IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 30, 2019, accepted May 28, 2019, date of current version October 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2920957

Automatic Left Ventricle Recognition, Segmentation and Tracking in Cardiac Ultrasound Image Sequences

WEI-YEN HSU
Department of Information Management, National Chung Cheng University, Minhsiung 621, Taiwan
Advanced Institute of Manufacturing with High-tech Innovations, National Chung Cheng University, Minhsiung 621, Taiwan
Center for Innovative Research on Aging Society (CIRAS), National Chung Cheng University, Minhsiung 621, Taiwan

e-mail: shenswy@gmail.com

This work was supported by the Ministry of Science and Technology, Taiwan, under Grant MOST105-2410-H-194-059-MY3 and the Center for Innovative Research on Aging Society (CIRAS) from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by Ministry of Education (MOE) in Taiwan, under Grant RCN0014.

ABSTRACT In this study, we propose a novel method incorporating faster region-based convolutional neural network and active shape model to automatically recognize, segment, and track the left ventricle in cardiac ultrasound image sequences, respectively. Ultrasound images typically contain noise and artifacts. The conventional filters cannot preserve the edges of image contours, and thus blurry images are often obtained. In this study, we propose an improved adaptive anisotropic diffusion filter to effectively reduce noise and reinforce image contours. In addition, because of the shape and appearance of the left ventricle vary considerably between adjacent images, conventional methods cannot automatically identify the position of the left ventricle or accurately segment them. A novel method that combines the faster region-based convolutional neural network with the active shape model is proposed to automatically recognize, segment, and track the left ventricle in cardiac ultrasound image sequences. Compared with four state-of-the-art approaches, the method proposed in this study can be applied to accurately segment and track the left ventricle in cardiac ultrasound image sequences. The proposed method produces the most satisfactory results in terms of visual presentation and segmentation quality based on four criteria.

INDEX TERMS Cardiac ultrasound image, left ventricle recognition, image segmentation and tracking, faster region-based convolutional neural network (Faster R-CNN).

I. INTRODUCTION

The heart is the power source of the human circulatory system and a vital organ in the human body. Heart dis-ease influences human health and lives, and coronary heart disease (CHD) is one of the leading causes of death worldwide. Therefore, understanding the operating mechanism and characteristics of the heart could facilitate the prevention and treatment of heart disease. In recent years, medical imaging equipment has undergone rapid development, and cardiac imaging technology has evolved from electronic beam computed tomography (CT) and one-dimensional echocardiography into multi-slice cardiac CT, cardiac magnetic resonance imaging (MRI), and three-dimensional ultrasound imaging [1]–[3]. These novel imaging methods can be used to present complex images of the heart's anatomy. Cardiac ultrasound images are extremely useful for assessing cardiac physiological indicators and diagnosing heart failure, CHD, and cardiomyopathy. In recent years, clinical research has focused on assessing cardiac beats, cardiac resynchronization therapy, and myocardial exercise. Cardiac ultrasound technology is easy to operate, reliable, and practical, and is a powerful noninvasive technique for comprehensive cardiac function assessment, research on ischemic heart disease, and cardiac resynchronization therapy. Therefore, cardiac ultrasound technology could have a wide range of clinical applications in the future.

Studies on heart functions have typically focused on analyzing the left ventricle. The left ventricle is responsible for systemic blood supply. We can obtain all of left ventricular end-diastolic volume, left ventricular end-systolic volume, left ventricular ejection fractions, and stoke volume through understanding changes in the left ventricle. It can facilitate the development of a quantitative method for preventing and

The associate editor coordinating the review of this manuscript and approving it for publication was Cristian A Linte.

treating cardiovascular diseases and reduce cardiovascular disease-related risk and mortality.

Two clinical image segmentation methods are the semi-automatic and automatic methods. A semi-automatic segmentation method is a kind of segmentation that needs the user to outline the rough region of interest with the mouse clicks and the algorithm is then applied so that the path that best fits the edge of the image is shown. The following two difficulties are involved in the semi-automatic image segmentation method: (1) because of a large amount of data, marking the position or contours of the left ventricle is tedious and time-consuming; and (2) observational variance occurs among observers [4]. It makes the automatic segmentation of left ventricular image sequence substantial in the accuracy of positions and contours. Therefore, the automatic image segmentation method is more appealing than the semi-automatic method [5], [6]. The automatic method comprises two steps: automatic determination of the position of the left ventricle and segmentation and tracking of the left ventricle [7]. Current position detection methods (e.g., scale-invariant feature transform (SIFT) [8] and the histogram of oriented gradients (HOG) [9]) can produce noise and artifacts in cardiac ultrasound images, and because left ventricular shape and appearance vary considerably, these methods cannot accurately identify the position of the left ventricle. Since 2012, deep learning has received substantial attention and is useful for examining images. Compared with spatial pyramid pooling in deep convolutional networks (SPP-net) [10], the faster region-based convolutional neural network (faster R-CNN) [11] can overcome temporal limitations because of regional calculation and an ideal detection rate [12].

Regarding image segmentation and tracking, the active shape model (ASM) requires a priori shape knowledge [13]. An a priori model is based on a series of points in an image, and principal component analysis (PCA) is typically used to describe the shape of the target to be segmented as a priori knowledge and express changes in the shape of the deformed model [14]. Compared with other image segmentation methods, the ASM can effectively segment a blurred border. Through training in use of the ASM, the burden of medical staff can be reduced. In addition, with a point distribution model (PDM) [15], [16], the geometric structure of the target is considered and parametric adjustment is limited according to training data. Accordingly, shape changes can be limited to a reasonable range. The ASM can overcome the high sensitivity of the active contour model to the initial contour position. In addition, the active contour model has the following weaknesses: its segmentation effect and speed de-pend on the initial contour position, and it can identify only the edge of a target without considering the geometric structure of the target. The ASM can overcome these weaknesses. The ASM has been applied extensively in facial recognition systems but seldom in cardiac ultrasound images.

