Neurocomputing 438 (2021) 84–93

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Interactive prostate MR image segmentation based on ConvLSTMs and GGNN

Zhiqiang Tian [a,*], Xiaojian Li [a], Zhang Chen [a], Yaoyue Zheng [a], Hongcheng Fan [a], Zhongyu Li [a], Ce Li [b], Shaoyi Du [c]

[a] School of Software Engineering, Xi'an Jiaotong University, Xi'an 710049, PR China
[b] College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, PR China
[c] Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710049, PR China

ARTICLE INFO

Article history:
Received 8 January 2020
Revised 21 April 2020
Accepted 8 May 2020
Available online 23 January 2021

Keywords:
Medical image segmentation
Gated graph neural network
Long short term memory
User interaction

ABSTRACT

Accurate segmentation of the prostate on magnetic resonance (MR) images plays an important role for prostate cancer diagnosis and treatment. Although many automated prostate segmentation methods have been proposed, the performance still faces several challenges, which includes large variability in prostate shape, unclear boundary, and complex intensity distribution. Therefore, the results obtained from the automated methods should be further refined by users to get a more accurate and reliable segmentation. In this paper, we propose an end-to-end interactive segmentation method to refine the automated results. A convolutional long short term memory (convLSTM) module and a gated graph neural network (GGNN) are presented in the proposed method for prostate segmentation in both automated and interactive manners. A boundary loss is proposed to train our model. We evaluated the proposed method on two public available datasets and one in–house dataset. Experimental results show that the proposed convLSTM module could obtain a DSC of 91.78% on the test dataset, which outperforms eight state-of-the-art methods. A further 1.5% improvements can be obtained by user interactions based on the GGNN. The segmentation time including user interactions and inference time was 2.3 min on average for segmenting one volume.

1. Introduction

Prostate cancer is one of the most common cancer diseases in the world and causes massive people deaths every year. An estimated 174650 new cases of prostate cancer will be diagnosed in the United States during 2019. 31620 deaths will occur from prostate cancer in 2019. Magnetic resonance imaging (MRI) has become a routine modality for prostate treatment planning and many other applications [1–4]. Accurate segmentation of the prostate and lesions from MRI is an important step in prostate cancer diagnosis and treatment. However, fully manual segmentation of each MR image is a tedious task. In addition, it is a time consuming and subjective work, which rely on the experience of the readers. Therefore, many prostate MR image segmentation methods have been proposed in recent years [5–7,4].

Although many deep learning based automated segmentation methods have achieved good performance for medical image segmentation [8–11], these methods are not robust and accurate enough for routine clinical use. To address the problem, semi-automatic and interactive segmentation methods are proposed [12,13,7], which incorporates minimal user input. These methods are becoming an attractive and reasonable choice to improve automatic segmentation results.

In this work, we propose a contour interaction-based segmentation method for prostate on MR images. The proposed method gains an initial segmentation by drawing a bounding box around the object firstly, which similar to GrabCut [14]. In contrast with GrabCut method that gets segmentation by drawing click/scribble on background and foreground regions respectively, we present a different interactive manner by dragging a wrong control point on the predicted prostate contour to a right location. The user can intervene and correct the wrong predicted control points whenever an inaccurate segmentation occurs. The pipeline of our proposed methods involves two techniques. First, convolutional long short-term memory [15] is adopted to predict the vertices location of the prostate contour sequentially, which is similar to the process of users delineating the prostate contour. Second, we

* Corresponding author.
E-mail address: zhiqiangtian@xjtu.edu.cn (Z. Tian).

https://doi.org/10.1016/j.neucom.2020.05.121
0925-2312/

exploit a gated graph neural network [16] to learn the interactive capability and get a refined contour.

To the best of our knowledge, this is the first work exploring convolutional LSTM and GGNN based interactive method for prostate segmentation on MR image. The proposed method is designed for 2D segmentation slice-by-slice. The contributions of the proposed method are listed as follows. (1) We propose an end-to-end segmentation method for prostate MR images both in automated and interactive manners. A convolutional LSTM and a gated graph neural network are introduced in the proposed method. The prostate contour can be automatically obtained based on the ConvLSTM. Then, the contour is refined by user interaction based on the GGNN. (2) A loss function consisting of cross entropy loss and boundary loss is proposed to improve the performance of the proposed method. (3) The proposed ConvLSTM module is able to improve the segmentation 1.0% in DSC compared with the state-of-the-art method on a prostate MR dataset. By adopting the GGNN, we can achieve a further 1.5% improvements upon the user refinement.

The remainder of the paper is organized as follows. In Section 2, the related works are reviewed. In Section 3, we present details of the proposed method. Experimental results are shown in Section 4. The paper is concluded in Section 5.

2. Related work

In the past decades, hand-crafted feature based and deep learning based methods have been widely used in the application of prostate segmentation [3,17,18,6]. According to whether there is a human interaction in the segmentation process, these methods can be divided into automatic segmentation method and interactive segmentation method.

Recently, many automatic methods have been proposed for segmentation task, which includes deformable model based [19,14], traditional machine learning based methods [20,21], and deep learning based methods [6,9,10,17,22,23,18]. For biomedical image segmentation task, level set algorithm gains more and more attentions because of the advantage of numerical computations [21]. In the level set methods [24], the curve includes the internal energy coming from the curve and the external energy coming from the data. The curve evolves iteratively by moving the descent of the level set energy. A large variety of traditional machine learning based segmentation methods have also been proposed. For example, Li et al. [25] presented an online learning and patient-specific classification method based on location-adaptive image context for prostate segmentation. Soumya et al. [26] proposed a supervised learning method based on decision forest to achieve a probabilistic representation of the prostate voxels. Similarly, Yang et al. [27] presented a 3D prostate segmentation method, which combines longitudinal image registration and machine learning method. Stephanie et al. [28] proposed a registration and machine learning-based automated segmentation method for subcortical and cerebellar brain structures.

For deep learning based methods, convolutional neural networks (CNNs) have achieved great success in both the computer vision and medical image analysis fields. Following this trend, many researchers utilize various CNNs for learning image feature representation in the application of medical image segmentation [6,9,10,17,22,23]. Fully convolutional networks (FCN) [29] is the first work to use CNNs for segmentation task. Inspired by FCN, researchers proposed lots of FCN-based algorithms for medical image segmentation [6,10,22,23]. For example, Tian et al. [18] proposed an end-to-end deep fully convolutional neural network to segment the prostate automatically, which is called PSNet. Ronneberger et al. [22] took the idea of the FCN and proposed a U-net architecture. It can successfully extract representative feature for medical segmentation task with a reasonable network depth. The architecture consists of a contracting path to capture context information and a symmetric expanding path to find object localization. Several works also focus on 3D architectures for volumetric medical image segmentation. Milletai et al. [23] presented a 3D V-Net architecture with the 3D convolution to perform volumetric medical image segmentation. Yu et al. [6] proposed a volumetric ConvNet to segment prostate on MR images. Zhu et al. [11] used an inter-slice correlation of recursive neural networks for automatic prostate MR image segmentation. Although these automatic segmentation methods could get good segmentation, it is still not accurate enough for routine clinical use.

