The subjects were divided into 15 training subjects and 30 test subjects, and each subject included two labeled volumes in the end-systole (ES) and end-diastole (ED) frames.

However, due to a small training dataset, the data augmentation is necessary for training deep learning model. To solve this problem, we employed an appropriate data augmentation method based on the conversion from 3D volume to 2D slices. As shown in Figure 1, the 3D volumes were sliced along the Z axis into 2D images, and the 3D mesh of ground truth was simultaneously sliced to match the corresponding 2D images. In this way, all the 2D images were used as the input, so that the number of input samples is increased by N times of the number of 3DE samples (N is the height of Z axis). In this way, we not only get more training samples, but also save the number of parameters on deep learning model (because we convert the input format from 3D to 2D).

Based on the generated 2D images datasets, some traditional data augmentation methods (such as rotation and resizing) can be adopted to achieve further data augmentation to get lots of training images for the following deep FCN training. Note that we copy the channel of an original 2D image three times to generate a 2D image with three channels to fit the input of the FCN for the following transfer learning technology. Because the weights, which are used for transfer learning, are from VOC dataset. The images of this VOC dataset are RGB images, and every image has three channels. Hence, the only way is to copy the channel of an original 2D image three times to make sure the same number of channels to make the transfer learning work well. If we modify the FCN to deal with a single channel, we are unable to transfer the weights properly from the pretrained FCN model on VOC datasets.

2.1.2. Transfer Learning Based on VOC Datasets.

Though the number of training samples was increased by the proposed augmentation method, training the deep FCN model from scratch is difficult. Inspired by some related researches, which adopt the transfer learning technology to overcome the limitation of datasets and avoid the overfitting problems [42–44], we transfer the weights from pretrained FCN model [33] on VOC datasets (which were collected from photo-sharing web site and include 1464 RGB images for pixel-level segmentation) [45] to initialize FCN model and then fine-tune it on augmented echocardiography datasets. To prove the transferability of the pretrained FCN model between the RGB and echocardiography datasets, we used the pretrained FCN model on VOC datasets to predict the pixel classes of the 2D images from 3DE directly. Figure 2 shows the 21 feature maps (output of the last layer in FCN) and the final output result of FCN model with pretrained weights on VOC datasets. We can see that the regions of LV in 2D images have higher response values and can be extracted as a single class. This result proves that the pretrained model has transferability.

2.1.3. Fully Convolutional Networks Using Features Fusion.

We adopted the FCN in [33] as our basic networks structure and modified the number of channels according to the
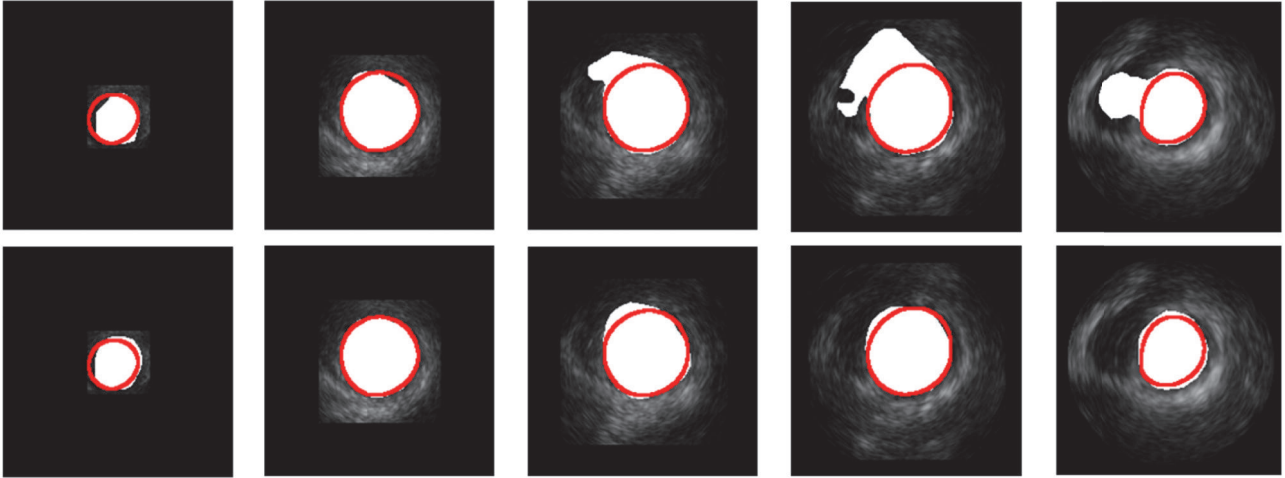
FIGURE 3: Comparing ground truth (red circles) with the segmentation results of original FCN (first row) and improved FCN (second row) of Patient 15 (which was used as validation subject) from the CETUS benchmark [3]. The slices from left to right correspond to the LV parts from apex to base.

number of classes (from 21 to 2) in all deconvolution layers. The 21 classes are aeroplane, bird, bicycle, boat, bottle, bus, car, cat, chair, cow, dining table, dog, horse, motorbike, person, potted plant, sheep, sofa, train, TV/monitor, and background [45]. The 2 classes are LV and background. We used the pretrained models on VOC datasets to initialize all the convolutional layers of the FCN and fine-tuned the whole networks on the adopted echocardiography datasets. The details about networks training are shown in Section 2.1.5. The first row of Figure 3 shows the LV segmentation results using the FCN model trained through transfer learning (using pretrained models), which shows the potential of FCN based on transfer learning. However, to achieve more accurate segmentation, we proposed the new structure to improve the coarse segmentation accuracy.

Inspired by the feature fusion strategy (it is also called skip structure) [33, 46–48] and residual networks [35], we proposed a new feature fusion method with residual connections. The new proposed method is based on an assumption that the feature fusion strategy can combine the feature maps in different levels (the low level in the bottom layers and the high level in the top layers), which are complementary to each other to boost the segmentation accuracy. Additionally, we employed the residual module with identity mapping [49] to encode the mid-layers features, based on the characteristic that the residual networks enable effective backward propagation of the gradient through the identity mapping and achieve fast convergence as well as good feature representation [50]. This characteristic is very important for training relatively complex convolutional neural networks (CNN). As shown in Figure 4, the parts with dotted lines denote the added streams and modules on the original FCN in [33]. In brief, we added two streams from the first two pooling layers (pooling 1 and pooling 2) to the final deconvolution layers, and the residual modules are added to every skip stream to improve the ability of feature representation. The adopted residual module includes 4 convolutional layers, and each layer includes 2

kernels with $3*3$ kernel size following by the BN layer [51] and Relu activation layer (rectified linear units f(x)=max(0, x)). Additionally, because only the LV regions and background were classified in our study, hence, in all deconvolution layers we employed two deconvolution kernels to generate the final two probability feature maps instead of 21 feature maps in the original FCN.