In this study, we integrated the faster R-CNN and ASM to automatically segment and track the left ventricle of cardiac ultrasound image sequences. We used the faster R-CNN to extract and recognize the characteristics of left ventricular images. Subsequently, we used the proposed improved ASM to automatically segment and track left ventricular images.

II. MATERIALS AND METHODS

Figure 1 presents the procedures of this study. Noise in images (including training and testing data) was effectively reduced and image contours were reinforced through image enhancement and smoothing by using the pro-posed improved adaptive anisotropic diffusion filter. Subsequently, the faster R-CNN was used to automatically determine the position of the left ventricle and the proposed improved ASM was employed to correctly segment and track the left ventricle. Following training, we implemented these procedures to test the image sequence and achieve the goal of accurate automatic segmentation and tracking. Finally, we used the Sørensen–Dice coefficient [17], mean absolute deviation (MAD) [18], and Hausdorff distance (HD) [19] to assess the results.

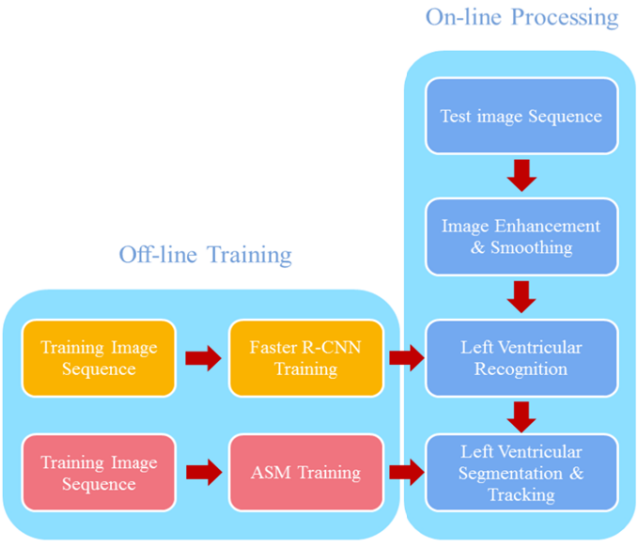


FIGURE 1. Flowchart of the proposed method.

In on-line processing, the steps of the proposed method that is performed in cardiac ultrasound test image sequences (N image frames) are described in detail as follows:

1) Noise of test image frame 1 is reduced by using the proposed improved adaptive anisotropic diffusion filter.
2) Initial position of the left ventricle in image frame 1 is determined automatically by means of the trained parameters of faster R-CNN.
3) Final contour of the left ventricle in image frame 1 is obtained by iteratively using the trained parameters of the proposed improved ASM.
4) For image frame $i = 2$ to N,
5) Noise of test image frame i is reduced by the improved adaptive anisotropic diffusion filter.

6) Initial position of the left ventricle in image frame i is determined automatically by means of the trained parameters of faster R-CNN.
7) Initial contour of the left ventricle in image frame i is obtained from final contour of that in image frame i-1 and initial position of that in image frame$i$.
8) Final contour of the left ventricle in image frame i is obtained by iteratively using the trained parameters of the proposed improved ASM.
9) End

A. MATERIALS

The materials used in this study were a series of cardiac ultrasound images provided by physicians in the cardiology department of Mackay Memorial Hospital in Taipei City. The format was digital imaging and communications in medicine. In the experiment, 30 series of ultrasound images were employed, each comprising 1500–5000 images. Each image was 636 × 434 pixels and contained seven standard perspectives. This study focused on the left ventricle; therefore, we processed only the apical 4-chamber view of each left ventricular image (Fig. 2).

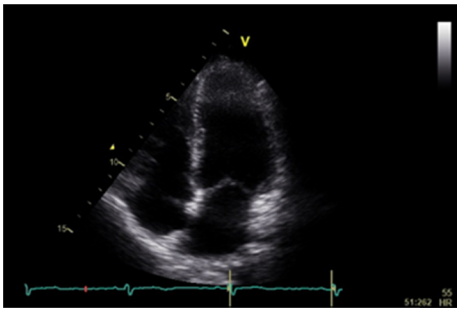
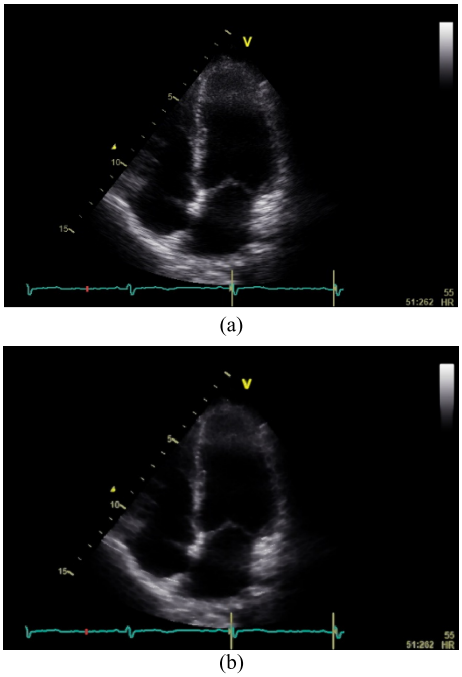


FIGURE 2. Apical 4-chamber view of a cardiac ultrasound image.

Of the 30 series of images, 10 served as the training set, 5 served as the verification set, and 15 served as the testing set. From the training and verification sets, 300 images including ventricular images that could be easily or not easily recognized were randomly selected for training. From the testing set, 300 images with a constant interval (including multiple cardiac cycles) were selected for image segmentation and used for cardiac tracking. This method was compared with several state-of-the art approaches.