For interactive segmentation methods, it provides an effective manner where a human and a machine collaborate to get a more accurate segmentation result. Many interactive image segmentation algorithms have been proposed [14,12,30,13,31]. GrabCut was [14] proposed for interactive object segmentation based on the graph cut algorithm [8]. The GrabCut uses an easy interactive way that only needs the user drawing a bounding box around the interesting object region. A scribble-based interactive graph cuts [19] method was proposed by Boykov. A max-flow/min-cut algorithm was used to provide a global optimal solution for final segmentation result. With the success of using deep learning methods in automatic segmentation manner, more and more interactive segmentation methods using CNNs have been proposed. For instance, Wang et al. [13] proposed an interactive medical image segmentation method by adding an image-specific adaptation model for CNNs to get segmentation result. Lin et al. [32] developed an interactive algorithm to train CNN for segmentation supervised by scribbles. Wang et al. [33] proposed an interactive segmentation method that adopts geodesic distance transforms of scribbles as a channels of CNN. Rajchl et al. [34] combined Grab-cut and CNN for medical image segmentation. Papadopoulos et al. [35] proposed an extreme-point based interactive segmentation method, which allows the users to click the top, right-most, bottom and left-most extreme points of an object for getting the segmentation result. These bounding-box-based and scribble/click-based interactive methods treat segmentation as a pixel-wise labeling problem that needs to classify each pixel. Comparing with these interactive segmentation methods, the proposed contour-based method does not need to classify each pixel of the image, but only predicts the vertices on the contour. In addition, the contour-based interactive manner can directly obtain the final accurate boundaries by correcting erroneously predicted vertices.

Castrejon et al. [36] proposed a contour-based interactive segmentation method. An object boundary is used to present the segmentation result. The boundary is refined by dragging the wrong predicted boundary to their correct locations. Acuna et al. [37] proposed an efficient interactive annotation method that refers to as polygon-RNN++ based on the GGNN. This is a 2D segmentation method in a contour-based interactive manner. Wang et al. [38] first applied GGNN to handle the interactive 3D segmentation. The GGNN was used to propagate the user interaction to the 3D neighboring nodes. These two methods used a region-based loss function for contour-based interactive segmentation. In contrast, we proposed a contour-based loss (boundary loss) for contour-based segmentation. In addition, the polygon-RNN++ [37] consists of four main modules, including recurrent neural network, reinforcement learning, evaluator network, and graph neural network. These four modules were trained separately that is difficult to get a global optimization. In contrast, the proposed method only contains two main modules that was trained end-to-end.

In this paper, we propose a ConvLSTM and a GGNN based interactive segmentation method to predict the prostate contour in an end-to-end manner. We denote prostate contour as a series of

connected sequential vertices by imitating manual delineation from radiologists. The LSTM enables our model to capture spatial relation between neighboring vertices as what radiologist does during delineation. It is suitable for the convolutional LSTMs to segment prostate by predicting the vertices of the prostate contour sequentially. These vertices are connected with each other to form a graph. The GGNN is good at processing graph data [39]. Therefore, the GGNN is adopted to refine the locations of these vertices on the prostate contour.

3. Method

3.1. Overview of the proposed method

In this study, the proposed method is designed for 2D segmentation slice-by-slice. Prostate segmentation is obtained by predicting the vertices of the prostate contour sequentially. The vertices prediction of the prostate contour is considered as a classification task. The proposed method continues its automatic prediction and interactive correction iteratively by moving the erroneously predicted contour vertices to their right locations. Several cascaded convolutional LSTM [15] layers are used to predict the location of each contour vertex step by step, which is called ConvLSTMs module. The input of ConvLSTMs module is a $28 \times 28 \times 128$ CNN feature map, which is produced by an atrous multi-scale feature encoder. In our multi-scale feature encoder, a modified ResNet101 [40] with atrous convolutional (AC) operation, a SE block [41], and a skip-layer architecture are used to get an effective output resolution and a multi-scale feature representation. In particular, we also use a location block in the ConvLSTMs module. It aims to get a precise presentation for a vertex of prostate contour, which is a one-hot encoding with size of $28 \times 28 \times 1$. In addition, we use a GGNN module to gain user interactive ability and a high output resolution. Our GGNN module consists of a propagation block with a gated recurrent unit (GRU) [42] and a prediction block. The GGNN has two inputs, which are the corrected vertices of the contour produced by the ConvLSTMs module and a $112 \times 112 \times 256$ multi-scale feature produced by the multi-scale feature encoder. In our implementation, we mimic a user dragging an inaccurate vertex to the true location iteratively. In each interactive step, we only let the $2k$ neighbor vertices of current corrected vertex be predicted. The overview of the proposed method is shown in Fig. 1.

3.2. CNN encoder

Inspired by ResNet-101, a multi-scale CNN feature encoder is proposed, which is shown in Fig. 2. Different with ResNet-101, the fully connected layers and the last average pooling layer are removed in our method. In addition, the convolution operations in the last two residual layers of ResNet-101 are replaced by atrous convolution with 2, 4 dilated rates respectively. A squeeze-and-excitation (SE) block is also added at each residual layer, which enables the CNN encoder to get more important feature by learning a channel-wise weight. Four 3×3 convolution layers are adopted after four layers, which are the first 7×7 convolution layer, the first residual layer, the second residual layer, and the last residual layer with AC, respectively. In order to get multi-scale spatial feature, the output features of four 3×3 convolutions are concatenated as the input of ConvLSTMs module. Three bilinear upsampling operations are performed to make them have same spatial size. Then, the concatenated feature are fed into two consecutive 3×3 convolution layers. Finally, the output feature of the two consecutive 3×3 convolution layers is concatenated with the output of a pyramid scene parsing network (PSPNet) module [43], which could yield an accurate segmentation result. In our experiment, the pyramid scene parsing network follows the last residual block. The final concatenated $28 \times 28 \times 128$ feature will be fed into the GGNN module. The sizes of all feature maps are also shown in Fig. 2.

3.3. ConvLSTMs module

Convolutional LSTM is useful for getting the spatial contextual information of a sequential input data and predicting vertices by applying linear and non-linear functions, which could carry history information. In this work, the ConvLSTM is used as a decoder to make a coherent prediction of the vertices of the prostate contour. A two-layer ConvLSTM is performed to output a vertex at each time step. The ConvLSTMs module would be performed in T time steps. T can be changed during train stage and test stage. The predictions of $t-1$ and $t-2$ time steps, the hidden state of $t-1$ time step, and the CNN feature are concatenated as the input of the t time step to make a prediction.

Each time step is followed by a location block. The location block consists of a 1×1 convolution, a ReLu non-linear function, and a softmax function. The output of location block is a

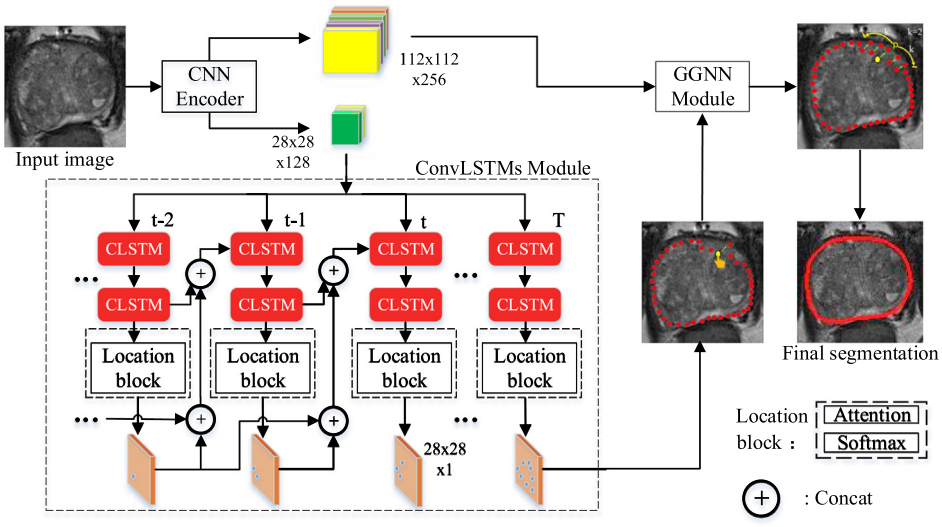


Fig. 1. Overview of the proposed method.

