This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2794552, IEEE Journal of Biomedical and Health Informatics

Left Atrial Appendage Segmentation Using Fully Convolutional Neural Networks and Modified Three-dimensional Conditional Random Fields

Cheng Jin, Jianjiang Feng*, *Member, IEEE*, Lei Wang, Heng Yu, Jiang Liu, Jiwen Lu, *Senior Member, IEEE* and Jie Zhou, *Senior Member, IEEE*

***Abstract*—Thrombosis has become a global disease threatening human health. The left atrial appendage (LAA) is a major source of thrombosis in patients with atrial fibrillation (AF). Positive correlation exists between LAA volume and AF risk. LAA morphology has been suggested to influence thromboembolic risk in AF patients and to help predict thromboembolic events in low-risk patient groups. Automatic segmentation of LAA can greatly help physicians diagnose AF. In consideration of the large anatomical variations of the LAA, we proposed a robust method for automatic LAA segmentation on computed tomographic angiography (CTA) data using fully convolutional neural networks (FCNs) with three-dimensional (3-D) conditional random fields (CRFs). After manual localization of ROI of LAA, we adopted the FCN in natural image segmentation and transfered their learned models by fine-tuning the networks to segment each two-dimensional (2-D) LAA slice. Subsequently we used a modified dense 3-D CRF that accounts for the 3-D spatial information and larger contextual information to refine the segmentations of all slices. Our method was evaluated on 150 sets of CTA data using five-fold cross-validation. Compared with manual annotation, we obtained a mean dice overlap of 94.76% and a mean volume overlap of 91.10% with a computation time of less than 40 s per volume. Experimental results demonstrated the robustness of our method in dealing with large anatomical variations and computational efficiency for adoption in daily clinical routine.**

***Index Terms*—deep learning, fully convolutional neural networks (FCNs), left atrial appendage (LAA), LAA closure surgery segmentation, 3-D conditional random field (CRF).**

I. Introduction

LEFT atrial appendage (LAA) is a finger-like extension from the main body of the left atrium (LA) with unique embryonic origin, anatomic structure and physiological function [1]. It has the functions of active contraction and secretion, which are of great significance to relieve the pressure of the LA and to ensure the filling of the left ventricle (LV) (Fig. 1 a) [2], [3]. In cardiovascular disease, the LAA is the main body of thrombosis because of its special anatomical and functional features, and these thrombi mainly attach to the ridge between the LAA and the left superior pulmonary vein (LPSV) (Fig. 2) [4], [5]. Nearly 90% of the thrombi are responsible for stroke [6] in non-valvular atrial fibrillation (AF).

In virtue of LAA occlusion [7], [8] and catheter radiofrequency ablation surgeries [9], [10], we can prevent and treat these diseases. In LAA occlusion surgery, the closer is implanted in the neck of the LAA to block blood flow between the LA and the LAA for minimizing the risk of thrombosis. The characteristics of the LAA neck must be accurately obtained before surgery (Fig. 1 c). In catheter radiofrequency ablation surgery, the catheter is led through the vein or arterial blood vessels into the specific parts where the release of radiofrequency currents lead to local endocardium and subendocardial myasthenia necrosis, thus blocking the abnormal conduction beam and origin causing rapid arrhythmia abnormality. The specific parts mainly scatter on the wall of the LA or LAA, which must be mapped accurately in this interventional technology (Fig. 1 d). To sum up, it is important for physicians to have a precise grasp of the surface morphology and structure of LAA for these surgeries, such as the following:

- The distance between the left superior pulmonary vein (LSPV) and the LAA body.
- The long and short diameters of the intersecting surface of the LAA neck, the thickness and length of the LAA neck, an extension direction of the LAA neck body.
- The number, shape and size of the LAA lobes.

The vigorous development of modern medical imaging provides us many favorable conditions for observation and study of the LAA. For instance, computed tomography (CT) is currently one of the most advanced diagnostic technologies for cardiovascular diseases. Analysis of CT data is well recognized to be greatly useful for physicians to have a grasp of the precise anatomic structure of the LAA in advance for the treatment of thrombosis (Fig. 1 b). However, manual segmentation of the LAA based on CT data is slow and time-consuming, and interpretation greatly vary among expert users. Thus an approach for accurate automated segmentation of the LAA should be proposed.

After manual localization of ROI of LAA, we proposed a robust method for automatic LAA segmentation on computed tomographic angiography (CTA) data using fully convolutional

Manuscript received August 08, 2017; revised November 20, 2017; accepted January 07, 2018. This work is supported by the National Natural Science Foundation of China under Grants 61622207. Asterisk indicates corresponding author.

C. Jin is with the Department of Automation, Tsinghua University, State Key Lab of Intelligent Technologies and Systems, and Tsinghua National Laboratory for Information Science and Technology (TNList), and Qiqihar Medical University, China. (e-mail: jin-c12@mails.tsinghua.edu.cn).

The other authors are with the Department of Automation, Tsinghua University, State Key Lab of Intelligent Technologies and Systems, and Tsinghua National Laboratory for Information Science and Technology (TNList), Beijing 100084, China (e-mail: {jfeng, jzhou and lujiwen}@mail.tsinghua.edu.cn. {w-l14, h-yu14 and liujiang15}@mails.tsinghua.edu.cn.)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2794552, IEEE Journal of Biomedical and Health Informatics

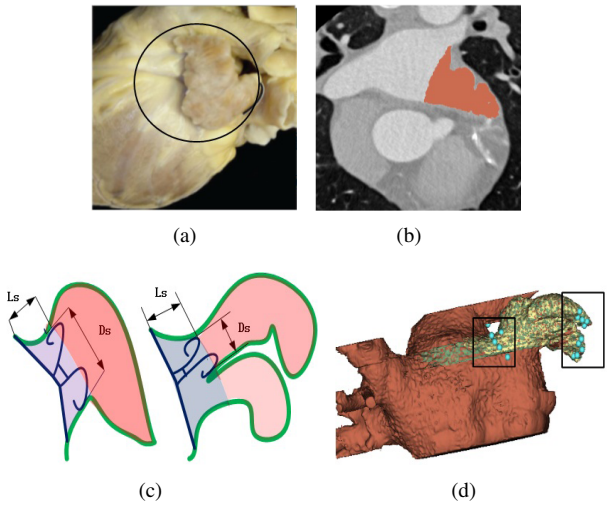


Fig. 1. LAA. (a) The LAA marked by the circle. (b) An axial CT slice of the LAA marked by orange. (c) LAA occlusion, the lobes are marked light red, the neck is marked light blue, and the closure is marked navy blue. Ls refers to the length of the sheath, Ds refers to the diameter of the sheath. (d) Catheter ablation, the dots in the black boxes represent sites of delivery of radio frequency energy.

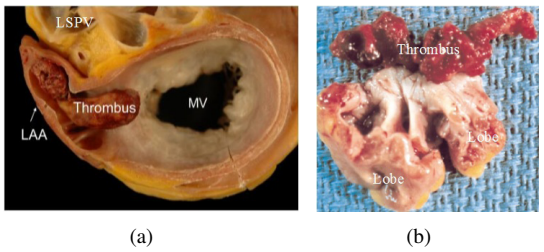


Fig. 2. LAA thrombi. (a) Adjacent tissues [11]. (b) Thrombi of the removed appendages.

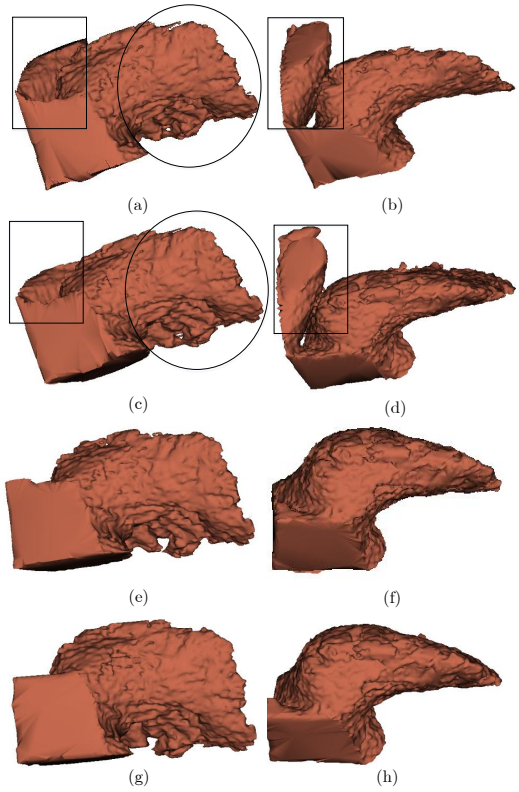


Fig. 3. The 3-D models (shown in two different viewpoints) of three LAA segmentation algorithms and manual segmentation for a dataset, the LSPV is marked by the rectangle, and the lobes is marked by the circle. (a), (b) 3-D region-growing approach. (c), (d) 3-D maximally connected component approach. (e), (f) Our proposed approach. (g), (h) Manual segmentation.

neural networks (FCNs) and three-dimensional (3-D) conditional random fields (CRFs). We adopted the extant FCN in natural image segmentation and transfered their learned models by fine-tuning the networks to segment each 2-D LAA slice. Subsequently we used a dense 3-D CRF that accounts for the 3-D spatial information and larger contextual information to refine the segmentations of all slices. Our method was evaluated on 150 sets of CTA data by using five-fold cross-validation. Referring to manual annotation, we obtained a mean dice overlap of 94.76% and a mean volume overlap of 91.10% with computation time less than 35 s per volume. Experimental results demonstrated the robustness of our method when dealing with large anatomical variations and computational efficiency for adoption in daily clinical routine.