B. IMAGE ENHANCEMENT AND SMOOTHING USING IMPROVED ADAPTIVE ANISOTROPIC DIFFUSION FILTER

The original ultrasound images contained noise and artifacts. Following convention, a filter based on a fuzzy method (e.g., Gaussian filter, median filter, mean filter) was employed to eliminate all noise. However, although filters can be used to eliminate noise and artifacts, they cannot preserve image contours. In this study, we proposed an improved adaptive anisotropic diffusion filter based on an originally anisotropic diffusion method [20], [21] to effectively reduce noise and simultaneously enhance image contours. In the improvement, adjacent areas are enhanced and considered from original four-neighborhood expanding to eight-neighborhood, and different weightings, including α and β parameters, are then applied to four-neighborhood and diagonal-neighborhood respectively to effectively distinguish the border from the background. Finally, smoothing (α weighting parameter) and sharpening (β weighting parameter) mechanisms are incorporated to form a novel diffusion model, as shown in Fig. 3. When the improved adaptive anisotropic diffusion filter was used to process an image, the image contours became clear. In addition, noise in the image was effectively reduced.



(a)

(b)

FIGURE 3. Results of image enhancement and smoothing: (a) original image; (b) enhanced and smoothed image with improved adaptive anisotropic diffusion filter.

C. LEFT VENTRICLE RECOGNITION BASED ON THE FASTER R-CNN

Left ventricle shape and appearance varied considerably in this study, and the conventional method cannot be used to automatically identify the position of the left ventricle. Therefore, the faster R-CNN was used to automatically and accurately determine the position. The faster R-CNN was composed of a region proposal network (RPN) and fast R-CNN testing network. Fig. 4 presents the steps in the training stage. The RPN network and faster R-CNN testing network were jointly trained, as shown in Fig. 5.

1) PRETRAINING THE CNN MODEL

The RPN and faster R-CNN are initialized with a network pretrained on Image-net [22]. Typically, the ZFnet network [23] and visual-geometry-group 16 network [24]

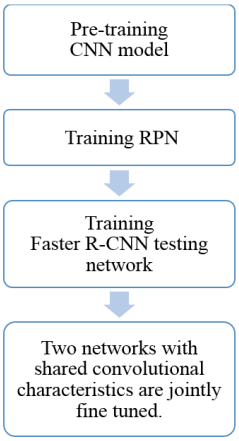


FIGURE 4. Flowchart of faster R-CNN training for specific object recognition.

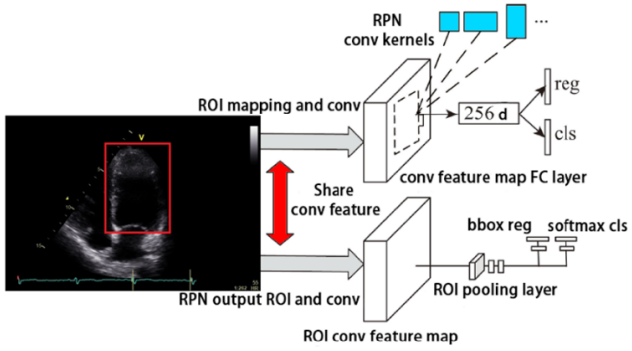


FIGURE 5. Joint training of RPN and faster R-CNN.

are adopted; however, because of concern of the computations in this study, only the ZFnet network was employed. Fig. 6 presents the framework of the ZFnet network including five convolutional layers, where two-headed arrow represents the number of channels, i.e. the number of convoluting with different filters, and contrast norm. means contrast normalized. The convolutional layers were connected to the max pool to reduce the number of parameters and calculation complexity. The feature layer and softmax layer were connected with each other and positioned behind the convolutional layers. The fifth convolutional layer in the ZFnet network contained 256 channels and was called the feature map, which was the deep-layer convolutional feature of an input image. Deep-layer features were similar in the same types of objects but varied considerably between different types of objects. Therefore, objects on the feature map were dividable. Specific layers were added to the output of the ZFnet network to obtain the RPN and faster R-CNN. These specific layers can be used to extract an area containing a target from an input image and to calculate the related probability.

2) TRAINING RPN

In this study, when the left ventricle image was used to train the RPN, we adopted the pretrained ZFnet model to initialize

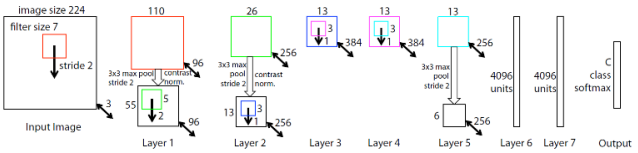


FIGURE 6. Eight layers of the ZFnet network.

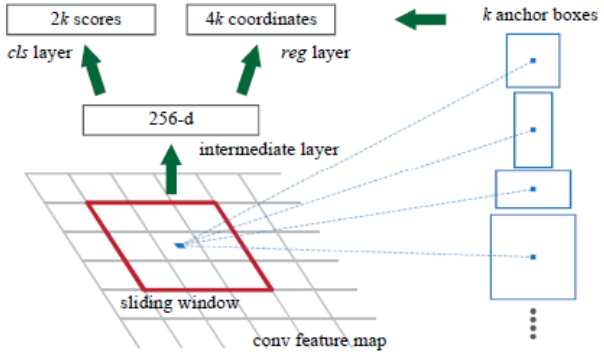


FIGURE 7. RPN framework.

the RPN. Subsequently, a multi-task loss function combining the losses of classification and bounding box regression and a back-propagation algorithm were used to fine tune the RPN, for which an arbitrary image size served as an input and a series of outputs served as regional proposals that likely contained a target. As shown in Fig. 7, behind the fifth-layer feature map of the ZFnet network (CONV5), a small convolutional layer was added. For each position on the feature map, the small convolutional layer was used for convolutional computation. Regarding the anchors in each position, each of three sizes (128^2, 256^2, and 512^2) was combined with each of three ratios (1:1, 1:2, and 2:1) to form nine kinds of anchors, which were employed to predict which position contained a target and provide accurate regional proposals [12]. Following convolution, a 256-dimensional vector for a single position was obtained. This vector reflected the deep-layer characteristics of the anchors in that position and was used to predict the probabilities of which anchors in that position were targets or background.