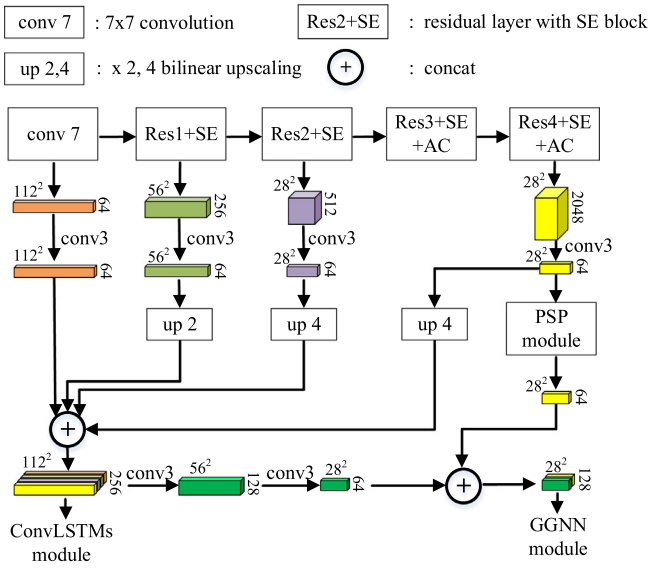


Fig. 2. Multi-scale CNN feature encoder.

$28 \times 28 \times 1$ one-hot encoding feature, which shows the location of the predicted vertex at time step t.

3.4. GGNN

The GGCNN contains two blocks, which are propagation block and prediction block. The propagation block is used to propagate the information between vertices, and the prediction block is applied for predicting the location of vertex. The propagation block consists of two weight blocks. Each weight block has three parallel channels, which is shown in Fig. 3. Each channel contains two FC layers and a *tanh* activation function. In addition, a gated recurrent unit is used to update the incorporated information based on previous predicted vertices. The output of the propagation block is fed into the gated recurrent unit. The propagation block [16] can be formulated as follows.

$$h_v^0 = \left[x_v^T, 0\right]^T \tag{1}$$

$$a_v^t = A_v^T \left[h_1^{(t-1)T} \ldots h_{|V|}^{(t-1)T}\right]^T + b \tag{2}$$

$$h_v^t = f_{GRU}\left(h_v^{t-1}, a_v^t\right) \tag{3}$$

The matrix A presents how nodes of a graph communicate with each other. h_v^t is the hidden state, where t denotes the time-step and x_v is the initial state of a node v. As for the prediction block, the graph-level output for node v can be defined as follows.

$$\mathrm{h}_v = \tanh\left(\sum_{v \in V} \sigma\left(f_1\left(concate\left(h_v^T, x_v\right)\right)\right) \odot tanh\left(f_2\left(concate\left(h_v^T, x_v\right)\right)\right)\right) \tag{4}$$

where the sigmoid function $\sigma(*)$ decides which node is relevant to the current graph-level task. f_1 and f_2 are either a neural network or a linear function that output real-valued predicted vectors. The *tanh* function can be also replaced with an identity. In our implements, a FC layer followed by an activation function and another FC layer to form the prediction block.

In this study, the number of vertices was fixed as 30 for all slices. Also, the vertex number does not change from the convLSTM to GGNN output. We have tried different numbers of vertices for presenting the prostate contour. The experimental results show that the accuracy have a negligible change when the number of vertices varies from 25 to 40.

3.5. Loss function

To better evaluate the similarity between ground-truth contour and the predicted contour, a new loss function consisting of a cross entropy loss function L_{ce} and a boundary loss function [44] L_{bl} is proposed. The cross entropy loss L_{ce} is computed as follows.

$$L_{ce}(\theta) = -\sum_y y \log \sigma_\theta(x) + (1-y)\log\left(1-\sigma_\theta(x)\right) \tag{5}$$

where $\sigma_\theta(x)$ is the output of the a deep neural network, x is the input of the network, while y represents the ground truth. θ denotes the trainable parameters of the neural network. The boundary loss L_{bl} is defined as follows.

$$L_{bl}(\theta) = \int_I \varphi_{dist}(C_G(p), p) softmax(\sigma_\theta(p)) dp \tag{6}$$

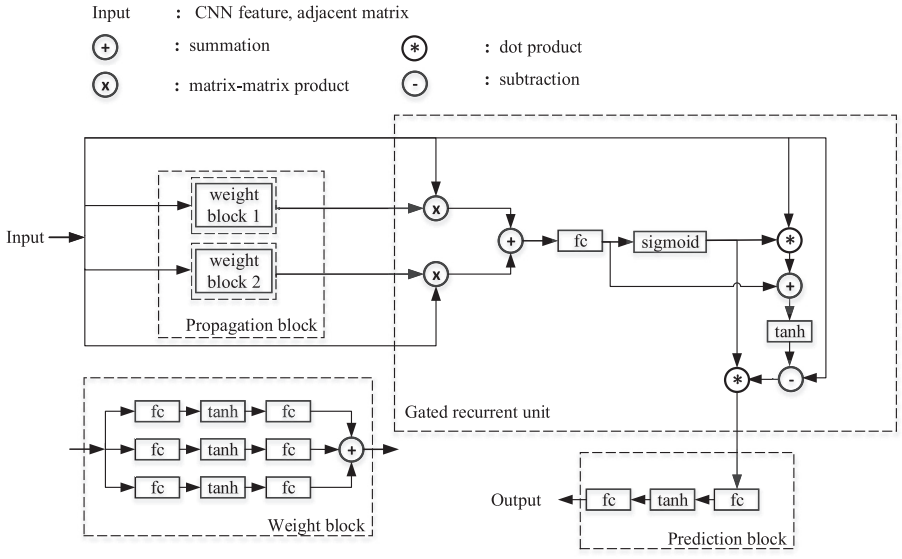


Fig. 3. Diagram of the gated graph neural network. The *fc* represents the fully connected layer.

$\varphi_{dist}(C_G(p), p)$ is a signed distance matrix between point $p \in I$ and the nearest point $C_G(p)$ on the ground-truth contour C_G, which is constructed by the distance transform of the ground-truth contour. $\varphi_{dist}(C_G(p), p) = -\|p - G(p)\|$ if p in the ground-truth region G and $\varphi_{dist}(C_G(p), p) = \|p - G(p)\|$ otherwise. I denotes an image. The diagram of the boundary loss is shown in Fig. 4. In training stage, the L_{total} is used as final loss function, which is defined as follows.

$$Ł_{total}(\theta) = \lambda L_{ce}(\theta) + (1 - \lambda) L_{bl}(\theta) \tag{7}$$

where λ is a weight between L_{ce} and L_{bl}. We set it as 0.3 in our experiments.

