Automated Segmentation of Whole Cardiac CT Images based on Deep Learning

Rajpar Suhail Ahmed, Dr. Jie Liu*
School of Computer & Information Technology
Beijing Jiaotong University,
Beijing, China 100044

Muhammad Zahid Tunio
Department of Computer Systems Engineering
Dawood University of Engineering and Technology
Karachi, Pakistan

***Abstract*—Segmentation of the whole-cardiac CT image sequence is the key to computer-aided diagnosis and study of lesions in the heart. Due to the dilation, contraction and the flow of the blood, the cardiac CT images are prone to weak boundaries and artifacts. Traditional manual segmentation methods are time-consuming and labor-intensive to produce over-segmentation. Therefore, an automatic cardiac CT image sequence segmentation technique is proposed. This technique was employed using deep learning algorithm to understand the segmentation function from the ground truth data. Using the convolution neural network (CNN) on the central location of the heart, filtering ribs, muscles and other contrasting contrast are not an obvious part of the removal of the heart area. Staked denoising auto-encoders are used to automatically deduce the contours of the heart. Therefore, nine cardiac CT image sequence datasets are used to validate the method. The results showed that the algorithm proposed in this paper has best segmentation impact to such cardiac CT images which have a complex background, the distinctness between the background and the target area which is not obvious; and the internal structure diversification. It can filter out most of the non-heart tissue part, which is more conducive to the doctor observing patient's heart health.**

Keywords*—Cardiac CT; segmentation; deep learning; automatic location; contour inference

I. Introduction

Cardiac disease is the leading cause of death in Human. The heart as a substantial organ, having only Imaging data to understand the location of its internal lesions, is an effective means of noninvasive diagnosis of cardiac disease [1]. Cardiac CT image sequence segmentation is the key to the diagnosis of cardiac diseases using CT images. Therefore, improvement of the segmentation accuracy of whole-heart CT image sequences has got a major concentration of cardiac disease research [2].

Heart and other soft tissue images are hard to get segmented because of its target and background contrast are small as compared to a large scale of noise. The traditional approaches mainly obtain the edge points of the target manually by the anatomical knowledge and experience by the clinician [3]. These methods have a high accuracy but time-consuming, labor-intensive, and the doctor's experience surely affect the accuracy of the segmentation strongly. With the development of computer and image processing technology, the researchers have developed a series of methods which semi-automatically extract the target through input of specified calculation parameters and human-computer interaction. These methods can combine a prior knowledge such as anatomical knowledge, which let it be more popular. In recent years, some full automatic segmentation methods were developed, which do not need to enter any parameters and interaction, make the segmentation more efficient. However, the accuracy of the segmentation cannot be ensured and the calculation is extremely expensive [4].

In order to increase the precision and segmentation efficiency regarding the automatic segmentation algorithm, a novel method based on deep learning used for whole cardiac CT image sequences segmentation is proposed here. The procedure is done in three stages: First, the convolution neural network technique used to locate the center of the heart. Then the stacked denoising auto-encoder was used to automatically deduce the contours of the heart. Finally, the Gaussian smoothing and cardiac segmentation were employed to further improve the efficiency of the segmentation. Experiments display that the preparation method based on deep learning has high segmentation efficiency and accuracy.

II. Methods

The Architecture of the prepared segmentation method is shown in Fig. 1. The segmentation method mainly includes the following three steps: 1) Region of interest consisting the heart is located using a convolutional neural network [5]-[9]. 2) The cardiac contour is inferred using staked denoising auto-encoders, which trained to delineate the heart [10]-[16]. 3) Smoothing cardiac contour using Gaussian Smoothing Filter and using smoothed cardiac contour image to segment cardiac CT image [17]-[19]. Each step is individually trained during an offline training process to obtain its optimum values of parameter. After training, the automatic segmentation to learn deep learning network employee is used. This way, three-steps are described below.

A. Automatic Location

CT cardiac tissues of the heart and chest raw image sequence usually contain tissues. In order to reduce the impact of the surrounding tissue on the segmentation while improve the accuracy of segmentation, the first step of the algorithm is to locate the region of interest (ROI) containing the heart and filter out ribs, vertebrae and other interference. Fig. 2 shows the diagram of the convolution neural network to locate the specific location of the heart of the process.

*National Natural Science Foundation of China under Grant Nos. 81571836.

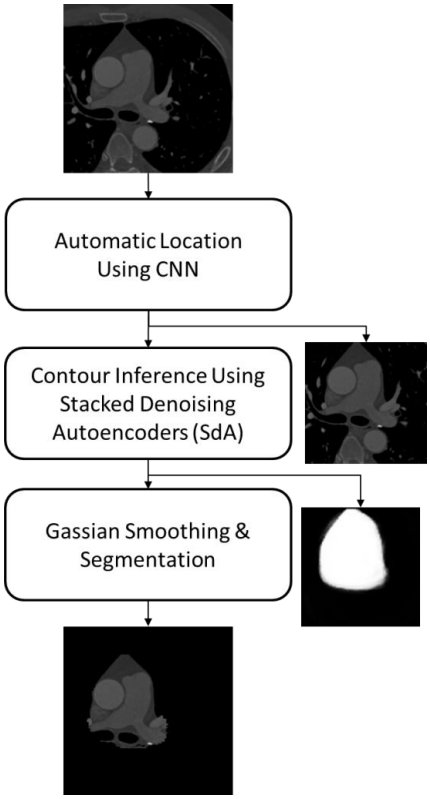


Fig. 1. The architecture of the developed algorithm.

