Interactive Medical Image Segmentation via A Point-Based Interaction

Jian Zhang, Yinghuan Shi, Jinquan Sun, Lei Wang, Luping Zhou, Yang Gao, Dinggang Shen

PII: S0933-3657(20)31263-X

DOI: https://doi.org/10.1016/j.artmed.2020.101998

Reference: ARTMED 101998

To appear in: *Artificial Intelligence In Medicine*

Received Date: 24 April 2020

Revised Date: 5 October 2020

Accepted Date: 23 November 2020

Please cite this article as: Jian Zhang, Yinghuan Shi, Jinquan Sun, Lei Wang, Luping Zhou, Yang Gao, Dinggang Shen, Interactive Medical Image Segmentation via A Point-Based Interaction, *<![CDATA[Artificial Intelligence In Medicine]]>* (2020), doi: https://doi.org/

Interactive Medical Image Segmentation via A Point-Based Interaction

Jian Zhang[a,*], Yinghuan Shi[a,b,*], Jinquan Sun[a], Lei Wang[c], Luping Zhou[d], Yang Gao[a,b,**], Dinggang Shen[e,f,**]

[a] State Key Laboratory for Novel Software Technology, Nanjing University, China
[b] National Institute of Healthcare Data Science, Nanjing University, China
[c] School of Computing and Information Technology, University of Wollongong, Australia
[d] School of Electrical and Information Engineering, University of Sydney, Australia
[e] Shanghai United Imaging Intelligence Co., Ltd., China
[f] Department of Artificial Intelligence, Korea University, Republic of Korea

ARTICLE INFO

Article history:

Keywords: Point-based interaction, Sequential patch learning, Medical image segmentation

ABSTRACT

Due to low tissue contrast, irregular shape, and large location variance, segmenting the objects from different medical imaging modalities (*e.g.*, CT, MR) is considered as an important yet challenging task. In this paper, we present a novel method for interactive medical image segmentation with the following merits. (1) Our design is fundamentally different from previous pure patch-based and image-based segmentation methods. We observe that during delineation, the physician repeatedly check the intensity from area inside-object to outside-object to determine the boundary, which indicates that *comparison in an inside-out manner is extremely important*. Thus, we innovatively model our segmentation task as learning the representation of the bi-directional sequential patches, starting from (or ending in) the given central point of the object. This can be realized by our proposed ConvRNN network embedded with a gated memory propagation unit. (2) Unlike previous interactive methods (requiring bounding box or seed points), we only ask the physician to merely click on the rough central point of the object before segmentation, which could simultaneously enhance the performance and reduce the segmentation time. (3) We utilize our method in a multi-level framework for better performance. We systematically evaluate our method in three different segmentation tasks, including CT kidney tumor, MR prostate, and PROMISE12 challenge, showing promising results compared with state-of-the-art methods.

1. Introduction

Accurate segmentation of different objects from medical imaging data (*e.g.,* MR, CT) is normally believed as one of the most significant steps for clinical treatment. However, traditional segmentation is often performed manually by the physician, which is extremely time-consuming. Also, inaccurate manual segmentation could not be avoided due to factors like fatigue. Moreover, different physicians sometimes provide different segmentation according to her/his own experience. Thus, automatic or semi-automatic segmentation methods to quickly and precisely obtain the object boundary are in urgent demand.

As deep learning goes popular in image segmentation tasks Chen et al. (2016); Long et al. (2015); Zhao et al. (2017), lots of related attempts have been conducted for medical image segmentation. Ciresan et al. (2012) trained a patch-based network to predict the class label of the central point in each patch and won the EM segmentation challenge at ISBI 2012 by a large margin. Havaei et al. (2017) used a two-pathway deep CNN to exploit both local features and global contextual features to improve segmentation. To achieve good localization and the use of context, Ronneberger et al. (2015) used skip connection to recover details lost by deep layers, which was generalized to other medical image segmentation tasks Çiçek et al. (2016); Milletari et al. (2016); Yu et al. (2017) with promising results.

Different from these deep learning methods, during manual delineation, the physician first locates the object roughly and keeps the appearance in mind, then repeatedly checks the intensity from the area inside the object to outside the object to delineate the boundary precisely especially when the images have blurry boundary. This reveals that the inside-out comparison is important for medical image segmentation. Unfortunately, existing deep learning-based methods directly classify

[*] Jian Zhang and Yinghuan Shi are the first co-authors.