3.6. Interactive segmentation

In order to mimic the processing of user interaction, the incorrect predicted vertex is moved to its correct location (ground truth). In this paper, the network is trained to predict the $2k$ (k neighbors on either side) neighbor interaction points of current moved vertex by predicting the clockwise k neighbors firstly and then counter-clockwise.

4. Experimental results

4.1. Dataset and evaluation metrics

In our experiments, 140 subjects of prostate MRI were used for training. These subjects are from three data sets, which are PROMISE12 (50 subjects), ISBI2013 (49 subjects), and in–house (41 subjects) data sets. The subjects are transversal T2-weighted MR images, which are scanned at 1.5 T and 3.0 T. The voxel size varies from 0.4 mm to 1 mm. The size of the transverse images is from 320×320 to 512×512. A windowed *sinc* interpolation is used to get isotropic volume for each case. 30 test subjects from PROMISE12 including the ground truths were used to evaluate the proposed method.To quantitatively evaluate the proposed method, four metrics were used, which are Dice similarity coefficient (DSC), relative volume difference (RVD), Hausdorff distance (HD) and average symmetric surface distance (ASD). The first two metrics are region-based and the last two metrics are distance-based. The DSC is used to evaluate the fraction of coverage between the prediction and ground truth, which is calculated as follows,

$$DSC = \frac{2|R_{gt} \cap R_{pre}|}{|R_{gt}| + |R_{pre}|} \times 100\%, \tag{8}$$

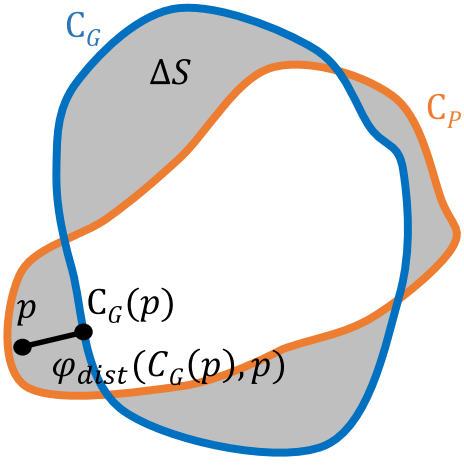


Fig. 4. Diagram of the boundary loss function. The C_G and C_P are the contours of the ground truth and the prediction. ΔS denotes the region between C_G and C_P. $\varphi_{dist}(C_G(p), p)$ is a signed distance matrix between point p and its nearest point $C_G(p)$ on the ground-truth contour C_G.

where R_{gt} and R_{pre} are respectively the prostate regions of ground truth and prediction. The operator $|*|$ represents the number of pixels in a region. The RVD is used to evaluate the prediction whether method tends to be under-segmentation or over-segmentation, which is defined as the following equation.

$$RVD = \frac{|R_{pre}| - |R_{gt}|}{|R_{gt}|} \times 100\%. \tag{9}$$

The HD is used to measure the Hausdorff distance between prediction and ground truth, which is defined as follows,

$$HD = \max\left(\max_{i \in B_{pre}}\left(\min_{j \in B_{gt}}(d(i,j))\right), \max_{j \in B_{gt}}\left(\min_{i \in B_{pre}}(d(i,j))\right)\right), \tag{10}$$

where B_{gt} represents the boundary of ground truth, B_{pre} represents the boundary of prediction. $d(i,j)$ is Euclidean distance between pixel i and pixel j. The ASD is calculated as follows,

$$ASD = \frac{1}{|B_{pre}| + |B_{gt}|} \times \left(\sum_{i \in B_{pre}} d(i, B_{gt}) + \sum_{j \in B_{gt}} d(j, B_{pre})\right), \tag{11}$$

where $d(*)$ presents the distance from a point to a boundary.

4.2. Implementation details

The proposed method was mainly implemented in Python language. A deep learning framework PyTorch was used to implement the proposed method. The codes ran on a platform of Ubuntu with 2 GPUs of NVIDIA GTX 1080 Ti. The proposed model was trained using the Adam optimizer. The initial learning rate was set as $2e^{-6}$ and decreased by a weight decay of 0.1 every two epochs. The batch size was set as 6.

4.3. Qualitative results

The qualitative results obtained from ConvLSTM module on six prostate volumes are shown in Fig. 5. The red curves are the predicted prostate contours from the proposed method, while the blue curves are the manually labeled ground truth. These images have different prostate sizes and shapes, which shows the robustness of our method for different prostate MR images. It also shows that the proposed method could get satisfactory segmentation results for prostate MR images.

4.4. Ablation study

We conducted an ablation study to evaluate the contributions of the AC, SE block, PSP module, and BL to the overall performance of the proposed method. The ablation results are shown in Table 1. When only atrous convolution (AC) is used, the DSC increases $0.76\% \sim 2.16\%$ for the whole-gland and three sub-regions, which are the apex, mid-gland, and base regions. When the SE block is added, the DSC improves 0.17% for the whole-gland. The PSP brings another 0.57% improvement in terms of DSC. The BL is found to be useful and finally makes the DSC increase to 91.78% for the whole-gland. From the table, we can see that the proposed method (Baseline + AC + SE + PSP + BL) has the highest DSC for the whole-gland and three sub-regions. Especially for the mid-gland, it achieves the highest DSC of 93.99%. For the whole-gland, the segmentation results are also the best in terms of RVD, HD, and ABD. These results of the ablation experiment shows that the performance can be improved by using AC, SE, PSP, and BL together or separately.

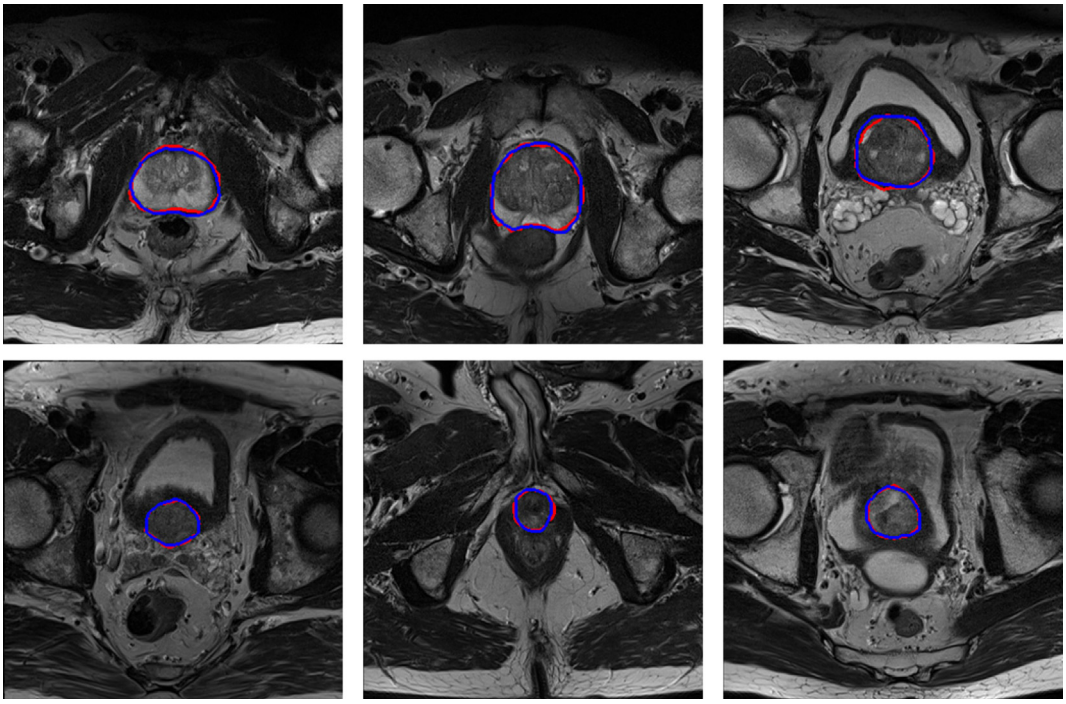


Fig. 5. Qualitative results based on automatic segmentation. The red curves are the predicted prostate contours, while the blue curves are the manually labeled ground truth.