In order to reduce the complexity of the algorithm and computational complexity, the original image size of 512×512 is down-sampled to 64×64, which is used as the input data of the neural network.

Then, the filter $F_l \in \Re^{11\times11}$, $b_0 \in \Re^{100}$ is employed to convolution operation with the input image to obtain a convolution characteristic map. I signify an input grayscale image having a size of 64×64, $I[i,j]$ represents the gray value of the point with the coordinate $[i,j]$ in the image. The convolution characteristic is calculated by this formula $C_l[i,j] = f(Z_l[i,j])$, where

$$Z_l[i,j] = \sum_{k_1=1}^{11}\sum_{k_2=1}^{11} F_l[k_1,k_2]\, I[i+k_1-1,\, j+k_2-1] + b_0[l] \quad (1)$$

In the formula $1\le i,j\le 54$, $l=1,...,100$. This results in 100 convolved features $Z_l \in \Re^{54\times54}$. Here, $x[i]$ signify the ith number of the vector x and $X[i,j]$ signify the number of ith row and jth column of the matrix X.

Next, the convolution feature map is sub-sampled by maximum pooling, and each pooled feature value calculates the maximum of the adjacent overlapping.

7×7 size convolution map area is calculated through (2).

$$P_l[i_1,j_1] = \max\Big(\sum_{i=7i_1-6}^{7i_1}\sum_{j=7j_1-6}^{7j_1} C_l[i,j]\Big) \quad (2)$$

For $1\le i_1,j_1\le 8$. The result in 100 reduced-resolution features, where $P_l \in \Re^{8\times8}, l=1,...,100$.

Finally, the pooled feature expands to the vector $p \in \Re^{6400}$, a logical layer that associates with cell output 1024. 32 x 32 size picture of the specific region of interest, logical regression layer will create a mask.

The output of the formula is $y_c = f(w_1 p + b_1)$, where $w_1 \in \Re^{1024\times8100}$ and $b_1 \in \Re^{1024}$ are trainable matrices.

Because the size of the original input CT image is 512×512, so the output mask image is up-sampled to the size of the original image, and finally, according to the mask image. To produce the ROI image for the next step, and the size is 400×400. It is necessary to train the network parameters to confirm the accuracy of the location.

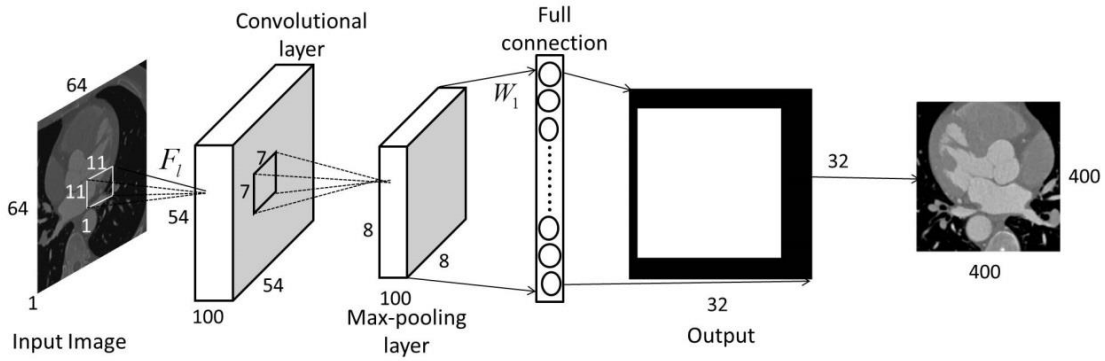


Fig. 2. The architecture of the automatic location for the heart in CT dataset.

The training and optimization of the parameters in location operation will introduce as follows:

1) Training Convolutional Network

The CNN training mainly focuses on obtaining the parameters $F_l, l=1,...,100$ b_0 of the filter, and the optimal values of the other parameters w_1, b_1. In the case of large enough data sets, these parameters can obtain through the training of the ideal parameter values; therefore, in the normal case, the parameters are randomly initialized and then trained to obtain the desired parameter values. When using the unsupervised method of hierarchical pre-training depth network weights, random noise is introduced into the visual layer (e.g. the input layer of data) of the network, which is called denoising auto-encoder (dA). As shown in Fig. 3, the random noise is added to the sample x from the QD distribution instead of adding the Gaussian noise, the input value (as many as half of them) is cleared to zero with a certain probability, this is known as stochastic corrosion process. This successfully solves the problem of few data sets and avoids the resulting overfitting. At here the stochastic corrosion process ratio is 0.5. As shown in Fig. 4, the number of input and output units of dA is 121 and the number of hidden layer units is 100. Randomly select $N_1=10^4$ blocks of size 11×11 from the input image sequence, each block is expanded into a vector $x^{(i)}\in\Re^{121}$, $i=1,...,N_1$ and is used as input to the dA.