3) TRAINING THE FASTER R-CNN TESTING NETWORK

The method for training the faster R-CNN testing network was based on regional proposals, and the pretrained ZFnet model was used to initialize the testing network. To extract the features of the five-layer convolutional network from an input image, the deep-layer features in CONV5 were extracted. All features in all 256 channels were concatenated into a 4096-dimensional feature vector called the FC6 feature layer. Another 4096-dimensional feature layer was added to form the FC7 feature layer, which was subsequently connected to the FC6 feature layer. The FC7 feature layer was used to predict the probability of a candidate region belonging to a certain category and the position of the border of a target.

Previously marked information was employed to fine tune the testing network through a back-propagation algorithm.

4) TWO NETWORKS THAT SHARED CONVOLUTIONAL FEATURES AND WERE FINE TUNED

The testing network obtained in the previous stage was employed to initialize the RPN, which shared the same convolutional layer indicated by the red double arrows in Fig. 5. The other part of the RPN was fine tuned. To correspond with the testing network, this part was called the FC layer of the RPN network. The two networks shared a deep convolutional layer. Finally, the shared convolutional layer was fixed and the FC layer of the faster R-CNN was fine tuned to form a joint network. Fig. 8 illustrates these steps.

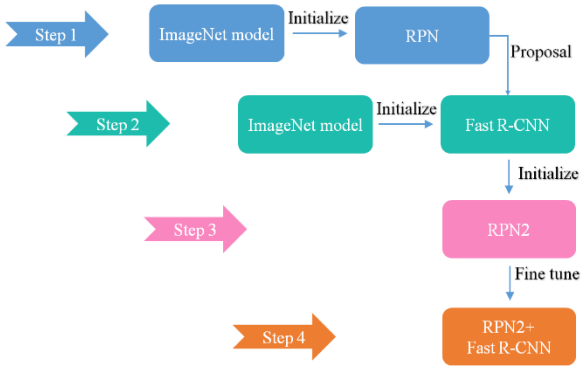


FIGURE 8. Cross-training of the joint network.

5) TRAINING THE FASTER R-CNN TESTING NETWORK

After training, the two networks used the identical five-layer convolutional neural network. Therefore, the testing and recognition processes were completed by performing a series of convolutional computations to reduce the amount of time required by the regional proposal steps. Figure 9 presents the testing and recognition processes. The steps are described as follows.

1) A convolutional computation series was performed on an entire image to obtain the feature map (i.e., CONV5).
2) The RPN was employed to produce numerous candidate regional proposals on the feature map.
3) Nonmaximum suppression (NMS) was performed on the candidate regional proposals and the first 300 proposals that obtained the highest scores were preserved.
4) The features in the candidate regions on the feature map were extracted to form a high-dimensional feature vector. The testing network was employed to calculate the score of each category and predict the border position of the optimal target.

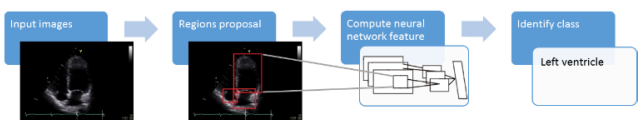


FIGURE 9. Testing and recognition processes.

D. LEFT VENTRICLE SEGMENTATION AND TRACKING BASED ON ACTIVE SHAPE MODEL

After the left ventricle position had been automatically identified, we used the ASM to correctly segment and track the left ventricle. The ASM requires a priori shape knowledge [13]. Compared with other image segmentation methods, the ASM can effectively segment a blurred border. With pretraining, the burden of medical staff can be reduced. In addition, the PDM [15], [16] considers the geometric structure of a target and limits parametric adjustment based on training data. Accordingly, shape change can be limited to a reasonable range. The ASM overcomes the high sensitivity of active contour model to the initial position of a contour and other disadvantages of active contour model [25], namely that its segmentation effect and speed depend on the initial contour position and it identifies only the edge of a target without considering the target's geometric structure.

The ASM method has been applied extensively in facial recognition systems and also in cardiac ultrasound images. In this study, we proposed an improved ASM method to segment the left ventricle and verified the segmentation effect of this method.

1) ESTABLISHING A SHAPE MODEL

A shape model was employed to train a series of images as well as an average shape, transformation matrix, and model constraints. Three steps were required: drawing feature points, alignment, and PCA.

First, for each of cardiac ultrasound images in the training set, feature points were drawn to obtain the shape of the left ventricle image (Fig. 10). Each feature point in each image was coded with a unique sequence number. The feature points on each image in the training set formed a shape represented by a 2n × 1 vector, expressed as follows:

$$x = (x_1, \ldots, x_n, y_1, \ldots, y_n)^T \tag{1}$$

where $x_1, \ldots, x_n$ denote the x coordinates of a series of feature points, $y_1, \ldots, y_n$ denote the y coordinates of the series, and n denotes the number of sampled points on the contour.

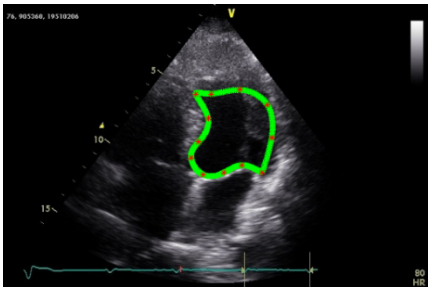


FIGURE 10. Feature points of the left ventricle.

To determine the statistical characteristics of the coordinates of model points in the training set and to render the corresponding points among various samples comparable, the influences of sample size and position must be eliminated by the alignment. In other words, for each point, the distribution vector has rotational invariance. A shape can be aligned with another shape, i.e. the vectors of two shapes (x_i and x_j), through rotation, zooming in or out, or displacement to minimize the difference between two shapes [26], [27]. Finally, the minimum distance between the vectors of the two shapes can be obtained. Following alignment, the feature points become concentrated.