In the next section, we review related work on LAA segmentation and recent approaches to the segmentation of medical images using deep neural networks.

II. Related work

Owing to its small yet highly variable structure, LAA segmentation is a challenging job using CT data (Fig. 1), and few people have focused on this problem. Grasland-Mongrain *et al*. [12], [13] started from the segmentation of the LA [14], and then grew the LAA out by shape-constrained and inflation deformable models. It has some weakness in segmenting the complex tip of the LAA [12], [13] due to the constraints arising from the shape model. Zheng *et al*. [15] exploited a multi-part model to segment the whole left atrium (LA) including the LAA on C-arm CT. They used only a smooth mesh to enclose all LAA lobes roughly, mainly because the LAA has many small curly lobes but their boundaries are ordinarily not clear on C-arm CT.

The LAA has complex tips and lobes, and its precise segmentation is important for modeling LAA morphology. Therefore, non-model-based approaches seem to be suitable for this task since they are mainly driven by the image to be segmented without explicit shape constraints. We have applied two widely used non-model-based approaches, 3-D region growing [16] and 3-D maximally connected component [17], to segment the LAA in the bounding box (which will be explained in Section 3.1), but the segmentation results have the following problems (Fig. 3):

- Incorrectly including the left superior pulmonary vein (LSPV) which is adjacent to the LAA.
- The crimping or adhesion of the LAA lobes.

Graph cut [18] is an inference approach used in MRF models. It is a widely used segmentation approach that is suitable for cases with larger differences between the target and the background in the medical images. However this method is laborious because it requires the user to label the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2794552, IEEE Journal of Biomedical and Health Informatics

foreground and background seeds. In addition the parameter λ in graph cut has a great impact on the final result, and it is a challenging task to determine an optimal λ value. In fact, a single λ value cannot guarantee good results for all image conditions. Besides, it cannot avoid susceptibility to noise. In fact the parameter λ can be automatically learned with structured output learning methods. Instead of finding the optimal λ value, some work [9,10] also generated a pool of proposals using parametric max-flow/min-cut solver [11]. Parametric max-flow can solve max-flow/min-cut problems with a set of λ values while max-flow in graph cut just solves the problem with a single λ value. Our group has [19] proposed an LAA segmentation algorithm based on parametric graph cut, and achieve a good segmentation result. However, the running time of this approach is long because it needs to solve a number of 2-D graph-cut problems to generate hundreds of segmentation proposals for each slice of LAA.

In recent years, the Convolutional Neural Networks (CNNs) have also been adopted to the problem of medical image analysis. However, for the application of medical images, the primary bottleneck is the lack of data. Training data of CNNs needs to have precise class labels, but manually labelling volume data with the scale of Imagenet [20] is unimaginable. The results using the traditional method (CNNs) have some problems as follows [21]:

1) reduced feature resolution (repeated combination of max-pooling and down-sampling);
2) existence of objects at multiple scales;
3) reduced localization accuracy due to CNN invariance.

Thus, according to these problems, the DeepLab [22] has the following optimization:

1) An atrous convolution is created to represent unsampling, which can be either enlarge parameters or computational quantities to maintain the size and scale of the image;
2) The image is put into the same size forcibly by the traditional method. However, this can cause some features to distort or disappear. Atrous spatial pyramid pooling (ASPP) is used to solve this problem motivated by spatial pyramid pooling, which is used to remove the constraints of the fixed size of the network;
3) To solve the invariance of DCNN, some methods such as skip layers [21] can be adopted, but the author uses a fully connected conditional random field (CRF). To a certain extent, the feature map obtained with CNN is smooth enough, so short-range CRF does not work much. Thus the fully connected CRF with global information is used. This is the first reason why we choose full connection CRF as postprocessing.

DeepLab uses the results of FCN as a unary potential function of CRF. It concludes coarse segmentation in front and fine segmentation in the post. Shuai Zheng $et\ al$ [23] also introduced a new form of CNN that combines the strengths of CNN and CRF-based probabilistic graphical modeling. They formulate mean-field approximate inference for the CRF with Gaussian pairwise potentials as recurrent neural networks, Their system fully integrates CRF modelling with CNN, making it possible to train the whole deep network end to end with the usual back-propagation algorithm, avoiding off line postprocessing methods for object delineation.

Designing a strong feature representation is conclusive for semantic segmentation. We believe it is advantageous that fine segmentation is only located at a later stage, avoiding the commitment to premature decisions. This is the second reason why we choose full connection CRF as postprocessing with the abandonment of CRF as the RNN network.

In medical image analysis, Huazhu Fu $et\ al$ [24] used the FCN + CRF method for retinal vessel segmentation and obtained good results. Neeraj Dhungel $et\ al$ [25] explored DCNN and deep belief networks (DBNs) for the segmentation of breast masses. The structured prediction models present the results of two methods the CRF and structured support vector machine (SSVM) -based method. In the CRF, Neeraj Dhungel $et\ al$ adopt the inference algorithm based on tree re-weighted belief propagation with truncated fitting training. In the SSVM, the inference is based on graph cut with maximum margin training. They also show that the CRF model is signicantly faster than SSVM in terms of training and inference time when combined with deep learning potential functions. Kamnitsas $et\ al.$ [26] presented a multi-scale 3-D CNN with fully connected CRF for accurate brain lesion segmentation. Fully 3-D CNNs produce accurate soft segmentation maps, but induce an increased number of parameters and significant memory and computational requirements. One of the reasons that discourage the use of 3-D CNNs is the training procedure, which requires massive amounts of memory and training data to store the model and allow for good convergence and generalization respectively. Combining segmentation of the liver and its lesions, Christ $et\ al.$ [27] trained and cascaded two fully convolutional neural networks (FCNs). First, they trained using an FCN to segment the liver as region of interest (ROI) input to the next FCN. Second, the FCN then segments the lesions from the predicted ROI of step 1. Furthermore they refine the segmentations of the cascaded fully convolutional neural networks (CFCNs) using a dense 3-D CRF that accounts for both spatial coherence and appearance. This paper also suggests that volumetric FCN actualization with 3-D convolutions is intensively constraint by GPU hardware and available video random access memory (VRAM). In addition, the spatial resolution of each direction of CT volumes (e.g. 0.32-0.38 mm in x, y and 0.35-0.43 mm in z in our cardiac CT volumes) makes it more cumbersome to train discriminative 3-D filters.

We make some improvements to the related methods which produce much better results. To prevent "over fitting" of many parameters in FCN, image enhancement is needful to rely on large-scale, high-quality training data. Thus we first use multi-scale retinex with color restoration (MSRCR) [28], [29] to enhance the CT image. Then we transform a single-channel grayscale image into a 3-channel RGB image, which improves the resolution of extracting local features (boundary, orientation, etc.) by increasing the number of convolutions [30]. Special discussion is in Section III-A. We have also tried to address two deficiencies of fully connected CRF: First, some deviation remains in the pre-set weight of Gaussian function;

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2794552, IEEE Journal of Biomedical and Health Informatics

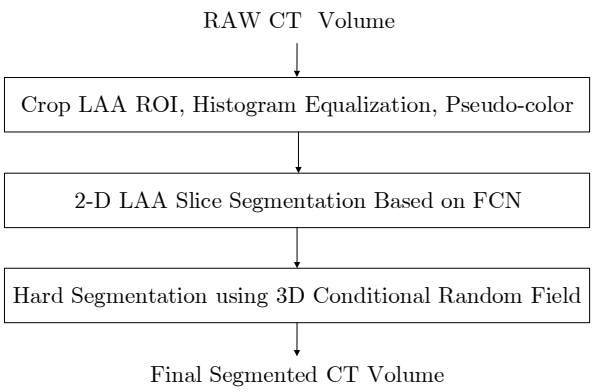


Fig. 4. Overview of the proposed LAA segmentation pipeline.