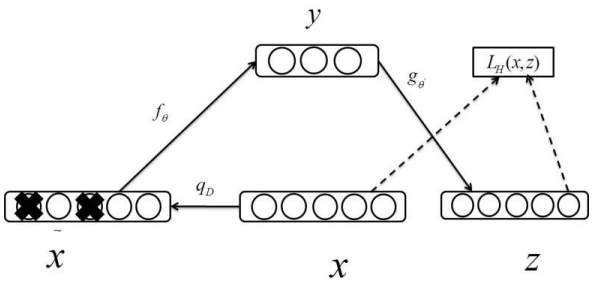


Fig. 3. The architecture of denoising auto-encoder.

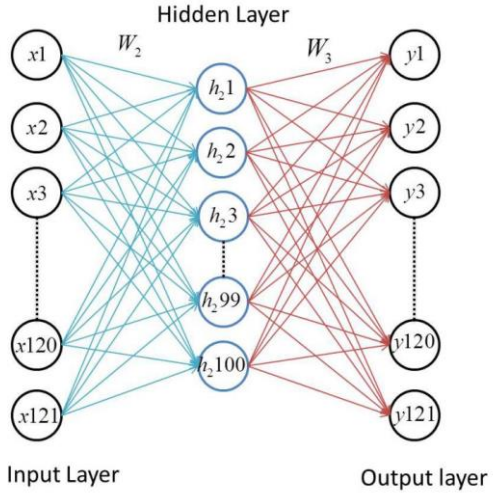


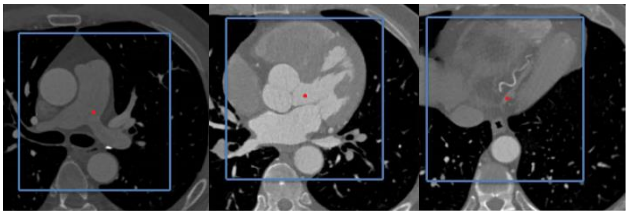
Fig. 4. Train denoising auto-encoder is to initialize filters.

The weight parameter matrix between input layer and hidden layer is $W_2\in\Re^{100\times121}$, and the weight parameter within the implicit layer and the output layer is $W_3\in\Re^{121\times100}$. Hidden layer and output layer are calculated through the formula $a_2^{(i)}=f\left(w_2x^{(i)}+b_2\right)$ and $y^{(i)}=f\left(w_3a_2^{(i)}+b_3\right)$, here the data $x^{(i)\prime}$ and $a_2^{(i)\prime}$ are the stochastic corruption process of $x^{(i)}$ and $a_2^{(i)}$. The sigmoid activation function is $f(x)=1/\left(1+e^{-x}\right)$, $b_2\in\Re^{100}$, $b_3\in\Re^{121}$ is the bias vector. The purpose of the dA is to recompose $x^{(i)}$ at the output layer over the output of the hidden layer. Therefore, the input data is used as the mark data without requiring the actual mark data. The dA parameter optimization is achieved through the low-cost function.

$$J(w_2,b_2)=\frac{1}{2N_1}\sum_{i=1}^{N_1}\left|y^{(i)}-x^{(i)}\right|^2 + \frac{\lambda}{2}\left(\|w_2\|^2+\|w_3\|^2\right)+\beta\sum_{j=1}^{k}KL\left(\rho\cdot\hat{\rho}_j\right) \quad (3)$$

The first term is the direct mean square error within the output $y^{(i)}$ and the ideal output $x^{(i)}$. Moreover, in order to prevent overfitting, the second terminology is added; a weight attenuation term whose purpose is to reduce the magnitude of the weights. At the same time in order to be able to learn more features from the input data, the hidden unit added sparse constraints. In this way, a dA network is built. Here, KL divergence limits the activation mean $\hat{\rho}_j=\left(1/N_1\right)\sum_{i=1}^{N_1}a_2^{(i)}[j]$, $j=1,...,100$ of the hidden layer, and ρ_j is usually small in order to maintain equilibrium with the discrete parameter ρ.

The weight attenuation coefficient λ and the discrete coefficient controls the relative of the three conditions is the importance of the cost function. The optimal values are $\lambda=10^{-4}$, $\rho=0.1$, and $\beta=3$. After the dA is trained, W_2 is the initial value of $F_l\in\Re^{11\times11}$, $l=1,...,100$ and $b_0=b_2$ for the next step.



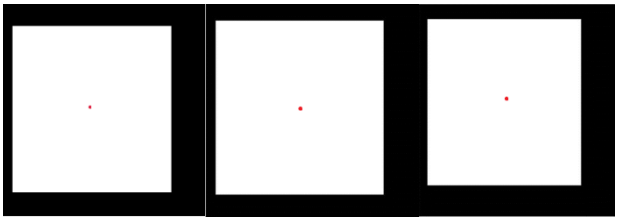


Fig. 5. Input images (top) and the use of automatic location for training related binary mask image (bottom).

The feed-forward operation, then (1) and (2) through to the output layer is performed, and the output layer is already trained through the minimal cost function.

$$J\left(w_1,b_1\right)=\frac{1}{2N_2}\sum_{i=1}^{N_2}\left|y_c^{(i)}-I_{roi}^{(i)}\right|^2+\frac{\lambda}{2}\left(\left\|w_1\right\|^2\right) \qquad (4)$$

Where I_{roi} is described data matching to the *i*th input image and N_2 is the number of training images? The described data of the output layer is a binary mask image which generates through the described manual. The binary mask is a black background and the white foreground image matching to the ROI.

