

A Framework for Left Atrium Segmentation on CT Images with Combined Detection Network and Level Set Model

Yashu Liu¹, Kuanquan Wang^{1*}, Gongning Luo¹, Henggui Zhang^{1,2}

¹Harbin Institute of Technology, Harbin, China

²University of Manchester, Manchester, UK

Abstract

In this paper, we proposed a framework for left atrium (LA) segmentation on CT with combined detection network and level set model. The proposed framework consists of two steps. Firstly, we trained a Faster RCNN to generate location boxes for LA. The obtained location box can remove unrelated regions to reduce the interference of background and similarity tissues. Secondly, we utilized a self-adapted threshold on the location box to get the initialization for the level set model, which is nearer the LA and more robust than the random and fixed initialization. Then we proposed a 3D level set model with a new edge indicator based on DRLSE for the final LA segmentation. This edge indicator incorporated both numerical and direction information of the data gradient. Hence, the proposed level set model can guide the contour to the correct boundary when there are many boundaries surrounded the object. The framework was trained and evaluated on MICCAI 2013 LA segmentation challenge. The proposed segmentation method achieved the Dice score of 86.46%. Comparing to the original DRLSE, it achieved a 2.72% improvement on the Dice score.

1. Introduction

Cardiac Computer Tomography (CT) is a popular physical inspection to diagnose cardiac disease. Atrial fibrillation (AF) is a common and risky sustained cardiac arrhythmia disease, which occurs on left atrial (LA) [1]. A credible mechanism of AF is structure remodeling of LA [2]. Therefore, understanding the LA shape is crucial for cure AF. Moreover, LA segmentation is the basic process for the understanding. However, the small, complex structure and individual differences of LA increase the difficulty for LA segmentation. In addition, there is the problem of similarity intensity distribution with other tissues. Due to these challenges, the LA segmentation in clinical still relies on manual segmentation. However, manual segmentation is tedious and subjective. In this paper, we proposed a framework for LA segmentation on

CT.

Due to accurate location can remove the most unrelative background, location is always an efficient process for accelerating small objects segmentation. Luo et al. [3] utilized a combination of intersecting views and atlas-based method to locate the left ventricle. Ammar et al. [4] located the LA by an empirical threshold method. Jia et al. [5] located the LA through drawing polygons manually. These location methods are empirical and subjective. Inspired by the successful Faster RCNN [6] on natural images, we utilized it to automatic locate the LA. The result indicated that it can provide a more efficient and accurate location of LA.

Even though LA is a very small tissue in a CT volume, it is a very complex structure and has great difference among patients. It consists not only of LA cavity, but also of LA appendage, and tight pulmonary veins. Hence, we utilized a level set method for LA segmentation based on the former location step. The level set method is well known as the prominent ability to handle complex topology change easily and provide smooth contours and its efficiency on medical image segmentation has been evaluated on many literatures [7,8]. In this kind of method, the contour of segmentation is implicitly represented as a zero-level set in a higher dimensional function, which is called level set function. It converts the segmentation problem into a function minimization problem. The general level set function is represented in Equation 1:

$$\frac{\partial \varphi}{\partial t} = F |\nabla \varphi|, \quad (1)$$

where φ is the evolution function, which should be a signed distance function (SDF) during the evolution; F is the speed function to control the speed of the contour evolving. There will be a re-initialization process among the evolution in order to keep φ as an SDF. However, re-initialization is very time-consuming. Li [9] proposed a DRLSE with a two-term penalize item to overcome the re-initialization problem. It decreases the computation complexity and improving efficiency. The evolution function is shown in Equation 2:

$$\frac{\partial \varphi}{\partial t} = \mu \operatorname{div}(d_p(|\nabla \varphi|) \nabla \varphi) + \lambda \delta_\epsilon(\varphi) \operatorname{div}\left(g \frac{\nabla \varphi}{|\nabla \varphi|}\right) + \alpha g \delta_\epsilon(\varphi), \quad (2)$$

where $\mu > 0, \lambda > 0$, and $\alpha \in \mathbb{R}$ are coefficients for the three items. The first item is the penalize item, named distance regularization energy. The second and third items are corresponding to the length item and area item. The regularization energy utilized $d_p(|\nabla\phi|) = 1 - \frac{I}{|\nabla\phi|}$ as a

metric to overcome the re-initialization problem. It will always keep the function ϕ closing to an SDF in $\Omega \in \mathbb{R}^2$.