Second, since the measure function uses only the distance size without the directional information in space, different directions may also lead to different degrees of interaction. Therefore, we define a new and simple class-adaptive smooth factor β. Special discussion is in Section III-C. Factor β is automatically calculated from the classification probabilities of FCN and is anisotropic for each class. The experimental results show that the adaptive algorithm can effectively segment noisy images and image detail information is revealed.

III. Method

In this paper, we present a method for automatic LAA segmentation on CTA data using FCN and 3-D CRFs. To segment each 2-D LAA slice (the reason why we prefer slices to volume will be explained in Section 3.1), we first adopted the extant FCN in natural scene segmentation and transfered their learned models by fine-tuning the networks to our task. After FCN generated soft probability maps, we used 3-D CRFs which capture the spatial information across slices to produce the final hard segmentation labels. The entire framework is depicted in Fig. 4 and the details of our method will be described in the following sections.

A. Preprocessing

The LAA is a little portion of the whole CT volume (shown in Fig. 5 a), typically with size of $100\times100\times60$ while the size of the whole CT volume is $512 \times 512 \times 300$. Training FCN with the whole CT volume is quite difficult because of the limited VRAM [31] and the extremely high class imbalance, even if additional weighting factors are introduced [21]. Thus our method would only need to ensure four fiducial points to obtain a bounding box to encase the LAA. All the algorithms proposed in this paper will be carried out inside this ROI. As shown in Fig. 5 b, c, d, one fiducial point is on the initial axial slice where the LAA emerges, two correspond to diagonal corners of the rectangle which encloses the LAA closely on the mid-axial slice in which the LAA area is normally the largest, and one is on the last axial slice where the LAA disappears. Then we increased the volume contrast by histogram equalization.

Data preprocessing is essential for the invariance and robustness properties required to the network when only a few

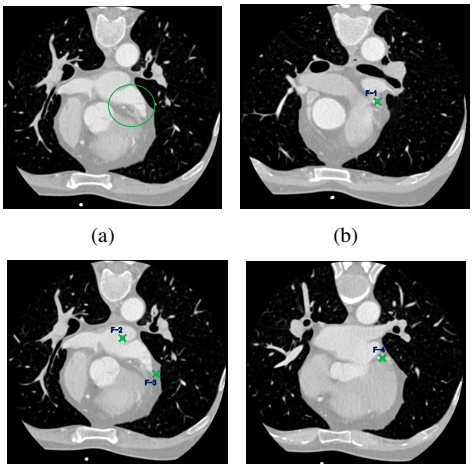


Fig. 5. The ROI of LAA defined by four fiducial points. (a) The LAA marked by the circle. (b) One fiducial point on the initial axial slice where the LAA emerges. (c) Two correspond to diagonal corners of the rectangle which encloses the LAA closely on the mid-axial slice in which the LAA area is normally the largest. (d) One fiducial point on the end of axial slice where the LAA will soon be gone.

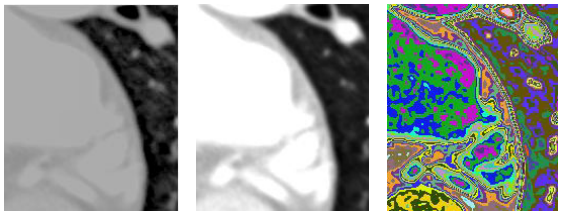
TABLE I
TRANSFORMATION FROM GRAY LEVEL TO COLOR

Input gray level	Output color	R	G	B
$0 \sim 31$	blue	0	0	255
$32 \sim 63$	green	0	255	0
$64 \sim 95$	light blue	193	210	240
$96 \sim 127$	purple	141	75	18
$128 \sim 159$	red	255	0	0
$160 \sim 191$	orange	250	140	53
$160 \sim 191$	yellow	255	255	0
$160 \sim 191$	light yellow	227	207	89

training samples are available. It is necessary to rely on large-scale, high-quality training data to prevent "over fitting" of many parameters in CNNs. Thus we first used a method called multi-scale retinex with color restoration (MSRCR) [28], [29] to enhance the CT image (Fig. 6 b). Then we converted the gray-level images to pseudo color images using the MATLAB function $label2rgb$ with color palette, its transformation shown in Table I. The purpose was to transform a single-channel grayscale image into a 3-channel RGB image (Fig. 6 c), which improves the resolution of extracting local features (boundary, orientation, etc.) by increasing the number of convolutions [30]. Specifically in Fig. 7, the first layer convolution filters of CNNs are visualized. We notice that somewhat blurry filters are learned with random initialization from the original CT images (Fig. 7 a). However, many higher orders of contrast or boundary-preserving patterns (that enable capturing image appearance details) are evidently learned through the image enhancement (Fig. 7 b), which significantly improves the performance of our segmentation algorithm.

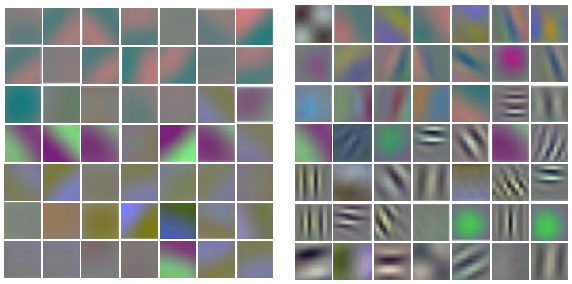
It is generally well known that successful training of deep networks requires a huge number of annotated samples. However for medical tasks, very few volume data are available to train a powerful deep network. To obtain sufficient training data, we would utilize all 2-D axial slices instead of volumes to train FCN, with more than 5000 slices available from our

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2794552, IEEE Journal of Biomedical and Health Informatics



(a) (b) (c)

Fig. 6. An example of image enhancement for an axis LAA CT slice. The details of the dark area can be effectively reflected. (a) Original. (b) Multiscale retinex. (c) Multi-channel.



(a) (b)

Fig. 7. Visualization of the first-layer convolution filters of CNNs with random initialization from original CT images and image enhancement. Many higher orders of contrast or boundary-preserving patterns are evidently learned through image enhancement. (a) Learned from original images. (b) Learned from the images enhancement.

data set. The axial slices were selected because they are not suitable for the study of LAA morphology for the other sagittal and coronal slices [32]; meanwhile the number of axial planes is relatively fewer.

B. 2-D FCN for soft probability maps

For each 2-D axial LAA slice, we first adopted the FCN to generate a soft probability map. The network architecture is shown in Fig. 8, which is adopted from the VGG 16-layer net. The VGG 16-layer net consists of five convolution groups with each group containing 2, 2, 3, 3, and 3 layers respectively and three fully connected layers. Each convolution layer is followed by a ReLU activation function and the max pooling layer is after each convolution group. The convolution layer is the primary building block to construct a CNN architecture. By stacking several layers on top of each other, we can extract a hierarchy of features. Each convolution layer can extract features into the hierarchy from its previous layer to the next layer connected. A single convolution layer takes input as a stack of input channels and generates output as several output channels, which we call feature maps. A feature map can be considered as a topologically arranged map, which is the response of a specific local nonlinear feature extractor identically applied to the spatial neighborhood of the input channels in a sliding window fashion [33]. Suppose each layer $l \in [1, L]$ contains C_l feature maps and the neurons of the mth feature map $\mathbf{FM}_l^m$ in the lth layer are arranged in a 2-D array, then the feature map is constituted as follows:

$$\mathbf{FM}_l^m = f\Big(\sum_{n=1}^{C_{l-1}} \mathbf{k}_l^{m,n} * \mathbf{FM}_{l-1}^n + b_l^m\Big) \,, \tag{1}$$

In Eq. (1), the feature map is computed by convolving each of the feature maps of the previous layer with a 2-D convolution kernel $\mathbf{k}_l^{m,n}$, adding a learned bias b_l^m and applying to a non-linear function f. The feature map $\mathbf{FM}_0^m$, input to the first layer, corresponds to the channels of the input images. All kernels are matrices of learned hidden weights $\mathbf{W}$. In addition, Eq. (1) can be viewed as the convolution of the 3-D concatenated kernel $\mathbf{k}_l^m = (\mathbf{k}_l^{m,1}, ..., \mathbf{k}_l^{m,C_{l-1}})$ and the concatenated former channels $\mathbf{FM}_{l-1} = (\mathbf{FM}_{l-1}^1, ..., \mathbf{FM}_{l-1}^{C_{l-1}})$. This intuitively expresses that the neurons of higher layers in the hierarchy combine the features extracted in preceding layers. Following the convolution operation, an element-wise non-linearity is applied to obtain features that are nonlinear transformations of input. In the VGG 16-layer net, ReLU activation is adopted as a nonlinear function f after each single convolution layer. After each convolution group, max pooling operation is done by taking the maximum neuron response over sub-windows within each feature map. This can be defined as follows:

$$\mathbf{FM}_{i,j}^m = \max_p \mathbf{FM}_{i+p,j+p}^m \,, \tag{2}$$

where p determines the size of the max pooling window. The pooling window can be overlapped or not. After the max-pooling operation, the size of the feature map is shrunk. This is decided by the pooling size p and the stride parameter s, which correspond to the horizontal and vertical increments by which the pooling window is arranged. This subsampling procedure is proved beneficial by introducing local translation invariance. The VGG 16-layer net has three fully connected layers and the output $\mathbf{y}$ of last fully connected layer is fed into a softmax function that produces the predicted probability $p_i(\mathbf{y}) = \exp(y_i)/\sum_{j=1}^{C} \exp(y_j)$ for class i. Note that $\mathbf{y}$ is a vector with length C equaling to the total number of class.