As shown in Fig. 5, the midpoint of the foreground corresponds to the midpoint of the heart contour, the binary mask is down-sampled to 32×32 and expanded into vectors I_{roi} for the network training.

Finally, the fine tuning of the entire network achieved is by minimizing the cost function represented by (5).

$$J\left(F_l,b_0,w_1,b_1\right)=\frac{1}{2N_2}\sum_{i=1}^{N_2}\left|y_c^{(i)}-I_{roi}^{(i)}\right|^2+\frac{\lambda}{2}\left(\left\|w_1\right\|^2+\sum_{l=1}^{100}\left\|F_l\right\|^2\right) \qquad (5)$$

The cost function can be minimized by back propagation aggregate, where $\lambda=10^{-4}$, the training method only runs once.

B. Contour Inference

Unlike the literature [6], to train and use the stacked dA to infer the contour of the heart is shown in Fig. 6. The Stacked dA has an input layer, two hidden layers, and an output layer. The sub-image obtained by the convolution network is down-sampled and expanded into a vector $x_s\in\Re^{6400}$ and feedback to the input layer. The output of the two hidden layers is calculated by $a_1=f\left(w_4x_s'+b_4\right)$ and $a_2=f\left(w_5a_1'+b_5\right)$. Finally, the output layer computes $y_s=f\left(w_6a_2'+b_6\right)$ to produce a binary mask image with a pixel value of 0 or 1, where the heart segment corresponds to 1. Here, the parameters $w_4\in\Re^{100\times6400}$, $b_4\in\Re^{100}$, $w_5\in\Re^{100\times100}$, $b_5\in\Re^{100}$, $w_6\in\Re^{6400\times100}$, $b_6\in\Re^{6400}$ are training matrices and vectors, respectively, obtained by the training program, the data x_s', a_1' and a_2' are the result of stochastic corruption process of data x_s, a_1 and a_2.

1) Training stacked-dA

Training divides into two steps, similar to previous convolution neural network training, include pre-training and fine-tuning. Because of the limited data available in the experiment, a layer-to-layer approach is used for pre-training parameters. This method effectively reduces the overfitting. In the layer-to-layer pre-training process, parameters w_4, w_5 of the staked-dA are achieved layer-to-layer with no described data. The parameter is achieved applying described data. This Valuation is as follows.

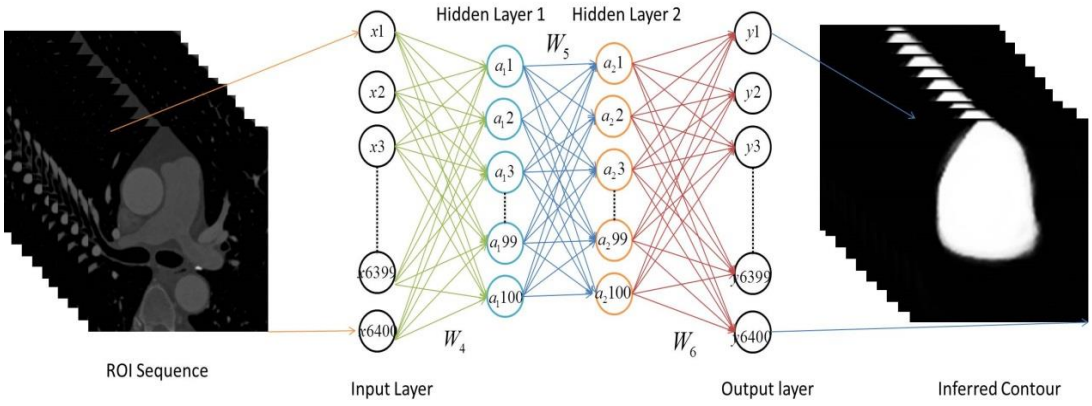


Fig. 6. Stacked dA for inferring the cardiac contour. An input is sub-image and the binary mask is an output.

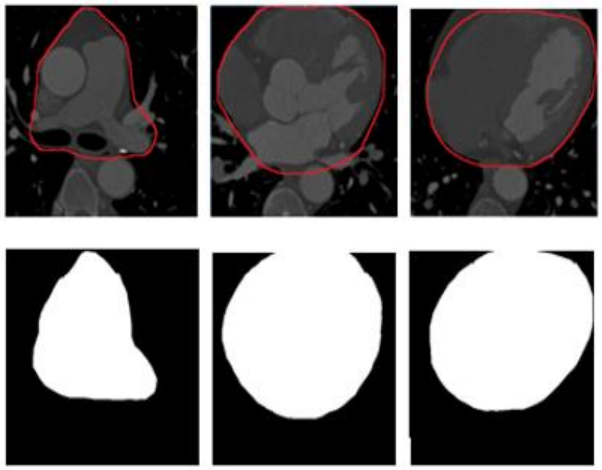


Fig. 7. Typical sub-image with manual heart segmentation and the matching binary mask applied training the stacked denoising auto-encoder.

Initially, an input layer and from the H_1 hidden layer is distinct from the stacked-dA by calculating the output layer of the similar size as an input layer to form a sparse dA (similar to Fig. 4). An unsupervised way to get the w_4 parameter is trained to sparse dA, as described in Section 2.1.1. The optimal parameter is set as $\lambda = 3\times10^{-3}$, $\rho = 0.1$, and $\beta = 3\times10^{-1}$.