$g = \frac{1}{I + |\nabla G_\sigma * I|^2}$ is the classical edge indicator function based on the image gradient, where G_σ is the Gaussian kernel with standard deviation σ . g is trending to 0 at the edge region with a sharp gradient, and it is trending to 1 at the flatten region with a smooth gradient. $\delta_\epsilon(\phi)$ is the Dirac delta function.

Even though this method has achieved a satisfying result in some cases. The edge indicator, g , just utilized the numerical value of the image gradient. The direction of the gradient is abandoned. Therefore, the initial contour of DRLSE must completely surround or be surrounded by the object. In this paper, we proposed a level set method based on a new edge indicator with both numerical and direction information of gradient for LA segmentation task.

The rest of this paper is organized as follows. We will introduce our method in section 2. Then the experiments and results are shown in section 3. At last, we conclude this paper in section 4.

2. Methods

Our LA segmentation framework is shown in Figure 1. As the figure shows, it has two main steps, location and segmentation. The two sub-steps for segmentation is Otsu [10] and 3D level set method. They are corresponding to initialization and final segmentation method.

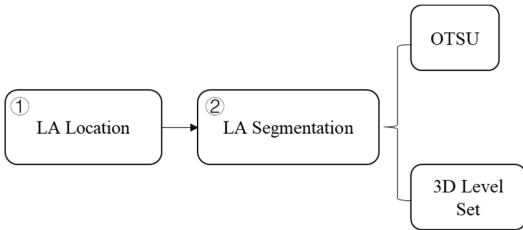


Figure 1. The pipeline of the proposed framework.

(1) Firstly, we trained a Faster RCNN to generate a location box for LA on the image level. The obtained location box can remove the most of unrelated regions to reduce the interference of background and similarity tissues.

(2) Secondly, we utilized a self-adapted threshold, Otsu, on the location box to get the initial contour for the level set model, which is nearer the LA and more robust than the

random and fixed initialization. And then, we utilized the proposed edge indicator to incorporate DRLSE for the LA segmentation on the 3D volumes.

2.1 LA Location

We utilized Faster RCNN as our location network. The pipeline of it is shown in Figure 2. It has achieved great success in natural image detection. It introduces a Region Proposal Network (RPN) that shares full-image convolutional features with Fast RCNN, which proposes a Region of Interest (RoI) layer streamlined the training process. This detection model improves detection accuracy while increasing the training speed.

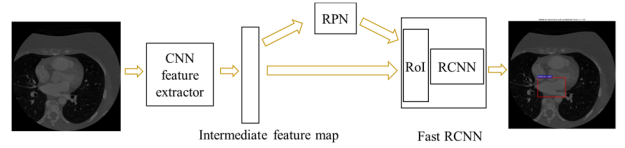


Figure 2. The pipeline of location network

The CNN feature extractor extracts intermediate feature maps, which are shared between RPN and Fast RCNN, from the input image. RPN generates region proposals and gives a probability of each proposal by the intermediate feature maps. Fast RCNN infers the category and refines the category-specific box for each proposal.

There are many CNN architectures for the CNN feature extractor, and deeper one is better in the natural image field. However, due to the lack of data and features, the very deep architecture is not suitable for medical image. For balancing the training time and accuracy, we selected ZF net [11] as our CNN feature extractor.

2.2 LA Segmentation

After we obtain the location box of the LA, we adopt Otsu threshold on the box to generate the initialization of level set method. The intensity of pixels, which are out of the location box, set to zero. Due to the box is mainly covered the LA. Otsu threshold can provide a satisfying initial contour. It is more accurate and efficient than random and manual initialization. Then we cropped the data volume and initialization volume with the minimum adjacent rectangle cuboid of the location box for LA segmentation.

Equation 3 is the proposed edge indicator g' .

$$g' = \frac{I}{I + k * |\nabla G_\sigma * I|^2}, \quad (3)$$

$$k = \delta_\epsilon(\nabla^2 I)$$

where ∇^2 is the Laplace operator. It will calculate the divergence of the image gradient. The absolute value of divergence describes the degree of divergence of the image

gradient. The positive value indicates that the gradient in this pixel is higher than the surroundings. The negative value indicates that the gradient in this pixel is lower than the surroundings. LA has an enhancement contrast than surroundings in the CT volumes. The divergence of the correct edge gradient should be positive. Moreover, the Dirac function $\delta_\epsilon(x)$ is zero when $x < 0$. Therefore, g' will be trending to zero at the bright edge region with positive divergence; it will be trending to one at the flatten and the dark edge region with negative divergence. Hence, our proposed level set method with the new edge indicator can obtain a more accurate for LA segmentation. In addition, the efficiency of it is verified in the next section.