Following shape alignment, PCA can be performed on the sample set [14] to obtain the main direction of the sample set in the space. After a shape model has been established, a shape can be expressed as follows:

$$x \approx \bar{x} + Pb \tag{2}$$

where P denotes the t-dimensional feature vector matrix obtained through PCA and b denotes the parameter for shape change representing the projection of an arbitrary heart in the feature space. To constrain the shape of left ventricle, the range of b is calculated as follows:

$$\left|b_j\right| \leq 3\sqrt{\lambda_j} \tag{3}$$

where λ_i denotes the variance of shape of left ventricle.

2) ESTABLISHING A PROFILE MODEL

To obtain the profile of each feature point, the grayscale distribution around a specific point on each image in the training set was examined. In addition, along the normal direction of a feature point, pixel-level grayscale values were sampled. Second, a profile model was established, where g_{ij} denotes the jth feature point of the ith sample and g_{ijk} denotes the kth grayscale value sampled. Therefore, the $(2k+1)$ grayscale values for g_{ij} can be expressed as follows:

$$g_{ij} = \left(g_{ij1}, g_{ij2}, \cdots g_{ij(2k)}, g_{ij(2k+1)}\right)^T \tag{4}$$

3) SEARCHING THE FEATURE POINTS

After the shape and profile models had been established, the parameters obtained from the models were used to search targets positioned among the images. Through the conventional ASM method, the minimum Mahalanobis distance was used to obtain the optimal matching points. However, this method is unstable and can only be used to search a target within a small range [16]. To overcome these problems, we proposed an improved ASM method, which is inspired and then improved from [16]. The matching function for the similarity between two feature points is proposed and expressed as follows:

$$f\,(g) = \sum_{i=0}^{k-1} g_i + \sum_{i=k}^{2k} (1 - g_i) \tag{5}$$

The optimal matching points were obtained by iteratively calculating and adjusting shape parameters. However, we obtained only optimal matching points as opposed to optimal matching shapes. In other words, we cannot guarantee that the ventricular shape obtained was the final ventricular shape. Therefore, the model must be constrained. We projected the searched shape onto a model space and constrained the range of the b value in the model to ensure that each feature point was the optimal feature point locally and overall.

The initial position of the searched target substantially influenced the ASM. If the searched position deviates from the searched target, the search processes are limited to a minimal local range, and this leads to search failure. Currently, for most ASM search algorithms, the initial search position is determined manually rather then automatically. In this study, we proposed that the faster R-CNN can be used to automatically recognize the position of the left ventricle. Accordingly, all images can be accurately and automatically segmented. The whole ASM based segmentation after obtaining the ROI from the Faster RCNN model is shown in Fig. 11.

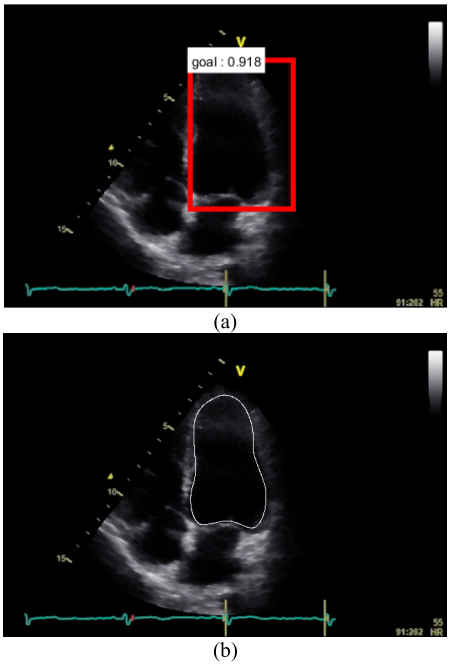


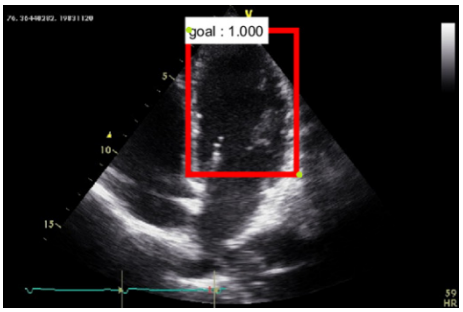
FIGURE 11. Whole ASM based segmentation after obtaining the ROI from the Faster RCNN model.

III. RESULTS AND DISCUSSION

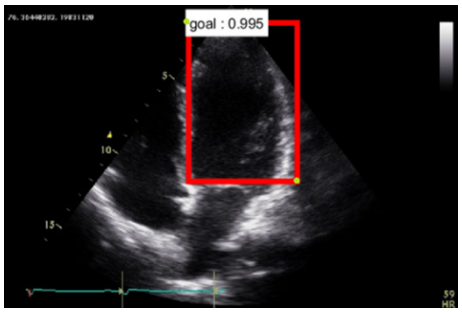
The results of left ventricular recognition and segmentation will be demonstrated in the following. In addition, we also show the performance analysis of proposed method with four assessment indicators and the computation time that it needs. Finally, we compare the proposed method with four state-of-the-art approaches to verify our performance.