Table 1
Quantitative results of ablation study on whole-gland and three subregions, which are apex, mid-gland, and base subregions, which are apex, mid-gland, and base subregions. Four metrics were used to evaluate the performance, including the DSC (%), RVD (%), HD (*mm*), and ASD (*mm*).

		DSC	RVD	HD	ABD
Whole-gland	Baseline	88.81	7.14	11.80	2.07
	Baseline + AC	90.49	−3.17	12.13	1.87
	Baseline + AC + SE	90.66	4.23	11.79	1.92
	Baseline + AC + SE + PSP	91.23	2.65	11.62	1.83
	Baseline + AC + SE + PSP + BL	**91.78**	**2.64**	**10.32**	**1.73**
Apex	Baseline	87.50	8.20	11.95	2.29
	Baseline + AC	88.26	−0.71	11.47	2.13
	Baseline + AC + SE	89.98	−3.39	10.06	2.05
	Baseline + AC + SE + PSP	90.08	−2.61	10.54	2.04
	Baseline + AC + SE + PSP + BL	90.26	−2.47	9.61	1.96
Mid-gland	Baseline	91.06	8.09	15.22	2.24
	Baseline + AC	93.11	−1.45	9.74	1.85
	Baseline + AC + SE	92.85	−2.62	10.05	1.90
	Baseline + AC + SE + PSP	93.47	−0.96	10.01	1.81
	Baseline + AC + SE + PSP + BL	93.99	−1.38	8.60	1.67
Base	Baseline	87.21	5.39	13.38	2.73
	Baseline + AC	89.37	−7.41	12.14	2.45
	Baseline + AC + SE	88.55	−7.01	13.15	2.61
	Baseline + AC + SE + PSP	89.51	−5.32	12.06	2.40
	Baseline + AC + SE + PSP + BL	90.44	−4.79	10.65	2.25

4.5. Quantitative results

To further evaluate the performance of the proposed method, we compared our approach with eight state-of-the-art methods, which are PSPNet [43], FCN [29], U-Net [22], V-Net [23], DeepLabV3+ [45], Grab-Cut [14], PolyRNN++ [37], and ExtremeCut [46]. The quantitative comparison results are shown in Table 2. For the whole-gland of prostate, the proposed method gets the highest DSC of 91.8% with the lowest standard deviation of 1.3%. Compared with eight state-of-the-art methods, the proposed method gets the lowest absolute value of RVD. In terms of the other two metrics HD and ABD, the proposed method also achieves the best performance with low standard deviation. For the other three subregions, we can see that the proposed method performs rather well in the apex and base subregions. For the middle-gland subregion, the proposed is slight lower than the U-net method but has a lower standard deviation than the U-net. The quantitative evaluation agrees with the conclusions from the qualitative results.

To further evaluate the proposed boundary (BD) loss function, we performed a comparison experiment with the active contour (AC) loss [47]. The AC loss was combined with our proposed plain model (without loss) for the comparison experiment. The comparison results are presented in Table 3. From the table, we can observe that the proposed loss function performs better than the AC loss.

Table 2
Comparison of the proposed method with eight state-of-the-art methods.

Whole		DSC	RVD	HD	ABD
	PSPNet	80.5 ± 7.2	7.4 ± 15.3	20.2 ± 14.3	2.5 ± 0.7
	FCN	82.4 ± 5.6	6.1 ± 10.6	19.6 ± 19.8	2.4 ± 0.7
	U-Net	84.7 ± 6.5	3.4 ± 8.0	15.9 ± 6.9	1.9 ± 0.4
	V-Net	85.3 ± 6.8	3.5 ± 8.8	16.8 ± 6.6	2.0 ± 0.5
	DeepLabV3+	86.5 ± 5.1	−6.2 ± 7.1	23.1 ± 19.1	2.2 ± 0.4
	Grab-Cut	87.0 ± 4.4	3.2 ± 6.5	17.2 ± 10.3	1.9 ± 0.5
	PolyRNN++	88.1 ± 3.4	2.7± 6.4	15.8±5.2	2.1±0.5
	ExtremeCut	90.8 ± 2.5	−3.4 ± 2.2	10.9 ± 2.2	1.9 ± 0.1
	Ours	**91.8** ± 1.3	**2.6** ± 3.8	**10.3** ± 4.1	**1.7** ± 0.4
Apex	PSPNet	73.9 ± 12.2	12.9 ± 25.4	19.0 ± 8.9	3.4 ± 1.2
	FCN	77.6 ± 13.2	15.9 ± 36.4	15.2 ± 8.2	3.1 ± 1.5
	U-Net	78.4 ± 13.6	13.6 ± 43.7	14.8 ± 7.0	2.5 ± 1.0
	V-Net	83.1 ± 11.2	1.1 ± 20.7	14.9 ± 9.5	2.4 ± 1.0
	DeepLabV3+	83.0 ± 8.5	−11.1 ± 18.1	13.7 ± 5.3	2.8 ± 0.9
	Grab-Cut	84.1 ± 6.7	7.1 ± 11.4	11.6 ± 5.1	2.6 ± 0.6
	PolyRNN++	87.5 ± 3.5	7.5 ± 8.4	12.7 ± 5.1	2.2 ± 0.7
	ExtremeCut	88.8 ± 3.2	−4.1 ± 5.1	10.0 ± 3.8	2.2 ± 0.2
	Ours	90.3 ± 2.2	−2.5 ± 6.6	9.6 ± 4.7	2.0 ± 0.5
Mid	PSPNet	89.9 ± 5.1	−1.3 ± 11.2	17.0 ± 13.2	2.7 ± 1.0
	FCN	92.1 ± 3.0	1.6 ± 11.7	15.8 ± 19.8	2.2 ± 0.9
	U-Net	94.8 ± 1.4	−1.3 ± 5.5	8.9 ± 4.4	1.5 ± 0.3
	V-Net	93.9 ± 1.8	2.5 ± 6.2	11.2 ± 5.4	1.7 ± 0.5
	DeepLabV3+	93.3 ± 2.3	−1.7 ± 7.9	11.7 ± 14.5	1.9 ± 0.4
	Grab-Cut	93.1 ± 2.3	2.5 ± 5.8	9.3 ± 4.0	1.9 ± 0.5
	PolyRNN++	92.9 ± 2.2	2.3 ± 5.0	9.7 ± 3.4	2.0 ± 0.5
	ExtremeCut	93.2 ± 2.2	−2.7 ± 2.4	8.9 ± 1.8	1.9 ± 0.2
	Ours	94.0 ± 1.0	−1.4 ± 1.0	8.6 ± 3.3	1.7 ± 0.4
Base	PSPNet	73.6 ± 10.6	−21.2 ± 20.0	22.5 ± 12.0	4.2 ± 1.7
	FCN	74.5 ± 12.7	19.1 ± 21.3	19.6 ± 10.2	3.6 ± 1.3
	U-Net	78.2 ± 14.1	12.5 ± 35.3	18.9 ± 10.9	3.0 ± 1.2
	V-Net	76.3 ± 16.1	23.0 ± 30.0	18.3 ± 8.8	3.1 ± 1.1
	DeepLabV3+	81.0 ± 11.5	−6.5 ± 21.2	21.8 ± 19.0	3.1 ± 1.4
	Grab-Cut	82.1 ± 9.5	5.5 ± 10.2	16.1 ± 11.2	2.8 ± 0.9
	PolyRNN++	87.2 ± 5.0	−3.8 ± 11.5	12.7 ± 5.6	2.5 ± 0.7
	ExtremeCut	89.7 ± 3.2	−4.1 ± 3.4	10.1 ± 2.9	2.4 ± 0.3
	Ours	90.4 ± 3.0	−4.8 ± 7.2	10.6 ± 4.3	2.2 ± 0.6

Table 3
The comparison results between the proposed boundary (BD) loss and the active contour (AC) loss functions.