2.1.4. Loss Function. For 2D echocardiography images, the size of LV foreground is usually much smaller than the size of background. Hence, the number of pixels in LV region and the number of pixels in background are heavily imbalanced. Here, given an image I and its ground truth y, a weighted cross-entropy loss function is used to balance LV region and background classification as follows:

$$
Loss = -a \sum_{i=1}^{Y+} \log P\left(y_i = 1 \mid I; w\right) - (1-a)
$$

$$
\cdot \sum_{i=1}^{Y-} \log P\left(y_i = 0 \mid I; w\right)
$$

(1)

where Y_+ and Y_- denote the numbers of pixels belonging to the foreground (LV region) and background, respectively, in the ground truth, a= $Y_-/(Y_++Y_-)$, P denotes the probability of predicted classification, and w means the weights of trained networks.

2.1.5. Network Training. To utilize the pretrained FCN model and the advantage of added residual modules well, we adopted the two-stage manner to train the proposed networks. Firstly, we transferred the pretrained FCN model to initialize the parameters in the whole networks except for the deconvolution layers and the added residual modules. And then, we fixed the parameters of the layers with pretrained initialization and trained the deconvolution layers and the added residual modules; in this way, the deconvolution layers
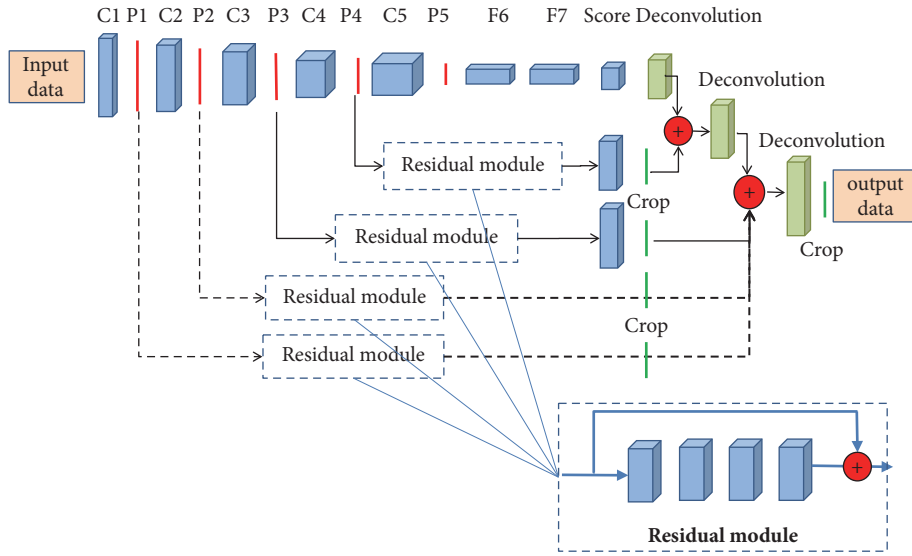
FIGURE 4: The structure of the improved FCN. Based on original FCN in [33], we added two streams from the first two pooling layers (pooling 1 and pooling 2) to the final deconvolution layers, and the residual modules are added to every skip stream to improve the ability of feature representation.

and the added residual modules can be relatively better pretrained for initialization. Finally, we fine-tuned the whole networks to get the model for LV coarse segmentation. The second row of Figure 3 shows the segmentation results of 2D slices from the apex to the base of LV using improved FCN based on features fusion and weighted cross-entropy loss function. Compared to the traditional FCN, the improved FCN can generate more accurate segmentation results. In some cases, the proposed method can discard the parts which are not belonging to the LV, and this is critical for the following fine segmentation step. And more experiments results will be shown in Section 3 to prove the superiority of the proposed network.

2.2. Fine Segmentation. Figure 3 displays the segmentation results in the coarse segmentation stage. Based on the coarse segmentation results, we proposed a fine segmentation method based on 3D initialization and the 3D snake model [52]. Then, we introduce the fine segmentation method in the aspects of the 3D initialization and improved 3D snake model.

2.2.1. Initialization. A good initialization is critical for LV segmentation on 3DE, which is low signal-noise ratio. Some similar works in [17, 38] have proved the feasibility of the 2D initialization based on deep learning for the segmentation task on MR images. Here we proposed an automatic initialization method on 3D space for LV segmentation on 3DE. In general, the initialization task consists of two subtasks: the center localization and the scale estimation of LV region. As shown in Figure 5, the improved FCN has advantages. It is able to get the relatively good segmentation results on 2D slices. To achieve the good center localization in 3D space, firstly, LV centers were estimated through averaging the coordinate values of all foreground pixels on coarse

segmentation results in every slice directly. However, as we can see in Figure 5 (the red curve), the computed centers were noncollinear in 3D space. This problem leads to the misalignment between the adjacent slices and reduce the quality of constructed 3D initialization model. Hence, to achieve more accurate center point localization, as shown in Figure 5, the widely used quadrature curve fitting method [17] was used to correct the estimated centers to avoid the misalignment.

The scale estimation of LV region is on the basis of the assumption that the shape of LV in a 2D slice approximates a circle. This assumption is beneficial to LV initialization, which has been proved in [4]. Hence the radiuses of the LV region were computed using the formula $R = (R_1 + \ldots + R_8)/8$, where the R denotes the average of radiuses in equal angles' samples as shown in Figure 5.

After the center localization and the scale estimation of LV region, we finally obtained some LV contours on slices along the Z axis. To achieve highly efficient model representation, the 3D mesh was reconstructed based on the resampling method. We equally resampled the contours according to the formula $S_i = B_{1+i*k}$, where S denotes the resampled contours set, S_i denotes the ith sampled contours along the Z axis, and k denotes the sampling interval, B denotes original contours set along the Z axis. Finally we reconstructed the initial 3D mesh model through the method in [53].

2.2.2. Active Deformable Model. Based on the initial 3D mesh, an improved 3D snake model was proposed to segment LV from 3DE. Analogous to two-dimensional snake [52], our improved 3D snake was also based on the minimization of energy function and was optimized by gradient descent. Although similar in some ways, an improved energy function is designed in our method. In the proposed 3D snake model,
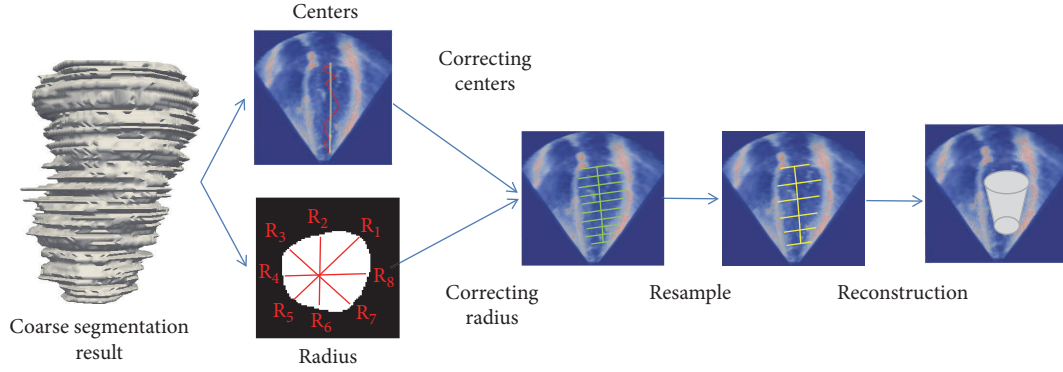
FIGURE 5: The pipeline of automatic 3D model initialization for active deformable model. Based on the segmentation result of improved FCN, we corrected the center and the scale of LV region and then reconstructed the 3D mesh based on the resampling method.