A. RESULTS OF LEFT VENTRICULAR RECOGNITION

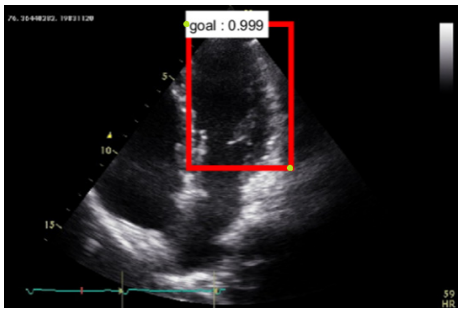
First, we presented the results of left ventricular recognition (Fig. 12), where the red bounding box denotes the identified ventricle, and the number shown by goal represents the



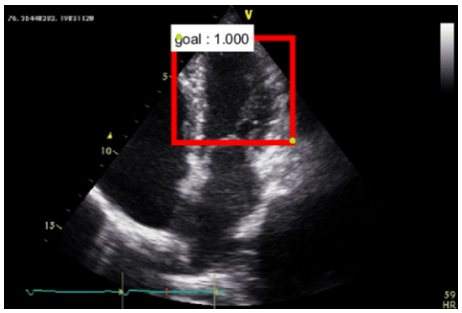
(a)



(b)



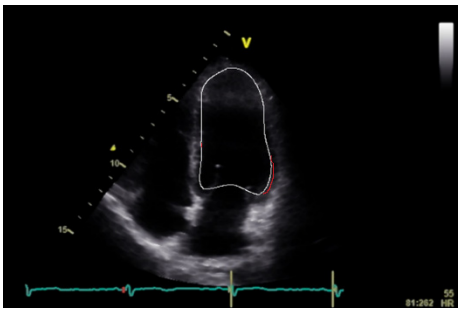
(c)

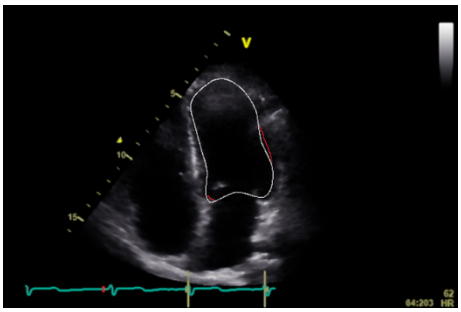


(d)

FIGURE 12. Left ventricular recognition under various circumstances.


(a)


(b)


(c)

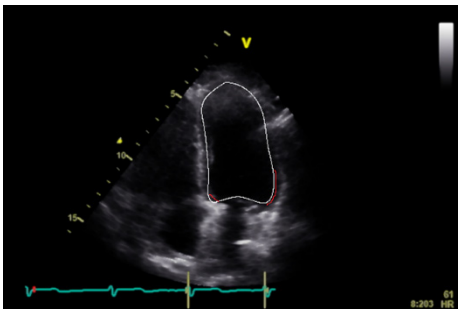

(d)

FIGURE 13. Left ventricular segmentation under various circumstances.

probability of the ventricle identified. As shown in Fig. 12(a), when the heat valve is open, the left ventricle and left atrium are connected to each other, thereby reducing the difference between grayscale values. Accordingly, if the conventional method is used, inaccurate recognition occurs. If our method is used, the left ventricle and its initial position can be accurately identified and segmented. Figure 12(b) shows an image of a closed heart valve, Fig. 12(c) shows an image of a broken left ventricle, and Fig. 12(d) shows an image where the left ventricle changes substantially. According to the results, regardless of how the ventricle changes, the proposed method can be employed to accurately recognize the left ventricle and identify its initial position.

B. RESULTS OF LEFT VENTRICULAR SEGMENTATION

Figure 13 presents the results of left ventricular segmentation. According to the results, for various patients and various cardiac phases, the proposed method can be used to accurately segment the contours of the left ventricle.

C. PERFORMANCE ANALYSIS

In this study, we used four assessment indicators to assess the effectiveness of the proposed method: Sørensen–Dice coefficient ("Dice" hereafter) [17], MAD [18], HD [19], Jaccard coefficient [32]. Dice is a function of similarity between two sets and is used to calculate the similarity between two sample sets, expressed as follows:

$$Dice\,(A, B) = \frac{2\,|A \cap B|}{|A| + |B|} \tag{6}$$

where A denotes the contour segmented by the researchers and B denotes the contour provided by a doctor. The value of Dice is between 0 and 1 (1 = the two contours completely overlap; 0 = the two contours do not overlap at all). A high Dice value denotes high similarity between the two sets. The MAD represents the mean of the average deviation value for Set A and the average deviation value for Set B, expressed as follows:

$$MAD = \frac{1}{2}\left\{\frac{1}{N_A}\sum\nolimits_{a\in A} d\,(a, B) + \frac{1}{N_B}\sum\nolimits_{b\in B} d\,(b, A)\right\} \tag{7}$$

A lower MAD value indicates higher similarity between the two contours. The HD is a point-feature-matching method that does not establish one-to-one correspondence between points and is used only to calculate the similarity (maximum distance) between two sets. Therefore, this method can be used to effectively process feature points. The HD between A and B is expressed as follows:

$$HD = \max\left\{\max_{a\in A} d\,(a, B)\,, \max_{b\in B} d\,(b, A)\right\} \tag{8}$$

A lower HD value indicates higher similarity between the two sets [19]. The calculations of the MAD and HD are not influenced by the number of values in the two sets. Therefore, the MAD and HD are efficient and accurate assessment indicators. Table 1 presents the segmentation effectiveness of the aforementioned three indicators (Dice, MAD, and HD) used to segment the cardiac ultrasound images of the 15 patients in the testing set. The results showed that all three indicators were effective and no substantial differences were observed among the patients. Therefore, the proposed method was highly accurate and stable.

D. COMPUTATION TIME

In this experiment, the device is with the CPU of AMD Phenom II X6 1055T Processor 2.8GHz, 8G RAM, and VGA card of NVIDIA GeForce GTX 960 (CUDA v6.5). Programming languages used are Matlab (version R2016b) and Microsoft Visual Studio (version 2013). The amount of execute time required for the proposed method is described as follows: The time to perform anisotropic diffusion for image enhancement and smoothing is 0.01 sec. When proposals are set to 300, the speed for recognizing a cardiac ultrasound image of the left ventricle is approximately 0.02 sec. on average. The speed for segmenting the left ventricle is 2–3 sec.. Overall, the amount of time required to process an ultrasound image is 2–3 sec. on average. Notably, left ventricular segmentation requires multiple iterations, and thus requires more time.

TABLE 1. Segmentation effectiveness.