	DSC	RVD	HD	ABD
Plain model + AC loss	90.7 ± 1.9	2.7 ± 4.1	12.4 ± 4.3	1.8 ± 0.4
Plain model + BD loss	91.8 ± 1.3	2.6 ± 3.8	10.3 ± 4.1	1.7 ± 0.4

In our experiment, the segmentation time including user interaction and inference time required by the proposed method was 12 s on average for segmenting one image. The segmentation time of Grab-cut was 55 s on average, while ExtremeCut was 8 s on average. Although ExtremeCut requires less time than our method, our method obtains a better accuracy. In addition, ExtremeCut cannot exactly correct the inaccurate boundaries based on its point-based interaction manner. In contrast, the proposed interactive segmentation method can directly obtain the final accurate segmentation by correcting erroneously predicted vertices.

We also investigated whether the proposed method works for a different segmentation task. A public fundus image dataset REFUGE [48] was adopted to further evaluate our method. In this experiment, 800 images were used for training, while 400 images were used for testing. The optic disc was segmented from the fundus image by the proposed method. Our method yields a DSC of 95.7±4.4%. The experimental result shows that the proposed method can be generalized to handle a different segmentation task and obtains a satisfactory result.

4.6. Interaction evaluation

Visual examples of how the proposed method is applied for interactive segmentation are shown in Fig. 6. Two automatic segmentations with mis-segmentations followed by interactive refinement are shown in Fig. 7.

To evaluate the performance of the proposed interactive segmentation method, the segmentation results obtained before and after user interaction are shown in Table 4. The average number of clicks is 4.59 in our experiment. From the table, we can see the interactive manner could yield better segmentation results compared with automated manner, which increases the DSC from 91.78% to 93.23% with only few interactions. Furthermore, the interactive manner also gets better performance in term of RVD, HD, and ABD.

4.7. Effect of the number of clicks

To evaluate the effect of the number of clicks, we performed a simulated experiment. In the experiment, we assume that the users would like to correct the worst predicted vertex. The predicted contour is compared with the ground truth to find the worst predicted vertex, which will be corrected by the simulated experiment. Therefore, we can mimic the processing of user interaction by moving the predicted incorrect vertex to its correct position. The experimental result is shown in Fig. 8. DSC is used in the experiment. The DSC became better with the increasing of the number of clicks, which indicates that the interaction is helpful

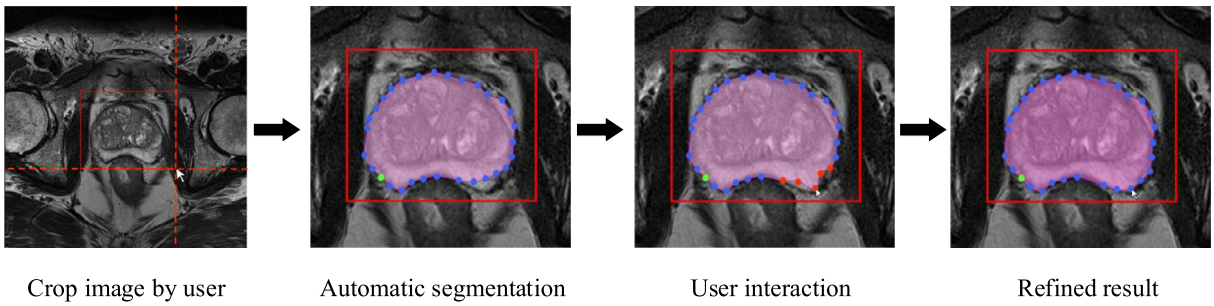


Fig. 6. Visual example of how the proposed method is applied for interactive segmentation. Red box is obtained based on the user cropping operation. The second image shows the automatic segmentation obtained by ConvLSTM module. Green vertex is the first predicted vertex. In the third image, user selects an erroneously predicted vertex and drags it to a proper location. Four neighboring vertices (red vertices) of the selected vertex are then re-predicted to new locations based on the GGNN. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

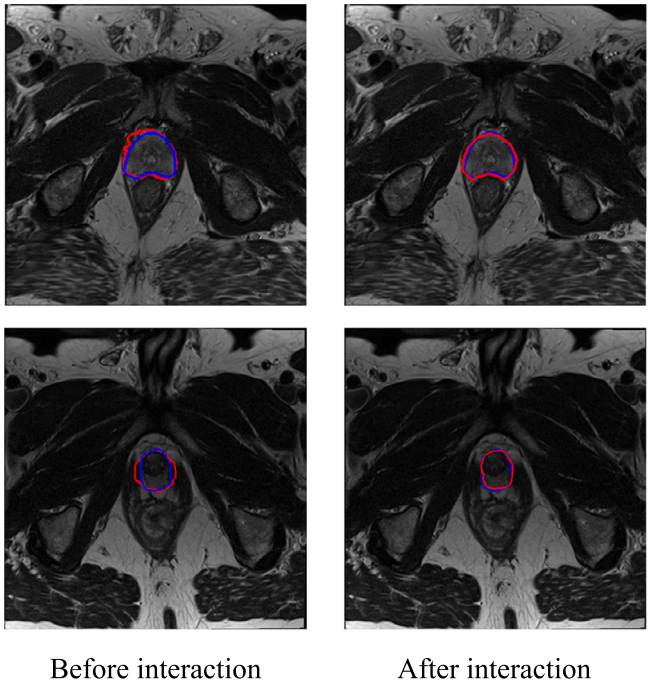


Fig. 7. The effect of user interactions. The red curves are the predicted prostate contours, while the blue curves are the ground truth.

for improving the prostate segmentation. When the number of the clicks reaches a certain value, the DSC has no increase anymore. Therefore, we should choose a suitable value for the number of the clicks. From the figure, we can see that 4 clicks are suitable for mid-gland, while 7 clicks are suitable for base, apex and whole-gland, respectively.

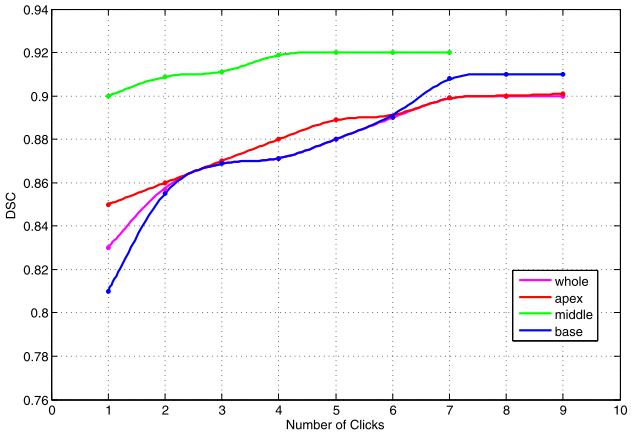


Fig. 8. Effect of the number of clicks.