As described in [21], our FCN is adopted from the VGG 16-layer net without the final classification layer, then turning all fully connected layers into convolutions. Besides a 1×1 convolution layer with channel dimension 2 is utilized to forecast probabilities of foreground or background at any positions of the coarse outputs, we upsample the coarse outputs to a pixel-dense probability map by a deconvolution layer. The upsampling can be carried out in-network for end-to-end learning by back propagation by using the pixel-wise loss. This is the FCN-32s architecture with the 32 pixel stride at the final output layer. Our model is pre-trained on contemporary VGG network, and then we fine-tune their learned representations on the FCN-32s by using our own data set. Usually FCN-32s can maintain the primary contours and structures but limit the scale of detail in the prediction. To address this problem, skips [34] are added that combine final output layer with former layers that have finer strides. As shown in Fig. 8, this operation turns into a DAG with edges that skip from lower layers to higher layers. By combining fine and coarse layers, we allow the local prediction of network while respecting the global structure. We add a 1×1 convolution layer after

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2794552, IEEE Journal of Biomedical and Health Informatics

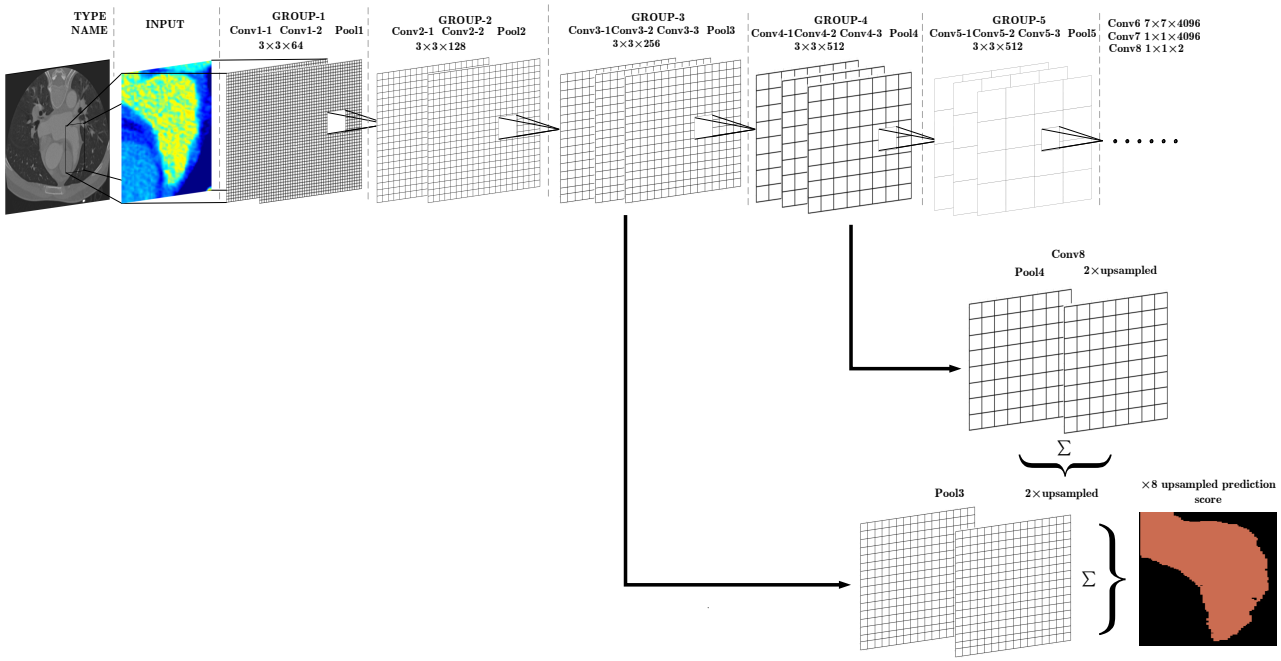


Fig. 8. Details of the LAA FCN configuration. It combines abstract, high-level semantic features with appearance, low-level spatial information. Each layer is displayed in the grids of the corresponding density. Row 1. FCN-32s. Row 2. FCN-16s. Row 3. FCN-8s.

pool4 layer to generate extra class predictions. By adding these predictions with the output computed on top of conv7 which is calculated by connecting a $2\times$ upsampling layer, we first divide the output stride of 16 pixels to predict the segmentation. The $2\times$ upsampling is initialized as bilinear interpolation, but the parameters are allowed to be learned by back propagation. Finally we upsample the predictions of stride 16 back to the image size. This net is the FCN-16s. FCN-16s is initialized with the parameters of the former FCN-32s and then learns end to end. We zero-initialize the new parameters acting on pool4 in order that the network starts with the same predictions as FCN-32s. We continue in this fashion by leveraging predictions from the pool3 layer which capture finer information to build the FCN-8s net. This is done by linking the pool3 layer to the DAG net in the same way as the linking of pool4 above. In our experiment, adding more lower layers cannot improve the segmentation accuracy any further. Thus we do not keep fusing even lower layers. The FCN-8s network is used as our final network because it takes advantage of coarse, high-level information and fine, low-level information.

The loss function is calculated by a pixel-wise softmax over the final output channels combined with the cross entropy. With softmax defined as:

$$p_{i,j}^k = \exp(a_{i,j}^k) / \sum_{k=1}^{K} \exp(a_{i,j}^k) \ , \tag{3}$$

where $a_{i,j}^k$ represents the activation in feature channel k at the pixel position $(i,j) \in \Omega$ with $\Omega \in \mathbb{Z}^2$. K is the total number of class and $p_{i,j}^k$ denotes the probability of pixel (i,j) belonging to the class k. Particularly, $K = 2$ in our task. Then the cross entropy loss penalizes the deviation of $p_{i,j}^k$ from ground truth label $\hat{p}_{i,j}^k$ at each position.

$$L = -\sum_{(i,j)\in\Omega} w_{i,j}[\hat{p}_{i,j}^f \log(p_{i,j}^f) + (1-\hat{p}_{i,j}^f)\log(1-p_{i,j}^f)] \ . \tag{4}$$

Note that $p_{i,j}^f$ is the foreground probability and $\hat{p}_{i,j}^f$ represents the ground truth. $w_{i,j}$ is the same weight as [21] to balance the classes.

Training The network is trained with all 2-D LAA slices and their corresponding segmentation maps using the stochastic gradient descent (SGD) implemented by MXNet [35] with GPU acceleration. First FCN-32s net is fine-tuned by transferring the VGG 16-layer learned model to our task. Then FCN-16s net and FCN-8s are trained sequentially based on 32s and 16s nets respectively. In the experiment, learning rate is chosen to be 0.0001; the momentum, 0.95; weight decay, 0.0005; and epoch, 50. Using a single Tesla K80 GPU it took 6 hours to completely train our FCN-8s net.

C. *Modified 3-D CRF for hard segmentation labels*

Volumetric FCN implementation with 3-D convolution kernels is limited by the available VRAM [31] and the extremely high class imbalance. In our data set, the class imbalance is as follows: The LAA has different complex morphologies named cactus, chicken wing, windsock, and cauliflower with volumes varying from 1 to 19 ml. We have the LAA CT data of 150 patients. The cauliflower-shaped LAA accounts for only 5% of these data, but these patients with AF more likely had prior stroke[36]. Compared with cauliflower, chicken wings accounted for 48% (more likely to remain stroke-free), cactus accounted for 30%, and windsock accounted for 17%. Furthermore, the anisotropic resolution of medical volumes (

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2794552, IEEE Journal of Biomedical and Health Informatics

e.g. the dimension of approximately 0.35 mm in xy and 0.5 mm in z in our data set) makes the training of discriminative 3-D filters complicated. Hence we decided to leverage 3-D CRF as proposed in [27] to utilize the spatial information across axial slices. Besides this, we found that 3-D CRF can capture even more global context in xy plane to correct some spurious outputs with small isolated regions or holes in the predictions that which are caused by local minima in training and noise in the input images.