The training input and output data of reserve dA are sub-images of the size 400×400 drawn out from the full-size training data, image midpoint matching to the center of the heart. The input image is down-sampled to the size of 80×80 to be compatible with an input size 6400 of the stacked-dA. Once training of reserve dA is complete, it's an output layer that is deserted. An output of the dA hidden layer unit is used as input to the next hidden layer H_2.

Next, separate the H_1 and H_2 from the stacked dA and add a layer of the same size as the H_2 output to compose a sparse dA. In addition, the next sparse dA is used to get w_5, with no described data. The input data is same as the input data of the previous dA hidden layer.

An output of the last hidden layer is the input of the last layer, it is trained in the unsupervised mode to obtain w_6. The cost function used to train the last layer is calculated as follows:

$$J\left(w_6, b_6\right) = \frac{1}{2N_2}\sum_{i=1}^{N_2}\left|y_s^i - I_{lv}^i\right|^2 + \frac{\lambda}{2}\left\|w_6\right\|^2 \qquad (6)$$

Where, $I_{lv}^i \in \Re^{6400}$ is the described dataset matching to the *i*th image. The described dataset is a binary mask image generated by hand segmentation. Fig. 7 depicts three input images and a matching binary mask image for the training of the stacked-dA. It should note that the binary mask image is expanded into the vector I_{lv} during optimization.

The purpose of layer-by-layer pre-training is to obtain the appropriate initial values for parameters w_4, w_5, w_6. Finally, the complete architecture is fine-tuned through reducing the cost function as shown in (2).

$$J\left(w_{4,}w_5, w_6, b_{4,}b_5, b_6\right) = \frac{1}{2N_2}\sum_{i=1}^{N_2}\left|y_s^i - I_{lv}^i\right|^2 + \frac{\lambda}{2}\left(\left\|w_4\right\|^2 + \left\|w_5\right\|^2 + \left\|w_6\right\|^2\right) \qquad (7)$$

The minimization process of the cost function uses a supervised back-propagation algorithm. The automatic detection, as well as the training process, perform only once. The optimal parameter is $\lambda = 10^{-4}$.

C. Smoothing and Segmentation

The final step of the algorithm is smoothing and segmentation, because the contour inference network output image size is 80×80. It has a clear jagged edge after the image is up-sampled to the size of 400×400. To resolve this issue, a two-dimensional Gaussian filtering algorithm is followed to smooth the contour inference image which is up-sampled to the size of 400×400, the algorithm formula is:

$$G\left(x, y\right) = \frac{1}{2\pi\sigma^2}e^{-\frac{\left(x^2+y^2\right)}{2\sigma^2}} \qquad (8)$$

Finally, a binary mask image with contour inference is obtained after high-pass filtering and binarization. The size of the smoothing filter is 10, after the experiment, when σ is 2.5, the smoothing effect is the best. Finally, the heart contours are registered by a high pass filter, and the segmentation result will be got.

III. EXPERIMENT

In the proposed method the training dataset from TaiZhou City People Hospital was used. The training dataset consisted of 9 groups, each group of dataset matching to a patient's cardiac CT images, the number of each group ranging from 241 to 344. The size of cardiac CT image is 512×512. The number of available experimental image of data is 2500.

In order to enhance the efficiency of the segmentation algorithm, the training dataset are linearly interpolated, and the image data is expanded to 5000 pieces. The described dataset consists of two parts; the described image is binary mask image. The first part is for convolution neural network training, the center of the region of interest corresponds to the cardiac center of the cardiac CT image. The second part is for stacked denoising auto-encoder training, the foreground white outline of the described image corresponds to the contour of the heart. The training is performed only once.

Considerable overfitting may happen in deep learning network due to the large number of parameters to be learned. Mainly focus a lot on the network to intercept the problem overfitting, to find the solution to this problem, we adopted layer-wise pre-training, l_2 regularization and denoising

constraints as detailed in Section 2.1 and 2.3. The challenge was the deficit of data and the use of layer-wise pre-training and denoising constraints is very useful. Hence, also kept the number of the layer in the location networks do not go before three and the number of the layer in the contours inferring do not go before four to make sure that the number of parameters is manageable. While training of the proposed method, cross-validation was accomplished through dividing the training dataset into 8 subjects for training and 2 subjects for validation and also early stopping to supervise and intercept overfitting. Furthermore, the training dataset is increased as described in the beginning. The hyperparameters of the networks, the number of layers and units, filter and pooling size, are persistent empirically throughout the cross-validation method.