the initialization 3D model obtained from the previous step was used as a 3D spatial regularization constraint for the proposed energy function. Compared to the regularization technology in [17, 38], the proposed 3D spatial regularization technology can appropriately limit the free degree of deformation, to avoid boundary leak between the LV and left atrium (LA) which is a common difficulty in the 3D LV segmentation tasks [14], because the LV and LA are adjacent and have similar pixels intensity. Additionally, the boundary leak due to the presence of papillary muscles can also be avoided.

The final 3D mesh of LV was got when the minimum of energy function was reached. The three-dimensional contour and energy function are defined as the following formulas, respectively:

$$X(u) = [x(u), y(u), z(u)],$$
$$u = (u^1, u^2) \in [0, 1][0, 1] \tag{2}$$

$$E = \int \frac{1}{2} \left(\alpha \left| \frac{\partial X}{\partial u} \right|^2 + \beta \left| \frac{\partial^2 X}{\partial u^2} \right|^2 \right) + \delta E_{ext}(X)$$
$$+ \eta E_{init}(X) \, du \tag{3}$$

where the first term and second term in (3) are internal energy and external energy, respectively. Same as the traditional snake model, X' and X'' are the first-order term and second-order term of X, respectively, and they are controlled by parameter α and β. The first-order term makes the mesh smooth and the second term makes the mesh continuous. The third term $E_{init}(X) = |X - X_{init}|^2$ was the spatial regularization term, in which X is the current mesh and X_{init} is the initial mesh. In this way, we not only can get a relatively finer mesh but also can avoid the absurd deformation due to boundary leak.

3. Evaluation and Results

3.1. Dataset. The proposed method has been tested by using the challenging CETUS benchmark datasets [3]. The ground truth is accessibly open for training set, but not for test set.

Hence the final evaluation results of test set are from the official online evaluation system.

However, to train and evaluate deep FCN model for coarse segmentation, we divided the original 15 training subjects into 12 training subjects (24 volumes) and 3 validation subjects (6 volumes). As described in Section 2.1, we converted the 30 volumes into 5362 2D images; furthermore we used the traditional data augmentation method to rotate and resize the initial 2D images for 10 times randomly to get final 53620 images, in which approaching 45160 and 8460 images were used for training and evaluating deep FCN, respectively.

3.2. Setting. In the coarse segmentation stage, the proposed FCN adopted the training setting similar to that in [33]. We set learning rate 10^{-4} in the type of linear decreasing after every 1000 iterations, the momentum 0.99, the batch size 256, and the max iteration number 10000, respectively. The networks did not use any regularization such as L1, L2 regularization and dropout. Additionally, the initialization of network weights is crucial because of that it directly affects the convergence speed and effectiveness. In order to solve this problem, the pretrained weights coming from FCNs (was trained by in VOC datasets) were used to initialize our network. As described in Section 2.1, we first pretrained the deconvolution layers and residual modules fixing the other layers for first 1000 iterations and then trained the whole networks during the other 9000 iterations. In the fine segmentation stage, to guarantee the repeatability of the shown experiment results, we set the parameters of improved 3D snake $k = 10$, $\alpha = 0.1$, $\beta = 0.2$, $\delta = 0.1$, and $\eta = 0.1$, respectively, though parameters are robust in the most of cases.

In our study, the proposed FCN was implemented based on the widely used Caffe framework [54]. The whole experiment was performed on NVIDIA Titan X GPU.

3.3. Metrics. We adopted official metrics to evaluate the proposed method [3], such as the mean surface distance (d_m), Hausdorff surface distance (d_h), and modified dice similarity index $(D*)$, using which we compared the accuracy

TABLE 1: Segmentation accuracy (pixel accuracy (acc), mean accuracy (mean acc), and mean IOU) of different types of FCN in the stage of coarse segmentation in validation set.

The type of networks	Acc	Mean acc	Mean IOU
FCN	0.874	0.853	0.57
U-net	0.854	0.79	0.59
U-net+T	0.91	0.83	0.75
FCN+T	0.9810	0.9678	0.8485
FCN+T+P$_{(3\&4)}$	0.984	0.969	0.867
FCN+T+P$_{(2\&3\&4)}$	0.989	0.973	0.889
FCN+T+P$_{(1\&2\&3\&4)}$	**0.991**	**0.981**	**0.918**

of our method with the ground truth from official CETUS benchmark [3]. S and S_t denote the surface from the proposed method and surface from ground truth, respectively. Mean surface distance measures the mean distance between S and S_t and it can be computed by

$$d_m = \left[\overline{d}\left(S, S_t\right) + \overline{d}\left(S_t, S\right) \right] \tag{4}$$

$$d_h = \left[\max\left(\min\left(S, S_t\right), \min\left(S_t, S\right)\right) \right] \tag{5}$$

Here, $\overline{d}(S, S_t)$ denotes the mean distance between every voxel from S and the closest voxel from S_t; $\overline{d}(S_t, S)$ is computed in analogous way. As shown in (5), the Hausdorff surface distance measures the local maximum distance between S and S_t. Modified dice similarity index measures the overlap of surface and is computed by $D* = 1 - 2(V \cap V_t)/(V + V_t)$. V and V_t denote the volume from proposed method and ground truth, respectively.