[**] Corresponding author: Yang Gao and Dinggang Shen. The work was supported by the National Key R&D Plan (2019YFC0118301) and NSFC (61673203).

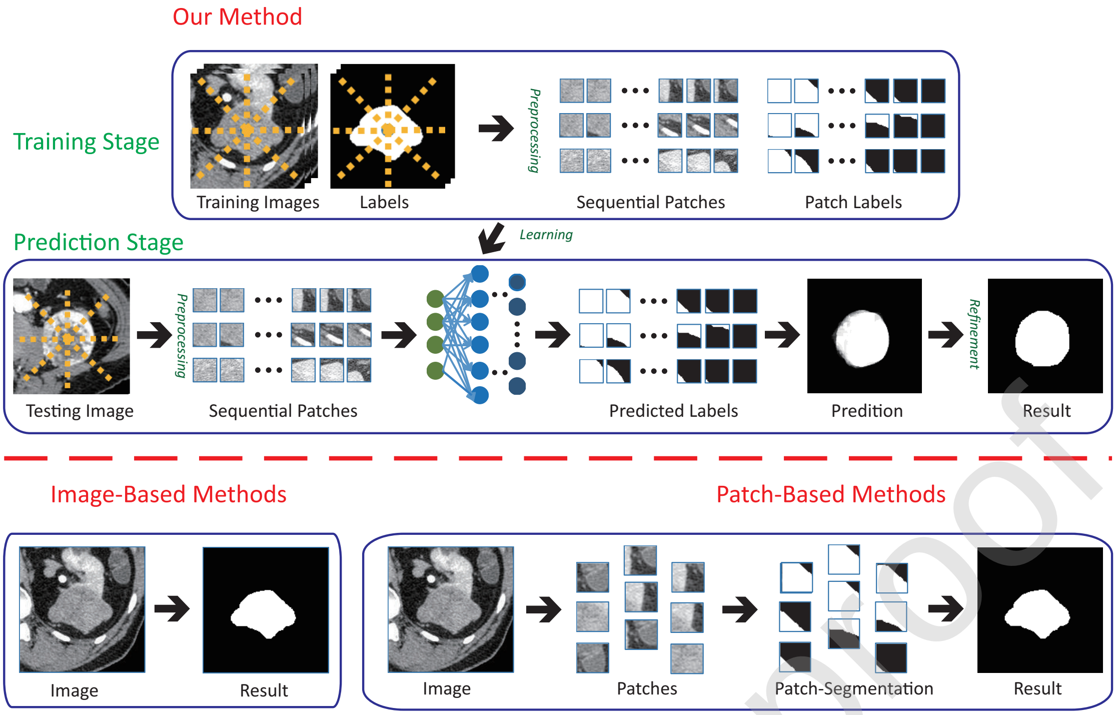

Fig. 1. Framework of the method for interactive segmentation and the difference of the method from previous methods. the method extracts sequential patches with 4 or 6 strides along 16 rays extending from the inside of the object to outside (Rays are the lines starting from the center of the object to the outside of the object, along which patches are extracted. For better visualization, only 8 rays and sparsely sampled patches are shown). A sequence of patches along a certain ray is viewed as a single training sample. The image-based method takes the whole image as input and outputs its corresponding prediction directly. The patch-based method takes each patch as an individual sample.

each pixel/voxel after several convolution and pooling operations, which usually leads to inferior segmentation result on the cases with low contrast and blurry boundary.