The performance of the segmentation algorithm is evaluated by its accuracy, regional statistical characteristics, and reliability [20]. The manual segmentation result is taken as the reference standard, and the performance of the segmentation algorithm can realize by calculating the segmentation algorithm accuracy. Accuracy is the degree of consistency between the segmentation results and the manual segmentation results [21], where expressing the accuracy of the segmentation by calculating the percentage of correct classified pixels as a percentage of the number of reference pixels [22]. The four measures of accuracy are True Positives (TP), False Positive (FP), True Negative (TN) and False Negative (FN), the matching percentages are TPR, FPR, FNR, and TNR. The correct rate (TTR) is calculated by (9).

$$TPR = \frac{TP}{T} \times \frac{100}{100}$$
$$FNR = \frac{FN}{T} \times \frac{100}{100}$$
$$FPR = \frac{FP}{I-T} \times \frac{100}{100} \quad (9)$$
$$TNR = \frac{TN}{I-T} \times \frac{100}{100}$$
$$TTR = (1 - \frac{(FP+FN)}{\mathrm{T}}) \times \frac{100}{100}$$

In the recent study, this program designed in MATLAB 2015 a, accomplished on a Dell OptiPlex 7020 machine, with Intel(R) Core(TM) i5-4150 CPU 3.5GHz, 16GB RAM, on a 64-bit Windows 10 platform. The process was training working on the training dataset and tested on the testing dataset, the testing dataset includes three sequences of cardiac CT images.

IV. Results

A. Illustrative Results

In order to determine the aspect of every step, the effect of CNN (heart location, Step 1) for three image segments close the base/middle and the point of the heart as shown in Fig. 8, illustrates the location results of the different cardiac segment. Fig. 9 illustrates the location center of a whole sequence of cardiac CT images.

The deep learning method (contours inference, Step 2) outcome for three image segments close the base/middle and the point of the heart is shown in Fig. 10.

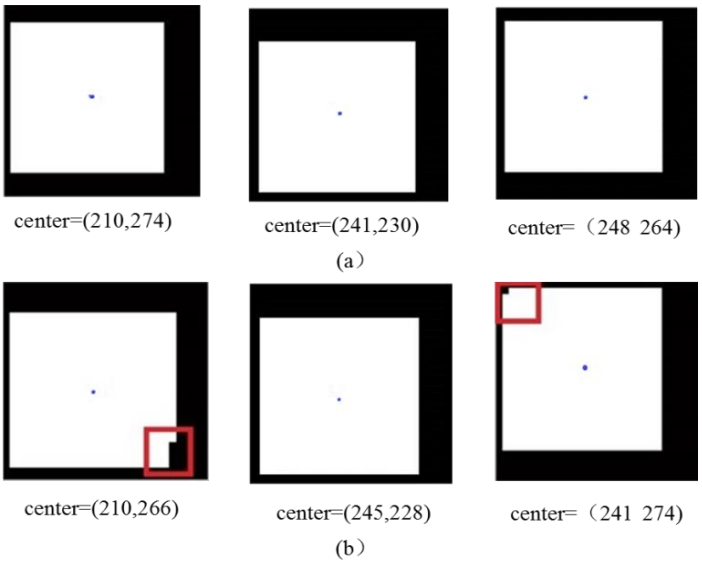


Fig. 8. CNN and manual location result from base to point of the heart.

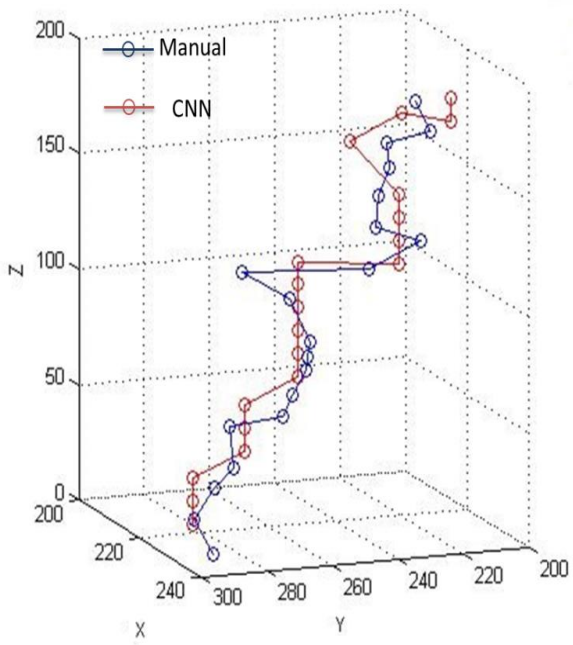


Fig. 9. Location centers from base to the apex of a whole cardiac CT image sequences.

The region growing (RG) segmentation algorithm is used as the contrast algorithm. Fig. 11 illustrates the original, region growing automated, and manual whole heart segmentation results for a typical cardiac CT dataset.

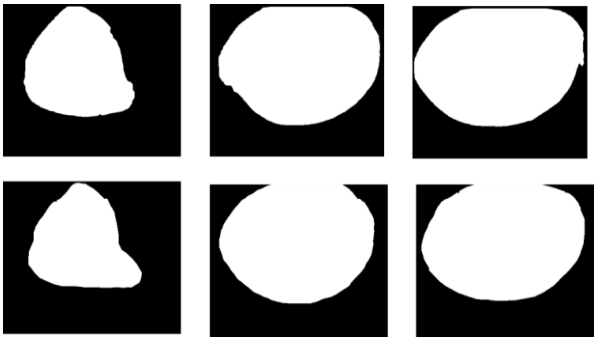


Fig. 10. Automatic (top) and manual (bottom) contours inference from base to the point of the hear.

From the base to Apex, in Fig. 11, original images (a), (b), (c) are shown in the first row, region growing segmentation results (d), (e), (f) are shown in the second row, automatic segmentation results (h), (i), (j) are shown in the third row, and the ground truth manual segmentations (k), (l), (m) are shown in the last row.