Given all slice-wise soft predictions by the FCN, we consider all these 2-D probability maps jointly in the 3-D CRF framework to achieve more structured segmentation labels. This final label assignment can be formulated as maximum a posteriori (MAP) inference. The best labeling is estimated by the following equation:

$$\mathbf{x}^* = \arg\min_{\mathbf{x}\in\mathcal{L}} E(\mathbf{x}) \ , \tag{5}$$

with $\mathbf{x}$ corresponding to the labels of all voxels and $\mathcal{L}$ denoting the label set. $E(\mathbf{x})$ represents the energy function induced according to Gibbs distribution:

$$E(\mathbf{x}) = \sum_{i\in\mathcal{V}} \varphi_u(x_i) + \sum_{(i,j)\in\mathcal{E}} \varphi_p(x_i, x_j) \ . \tag{6}$$

Note that we define the CRF on the complete graph $G = (\mathcal{V}, \mathcal{E})$ with vertices $i \in \mathcal{V}$ for all voxels belonging to the volume and each edge $e \in \mathcal{E} = \{(i,j),\ \forall i,j \in \mathcal{V},\ \text{s.t. } i < j\}$ between all vertices. $\varphi_u(x_i) = -\log P(x_i|I)$ is the unary potential with $P(x_i|I)$ representing the probabilistic output of the FCN. The pairwise potential between any pair of voxels is of the form:

$$\begin{aligned}\varphi_p(x_i, x_j) = \mu(x_i, x_j)[w_1 \exp(-\sum_{d=1}^{3} \frac{|x_{i,d} - x_{j,d}|^2}{2\sigma_1^2}) \\ +w_2 \exp(-\sum_{d=1}^{3} \frac{|x_{i,d} - x_{j,d}|^2}{2\sigma_2^2} - \sum_{c=1}^{C} \frac{|I_{i,c} - I_{j,c}|^2}{2\sigma_3^2})] \ ,\end{aligned} \tag{7}$$

$\mu(x_i, x_j)$ is the Potts function

$$\mu(x_i, x_j) = \begin{cases} 1 & \text{if } x_i \neq x_j \\ 0 & \text{if } x_i = x_j, \end{cases} \tag{8}$$

$|x_{i,d} - x_{j,d}|$ is the spatial distance of voxels i and j and $|I_{i,c} - I_{j,c}|$ is the intensity difference of them in the original volume. The configurable weights w_1 and w_2 can define the relative strength of the two potentials and we tune their effective ranges by the kernel size σ_1, σ_2 and σ_3. The generalized cross-validation leads to a good selection of these parameters w and σ [37].

To adaptively and effectively segment noisy images, we introduce a smoothing factor β,

$$\beta_{k,o,i}^{(t)} = \sqrt{\frac{\sum_{i\in\mathcal{V}} \sum_{m\in N_i}^{o} \left(p_k^i - p_k^m\right)^2}{N}}, \tag{9}$$

where k, o respectively represent the class and orientations of neighborhood. We define the neighborhood of a voxel as $N_i = \{1, 2, \cdots, N\}$ where N is the number of all voxel in the neighborhood. p_k^i, p_k^m respectively represent the probabilities of a voxel i and its neighbor m which belongs to the class k. They come from the probabilistic output of the FCN. The reason for this definition of β is that the probability of voxel is the refinement of image information, and the change of image information can be observed from the difference in probability between a voxel and its neighbors that belong to a certain class. It is an indicator denoting which component is activated. It embodies the spatial context such as probability of occurrence and relative orientation between labels [38], [39]. For instance, the lobes are always in the lower right of the main body of LAA, LAA lobes occur more often than the Left Circumflex Artery (LCX) close to them in LAA segmentation (Fig. 13). Furthermore,

$$\begin{aligned}\varphi_p(x_i, x_j) = (\beta_{k,o,i} + \beta_{k,o,j})\mu(x_i, x_j) \\ [w_1 \exp(-\sum_{d=1}^{3} \frac{|x_{i,d} - x_{j,d}|^2}{2\sigma_1^2}) \\ +w_2 \exp(-\sum_{d=1}^{3} \frac{|x_{i,d} - x_{j,d}|^2}{2\sigma_2^2} \\ -\sum_{c=1}^{C} \frac{|I_{i,c} - I_{j,c}|^2}{2\sigma_3^2})].\end{aligned} \tag{10}$$

Finally Eq. 5 is solved by the efficient mean field approximation algorithm in [27]. A random search algorithm is used to choose the weights and kernels.

In summary, compared with other approachs, our approach not only pays attention to the detailed segmentation in each single slice, but also takes full advantage of the continuity of topological relationships between slices for the postprocessing based on 3-D volume.

IV. Experiments and Results

The data used in this research were 150 cases of coronary CTA, obtained using Philip 256-iCT. The experimental results are evaluated using five-fold cross-validation. The scanned information of CT contains the element spacing (0.357422 mm, 0.357422 mm, 0.449951 mm) and dim size (512×512).

A. Accuracy of the segmentation results

This paper uses the manual segmentation results as the ground truth. To avoid the observer bias, manual segmentation is carried out by 4 experienced specialists collaborating using the module for segmentation of Slicer 4.6 which is an open source software platform for medical image processing. In the Segment Editor, the following features are applied: overlapping segments, display in both the 2-D and 3-D views, editing in 3-D views, creating segmentation by interpolating or extrapolating segmentation on a few slices, and editing on slices in any orientation. To obtain an estimate of the actual error, manual and automatic segmentation are performed in the same pre-set bounding box encasing the LAA. We conducted a detailed statistical analysis of the segmentation accuracy for these cases of data by comparing automatic and manual segmentation. It includes two aspects, one is similarity analysis based on volume measurement and the other is error analysis based on surface distance measurement. In the absence of a explanation, FCN+CRF represents FCN+CRF with β in this paper.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2794552, IEEE Journal of Biomedical and Health Informatics

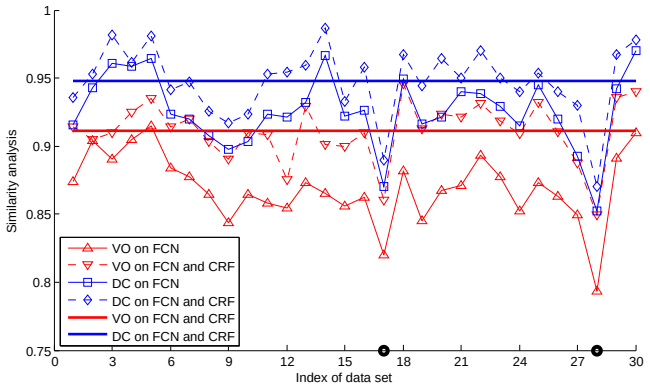


Fig. 9. Similarity analysis of volume overlap (VO) ratio and dice coefficient (DC) ratio in 30 cases of testing data.

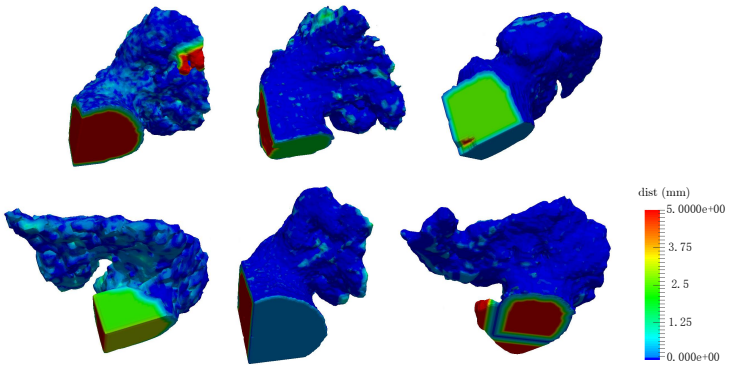


Fig. 10. The distribution of the surface distance error of the segmentation results of six different LAAs. The size of the error is proportional to the degree of color.

1) Similarity analysis: The similarity can be obtained by calculating the volume overlap (VO) ratio and dice coefficient (DC) ratio. The closer the two ratios are to 1, the closer the automatic results are to the manual. These two ratios are calculated as follows:

$$VO = \frac{|V_{seg} \bigcap V_{gt}|}{|V_{seg} \bigcup V_{gt}|}, \tag{11}$$

$$DC = \frac{2|V_{seg} \bigcap V_{gt}|}{|V_{seg}| + |V_{gt}|}, \tag{12}$$

where V_{gt} and V_{seg} denote the volume of the ground truth and the automatic segmentation respectively.