Besides, the performance of our method was also measured by official metrics, the modified correlation (corr∗), and standard deviation (std) in terms of end-diastolic volumes (EDV), end-systolic volumes (ESV), and ejection fractions (EF). EF is calculated by

$$EF = \frac{(EDV - ESV)}{EDV} \tag{6}$$

Additionally, to evaluate the accuracy of coarse segmentation, we adopted the traditional metrics for 2D images segmentation [33], the pixel accuracy (acc), mean accuracy (mean acc), and mean IOU. The IOU measures the region intersection over union for every class, and the mean IOU is the mean value of different classes of IOU [33]. The three metrics are calculated by

$$acc = \frac{\sum_i n_{ii}}{\sum_i t_i} \tag{7}$$

$$meanacc = \frac{\left(\sum_i n_{ii}/t_i\right)}{2} \tag{8}$$

$$meanIOU = \frac{(1/2)\sum_i n_{ii}}{\left(t_i + \sum_j n_{ji} - n_{ii}\right)} \tag{9}$$

where n_{ij} is the number of the pixels of class i predicted to class j; t_i denotes the pixels number of class i.
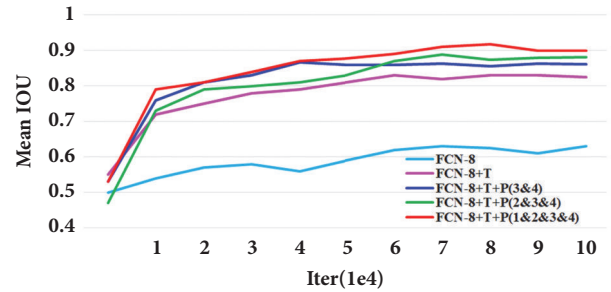

FIGURE 6: Different types of FCN's mean IOU curve along with increasing iterations in validation set.

3.4. Results. To prove the potential and superiority of the proposed coarse-to-fine framework, we present and discuss the experiment results in two aspects: the segmentation accuracy and clinical validation.

3.4.1. Segmentation Accuracy

(1) Performance of FCN in Coarse Segmentation. In order to better understand the importance of the improved FCN in our study, we compared the segmentation accuracy in terms of pixel accuracy, mean accuracy, and mean IOU. By using improved FCN, we obtained the initial segmentation results (coarse segmentation). To achieve clear expression and comparison, we define abbreviation T as the transfer learning based pretrained weights on VOC datasets; P_N denotes the added skip stream with residual modules from Nst pooling layers to the deconvolution layers. As shown in Table 1 and Figure 6, compared to FCN trained without pretrained model, the FCN trained with pretrained model from VOC datasets through transfer learning technology achieves significant improvement. These results prove the importance of the proposed transfer learning strategy. Additionally, as shown in Figure 6, with the increase of the training number, network using transfer learning and residual modules achieves higher accuracy and more rapid convergence than the FCN trained from scratch. Besides, we also can find that the proposed method achieves higher accuracy and faster convergence when it has more skip streams. The best results (pixel accuracy: 0.991; mean accuracy: 0.981; mean IOU: 0.918) were achieved by FCN+T+P (1&2&3&4) using the transfer learning and fused four skip streams. This

observation shows that the fusion of low-level features in the bottom layers and high-level features in the top layers can achieve satisfactory performance in the 3DE semantic segmentation application.

Besides, to evaluate the advantage of the proposed FCN, we conducted the same evaluation using the state-of-the-art medical image segmentation model (U-net [43]). For transfer learning on the U-net, we used the pretrained model provided by [43] to initialize the U-net and then fine-tuned it on the adopted training dataset for 9000 iterations. Besides, to fit the data input format of U-net, we adopted the original one-channel images as the input. As shown in Table 1, the adopted FCN model outperforms the U-net on the two cases (the model trained from scratch and the model trained based transfer learning). These results prove the superiority of the adopted FCN model. Additionally, the Refine-net [45] achieves the best performance on the natural segmentation task. However, the Refine-net is very deep and complex, which cannot fit the current data scale on 3DE. Hence, we do not try to evaluate the performance of Refine-net on LV segmentation task on 3DE.

(2) Performance of Fine Segmentation. In this section, the final fine segmentation performance is evaluated through the official online evaluation system [3], and we compared the segmentation results with state-of-the-art methods (the results of the five state-of-the-art methods come from the official reports in [3]) in terms of d_m, d_h, and $D*$. To show the superiority of the proposed initialization method based on the coarse segmentation using FCN, we conducted the experiments with varying initializations while keeping the traditional deformable model. As shown in Table 2, compared with the traditional initialization method [27], the proposed 3D model initialization method based on the FCN improves the final LV segmentation results. Besides, we also evaluated the proposed method. The evaluation and comparison results are shown in Table 3. We can see that the proposed method distinctly achieved the best segmentation results in mean surface distance and modified dice metric.

Figure 7 shows some segmented LV model from the improved FCN, the improved FCN with traditional snake, and the proposed coarse-to-fine method, respectively. We can see that the proposed method achieved the best segmentation results with the lowest surface distance range. For the improved FCN, due to some obvious leakages in the apex and shrinks in middle part, the segmentation results are relatively worse. The improved FCN with traditional snake can overcome the problem of leakages and shrinks in some extent. Our proposed method, which integrates improved FCN and 3D snake with regularization of constraint term, can further improve the segmentation results on the situations of leakages and shrinks.

3.4.2. Validation. Table 4 and Figure 8 show the correlation between the state-of-the-art segmentation methods (other fully automatic methods and our proposed method) and the ground truth, in terms of EDV, ESV, and EF. The modified correlation values of our proposed method on EDV, ESV, and EF are 0.018, 0.021, and 0.218, respectively, and the

corresponding standard deviations are 12 ml, 11.3ml, and 0.07. Table 3 suggests that our proposed method achieves the best performance in EDV and ESV comparing with the state-of-the-art methods in the aspect of EF correlation. The gap of EF correlation is because of the tendency of EDV's bias and ESV's bias are different, though EDV and ESV are more accurate. For example, we assume the ground truth of EDV and ESV are 100 ml and 60 ml. We also assume the first evaluated values of EDV and ESV are 105 ml and 65 ml, and the second evaluated values of EDV and ESV are 103 ml and 57 ml. Based on these values, we can obtain the EF of the ground truth: the first and second evaluated values are 0.4, 0.381, and 0.4466, respectively. Hence, though the accuracy of volume estimation is improved, the EF estimation accuracy is not improved. Furthermore, Figure 8 also displays the results of Bland-Altman analysis of the proposed method for EDV, ESV, and EF and the mean bias values are 1.41, 3.3, and -0.03, respectively. And the corresponding confidence intervals of EDV, ESV, and EF are [21.7, -24.5], [25.1, -18.6], and [0.13, -0.2]. These results indicate the high agreement between our method and ground truth and the big potential of clinical application.