B. Quantitative Results

Table I shows the evaluation of method prepared and region growing segmentation results, from which we can understand that the method prepared in this paper has a higher TTR (average 93.77%), TNR (average 98.15%) and lower FPR (average 1.84%); it is almost completely close to the manual segmentation.

Table II compare the running time of method proposed in this paper and RG spent in the segmentation. The average time of proposed method is 0.1244s, it is far less than the average time of RG (4.2385 s).

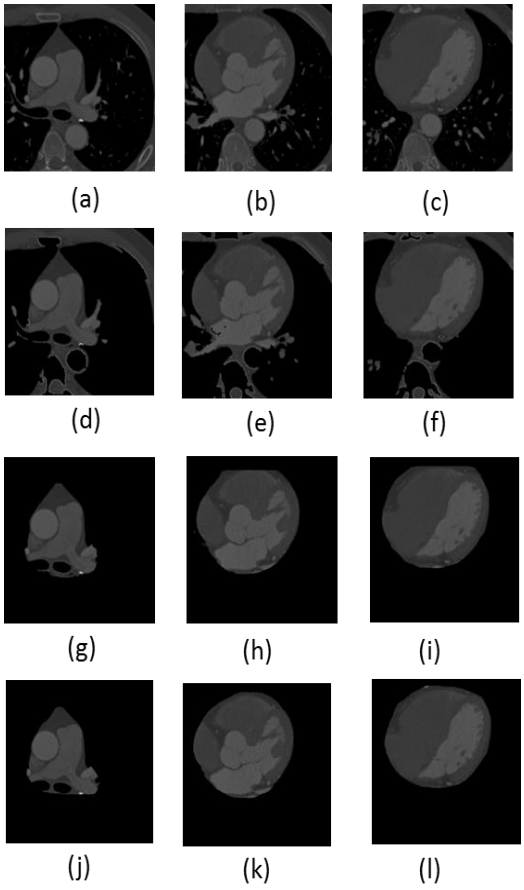


Fig. 11. Segmentation results of heart.

TABLE I. OBJECTIVE EVALUATION OF DEEP LEARNING AND REGION GROWING SEGMENTATION METHODS

Segment	Method	TPR (%)	FPR (%)	TNR (%)	FNR (%)	TTR (%)
Base	DL	98.67	1.30	98.69	1.32	92.23
	RG	99.43	25.03	74.97	0.56	-23.86
Mid	DL	97.78	1.53	98.46	2.22	94.8
	RG	99.56	23.29	76.71	0.43	54.26
Apex	DL	99.25	2.69	97.31	0.74	94.30
	RG	99.95	19.65	80.35	0.04	63.80
Average	DL	98.56	1.84	98.15	1.42	93.77
	RG	99.65	22.65	77.34	0.34	31.4

TABLE II. COMPARISON OF RUNNING TIMES

Segment	Base	Mid	Apex	Average
DL(s)	0.1227	0.1229	0.1276	0.1244
RG(s)	1.8297	4.6046	6.2820	4.2385

Approximated lapsed times of the training method were as follows: Training denoising auto-encoder to get filters: 123.4s, training convolutional neural network: 4.1 h, training stacked denoising auto-encoder: 3.2h. Once trained, the elapsed times of the heart segmentation in a typical CT image were as follows: ROI location (convolution, pooling, and the logistic regression):0.23s, contour inferring (stacked -dA):0.12s, and segmentation (smoothing and segmentation): 0.15s.

V. DISCUSSION AND CONCLUSION

In this research, we suggested and validated an automated heart segmentation process placed on deep learning algorithms. In this issue, cover the localization, contour inferring and segmentation tasks. Convolutional neural networks were preferred for the localization and extracting an ROI through which they are matched to special translation and variation in scale and pixels' density [23]. Also, cover a stacked dA for contour inferring as its simplicity in training and implementation still shown to become effective in different vision tasks [24]. Meaningfully, a perfect deep learning network is desired. However, this was impossible due to the limited number of training dataset. Hence, we kept the number of the layer in the location networks not to go beyond three and the number of the layer in the contours inferring not to go beyond four to make sure that the number of parameters are manageable. Also, employ the linear interpolation to enlarge the training dataset.

As seen in Fig. 3 and 4, the denoising auto-encoder and layer-wise pre-training were adopted in CNN training to

ensure the accuracy of localization. And the outcome of CNN in Fig. 8 and 9 also verify the effectiveness of the measures. The CNN and manual location of the heart are pretty close and the trend of location center is consistent, the average distance error is below 10 pixels. Although, there is still some error in Fig. 9, but the error is within the allowable range.

From Fig. 10, this could be recognized that the heart was accurately segmented from the base to point. In Fig. 11, the first row is the original cardiac CT image, the slices near the ground truth from a small base cane showed leakage. This position is one of the most difficult cases that due to the fuzziness of the heart border, contour tends to leak to surrounding areas in normal segmentation methods. In this study, the used staked-dA is perfect to solve this problem.

The quantitative results in Table I show that the method prepared in this study has a pretty high average accuracy (93.77%) of segmentation. Table II compare the average running time of method to the region growing algorithm, the result showed that the average running time of methods proposed is much faster, and can complete fully automatic segmentation tasks of whole heart CT image sequence in a short time.