Fig. 9 lists the calculation results of the VO and DC for 30 cases of testing data, which shows that the segmentation results of LAA have a high similarity between automatic and manual segmentation. The mean value of VO equals 91.10%, and the mean value of DC equals 94.76%. In two cases DC and VO were low (the circular marks in Fig. 9) because the original data were of low density resolution. At the same time we see that the improvement effect of adding 3-D CRF postprocessing is appreciable.

2) Error analysis: Surface distance measurement is to measure the minimum distance from each point of the surface of the automatic segmentation model to those of the manual segmentation model. This approach can evaluate the accuracy of the automatic segmentation result intuitively. The smaller the measured distance, the closer the automatic results are to the manual. Compared with the DC ratio, it can better represent the accuracy and where the error occurs, such as the interference from the left circumflex artery (LCX) around the lobes of the LAA when the DC ratio is commonly not reduced. The distance is calculated by

$$d_i = \mathrm{m}in(|x_{gd} - x_{seg}|). \tag{13}$$

Fig. 10 shows some cases of error distribution about the points of surface. The segmentation errors are less than 1.30 mm in most areas. However, in some areas, the segmentation accuracy declines slightly, such as the transition zone between the left atrium (LA) and the LAA. The reason is that the boundaries selected for manual segmentation differ from those

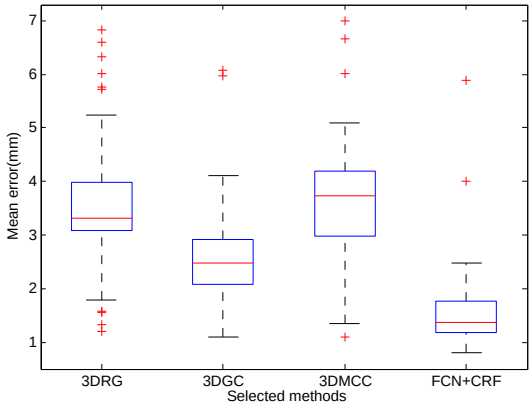
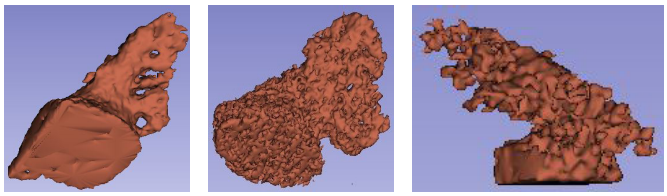


Fig. 11. The box-and-whisker plots of mean error of the LAA surface distance by each method. 3-D region-growing (3DRG), 3-D maximally connected component (3DMCC) and 3-D graph cut (3DGC) are shown.

for automatic segmentation. In addition, automatic segmentation has slight differences from manual segmentation in the areas of the LAA lobes and other parts because our approach just adds some postprocessings, the CRFs and 3-D region-growing method etc., which systematically explore morphological changes of the LAA and the topological relationship between adjacent slices are also integrated. We can pinch out or secondarily grow the corresponding tissues of the LAA but cannot do this for the manual labeling based on single visual judgments.

To verify the effect of FCN+CRF on improving the accuracy of LAA segmentation, we compared and conducted an error analysis of 3-D region growing (3DRG), 3-D maximally



(a) (b) (c)

Fig. 12. Some failed cases of the LAA segmentation.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2794552, IEEE Journal of Biomedical and Health Informatics

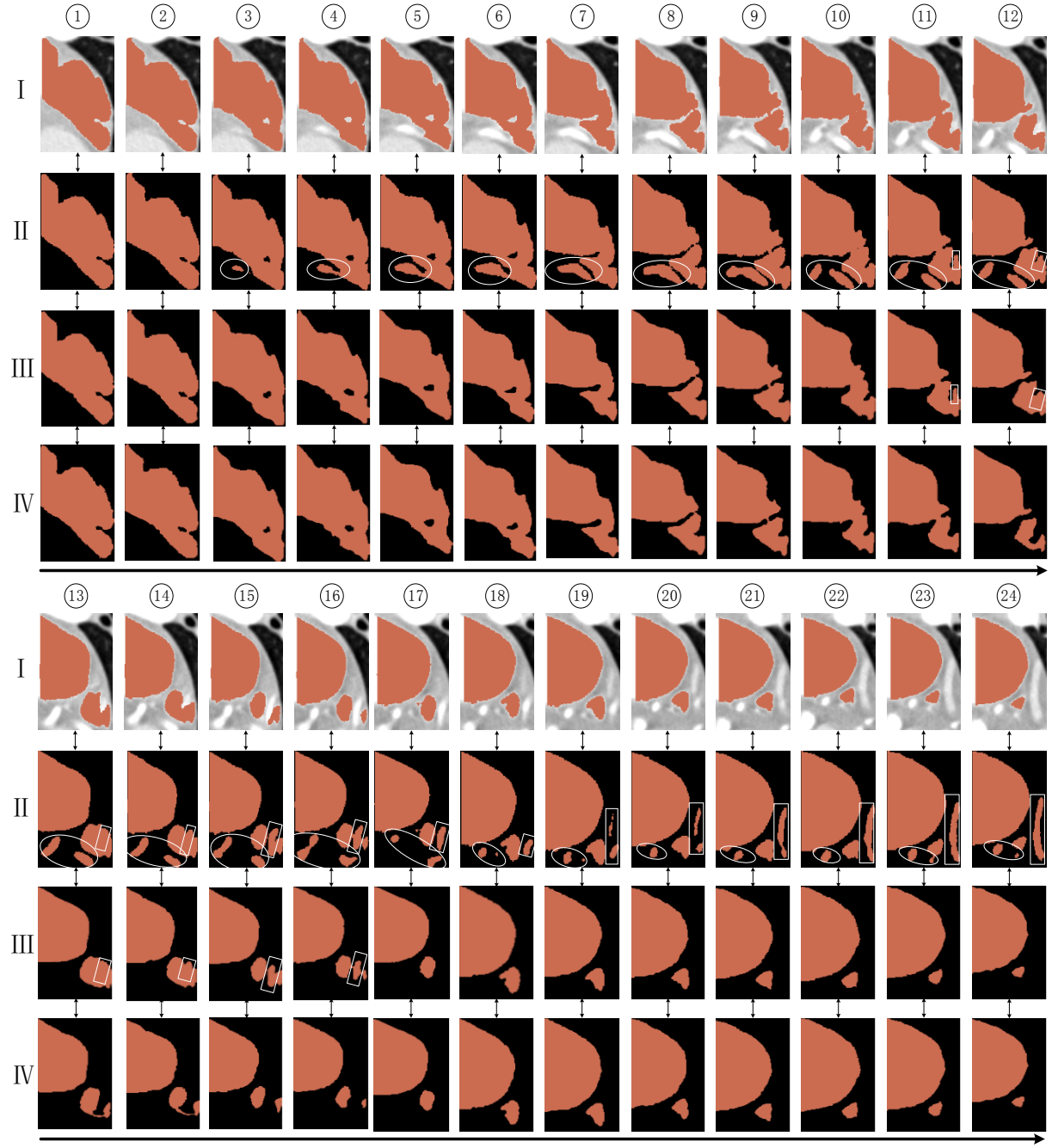


Fig. 13. The sporadic interferences of small LCXs patches located around the lobes of the LAA in 2-D slice. The LCXs are marked by the border. Row I. Manual segmentation. Row II. FCN. Row III. FCN+CRF without β. Row IV. FCN+CRF with β.

connected component (3DMCC) and 3-D graph cut (3DGC) segmentation by their surface distance in 150 cases as shown in Fig. 11. It shows that the mean error of FCN+CRF is grouped mainly between 1.1 and 1.7 mm and more concentrated than the other methods.