4. Discussion

In this paper, we proposed a fully automatic LV segmentation method, which is a coarse-to-fine framework based on the deep FCN and deformable model. In the case of insufficient training datasets, we successfully applied the transfer learning technology combined with the improved FCN for the coarse segmentation of 3DE. The experimental results suggested that the FCN model on the domain of nature image can be successfully transferred to the field of echocardiography images segmentation. This is due to the fact that the low-frequency features, such as edge information and texture information, of VOC images and CETUS images, are the same. Hence, the knowledge of low-frequency feature representations was successfully transferred from VOC images to CETUS images. Otherwise, the added skip streams with residual modules using identity mapping, which fuses different level features from different layers, can make contribution to rapid convergence and high accuracy. This is because the features are different from different layers. Adding the skip streams means that the network has more paths. Hence, adding the skip streams is able to add more different level features to improve the speed of convergence and the accuracy of segmentation. Furthermore, we used the modified loss function, which can solve the problem of data imbalance to improve the segmentation accuracy. As we can see in Figure 7 and Table 3, though the improved FCN improves the coarse segmentation performance and outperform the traditional LV segmentation method, the fine segmentation achieves even better LV segmentation performance. Hence, in Section 2.2, we proposed a 3D LV initialization method based on the coarse segmentation results. As we all know, a good initialization is critical for deformable segmentation model; the results in Table 2 also prove that the proposed 3D LV initialization outperforms the standard initialization method for LV segmentation on 3DE on the case of using the same deformable model.

TABLE 2: The segmentation results of different initialization methods with traditional snake on 3DE in test set.

Method	ED						ES					
	d_m		d_h		dice		d_m		d_h		dice	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
Standard initialization + traditional snake	4.9	1.7	15.9	5.1	0.23	0.18	5.8	1.35	13.5	7.9	0.35	0.15
FCN + traditional snake	4.6	1.58	13.8	4.7	0.29	0.17	5.5	1.67	12.9	7.2	0.38	0.19
FCN+T + traditional snake	3.8	1.34	11.7	4.5	0.26	0.15	4.9	1.27	11.4	5.8	0.34	0.16
FCN+T+P$_{(3\&4)}$ + traditional snake	3.35	0.89	9.8	4.2	0.23	0.13	3.48	0.91	10.1	4.9	0.30	0.14
FCN+T+P$_{(2\&3\&4)}$ + traditional snake	3.15	0.78	9.4	3.7	0.18	0.11	3.1	0.83	9.87	4.2	0.21	0.11
FCN+T+P$_{(1\&2\&3\&4)}$ + traditional snake	**2.9**	**0.7**	**8.9**	**3.1**	**0.13**	**0.008**	**2.8**	**0.75**	**9.3**	**3.9**	**0.16**	**0.05**

TABLE 3: The comparison results of the proposed methods and the five state-of-the-art segmentation methods (Barbosa et al. [28], Milletari et al. [30], van Stralen et al. [31], Smistad et al. [40], and Keraudren et al. [41]) on 3DE in test set.

Method	ED						ES					
	d_m		d_h		dice		d_m		d_h		dice	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
[28]	2.26	0.73	**8.10**	**2.66**	0.106	0.041	2.43	0.91	**8.13**	3.08	0.144	0.057
[30]	2.14	0.68	8.25	3.87	0.107	0.031	2.91	1.01	8.53	**2.30**	0.162	0.062
[31]	2.44	0.91	8.45	3.50	0.121	0.054	2.79	1.24	8.65	2.85	0.165	0.079
[40]	2.62	0.95	8.26	2.98	0.115	0.038	2.92	0.93	8.99	2.98	0.156	0.050
[41]	2.44	0.95	8.98	3.09	0.130	0.048	2.54	0.75	9.15	3.24	0.158	0.057
Standard initialization + improved snake	4.7	1.76	13.2	4.9	0.22	0.15	4.9	1.7	11.2	5.7	0.25	0.16
improved FCN	3.1	0.8	8.7	3.2	0.17	0.09	2.9	0.57	9.56	3.6	0.21	0.07
improved FCN+ level set	3.2	0.6	9.1	2.9	0.11	0.07	2.5	0.57	9.9	3.7	0.17	0.04
improved FCN+ improved snake	**2.03**	**0.41**	8.80	3.69	**0.098**	**0.0007**	**2.35**	**0.63**	9.09	3.42	**0.125**	**0.0008**

TABLE 4: The comparison with the state-of-the-art methods (Barbosa et al [28], Milletari et al. [30], van Stralen et al. [31], Smistad et al. [40], and Keraudren et al. [41]) on correlation between the segmentation results and the ground truth in test set.

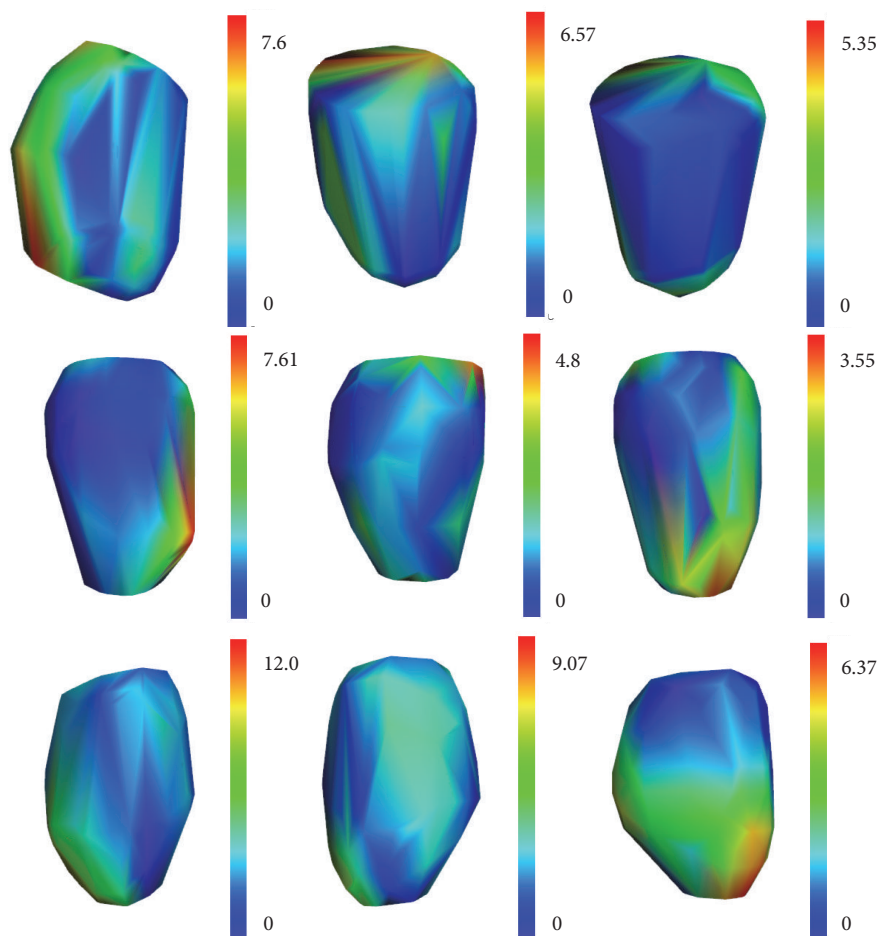| Method	EDV			ESV			EF		
	Corr* val	bias ml.	std ml.	Corr* val	bias ml.	std ml.	Corr* val	bias val	std val
[28]	0.035	-5	17.7	0.033	-6.8	13.9	**0.111**	**2.9**	5.2
[30]	0.047	5.1	19.0	0.040	-16.8	15.2	0.255	15.2	7.6
[31]	0.034	-15.4	16.0	0.036	-13.2	14.4	0.389	3.7	8.8
[40]	0.049	-10.1	19.4	0.036	-11.3	14.6	0.121	3.7	5.2
[41]	0.079	15.9	24.6	0.048	-6.2	16.6	0.281	12.1	10.6
The proposed method	**0.018**	**1.41**	**12**	**0.021**	**-3.59**	**11.3**	0.208	3.52	**0.07**

FIGURE 7: The segmentation results of three different subjects (which all are test subjects) of improved FCN (the first column), improved FCN with traditional snake (the second column), and our proposed method (the third column). The first row is ED model of subject 45, the second row is ED model of subject 40, and the third row is ES model of subject 30. The color map denotes the mean distance between result mesh and the corresponding reference mesh [3].