In summary, a novel segmentation process for a fully automatic cardiac CT is presented. The process places on deep learning algorithm broke down into automatic localization and inferring the heart contour. And the contour inferred was smoothing by the Gaussian filter. It could be recognized that the segmentation process prepared in this paper has better segmentation efficiency than the regional growth segmentation method and successfully filters out most of the interfering tissue; the result is the closest to the manual segmentation and can replace the manual segmentation and realize the Automatic Segmentation of cardiac CT image sequences.

In the future, other depth learning algorithms may be introduced to realize heart image segmentation. In the medical image field, the application of 4D images is the mainstream trend. The basic data of the 4D image is a large volume of information contained in the three-dimensional volume data. The existing segmentation algorithm can be extended to the three-dimensional space to directly segment the volume data.

ACKNOWLEDGMENTS

This work is partially supported by National Natural Science Foundation of China under Grant Nos. 81571836. Authors would also like to thank the Taizhou City People Hospital and Beijing Jiaotong University for giving us the opportunity to do some furthermore research on CT & MRI image feature evaluation.

REFERENCES

[1] Organization. W H. Global status report on non-communicable diseases 2014.[J]. Women, 2011, 47(26):2562-2563.

[2] S.Mendis, P.Puska, B.Norrving. Global Atlas on cardiovascular disease prevention and control.World Health Organization in collaboration with the World Heart Federation and the World Stroke Organization.2011, pp.3-18.

[3] J.Rogowska."Overview and fundamentals of medical image segmentation". Handbook of Medical Imaging Processing and Analysis. Pp. 69-85, 2000.

[4] N.R Pal,S.K.Pal,"A Review on Image Segmentation Techniques". Pattern Recognition,vol. 26, pp:1277-1294, 1993.

[5] Lecun Y, Kavukcuoglu K, Farabet C. Convolutional Networks and Applications in Vision[J]. 2010, 14(5):253-256.

[6] Szegedy C, Toshev A, Erhan D. Deep Neural Networks for object detection[J]. Advances in Neural Information Processing Systems, 2013, 26:2553-2561.

[7] Sermanet P, Eigen D, Zhang X, et al. Over Feat: Integrated Recognition, Localization and Detection using Convolutional Networks[J]. EprintArxiv, 2014.

[8] Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks[J]. Advances in Neural Information Processing Systems, 2012, 25(2):2012.

[9] Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 79(10):1337-1342.

[10] Bengio Y, Courville A, Vincent P. Representation Learning: A Review and New Perspectives[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2013, 35(8):1798-828.

[11] Vincent P, Larochelle H, Bengio Y, et al. Extracting and composing robust features with denoising auto-encoders[C]// International Conference. 2008:1096-1103.

[12] Bengio Y. Learning Deep Architectures for AI[J]. Foundations & Trends® in Machine Learning, 2009, 2(1):1-55.

[13] Baldi P, Guyon G, Dror V, et al. Auto-encoders, Unsupervised Learning, and Deep Architectures Editor: I[J]. Journal of Machine Learning Research, 2012.

[14] Deng L, Yu D. Deep Learning: Methods and Applications [J]. Foundations & Trends® in Signal Processing, 2013, 7(3):197-387.

[15] Vincent P, Larochelle H, Lajoie I, et al. Stacked Denoising Auto-encoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion.[J]. Journal of Machine Learning Research, 2010, 11(12):3371-3408.

[16] Avendi M R, Kheradvar A, Jafarkhani H. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI.[J]. Computer Science, 2016, 30:108-119.

[17] Xiao K, Ho S H, Salih Q. A Study: Segmentation of Lateral Ventricles in Brain MRI Using Fuzzy C-Means Clustering with Gaussian Smoothing[C]// Rough Sets, Fuzzy Sets, Data Mining and Granular Computing, International Conference, Rsfdgrc 2007, Toronto, Canada, May 14-16, 2007, Proceedings. DBLP, 2007:161-170.

[18] Xiao K, Ho S H, Salih Q. A Study: Segmentation of Lateral Ventricles in Brain MRI Using Fuzzy C-Means Clustering with Gaussian Smoothing [M]// Rough Sets, Fuzzy Sets, Data Mining and Granular Computing. Springer Berlin Heidelberg, 2007:161-170.

[19] Izquierdo M. E, Ghanbari M. Texture smoothing and object segmentation using feature-adaptive weighted Gaussian filtering[C]// Telecommunications Symposium, 1998. ITS '98 Proceedings. SBT/IEEE International. 1998:650-655 vol.2.

[20] Chalana V, Kim Y. A methodology for evaluation of boundary detection algorithms on medical images.[J]. IEEE Transactions on Medical Imaging, 1997, 16(5):642-52.

[21] Zhang Y J. A survey on evaluation methods for image segmentation ☆ [J]. Pattern Recognition, 1996, 29(8):1335-1346.

[22] Zhang Y. A Classification and Comparison of Evaluation Techniques for Image Segmentation[J]. Journal of Image & Graphics, 1996.

[23]]Bruna J, Zaremba W, Szlam A, et al. Spectral Networks and Locally Connected Networks on Graphs[J]. Computer Science, 2014.

[24] Vincent P, Larochelle H, Lajoie I, et al. Stacked Denoising Auto-encoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion[J]. Journal of Machine Learning Research, 2010, 11(12):3371-3408.