In this paper, inspired by the above observation, a novel interactive segmentation method is proposed for medical images, by incorporating point-based interaction and bi-directional sequential patch learning. In the method, the physician is first asked to take a few seconds to click on the rough central point of the object before segmentation (the point is named as the central point in this paper). With this click, an approximate location of the object will be initially determined. As observed in Fig. 1, there are many rays (based on a rough central point given by the physician), extending from the area inside of the object to outside. The method extracts patches with 4 or 6 strides along these rays resulting in several sequences of patches (it is named as sequential patches in this paper). Normally, in a certain sequence of patches extracted along a ray (the line starting from the center of the object to the outside of the object), with a very high probability, the first several patches along the rays are inside the object and the last several patches are outside the object. Also, the patches along the direction of the rays change sequentially. Thus, the feature representation of these sequential patches is learned to capture the shape and appearance changing from both the object-to-background and background-to-object directions, which is largely different from the previous pure patch-based and image-based segmentation methods (as shown in Fig.1). Following this idea, this paper innovatively models the segmentation task as learning the representation of the bi-directional sequential patches, according to the given central point. To be specific, a U-net likeRonneberger et al. (2015) model is built with sequential learning ability. With the help of designed bi-directional gated ConvRNN modules embedded in different feature levels, the model takes sequential patches as input, thus the receptive field of all the layers (even the shallower layers) could be increased. In other words, the receptive field increases via making the network goes wide, instead of deep. Besides, the bi-directional gated ConvRNN enables the model to capture spatial relation among adjacent patches and exploit more crucial details via inside-out comparison as what the physician does during segmentation. It is worth noting that the ConvRNN module in deeper layers could capture the spatial relation among patches in a general view, while the ConvRNN units in shallower levels will encode the detailed information for accurate segmentation. In brief, the work makes the following contributions:

- A simple point-based interactive method is proposed to improve the performance and meanwhile reduce the segmentation time. This point-based interaction could be naturally integrated with the state-of-the-art deep learning method.
- Compared to traditional deep learning segmentation methods, the method introduces inside-out comparison into the network, which has not been touched by previous works.
- This paper designs a simple yet effective ConvRNN mod-

ule. the model regards a sequence of patches as a single training sample and captures spatial relation among adjacent patches, which is different from patch-based methods (as shown in Fig. 1). Besides, the ConvRNN module could help the shallow network increase the receptive field to capture more context.

- The model is light (5.2MB), easy to train (3 hours on PROMISE12 dataset via Nvidia 1080), fast to test (80ms/slice on PROMISE12 testing dataset).

2. Related Work

Compared with natural image segmentation, medical image segmentation is required to face more challenges, *e.g.*, low tissue contrast, irregular shape, and large location variance, *etc*. Thus, traditional methods, specially developed for natural image segmentation, could not be directly applied to segment the medical objects. Previous segmentation methods can be roughly classified into two categories.

The first class is registration/deformable model-based methods. Yang et al. (2014) integrated the spatio-temporal information into a traditional registration-based method to perform effective 4-D MRI thoracic segmentation. Liao and Shen (2012) employed both patient-specific information and population by using anatomical feature selection and an online updating mechanism. In Zhuang et al. (2010), locally affine registration method (LARM) was proposed to provide the correspondence of anatomical substructures, and the free-form deformations with adaptive control point status were performed to refine local details. Mesejo et al. (2016) imposed the anatomical constraints during the model deformation procedure. Yan et al. (2010) extracted boundary as a robust shape prior to guide a promising segmentation. Gao et al. (2016) proposed to learn a displacement regressor which can provide a non-local external force for each vertex of the deformable model.

The second class is learning-based methods, which have received considerable attention in recent years, including the patch-based and image-based methods. Patch-based methods usually regard each patch as the input and predict its label as the output. For example, Li et al. (2015) designed a CNN model to predict the label of the input patch and combined the patch-level results to generate the final segmentation of liver tumors. Zhao and Jia (2016) integrated local and global region features into consideration and thus proposed a multi-scale convolution network to segment the brain tumor. To utilize the global context, as the image-based methods, whole image-based deep models have also been explored. U-net Ronneberger et al. (2015) is one of the representative image-based methods. Several variants have been also proposed to segment different objects in medical images Clark et al. (2017); Yu et al. (2017). Also, note that several learning-based detection methods are quite relevant to the segmentation task, *e.g.,* Yang et al. (2017); Wang et al. (2018). the method borrows the advantages from both image-based and patch-based methods, by modeling the sequential patches as aforementioned.