5. Conclusion

In this paper, we proposed an end-to-end interactive segmentation method, which consists of a convolutional long short term memory module and a gated graph neural network module. We evaluated the proposed method against eight state-of-the-art segmentation models on prostate MR image dataset. Experimental results demonstrate that the proposed method can get satisfactory results and achieve superior results compared with other state-of-the-art methods. In addition, we also evaluated the performance of the proposed interactive segmentation compared with the proposed automated segmentation method. The interactive segmentation could yield better result with few user interactions. In our future work, we will extend the proposed method to segment different organs from other medical image modalities.

The proposed method does not rely on a specific design of the backbone network. A more powerful backbone may improve the

Table 4
Evaluation of the interactive segmentation compared with the automated segmentation.

		DSC	RVD	HD	ABD
Before Interaction	Whole-gland	91.78	2.64	10.32	1.73
	Apex	90.26	−2.47	9.61	1.96
	Mid-gland	93.99	−1.38	8.60	1.67
(Without using GGNN)	Base	90.44	−4.79	10.65	2.25
After Interaction	Whole-gland	93.23	−1.89	9.87	1.57
	Apex	92.67	−1.63	8.58	1.64
	Mid-gland	94.26	−1.62	9.81	1.61
(With using GGNN)	Base	92.46	−2.51	10.90	1.95

performance of our method. Note that, before a backbone being adopted in the proposed method, the structure of the backbone should be improved to fit the proposed framework. Therefore, several improvements have been made in our method, e.g. atrous convolution, SE block, and PSPNet.

The proposed 2D GNN-based segmentation method can be generalized for the 3D segmentation task. One potential solution is that the predicted 3D segmentation mask can be considered as a 3D surface in a triangular mesh. The vertices of triangular mesh are considered as nodes in a 3D GNN model. Then, these nodes incorporated with 3D CNN features can be input to GNN for 3D segmentation task.

CRediT authorship contribution statement

Zhiqiang Tian: Conceptualization, Methodology, Writing - original draft. **Xiaojian Li:** Methodology, Writing - original draft, Software, Validation. **Zhang Chen:** Conceptualization, Methodology, Software. **Yaoyue Zheng:** Methodology, Validation, Software. **Hongcheng Fan:** Methodology, Validation. **Zhongyu Li:** Methodology, Validation. **Ce Li:** Methodology, Validation. **Shaoyi Du:** Methodology, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

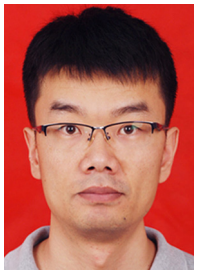
Acknowledgment

This work was supported in part by the NSFC Nos. 61876148 and 61866022, and funded by China Post-doctoral Science Foundation of No. 2018M631164 and the Fundamental Research Funds for the Central Universities of No. XJJ2018254.

References

[1] B. Fei, C. Kemper, D.L. Wilson, A comparative study of warping and rigid body registration for the prostate and pelvic mr volumes, Comput. Med. Imaging Graph. 27 (4) (2003) 267–281.
[2] B. Fei, J.L. Duerk, D.T. Boll, J.S. Lewin, D.L. Wilson, Slice-to-volume registration and its potential application to interventional mri-guided radio-frequency thermal ablation of prostate cancer, IEEE Trans. Medical Imaging 22 (4) (2003) 515–525.
[3] G. Litjens, O. Debats, J. Barentsz, N. Karssemeijer, H. Huisman, Computer-aided detection of prostate cancer in mri, IEEE Trans. Medical Imaging 33 (5) (2014) 1083–1092.
[4] W. Qiu, J. Yuan, E. Ukwatta, Y. Sun, M. Rajchl, A. Fenster, Prostate segmentation: an efficient convex optimization approach with axial symmetry using 3-d trus and mr images, IEEE Trans. Medical Imaging 33 (4) (2014) 947–960.
[5] Z. Tian, L. Liu, Z. Zhang, J. Xue, B. Fei, A supervoxel-based segmentation method for prostate mr images, Med. Phys. 44 (2) (2017) 558–569.
[6] L. Yu, X. Yang, H. Chen, J. Qin, P.A. Heng, Volumetric convnets with mixed residual connections for automated prostate segmentation from 3d mr images, in: Thirty-first AAAI conference on artificial intelligence, 2017.
[7] Z. Tian, L. Liu, Z. Zhang, B. Fei, Superpixel-based segmentation for 3d prostate mr images, IEEE Trans. Medical Imaging 35 (3) (2015) 791–801.
[8] Y. Boykov, G. Funka-Lea, Graph cuts and efficient nd image segmentation, Int. J. Computer Vision 70 (2) (2006) 109–131.
[9] D. Shen, G. Wu, H.-I. Suk, Deep learning in medical image analysis, Ann. Rev. Biomed. Eng. 19 (2017) 221–248.
[10] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. Van Der Laak, B. Van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88.
[11] Q. Zhu, B. Du, B. Turkbey, P. Choyke, P. Yan, Exploiting interslice correlation for mri prostate image segmentation, from recursive neural networks aspect, Complexity 2018 (2018).
[12] S.H. Park, Y. Gao, Y. Shi, D. Shen, Interactive prostate segmentation using atlas-guided semi-supervised learning and adaptive feature selection, Med. Phys. 41 (11) (2014) 111715.
[13] G. Wang, W. Li, M.A. Zuluaga, R. Pratt, P.A. Patel, M. Aertsen, T. Doel, A.L. David, J. Deprest, S. Ourselin, et al., Interactive medical image segmentation using deep learning with image-specific fine tuning, IEEE Trans. Medical Imaging 37 (7) (2018) 1562–1573.
[14] C. Rother, V. Kolmogorov, A. Blake, Grabcut: Interactive foreground extraction using iterated graph cuts, in: ACM transactions on graphics (TOG), Vol. 23, ACM, 2004, pp. 309–314.
[15] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, W.-C. Woo, Convolutional lstm network: A machine learning approach for precipitation nowcasting, in: Advances in neural information processing systems, 2015, pp. 802–810.
[16] Y. Li, D. Tarlow, M. Brockschmidt, R. Zemel, Gated graph sequence neural networks, arXiv preprint arXiv:1511.05493 (2015).
[17] Z. Tian, L. Liu, B. Fei, Deep convolutional neural network for prostate mr segmentation, Int. J. Computer Assisted Radiol. Surgery 13 (11) (2018) 1687.
[18] Z. Tian, L. Liu, Z. Zhang, B. Fei, Psnet: prostate segmentation on mri based on a convolutional neural network, J. Med. Imaging 5 (2) (2018) 021208.
[19] Y.Y. Boykov, M.-P. Jolly, Interactive graph cuts for optimal boundary & region segmentation of objects in nd images, in: Proceedings eighth IEEE international conference on computer vision. ICCV 2001, Vol. 1, IEEE, 2001, pp. 105–112.
[20] P.-H. Conze, V. Noblet, F. Rousseau, F. Heitz, V. De Blasi, R. Memeo, P. Pessaux, Scale-adaptive supervoxel-based random forests for liver tumor segmentation in dynamic contrast-enhanced ct scans, Int. J. Computer Assisted Radiol. Surgery 12 (2) (2017) 223–233.
[21] A. Hoogi, C.F. Beaulieu, G.M. Cunha, E. Heba, C.B. Sirlin, S. Napel, D.L. Rubin, Adaptive local window for level set segmentation of ct and mri liver lesions, Med. Image Anal. 37 (2017) 46–55.
[22] O. Ronneberger, P. Fischer, T. Brox, U-net, Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234–241.
[23] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision (3DV), IEEE, 2016, pp. 565–571.
[24] S. Osher, J.A. Sethian, Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations, J. Comput. Phys. 79 (1) (1988) 12–49.
[25] W. Li, S. Liao, Q. Feng, W. Chen, D. Shen, Learning image context for segmentation of prostate in ct-guided radiotherapy, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2011, pp. 570–578.
[26] S. Ghose, J. Mitra, A. Oliver, R. Marti, X. Lladó, J. Freixenet, J.C. Vilanova, D. Sidibé, F. Meriaudeau, A random forest based classification approach to prostate segmentation in mri, MICCAI Grand Challenge: Prostate MR Image Segmentation 2012 (2012) 125–128.
[27] X. Yang, B. Fei, 3d prostate segmentation of ultrasound images combining longitudinal image registration and machine learning, in: Medical Imaging 2012: Image-Guided Procedures, Robotic Interventions, and Modeling, Vol. 8316, International Society for Optics and Photonics, 2012, p. 83162O.
[28] S. Powell, V.A. Magnotta, H. Johnson, V.K. Jammalamadaka, R. Pierson, N.C. Andreasen, Registration and machine learning-based automated segmentation of subcortical and cerebellar brain structures, Neuroimage 39 (1) (2008) 238–247.
[29] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.
[30] N. Xu, B. Price, S. Cohen, J. Yang, T.S. Huang, Deep interactive object selection, in, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 373–381.
[31] J. Liew, Y. Wei, W. Xiong, S.-H. Ong, J. Feng, Regional interactive image segmentation networks, in: 2017 IEEE International Conference on Computer Vision (ICCV), IEEE, 2017, pp. 2746–2754.
[32] D. Lin, J. Dai, J. Jia, K. He, J. Sun, Scribblesup, Scribble-supervised convolutional networks for semantic segmentation, in, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 3159–3167.
[33] G. Wang, M.A. Zuluaga, W. Li, R. Pratt, P.A. Patel, M. Aertsen, T. Doel, A.L. David, J. Deprest, S. Ourselin, et al., Deepigeos: a deep interactive geodesic framework for medical image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 41 (7) (2018) 1559–1572.
[34] M. Rajchl, M.C. Lee, O. Oktay, K. Kamnitsas, J. Passerat-Palmbach, W. Bai, M. Damodaram, M.A. Rutherford, J.V. Hajnal, B. Kainz, et al., Deepcut: Object segmentation from bounding box annotations using convolutional neural networks, IEEE Trans. Medical Imaging 36 (2) (2016) 674–683.
[35] D.P. Papadopoulos, J.R. Uijlings, F. Keller, V. Ferrari, Extreme clicking for efficient object annotation, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 4930–4939.
[36] L. Castrejon, K. Kundu, R. Urtasun, S. Fidler, Annotating object instances with a polygon-rnn, in, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 5230–5238.
[37] D. Acuna, H. Ling, A. Kar, S. Fidler, Efficient interactive annotation of segmentation datasets with polygon-rnn++, in: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2018, pp. 859–868.
[38] X. Wang, L. Zhang, H. Roth, D. Xu, Z. Xu, Interactive 3d segmentation editing and refinement via gated graph neural networks, in: International Workshop on Graph Learning in Medical Imaging, Springer, 2019, pp. 9–17.