3. Results

3.1 Dataset

Our framework is trained and evaluated on the 2013 LA Segmentation Challenge [12] at MICCAI2013. This challenge released 30 pair CT and MRI volumes. Our framework is based on the 30 CT volumes. These volumes have the same transverse scope as (512, 512). The sagittal slices of each volume are different. We divided these volumes into 25 training volumes and 5 testing volumes for both location and segmentation stage. Notable, the location stage is processing on the 2D image level and the segmentation is performing on the 3D volume level.

For fine-tuning the Faster RCNN, we converted these volumes into 2D images according to the sagittal and the images are turned to pseudo-color in order to adopt the pre-trained parameters from nature dataset. Due to this is a segmentation challenge, there is no ready-made label for location. The label of Faster RCNN is obtained through the minimum adjacent rectangle of the segmentation label. During the training of the location model, the validation data is randomly split from the training data.

The metrics of location and segmentation are corresponding to IoU and Dice score. These metrics are both comparing the overlap between prediction and ground truth. In addition, the result will be better with a higher score. The calculating equations are shown below.

$$\begin{aligned} IoU &= \frac{1}{N} \sum \frac{L_i \cap G_{L_i}}{L_i \cup G_{L_i}} \\ Dice &= \frac{1}{N} \sum \frac{2 * |S_i \cap G_{S_i}|}{|S_i| + |G_{S_i}|} \end{aligned} \quad (4)$$

where N is the slices of each volume, L_i, S_i are corresponding to the result of location and segmentation for i^{th} slice. G_{L_i}, G_{S_i} are corresponding to the ground truth of location and segmentation for i^{th} slice.

3.2 Experiments

For selecting a correct LA location model, we evaluated three CNN feature extractors on time costing and IoU. The comparing result is shown in Table 1. ZF has five convolutional layers and two pooling layers. VGG16 is deeper than ZF, with thirteen convolutional layers and four pooling layers. VGG_M_1024 is a shortcut network of VGG16. Its structure is similar to VGG16, but with only five convolutional layers. All the feature extractors were pre-trained on VOC2007 and fine-tuned on the LA location dataset. From Table 1, we can observe that the performance of a deeper network is not better than the shallow network in this case. ZF network got a good balance on time spent and accuracy. Thus, we adopted the ZF as the CNN feature extractor to further tuning.

The whole location model was implemented on Caffe at NVIDIA Titan X GPU. The fine-tune parameters set as following. The initial learning rate was 1e-3 (descend 0.1 per 10,000 iterations), the momentum 0.9, the batch size 128 and the maximum iteration number 100,000, respectively.

Table 1. Comparison of three CNN feature extractors for LA location.

| CNN | Training Time (hour) | IoU |
|------------|----------------------|--------|
| ZF | 5.2 | 0.9041 |
| VGG_M_1024 | 2 | 0.840 |
| VGG16 | 12 | 0.7733 |

The final mean IoU on the five testing volumes is 0.5979. The recall and accuracy are 74% and 82.34%, respectively. The location model can detect the most of LA cross the whole slices of each volume. The main error rate is on the small object, such as the base and apex of LA. Even though the small object is detected, but the location box is too bigger than its real size.

The location box removed the most of irrelevant background and reduced the ratio of surrounding tissues, which have the similarity intensity of LA. We adopted Otsu on the box to automatic generating the initial contour for the level set model. Thus, the initial contour is very close to the boundary of LA. At last, we cropped the data volumes and initial volumes according to the minimum adjacent cuboid of LA location for the 3D level set model.

We compared the segmentation result between the original DRLSE and the proposed method based on the same initialization. The result is shown in Figure 3. The blue, red, and green contours are the ground truth, the segmentation results of the proposed method and the original DRLSE, respectively. The proposed method can provide a smoother contour than original DRLSE. However, it still has the shortage of distinguishing the boundary of LA and LV.

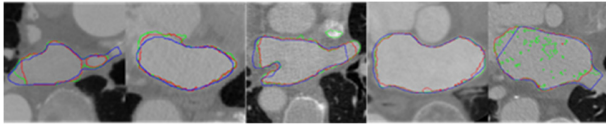


Figure 3. Examples of the segmentation results on five testing volumes. The blue, red, and green contours are corresponding to the ground truth, the results of the proposed method, and the original DRLSE.

Table 2 represents the final Dice score of the two models across the five testing volumes. Both of the models are based on the same initialization, which is generating through Otsu on the location boxes. The proposed method performed better than the original DRLSE on the most volumes. The average Dice score has a 2.72% improvement than DRLSE. This indicates the efficiency of our proposed edge indicator.