Moreover, from the perspective of interactive segmentation, the method belongs to the point-based interaction. Point-based interactive segmentation models have aroused increasing interest very recently. There are also several traditional methods, such as live-wire and level set. Live wire-based methods Grady (2008); Poon et al. (2008); Zewei et al. (2014) produce edge segmentation between two points provided by user, while level set-based methods Li et al. (2010) start from a specified contour to segment image by the evolution of level set function. However, how to effectively integrate point-based interaction with the state-of-the-art deep learning methods is still in its early stage. In Bearman et al. (2016), the points given by humans were regarded as weak supervision label for segmentation. Sun et al. (2017) focused on generating a prior map according to the given point to improve segmentation performance. Furthermore, the RNN has been introduced to image segmentation in recent years. Liang et al. (2016) separated natural images with clear boundaries into different disjoint superpixels according to raw RGB pixels, then employed RNN to capture relations among superpixels. Cai et al. (2017) and Zhu et al. (2018) used RNN to capture the relations among adjacent slices. It is claimed that the method is fundamentally different from these previous ones: the point in the method is just an indicator of the object; the method captures the inside-out intensity changing trends with overlapped patches (which may contain both foreground and background) to determine the blurry boundary. Besides, ConvRNN could be embedded into different levels of convnet instead of only the last few layers.

3. Framework and Preprocessing

The whole framework of the method is shown in Fig. 1, including the training and predictions stage. The training stage consists of only sequential patch learning, while the predictions stage contains the preprocessing, sequential patch segmentation, and result refinement.

In the training stage, the preprocessing step aims to extract the sequential patches for the following learning step. Since the ground truth is available during the training stage, for each image, the method first calculates its mass center (of the object) as the initialized point, and then extract the sequential patches along the rays extending from this point, which could fully capture the relation of continuous changing from the inside to outside. Afterward, these obtained sequential patches with their respective labels will be employed to train the proposed model, which will be elaborated in Section 4.

In the predictions stage, to segment a new coming testing image, at first, an initial central point in the testing image will be roughly clicked by the physician, and then the sequential patches (according to the central point) will be extracted. Afterward, these extracted sequential patches will be fed to the trained model to obtain the corresponding likelihood maps. Finally, these likelihood maps will be combined for final segmentation in the result refinement step.

Preprocessing. As shown in Fig. 1, the method extends rays (16 rays in experiment setting with better performance than the case of using 8 rays although more computationally expensive.) from the initial point. For each ray, the method then extracts a sequence of patches with their size as 32×32. Since the spatial

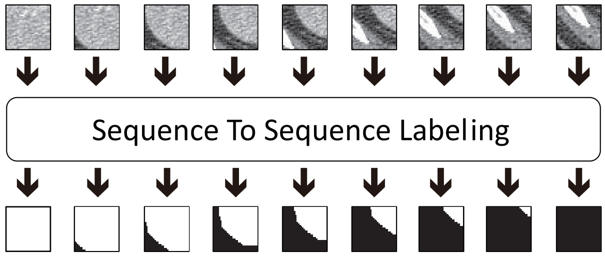

Fig. 2. Illustration of sequential patch-based segmentation problem. The input is a sequence of patches extracted from CT kidney and the output is the corresponding predictions.

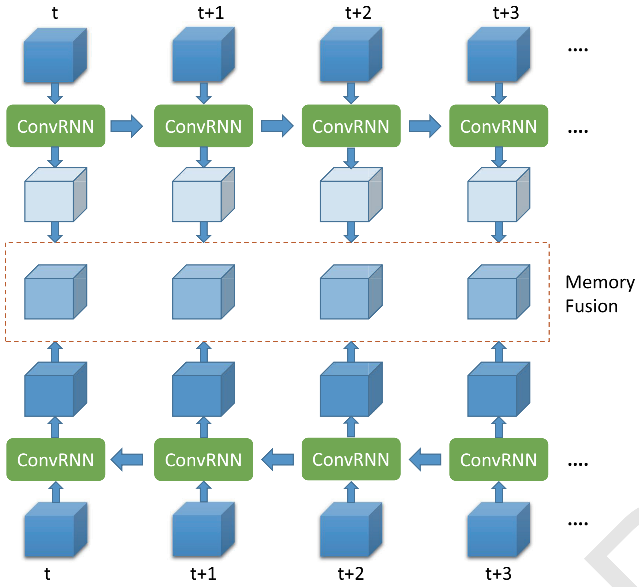

Fig. 3. Illustration of Bi-ConvRNN. The upper ConvRNN unit processes the top feature sequence, while the lower ConvRNN unit processes the bottom feature sequence in an inverse order. Then a Memory Fusion module is employed to fuse two feature sequences.

smoothness among these patches is important in the method, patches with 4 or 6 strides are extracted to guarantee that consecutive patches contain an overlap. Since it is found that the first several patches in a sequence are normally inside the object while the last several patches are normally outside the object, the boundary of the object could appear in the rest several patches in the middle. Therefore, the advantage of the patch extraction strategy is that these patches could globally cover the whole object and also reduce the redundant background regions at the same time.