Additionally, due to the fact that left atrium and left ventricle have the similar intensity value and the presence of papillary muscles, the leakages and shrinks often happen if using the traditional deformable model without the spatial regularization. In this aspect, the proposed 3D snake model with the spatial regularization can improve the final segmentation performance through limiting the free degree of deformation of 3D snake. The results in Section 3 also prove that the proposed method outperforms the state-of-the-art methods for this task on most of the measure indexes.

Though the coarse segmentation results cannot be directly used for clinical application, this attempt shows the potential of end-to-end FCN for LV segmentation task on 3DE. The main limitation of coarse segmentation using FCN is the insufficient training data, which results in that the LV segmentation task on 3DE by end-to-end FCN would not be possible in any time soon. In the future, if more 3DE datasets are accessibly open, the purely end-to-end FCN with superior performance for LV segmentation tasks may be possible and will generate surprising results.

Additionally, 3D snake was used and considered as a postprocessing for the fine segmentation, like conditional random fields (CRF) which is widely used as postprocessing of the natural image segmentation [55–57]. Compared to the purely end-to-end deep learning method, the coarse-to-fine framework has high interpretability [58, 59]. In this paper, the interpretability means that the final segmentation step can be represented by obvious formula, not just the end-to-end deep learning technology. The segmentation performance and model's interpretability are both important for practical clinical application. In the future, we will also study how to formulate the coarse-to-fine frameworks into an end-to-end optimization model to further improve the segmentation performance, at the same time maintaining the interpretability of the coarse-to-fine framework.

5. Conclusion

This paper proposed a new fully automatic method for LV segmentation of 3DE based on a coarse-to-fine framework. The proposed FCN, based on transfer learning and
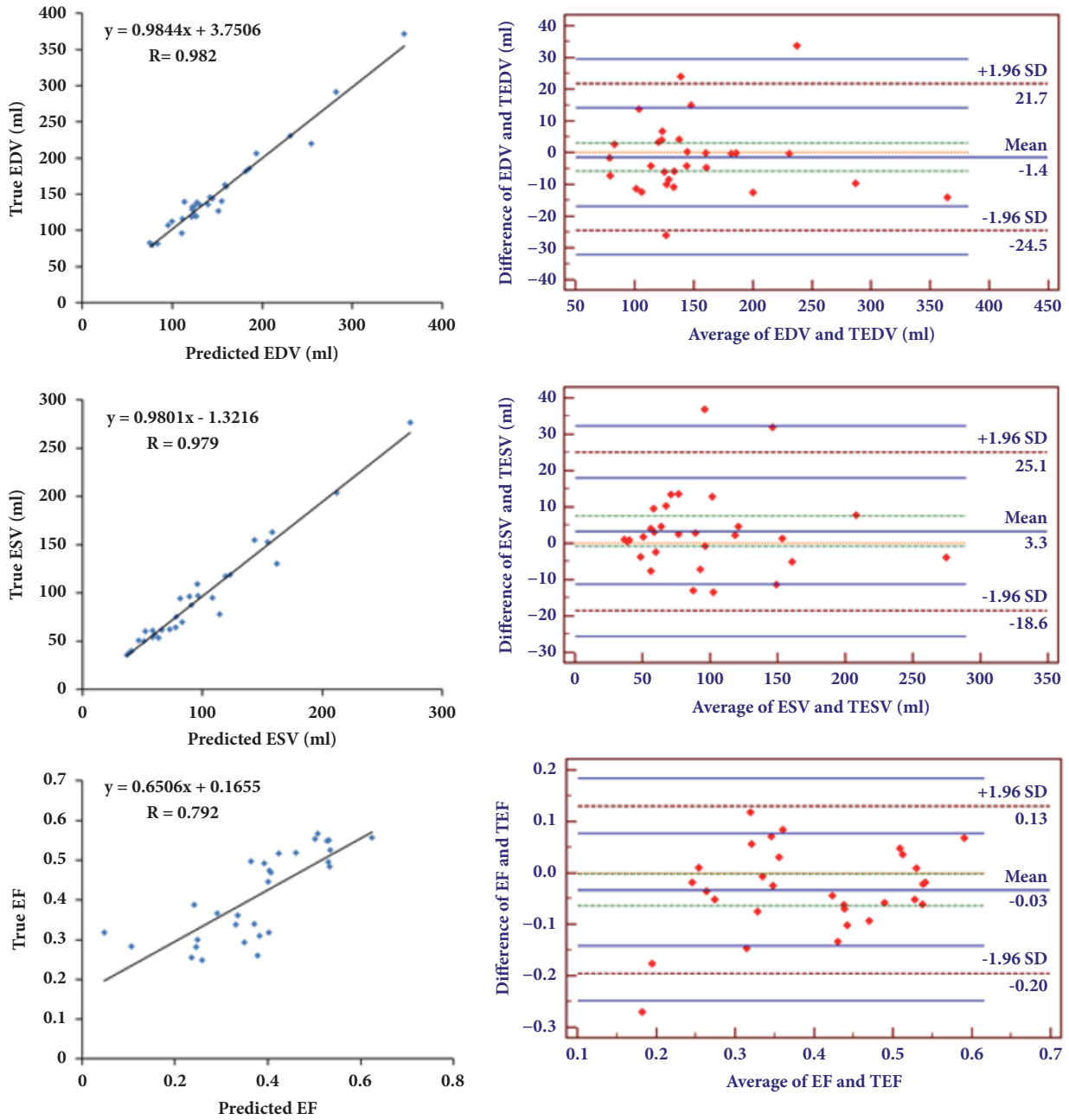
FIGURE 8: The correlation plots (the left column) and Bland-Altman analysis plots (the right column) for EDV, ESV, and EF on test subjects.

residual modules, was used for initial segmentation. From the initial segmentation, an initial model was automatically constructed. The initial model was then used as a constraint item in the energy function of the 3D snake for fine segmentation. The segmentation accuracy and clinical performance (it means the clinical measure results by the proposed method in a true clinical setting) have demonstrated that the proposed method is accurate and has outperformed the state-of-the-art methods on most of the clinical measure indexes. To our best knowledge, we were the first to use a FCN to address the 3DE segmentation problem. This attempt shows the potential of an end-to-end

FCN for LV segmentation of 3DE. Besides, the proposed fine segmentation method based on a deformable model not only further improves the segmentation performance, but also provides interpretability for the final segmentation results.