Table 2. The Dice score of the five testing volumes among DRLSE and the proposed level set model based on the same initialization.

| Volume ID | DRLSE(%) | Proposed(%) |
|-----------|----------|-------------|
| #1 | 90.17 | 88.39 |
| #2 | 79.69 | 81.45 |
| #3 | 71.20 | 84.96 |
| #4 | 89.81 | 87.49 |
| #5 | 87.84 | 89.99 |
| average | 83.74 | 86.46 |

4. Conclusions

In this paper, we proposed a framework for LA segmentation on CT. The framework was evaluated on the 2013 Left atrial segmentation challenge. It combined the Faster RCNN location model and a 3D level set segmentation model. At first, we fine-tuned Faster RCNN on the location dataset, which is generated from the segmentation dataset. Then we adopted Otsu to get the initialization for the level set model. The input of the level set is cropped from the original volumes according to the location model. The 3D level set model is based on DRLSE with a new edge indicator. The edge indicator combined both numerical and gradient information of data gradient. At last, the proposed framework obtained an 86.46% Dice score for LA segmentation on CT volumes. Comparing to the original DRLSE, our proposed method improved 2.72% on Dice score. Although the proposed level set method can guide the contour to evolve into the correct boundaries, it is still sensitive with the initialization. Moreover, this is a semi-automatic framework. In the future, a fully automatic and robust segmentation method will be studied.

Acknowledgments

This work was supported by the National Key R&D Program of China under Grant 2017YFC0113000.

References

- [1] M. Haissaguerre, P. Jais, DC. Shah, and et al., "Spontaneous initiation of atrial fibrillation by ectopic beats originating in the pulmonary veins," *N. Engl. J. Med.*, vol. 339, no. 10, pp. 659-666, Sep. 1998.
- [2] C. McGann, N. Akoum, A. Patel, E. Kholmovski, and et al., "Atrial fibrillation ablation outcome is predicted by left atrial remodeling on MRI," *Circ. Arrhythmia Elec.*, vol. 7, no. 1, pp. 23-30, Feb. 2014.
- [3] G. Luo, S. Dong, K. Wang, and et al., "Multi-views fusion CNN for left ventricular volumes estimation on cardiac MR images," *IEEE Trans. on Biomed. Eng.*, vol. 65, no. 9, pp. 1924-1934, Sept. 2018.
- [4] M. Ammar, S. Mahmoudi, M. A. Chikh, and A. Abbou, "Toward an automatic left atrium localization based on shape descriptors and prior knowledge," *LNCS*, vol. 8330, pp. 42-48, Sep. 2014.
- [5] S. Jia, L. Cadour, H. Cochet, and M. Sermesant, "STACOM-SLAWT Challenge: Left atrial wall segmentation and thickness measurement using region growing and marker-controlled geodesic active contour," in *Proc. STACOM*, Springer, 2016, pp. 211-219.
- [6] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal.*, vol. 39, no. 9, pp. 1137-1149, June 2017.
- [7] C. Liu, W. Liu, and W. Xing, "A weighted edge-based level set method based on multi-local statistical information for noisy image segmentation," *J. Vis. Commun. Image R.* vol. 59, pp. 89-107, Feb. 2019.
- [8] C. Ma, G. Luo, and K. Wang, "Concatenated and connected random forests with multiscale patch driven active contour model for automated brain tumor segmentation of MR images," *IEEE Trans. on Med. Image*, vol. 37, no. 8, pp. 1943-1954, Aug. 2018.
- [9] C. Li, C. Xu, C. Gui, and M. D. Fox, "Distance regularized level set evolution and its application to image segmentation," *IEEE Trans. on Image Processing*, vol. 19, no. 12, pp. 3243-3254, Dec. 2010.
- [10] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. on SMS*, vol. 9, no. 1, pp. 62-66, Jan. 1979.
- [11] M.D. Zeiler, R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. 13th Euro. Conf. Comput. Vis*, Springer, 2014, pp. 818-833.
- [12] C. Tobon-Gomez, A.J. Geers, J. Peters, and et al., "Benchmark for algorithms segmenting the left atrium from 3D CT and MRI datasets," *IEEE Trans. on Med. Image*, vol. 34, no. 7, pp. 1460-1473, July 2015.

Address for correspondence:

Kuanquan Wang.
Mailbox 332, Harbin Institute of Technology, Harbin 150001, China.
wangkq@hit.edu.cn