Patient	Dice (%)	MAD (mm)	HD (mm)
1	92.66±2.00	1.96±0.51	6.06±2.01
2	91.14±1.82	1.97±0.43	6.46±1.96
3	92.43±1.94	1.96±0.48	6.14±2.03
4	91.30±1.86	1.97±0.45	6.39±1.91
5	92.71±1.89	1.98±0.50	6.21±2.03
6	91.41±1.66	1.95±0.42	6.33±1.89
7	92.24±2.03	1.96±0.47	6.23±1.94
8	91.84±2.08	1.97±0.39	6.36±1.98
9	91.94±2.07	1.98±0.49	6.26±1.90
10	91.77±1.87	1.96±0.40	6.37±1.84
11	92.71±2.05	1.95±0.49	6.16±2.04
12	91.88±1.87	1.96±0.38	6.40±2.11
13	92.04±2.08	1.97±0.55	6.18±1.94
14	91.77±1.90	1.97±0.42	6.36±2.04
15	92.07±2.02	1.96±0.49	6.29±1.99
mean	91.88±1.99	1.97±0.48	6.38±1.99

E. COMPARISONS WITH FOUR STATE-OF-THE-ART APPROACHES

Table 2 shows the results of a comparison between the proposed method, including single-fold (listed in Table 1) and three-fold cross-validation, and four state-of-the-art approaches. In the approach proposed by Bosch *et al.* [28], the processing speed and HD were excellent but a large training set was required to train a model. In addition, temporal phases were employed to judge the position of the left ventricle; therefore, this method relied on a priori knowledge and was unstable. The approach proposed by Qin *et al.* [29] was excellent. The level-set segmentation method was unsupervised and did not require a training set. However, sparse matrix transformation was required to identify the right ventricle. Accordingly, a large training set and a large amount of processing time were required. Because a level set was used, the parameters needed to be readjusted for a dramatic cardiac change, which would lead to the results being unstable. In the approach proposed by Carneiro and Nascimento [30], the minimum training set containing 20 images was used to establish a model; however, this approach did not yield accurate segmentation results. A model based on a training set containing 496 images was eventually established. The approach proposed by Hansson *et al.* [31] was an unsupervised segmentation method, and thus training was not required, which resulted in considerable segmentation errors.

TABLE 2. Comparison results of proposed method and four state-of-the-art approaches.

Methods	Training Set Size (frame)	Computation Speed (fps)	Dice (%)	MAD (mm)	HD (mm)	Jaccard Coefficient
Bosch et al. [28]	1040	**2.67**	-	-	**3.3**	-
Qin et al. [29]	450	0.01	90.8 ±1.7	2.0 ±0.42	6.86 ±1.71	-
Carneiro & Nascimento [30]	496	0.2	-	**1.94 ±0.51**	-	0.83 (0.82, 0.84)
Hansson et al. [31]	0	0.3	-	2.58 ±0.85	-	-
Proposed Method	300+300	**0.3~0.5**	91.88 ±1.99	1.97 ±0.48	6.38 ±1.99	0.83 ±0.010
Proposed Method (Three-fold Cross-validation)	300+300	**0.3~0.5**	**92.13 ±1.87**	1.95 ±0.50	**6.29 ±2.01**	**0.86 ±0.007**

Overall, the method proposed in the present study can be used to rapidly and accurately process cardiac ultrasound images. Because of concern of the computations, the pretreatment processes were simple and the temporal costs were low. In addition, the proposed method was stable without a priori knowledge. Even if an image series contained imperfections, the proposed method could be used to accurately recognize and segment an image.

IV. CONCLUSION

In this study, we proposed a method that integrated the faster region-based convolutional neural network and active shape model to automatically recognize, segment and track left ventricle in cardiac ultrasound image sequences, respectively. Proposed improved adaptive anisotropic diffusion filters can effectively reduce noise and reinforce image contours. The proposed method can accurately and automatically segment and track the left ventricle in cardiac ultrasound image sequences. Compared with four state-of-the-art approaches, the application of the proposed method to segment and track a cardiac ultrasound image of the left ventricle produced more accurate results. It can effectively aid and facilitate the doctors to diagnose hearts beat irregularly. By using the proposed method to quantify the cardiac ultrasound images, we can also help the doctors to observe signs of heart disease in patients and perform professional medical evaluation.

In future work, we intend to apply the proposed method to analyze various medical (e.g., MRI images of the endocardium and epicardium in the left ventricle, images of lung and kidney tumors). We hope that this research can facilitate disease diagnoses based on medical images.

ACKNOWLEDGMENT

The author would like to thanks his student, Nai-En Chang, for help with parts of the materials and methods.

REFERENCES

[1] Y. Zhou, S. Giffard-Roisin, M. De Craene, S. Camarasu-Pop, J. D'Hooge, M. Alessandrini, D. Friboulet, M. Sermesant, and O. Bernard, "A framework for the generation of realistic synthetic cardiac ultrasound and magnetic resonance imaging sequences from the same virtual patients," *IEEE Trans. Med. Imag.*, vol. 37, no. 3, pp. 741–754, Mar. 2018.

[2] A. Elen, H. Choi, D. Loeckx, H. Gao, P. Claus, P. Suetens, F. Maes, and J. D'hooge, "Three-dimensional cardiac strain estimation using spatio–temporal elastic registration of ultrasound images: A feasibility study," *IEEE Trans. Med. Imag.*, vol. 27, no. 11, pp. 1580–1591, Nov. 2008.

[3] A. P. King, K. S. Rhode, Y. Ma, C. Yao, C. Jansen, R. Razavi, and G. P. Penney, "Registering preprocedure volumetric images with intraprocedure 3-D ultrasound using an ultrasound imaging model," *IEEE Trans. Med. Imag.*, vol. 29, no. 3, pp. 924–937, Mar. 2010.

[4] F. Sardanelli, M. Quarenghi, G. Di Lee, L. Boccaccini, and A. Schiavi, "Segmentation of cardiac cine MR images of left and right ventricles: Interactive semiautomated methods and manual contouring by two readers with different education and experience," *J. Magn. Reson. Imag.*, vol. 27, no. 4, pp. 785–792, Feb. 2008.