The clinical evaluation suggests that the proposed method has potential for clinical application and may lead to a wide use of FCN in ultrasound image segmentation tasks. In addition, the proposed framework is flexible and can be extended into other applications. In the future, more work about the only pure end-to-end deep learning technology or end-to-end learning framework which combines the deep learning

and postprocessing method (which has high interpretability) for segmentation tasks on 3DE will be studied.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] X. Huang, D. P. Dione, C. B. Compas et al., "Corrigendum to "Contour tracking in echocardiographic sequences via sparse representation and dictionary learning" [Med. Image Anal. 18 (2) (2014) 253–271]," *Medical Image Analysis*, vol. 22, no. 1, p. 21, 2015.

[2] X. Zhuang and J. Shen, "Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI," *Medical Image Analysis*, vol. 31, pp. 77–87, 2016.

[3] O. Bernard, J. G. Bosch, B. Heyde et al., "Standardized evaluation system for left ventricular segmentation algorithms in 3D echocardiography," *IEEE Transactions on Medical Imaging*, vol. 35, no. 4, pp. 967–977, 2016.

[4] C. Santiago, J. C. Nascimento, and J. S. Marques, "Automatic 3-D segmentation of endocardial border of the left ventricle from ultrasound images," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 1, pp. 339–348, 2015.

[5] S. Zhang and D. Metaxas, "Large-Scale medical image analytics: Recent methodologies, applications and Future directions," *Medical Image Analysis*, vol. 33, pp. 98–101, 2016.

[6] G. Carneiro, J. C. Nascimento, and A. Freitas, "The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 968–982, 2012.

[7] Y. Zhu, X. Papademetris, A. J. Sinusas, and J. S. Duncan, "A coupled deformable model for tracking myocardial borders from real-time echocardiography using an incompressibility constraint," *Medical Image Analysis*, vol. 14, no. 3, pp. 429–448, 2010.

[8] N. Lin, W. Yu, and J. S. Duncan, "Combinative multi-scale level set framework for echocardiographic image segmentation," *Medical Image Analysis*, vol. 7, no. 4, pp. 529–537, 2003.

[9] A. R. de Alexandria, P. C. Cortez, J. A. Bessa, J. H. da Silva Félix, J. S. de Abreu, and V. H. C. de Albuquerque, "PSnakes: A new radial active contour model and its application in the segmentation of the left ventricle from echocardiographic images," *Computer Methods and Programs in Biomedicine*, vol. 116, no. 3, pp. 260–273, 2014.

[10] T. Dietenbeck, D. Barbosa, M. Alessandrini et al., "Whole myocardium tracking in 2D-echocardiography in multiple orientations using a motion constrained level-set," *Medical Image Analysis*, vol. 18, no. 3, pp. 500–514, 2014.

[11] K. Saini, M. L. Dewal, and M. Rohit, "A fast region-based active contour model for boundary detection of echocardiographic images," *Journal of Digital Imaging*, vol. 25, no. 2, pp. 271–278, 2012.

[12] T. Dietenbeck, M. Alessandrini, D. Barbosa, J. D'hooge, D. Friboulet, and O. Bernard, "Detection of the whole myocardium in 2D-echocardiography for multiple orientations using a geometrically constrained level-set," *Medical Image Analysis*, vol. 16, no. 2, pp. 386–401, 2012.

[13] V. Zagrodsky, V. Walimbe, C. R. Castro-Pareja, J. X. Qin, J.-M. Song, and R. Shekhar, "Registration-assisted segmentation of real-time 3-D echocardiographic data using deformable models," *IEEE Transactions on Medical Imaging*, vol. 24, no. 9, pp. 1089–1099, 2005.

[14] L. Yu, Y. Guo, Y. Wang, J. Yu, and P. Chen, "Segmentation of fetal left ventricle in echocardiographic sequences based on dynamic convolutional neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 8, pp. 1886–1895, 2017.

[15] Y. Yu, S. Zhang, K. Li, D. Metaxas, and L. Axel, "Deformable models with sparsity constraints for cardiac motion analysis," *Medical Image Analysis*, vol. 18, no. 6, pp. 927–937, 2014.

[16] J. Liu, X. Zhuang, L. Wu et al., "Myocardium segmentation from DE MRI using multi-component gaussian mixture model and coupled level set," *IEEE Transactions on Biomedical Engineering*, 2017.

[17] 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," *Medical Image Analysis*, vol. 30, pp. 108–119, 2016.

[18] S. Zhang, Y. Zhan, and D. N. Metaxas, "Deformable segmentation via sparse representation and dictionary learning," *Medical Image Analysis*, vol. 16, no. 7, pp. 1385–1396, 2012.

[19] H. C. van Assen, M. G. Danilouchkine, A. F. Frangi et al., "SPASM: a 3D-ASM for segmentation of sparse and arbitrarily oriented cardiac MRI data," *Medical Image Analysis*, vol. 10, no. 2, pp. 286–303, 2006.

[20] G. Luo, D. Sui, K. Wang, and J. Chae, "Neuron anatomy structure reconstruction based on a sliding filter," *BMC Bioinformatics*, vol. 16, no. 1, 2015.

[21] S. C. Mitchell, J. G. Bosch, B. P. F. Lelieveldt, R. J. Van der Geest, J. H. C. Reiber, and M. Sonka, "3-D active appearance models: Segmentation of cardiac MR and ultrasound images," *IEEE Transactions on Medical Imaging*, vol. 21, no. 9, pp. 1167–1178, 2002.

[22] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D. N. Metaxas, and X. S. Zhou, "Towards robust and effective shape modeling: sparse shape composition," *Medical Image Analysis*, vol. 16, no. 1, pp. 265–277, 2012.

[23] J. O. F. Hansegård and I. S. Rabben, "Real-time active shape models for segmentation of 3D cardiac ultrasound," in *Proceeding of the International Conference on Computer Analysis of Images and Patterns*, 2007.

[24] J. Hansegård, S. Urheim, K. Lunde, and S. I. Rabben, "Constrained active appearance models for segmentation of triplane echocardiograms," *IEEE Transactions on Medical Imaging*, vol. 26, no. 10, pp. 1391–1400, 2007.

[25] M. Ma, M. van Stralen, J. H. C. Reiber, J. G. Bosch, and B. P. F. Lelieveldt, "Model driven quantification of left ventricular function from sparse single-beat 3D echocardiography," *Medical Image Analysis*, vol. 14, no. 4, pp. 582–593, 2010.