At the same time, by comparison with that of Wang *et al.*[19], our FCN is implemented based on the MXNet [35] with GPU acceleration. It takes about 8 hours to completely train the FCN-8s on a NVIDIA Tesla K80 GPU using cuDNN v5.0. The segmentation of a LAA volume takes less than 35 s using GPU acceleration, with about 25 s for FCN segmentaion and 15 s for 3-D CRF. Our method finally obtains a mean dice overlap of 94.76%, which is comparable with the result of 95.21% in [19]. By contrast the method proposed by [19] requires more than 3.5 minutes to completely segment a LAA because this approach needs to solve a number of 2-D CRF problems to generate hundreds of proposals for each slice of LAA. FCN+CRF is more accurate and robust in LAA segmentation, but the accuracy of the automatic segmentation algorithm is dispersive in some cases as shown in Fig. 12, which is mainly due to generating mild motion artifacts when the CT images were collected or the contrast medium is not completely filled in the LAA. These regions are also difficult to segment manually.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2794552, IEEE Journal of Biomedical and Health Informatics

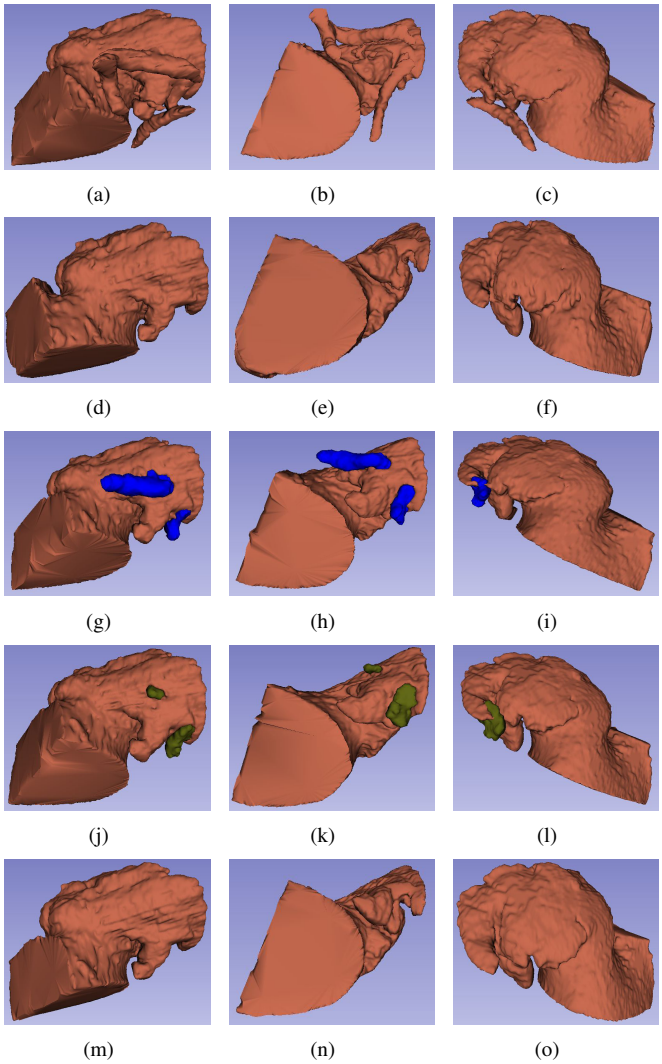


Fig. 14. Comparison of the 3D results using relevant segmentation approaches. (a), (b), (c) Initial states. (d), (e), (f) Manual segmentation. (g), (h), (i) FCN, the blue area represents the LCX. (j), (k), (l) FCN+CRF without β, the green area represents the LCX. (m), (n), (o) FCN+CRF with β.

B. Representative examples

More examples are given in this section to show the advantage of the proposed algorithm. Experimental results (Fig. 13) indicate that the FCN+CRF can segment LAA more precisely and even eliminate the interference of the LCX located around the lobes of the LAA, while FCN alone performs poorly. Specifically, the FCN alone has obvious filtering effects on the LCX with tubular features, but some large LCX patches located around the lobes are often misjudged as the LAA, as shown in ③, ④, $\cdots$, ㉓, and ㉔ of Fig. 13.II. The 3-D models are shown in (g), (h), and (i) of Fig. 14; The FCN+CRF without β can filter out some isolated LCX patches located around the lobes but cannot solve the adhesion of the LCX and lobes as shown in ③, ④, $\cdots$, ⑨, ⑩ and ⑰, ⑱, $\cdots$, ㉓, and ㉔ of Fig. 13.III. The 3-D models are shown in (j), (k), and (l) of Fig. 14. Furthermore the FCN+CRF with β properly solves these problems as shown in ⑪, ⑫, $\cdots$, ⑮, and ⑯ of Fig. 13.III. The 3-D models are shown in (m), (n), and (o) of Fig. 14.

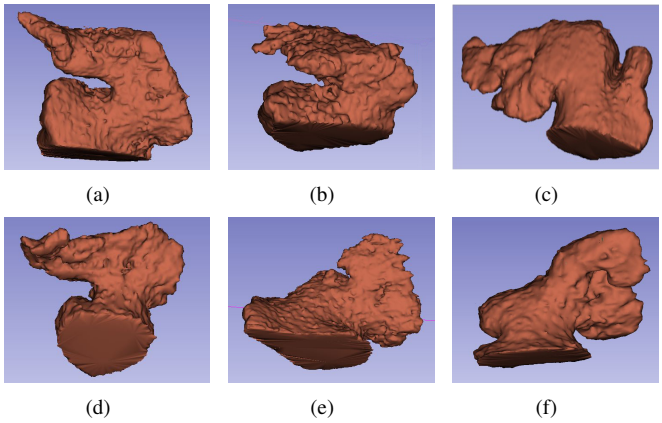


Fig. 15. Some segmentation results of the special LAA shapes.

Some segmentation results of the special LAA shapes selected from the 150 cases of data using the FCN+CRF are shown in Fig. 15. The LAA promptly flexes its body at the opening and reversely grows in Fig. 15 (a), (b) and (d), which brings great difficulties to implant the closure; the neck of the LAA is thick, the lobes are plump and laterally growing and the angle with LSPV becomes larger in Fig. 15 (c); and the neck of LAA is incrassate and the trend of its lobes is serpentine in Fig. 15 (e), (f).

V. Conclusion

We proposed a robust method for automatic LAA segmentation on CTA data using FCNs and 3-D CRFs. We adopted the contemporary FCN in natural image segmentation and transfered their learned models by fine-tuning the networks to segment each 2-D LAA slice. Subsequently we used a dense 3-D CRF that accounts for the 3-D spatial and larger contextual information to refine the segmentations of all slices. Our method was evaluated in 150 cases of CTA data using five-fold cross-validation. Compared with manual annotation, automatic segmentation allowed us to obtain a mean dice overlap of 94.76% and mean volume overlap of 91.10% with computation time of less than 40 s per volume. Experimental results demonstrated the robustness of our method when dealing with large anatomical variations and computational efficiency for adoption in daily clinical routine.

This systematic approach can also be applied to the problem of other chambers such as the LA which has complex shapes. However one of the shortcomings of the current approach is the need to manually mark the ROI of segmentation. We are attempting to develop an automatic positioning algorithm of LAA in CT images.

References

[1] N. M. Alsaady, O. A. Obel, and A. J. Camm, "Left atrial appendage: structure, function, and role in thromboembolism," *Heart*, vol. 82, no. 5, pp. 547–554, 1999.

[2] A. Alissa, Y. Inoue, J. Cammin, Q. Tang, S. Nazarian, H. Calkins, E. K. Fishman, K. Taguchi, and H. Ashikaga, "Regional function analysis of left atrial appendage using motion estimation CT and risk of stroke in patients with atrial fibrillation," *European Journal of Echocardiography*, vol. 17, no. 7, pp. 788–796, 2016.

[3] K. Fukushima, N. Fukushima, K. Kato, K. Ejima, H. Sato, K. Fukushima, C. Saito, K. Hayashi, K. Arai, T. Manaka *et al.*, "Correlation between left atrial appendage morphology and flow velocity in patients with paroxysmal atrial fibrillation," *European Journal of Echocardiography*, vol. 17, no. 1, pp. 59–66, 2016.

[4] H. Oe, Y. Ohno, T. Yamanaka, N. Watanabe, and H. Ito, "Biatrial appendage thrombi in a heart failure patient with sinus rhythm detailed assessment by real-time 3-dimensional transesophageal echocardiography," *Circulation*, vol. 133, no. 1, pp. e1–e4, 2016.

[5] J. P. Piccini and L. Fauchier, "Rhythm control in atrial fibrillation." *The Lancet*, vol. 388, no. 10046, pp. 829–840, 2016.

[6] G. J. Hankey, "Stroke," *The Lancet*, vol. 389, no. 10069, pp. 641–654, 2016.

[7] V. Y. Reddy, H. Sievert, J. L. Halperin, S. K. Doshi, M. Buchbinder, P. Neuzil, K. C. Huber, B. Whisenant, S. Kar, V. Swarup *et al.*, "Percutaneous left atrial appendage closure vs warfarin for atrial fibrillation a randomized clinical trial," *JAMA*, vol. 312, no. 19, pp. 1988–1998, 2014.

[8] R. P. Whitlock, J. S. Healey, and D. R. Holmes, "Left atrial appendage occlusion debate revisited," *Circulation*, vol. 131, no. 8, pp. 756–761, 2015.

[9] R. Weerasooriya, P. Khairy, J. Litalien, L. Macle, M. Hocini, F. Sacher, N. Lellouche, S. Knecht, M. Wright, I. Nault *et al.*, "Catheter ablation for atrial fibrillation," *Journal of the American College of Cardiology*, vol. 57, no. 2, pp. 160–166, 2011.