[39] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, P.S. Yu, A comprehensive survey on graph neural networks, arXiv preprint arXiv:1901.00596 (2019).
[40] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
[41] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7132–7141.
[42] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using rnn encoder-decoder for statistical machine translation, arXiv preprint arXiv:1406.1078 (2014).
[43] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2881–2890.
[44] Y. Boykov, V. Kolmogorov, D. Cremers, A. Delong, An integral solution to surface evolution pdes via geo-cuts, in: European Conference on Computer Vision, Springer, 2006, pp. 409–422.
[45] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European conference on computer vision (ECCV), 2018, pp. 801–818.
[46] K.-K. Maninis, S. Caelles, J. Pont-Tuset, L. Van Gool, Deep extreme cut: From extreme points to object segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 616–625.
[47] X. Chen, B.M. Williams, S.R. Vallabhaneni, G. Czanner, R. Williams, Y. Zheng, Learning active contour models for medical image segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 11632–11640.
[48] J.I. Orlando, H. Fu, J.B. Breda, K. van Keer, D.R. Bathula, A. Diaz-Pinto, R. Fang, P.-A. Heng, J. Kim, J. Lee, et al., Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs, Med. Image Anal. 59 (2020) 101570.



Zhiqiang Tian is an associate professor at Xi'an Jiaotong University. He received the B.S. degree in Automation Control from the Northeastern University in 2004, the M.S. and Ph.D. degrees in Control Science and Engineering from Xi'an Jiaotong University in 2007 and 2013, respectively. He was a postdoctoral fellow in the Department of Radiology and Imaging Sciences of Emory University from 2014 to 2017. His research interests are image/video processing, computer vision, multimedia, and medical image analysis.

Xiaojian Li is a master student at the School of Software Engineering in Xi'an Jiaotong University. He received the B.S. degree in software engineering from the Beihua University in 2017. His research interests include computer vision, machine learning, and medical image analysis.

Zhang Chen is a master student at the School of Software Engineering in Xi'an Jiaotong University. He received the B.S. degree in software engineering from Xi'an Jiaotong University in 2018. His interests include semantic segmentation, medical image analysis.

Yaoyue Zheng is a master student at the School of Software Engineering in Xi'an Jiaotong University. She received the B.S. degree of software engineering in Xi'an Shiyou University in 2018. Her research interests include machine learning, computer vision and medical image analysis.

Hongcheng Fan is undergraduate and majors in software engineering in Xi'an Jiaotong University. Her research interests include computer vision, machine learning, and medical image analysis.

Zhongyu Li received the BE and ME degree from Xi'an Jiaotong University, China and the Ph.D. degree in computer science from the University of North Carolina at Charlotte, United States in 2012, 2015 and 2018, respectively. Currently, he is an assistant professor in the School of Software Engineering at Xi'an Jiaotong University. His research interests include computer vision and medical image analysis.

Ce Li received his Ph.D. degree in pattern recognition and intelligence system from Xi'an Jiaotong University, China in 2013. He is a professor at the College of Electrical and Information Engineering, Lanzhou University of Technology. His research interests include computer vision and pattern recognition.

Shaoyi Du received B.S. degrees both in computational mathematics and in computer science, M.S. degree in applied mathematics and Ph.D. degree in pattern recognition and intelligence system from Xi'an Jiaotong University, China in 2002, 2005 and 2009 respectively. He worked as a postdoctoral fellow in Xi'an Jiaotong University from 2009 to 2011 and visited University of North Carolina at Chapel Hill from 2013 to 2014. He is currently a professor of Institute of Artificial Intelligence and Robotics in Xi'an Jiaotong University. His research interests include computer vision, machine learning and pattern recognition.