[26] D. Barbosa et al., "Fast and fully automatic 3d echocardiographic segmentation using B-spline explicit active surfaces," in *Proceedings of the 2012 9th Ieee International Symposium on Biomedical Imaging (Isbi)*, pp. 1088–1091.

[27] D. Barbosa, B. Heyde, M. Cikes et al., "Real-time 3D interactive segmentation of echocardiographic data through user-based deformation of B-spline explicit active surfaces," *Computerized Medical Imaging and Graphics*, vol. 38, no. 1, pp. 57–67, 2014.

[28] D. F. D. Barbosa, J. Dhooge, and O. Bernard, "Fast tracking of the left ventricle using global anatomical affine optical flow and local recursive block matching," in *Proceedings of the MICCAI Challenge on Echocardiographic Three-Dimensional Ultrasound Segmentation (CETUS)*, MIDAS Journal, Boston, MA, USA, 2014.

[29] G. Carneiro and J. C. Nascimento, "Combining multiple dynamic models and deep learning architectures for tracking the left ventricle endocardium in ultrasound data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2592–2607, 2013.

[30] M. Y. F. Milletari, N. Navab, and S. Ahmadi, "Left ventricle segmentation in cardiac ultrasound using hough-forests with implicit shape and appearance priors," in *Proceedings of the MICCAI Challenge on Echocardiographic Three-Dimensional Ultrasound Segmentation (CETUS)*, MIDAS Journal, Boston, MA ,USA, 2014.

[31] A. H. M. van Stralen, K. Leung, G. van Burken, and J. Bosch, "Segmentation of multi-center 3d left ventricular echocardiograms by active appearance models," in *Proceedings MICCAI Challenge on Echocardiographic Three-Dimensional Ultrasound Segmentation (CETUS)*, MIDAS Journal, Boston, MA, USA, 2014.

[32] G. Wang, S. Zhang, H. Xie, D. N. Metaxas, and L. Gu, "A homotopy-based sparse representation for fast and accurate shape prior modeling in liver surgical planning," *Medical Image Analysis*, vol. 19, no. 1, pp. 176–186, 2015.

[33] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015*, pp. 3431–3440, USA, June 2015.

[34] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '15)*, pp. 1–9, Boston, Mass, USA, June 2015.

[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '16)*, pp. 770–778, Las Vegas, Nev, USA, June 2016.

[36] G. Luo, S. Dong, K. Wang, W. Zuo, S. Cao, and H. Zhang, "Multi-views fusion CNN for left ventricular volumes estimation on cardiac MR images," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 1924–1934, 2018.

[37] G. Carneiro and J. C. Nascimento, "Multiple dynamic models for tracking the left ventricle of the heart from ultrasound data using particle filters and deep learning architectures," in *Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2010*, pp. 2815–2822, USA, June 2010.

[38] 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," *Medical Image Analysis*, vol. 35, pp. 159–171, 2017.

[39] S. Dong, G. Luo, G. Sun, K. Wang, and H. Zhang, "A left ventricular segmentation method on 3D echocardiography using deep learning and snake," in *Proceedings of the 43rd Computing in Cardiology Conference, CinC 2016*, pp. 473–476, Canada, September 2016.

[40] E. L. F. Smistad, "Real-time tracking of the left ventricle in 3d ultrasound using Kalman filter and mean value coordinates," in *Proceedings of the MICCAI Challenge on Echocardiographic Three-Dimensional Ultrasound Segmentation (CETUS)*, MIDAS Journal, Boston, MA, USA, 2014.

[41] O. O. K. Keraudren, W. Shi, J. Hajnal, and D. Rueckert, "Endocardial 3d ultrasound segmentation using autocontext random forests," in *Proceedings of the MICCAI Challenge on Echocardiographic Three-Dimensional Ultrasound Segmentation (CETUS)*, pp. 41–48, MIDAS Journal, Boston, MA, USA, 2014.

[42] H. Shin, H. R. Roth, M. Gao et al., "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[43] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14)*, pp. 580–587, Columbus, Ohio, USA, June 2014.

[44] R. Girshick, "Fast R-CNN," in *Proceedings of the 15th IEEE International Conference on Computer Vision (ICCV '15)*, pp. 1440–1448, December 2015.

[45] M. Everingham, L. van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (VOC) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.

[46] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention and Medical Image Computing and Computer-Assisted Intervention (MICCAI '15)*, vol. 9351 of *Lecture Notes in Computer Science*, pp. 234–241, November 2015.

[47] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: fully convolutional neural networks for volumetric medical image segmentation," in *Proceedings of the 4th International Conference on 3D Vision (3DV '16)*, pp. 565–571, IEEE, October 2016.

[48] G. M. A. Lin, C. Shen et al., "RefineNet: Multi-Path Refinement Networks with Identity Mappings for High-Resolution Semantic Segmentation," arXiv:1611.06612, 2016.

[49] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proceedings of the European Conference on Computer Vision*, pp. 630–645, 2016.

[50] A. Veit, M. Wilber, and S. Belongie, "Residual networks behave like ensembles of relatively shallow networks," in *Proceedings of the 30th Annual Conference on Neural Information Processing Systems, NIPS 2016*, pp. 550–558, Spain, December 2016.

[51] S. S. C. Ioffe, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," arXiv:1502.03167, 2015.

[52] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[53] L. Liu, C. Bajaj, J. O. Deasy, D. A. Low, and T. Ju, "Surface reconstruction from non-parallel curve networks," *Computer Graphics Forum*, vol. 27, no. 2, pp. 155–163, 2008.

[54] Y. Jia, E. Shelhamer, J. Donahue et al., "Caffe: convolutional architecture for fast feature embedding," in *Proceedings of the ACM International Conference on Multimedia*, pp. 675–678, ACM, Orlando, Fla, USA, November 2014.

[55] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2018.

[56] S. Zheng, S. Jayasumana, B. Romera-Paredes et al., "Conditional random fields as recurrent neural networks," in *Proceedings of the 15th IEEE International Conference on Computer Vision, ICCV 2015*, pp. 1529–1537, Chile, December 2015.

[57] G. Lin, C. Shen, A. Van Den Hengel, and I. Reid, "Efficient piecewise training of deep structured models for semantic segmentation," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016*, pp. 3194–3203, USA, July 2016.

[58] J. Wen, Y. Xu, Z. Li, Z. Ma, and Y. Xu, "Inter-class sparsity based discriminative least square regression," *Neural Networks*, vol. 102, pp. 36–47, 2018.

[59] J. Wen, B. Zhang, Y. Xu, J. Yang, and N. Han, "Adaptive weighted nonnegative low-rank representation," *Pattern Recognition*, vol. 81, pp. 326–340, 2018.