[5] V. Tirronen, F. Neri, T. Kärkkäinen, K. Majava, and T. Rossi, "An enhanced memetic differential evolution in filter design for defect detection in paper production," *Evol. Comput.*, vol. 16, no. 4, pp. 529–555, Dec. 2008.

[6] A. Caponio, F. Neri, and V. Tirronen, "Super-fit control adaptation in memetic differential evolution frameworks," *Soft Comput.*, vol. 13, nos. 8–9, p. 811, Jul. 2009.

[7] I. Silva, J. F. Z. Sanches, and A. G. Almeida, "Toward a fully automatic left ventricle segmentation using cine-MR images," Ph.D. dissertation, Instituto Superior Tecnico, Univ. Tecnica de Lisboa, Lisbon, Portugal, 2008.

[8] P. C. Ng and S. Henikoff, "SIFT: Predicting amino acid changes that affect protein function," *Nucleic Acids Res.*, vol. 31, no. 13, pp. 3812–3814, Jul. 2003.

[9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, vol. 1, no. 1, pp. 886–893.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Jul. 2015.

[11] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1440–1448.

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.

[13] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Comput. Vis. Image Understand.*, vol. 61, no. 1, pp. 38–59, Jan. 1995.

[14] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, nos. 1–3, pp. 37–52, Aug. 1987.
[15] G. Thimm and J. Luettin, "Optimal parameterization of point distribution models," IDIAP, Martigny, Switzerland, EPFL Rep. 82460, Aug. 1998.
[16] B. van Ginneken, A. F. Frangi, J. J. Staal, B. M. ter Haar Romeny, and M. A. Viergever, "Active shape model segmentation with optimal features," *IEEE Trans. Med. Imag.*, vol. 21, no. 8, pp. 924–933, Aug. 2002.
[17] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, Jul. 1945.
[18] X. Huang, D. P. Dione, C. B. Compas, X. Papademetris, B. A. Lin, A. Bregasi, A. J. Sinusas, L. H. Staib, and J. S. Duncan, "Contour tracking in echocardiographic sequences via sparse representation and dictionary learning," *Med. Image Anal.*, vol. 18, no. 2, pp. 253–271, Feb. 2014.
[19] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 9, pp. 850–863, Sep. 1993.
[20] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.
[21] S.-M. Chao and D.-M. Tsai, "An anisotropic diffusion-based defect detection for low-contrast glass substrates," *Image Vis. Comput.*, vol. 26, no. 2, pp. 187–200, Feb. 2008.
[22] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, Jun. 2014, pp. 580–587.
[23] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2014, pp. 818–833.
[24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Sep. 2014, pp. 1–14.
[25] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 321–331, Jan. 1988.
[26] F. L. Bookstein, "Landmark methods for forms without landmarks: Morphometrics of group differences in outline shape," *Med. Image Anal.*, vol. 1, no. 3, pp. 225–243, Apr. 1997.
[27] T. F. Cootes and C. J. Taylor, "Statistical models of appearance for computer vision," Imag. Sci. Biomed. Eng., Univ. Manchester, Manchester, U.K., Tech. Rep., 2014.
[28] J. G. Bosch, S. C. Mitchell, B. P. F. Lelieveldt, F. Nijland, O. Kamp, M. Sonka, and J. H. C. Reiber, "Automatic segmentation of echocardiographic sequences by active appearance motion models," *IEEE Trans. Med. Imag.*, vol. 21, no. 11, pp. 1374–1383, Nov. 2002.
[29] X. Qin, Z. Cong, and B. Fei, "Automatic segmentation of right ventricular ultrasound images using sparse matrix transform and a level set," *Phys. Med. Biol.*, vol. 58, no. 21, p. 7609, Oct. 2013.
[30] G. Carneiro and J. C. Nascimento, "Combining multiple dynamic models and deep learning architectures for tracking the left ventricle endocardium in ultrasound data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2592–2607, Nov. 2013.
[31] M. Hansson, S. S. Brandt, J. Lindström, P. Gudmundsson, A. Jujić, A. Malmgren, and Y. Cheng, "Segmentation of B-mode cardiac ultrasound data by Bayesian probability maps," *Med. Image Anal.*, vol. 18, no. 7, pp. 1184–1199, Oct. 2014.
[32] P. Jaccard, "The distribution of the flora in the alpine zone. 1," *New Phytologist*, vol. 11, no. 2, pp. 37–50, Feb. 1912.
[33] 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 Trans. Image Process.*, vol. 21, no. 3, pp. 968–982, Mar. 2012.
[34] O. Bernard, J. G. Bosch, B. Heyde, M. Alessandrini, D. Barbosa, S. Camarasu-Pop, and P. M . Jodoin, "Standardized evaluation system for left ventricular segmentation algorithms in 3D echocardiography," *IEEE Trans. Med. Imag.*, vol. 35, no. 4, pp. 967–977, Apr. 2016.
[35] F. Milletari, M. Yigitsoy, N. Navab, and S. Ahmadi, "Left ventricle segmentation in cardiac ultrasound using hough-forests with implicit shape and appearance priors," Comput. Aided Med. Procedures, Technische Univ. München, Munich, Germany, Tech. Rep., 2014.

WEI-YEN HSU received the Ph.D. degree from the Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan, Taiwan, in 2008. He is currently a Professor with the Department of Information Management, National Chung Cheng University, Chiayi, Taiwan. He was a recipient of the Young Scholar Award from Taipei Medical University and National Cheng Kung University, in 2011 and 2013, respectively. He also received the Outstanding Research Award from National Cheng Kung University, in 2019. He has been a Founding Member of the Brain-Computer Interface (BCI) Society, since 2015, and is also the Academic Editor of Medicine journal and Associate Editor of BMC *Medical Informatics and Decision Making* journal.