[10] G. Lee, P. Sanders, and J. M. Kalman, "Catheter ablation of atrial arrhythmias: state of the art," *The Lancet*, vol. 380, no. 9852, pp. 1509–1519, 2012.

[11] D. Feng, W. D. Edwards, J. K. Oh, K. Chandrasekaran, M. Grogan, M. W. Martinez, I. I. Syed, D. A. Hughes, J. A. Lust, A. S. Jaffe *et al.*, "Intracardiac thrombosis and embolism in patients with cardiac amyloidosis," *Circulation*, vol. 116, no. 21, pp. 2420–2426, 2007.

[12] P. Grasland-Mongrain, J. Peters, and O. Ecabert, "Combination of shape-constrained and inflation deformable models with application to the segmentation of the left atrial appendage," in *Proceedings of International Symposium on Biomedical Imaging*, 2010, pp. 428–431.

[13] P. Grasland-Mongrain, "Segmentation of the left atrial appendage from 3D images". Master Thesis. ENS Cachan. 2009.

[14] O. Ecabert, J. Peters, H. Schramm, C. Lorenz, J. von Berg, M. J. Walker, M. Vembar, M. E. Olszewski, K. Subramanyan, G. Lavi *et al.*, "Automatic model-based segmentation of the heart in CT images," *IEEE Transactions on Medical Imaging*, vol. 27, no. 9, pp. 1189–1201, 2008.

[15] Y. Zheng, D. Yang, M. John, and D. Comaniciu, "Multi-part modeling and segmentation of left atrium in C-arm CT for image-guided ablation of atrial fibrillation," *IEEE Transactions on Medical Imaging*, vol. 33, no. 2, pp. 318–331, 2014.

[16] T. Y. Law and P. A. Heng, "Automated extraction of bronchus from 3-D CT images of lung based on genetic algorithm and 3-D region growing," in *Proceedings of International Society for Optical Engineering*, 2000, vol. 3979, pp. 906–916.

[17] R. Lumia, "A new three-dimensional connected components algorithm," *Computer Vision Graphics and Image Processing*, vol. 23, no. 2, pp. 207–217, 1983.

[18] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient n-D image segmentation," *International Journal of Computer Vision*, vol. 70, no. 2, pp. 109–131, 2006.

[19] L. Wang, J. Feng, C. Jin, J. Lu, and J. Zhou, "Left atrial appendage segmentation based on ranking 2-D segmentation proposals," in *Proceedings of MICCAI Workshop on Statistical Atlases and Computational Models of the Heart (STACOM)*, Springer, 2016, pp. 21–29.

[20] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

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

[22] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *arXiv preprint arXiv:1606.00915*, 2016.

[23] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. Torr, "Conditional random fields as recurrent neural networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1529–1537.

[24] H. Fu, Y. Xu, D. W. K. Wong, and J. Liu, "Retinal vessel segmentation via deep learning network and fully-connected conditional random fields," in *Proceedings of International Symposium on Biomedical Imaging*, 2016, pp. 698–701.

[25] N. Dhungel, G. Carneiro, and A. P. Bradley, "Deep learning and structured prediction for the segmentation of mass in mammograms," in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2015, pp. 605–612.

[26] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation," *Medical Image Analysis*, vol. 36, pp. 61–78, 2017.

[27] P. F. Christ, M. E. A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. DAnastasi *et al.*, "Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields," in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2016, pp. 415–423.

[28] Z.-u. Rahman, D. J. Jobson, and G. A. Woodell, "Multi-scale retinex for color image enhancement," in *Proceedings of International Conference on Image Processing*, vol. 3, 1996, pp. 1003–1006.

[29] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing*, vol. 6, no. 7, pp. 965–976, 1997.

[30] H. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. J. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[31] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2013, pp. 246–253.

[32] Y. Wang, L. D. Biase, R. Horton, T. Nguyen, P. Morhanty, and A. Natale, "Left atrial appendage studied by computed tomography to help planning for appendage closure device placement," *Journal of Cardiovascular Electrophysiology*, vol. 21, no. 9, pp. 973–982, 2010.

[33] H. Mohammad, D. Axel, W.-F. David, B. Antoine, C. Aaron, B. Yoshua, P. Chris, J. Pierre-Marc, and L. Hugo, "Brain tumor segmentation with deep neural networks," *Medical Image Analysis*, vol. 35, pp. 18–31, 2017.

[34] C. M. Bishop, "Pattern recognition," *Machine Learning*, vol. 128, pp. 1–58, 2006.

[35] "Flexible and efficient library for deep learning," http://mxnet.io/index.html.

[36] J. Kosiuk, S. Nedios, J. Kornej, E. Koutalas, L. Bertagnolli, S. Rolf, A. Arya, P. Sommer, D. Husser, G. Hindricks *et al.*, "Impact of left atrial appendage morphology on peri-interventional thromboembolic risk during catheter ablation of atrial fibrillation," *Heart Rhythm*, vol. 11, no. 9, pp. 1522–1527, 2014.

[37] P. Craven and G. Wahba, "Smoothing noisy data with spline functions," *Numerische Mathematik*, vol. 31, no. 4, pp. 377–403, 1975.

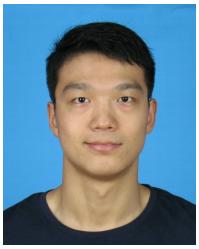
[38] C. Nikou, N. Galatsanos, and A. Likas, "A class-adaptive spatially variant mixture model for image segmentation," *IEEE Transactions on Image Processing*, vol. 16, no. 4, pp. 1121–1130, 2007.

[39] Z. Liu, X. Li, P. Luo, C.-C. Loy, and X. Tang, "Semantic image segmentation via deep parsing network," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1377–1385.

Cheng Jin received the M.E. degree from Harbin Engineering University. From 2005 to 2012, he was a professor in Harbin Engineering University and Qiqihar Medical University, China. Since 2012, He has been a Ph.D. candidate at the Department of Automation, Tsinghua University, Beijing, China. His research interests include medical image processing and computer vision.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2794552, IEEE Journal of Biomedical and Health Informatics

Jianjiang Feng is an associate professor in the Department of Automation at Tsinghua University, Beijing. He received the B.S. and Ph.D. degrees from the School of Telecommunication Engineering, Beijing University of Posts and Telecommunications, China, in 2000 and 2007. From 2008 to 2009, he was a Post Doctoral researcher in the PRIP lab at Michigan State University. He is an Associate Editor of Image and Vision Computing. His research interests include fingerprint recognition and computer vision.



Lei Wang received the B.E. degree from the Department of physics, Tsinghua University, Beijing, China, in 2014. He is currently a M.E. candidate candidate at the Department of Automation, Tsinghua University, Beijing, China. His research interests include medical image processing and computer vision.

Heng Yu He is currently a B.E. candidate at the Department of Automation, Tsinghua University, Beijing, China. His research interests include image processing and deep learning.

Jiang Liu He is currently a B.E. candidate at the Department of Automation, Tsinghua University, Beijing, China. His research interests include image processing and deep learning.

Jiwen Lu received the B.Eng. degree in mechanical engineering and the M.Eng. degree in electrical engineering from the Xi'an University of Technology, Xi'an, China, and the Ph.D. degree in electrical engineering from the Nanyang Technological University, Singapore, respectively. He is currently an Associate Professor with the Department of Automation, Tsinghua University, China. His research interests include computer vision, pattern recognition, and machine learning, where he has authored/co-authored over 130 scientific papers in these areas. He serves an Associate Editor for Pattern Recognition Letters, Neurocomputing, and IEEE Access. He was a recipient of the Best Student Paper Award from Pattern Recognition and Machine Intelligence Association of Singapore in 2012, and the National 1000 Young Talents Plan Program in 2015, respectively. He is a senior member of the IEEE.

Jie Zhou received the BS and MS degrees both from the Department of Mathematics, Nankai University, Tianjin, China, in 1990 and 1992, respectively, and the PhD degree from the Institute of Pattern Recognition and Artificial Intelligence, Huazhong University of Science and Technology (HUST), Wuhan, China, in 1995. From then to 1997, he served as a postdoctoral fellow in the Department of Automation, Tsinghua University, Beijing, China. Since 2003, he has been a full professor in the Department of Automation, Tsinghua University. His research interests include computer vision, pattern recognition, and image processing. In recent years, he has authored more than 100 papers in peer-reviewed journals and conferences. Among them, more than 30 papers have been published in top journals and conferences such as the IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Image Processing, and CVPR. He is an associate editor for the International Journal of Robotics and Automation and two other journals. He received the National Outstanding Youth Foundation of China Award. He is a senior member of the IEEE.