4. Proposed method

4.1. Problem Analysis

With obtained sequential patches in hand, this paper redefines the sequential patch-based segmentation as the problem of spatial sequence labeling. Normally, a sequence of patches is denoted as $\mathbf{X}$ and thus $\mathbf{X}_t$ indicates the t-th patch in this sequence. The segmentation model takes patch $\mathbf{X}_t$ as input at t-th step and outputs the corresponding segmentation result $\mathbf{Y}_t$ based on its memory over previous patch segmentation and current input patch. The spatial relationship between adjacent patches are encoded in the memory of the segmentation model. Accordingly, the sequential patch-based segmentation can be formulated as a task to train a sequence labeling algorithm $\mathcal{H} : \mathbf{X} \mapsto \mathbf{Y}$, which can assign the most likely label to each patch in the length-K input sequence:

$$\widehat{\mathbf{Y}}_1,\cdots,\widehat{\mathbf{Y}}_K = \mathop{\arg\max}_{\mathbf{Y}_1,\cdots,\mathbf{Y}_K} p(\mathbf{Y}_1,\cdots,\mathbf{Y}_K|\mathbf{X}_1,\cdots,\mathbf{X}_K) \tag{1}$$

It is worth noting that there is a key difference between the sequential patch-based segmentation and the traditional sequence labeling (in natural language process Nguyen and Guo (2007); Ma and Hovy (2016)): as shown in Fig. 2, the output of each patch in a sequence is a 2D matrix instead of a scalar (in traditional sequence labeling). For an input patch with the size of 32 × 32, the number of its possible output pixels can be up to $2^{32\times32}$. Therefore, it is difficult for existing models to well deal with the learning problem due to this large output volume. With a high probability, the first several patches along the rays are inside the object and the last several patches are outside the object. Thus, the model more focuses on propagating label similarity from both ends to the middle patches, which could reduce the large output dimensionality and hence make the model more trainable.

4.2. Gated Memory Propagation Unit

Regarding that the goal here has been formulated as a sequence labeling task, it is a natural way to integrate this sequence labeling task into a recurrent neural network (RNN). As aforementioned, the patch extraction could guarantee the start (or end) patches to be foreground (or background), which could be treated as important knowledge for learning.

As analyzed, the problem is to propagate the label with high confidence from both the ends (*i.e.*, foreground, or background) to the middle patches whose corresponding labels are with low confidence. To tackle this problem, this paper proposes to use ConvRNN cells to build a gated propagation unit, which could capture the spatial relationship and propagate visual memory from the current patch to its next patch. In particular, given a sequence of patches as the input, the propagation unit needs to encode the input into representative features and meanwhile propagate the label memory from the current patch to the next patch. Regarding that the boundary of a medical object often shows low contrast, the propagation should be sensitive to slight intensity changing.

Therefore, this paper proposes a simple yet effective ConvRNN cell as follows:

$$\mathbf{r}_t = \sigma(\mathbf{W}_{xr} * \mathbf{X}_t + \mathbf{W}_{hr} * \mathbf{H}_{t-1} + b_r) \tag{2}$$

$$\mathbf{H}_t = \mathtt{relu}(\mathbf{W}_{xh} * \mathbf{X}_t + \mathbf{r}_t \circ \mathbf{H}_{t-1} * \mathbf{W}_{h\tilde{h}} + b_h) \tag{3}$$

where σ is the sigmoid activation function, $*$, $\circ$ indicate the convolutional operator and element-wise multiplication, respectively. All the input $\mathbf{X}_1,\cdots,\mathbf{X}_t$, hidden state $\mathbf{H}_1,\cdots,\mathbf{H}_t$, and reset gate $\mathbf{r}_t$ are 3D tensors in $\mathbb{R}^{P\times M\times N}$, where P is the number of channels, M is the number of rows, N is the number of columns in feature map, respectively. The most important

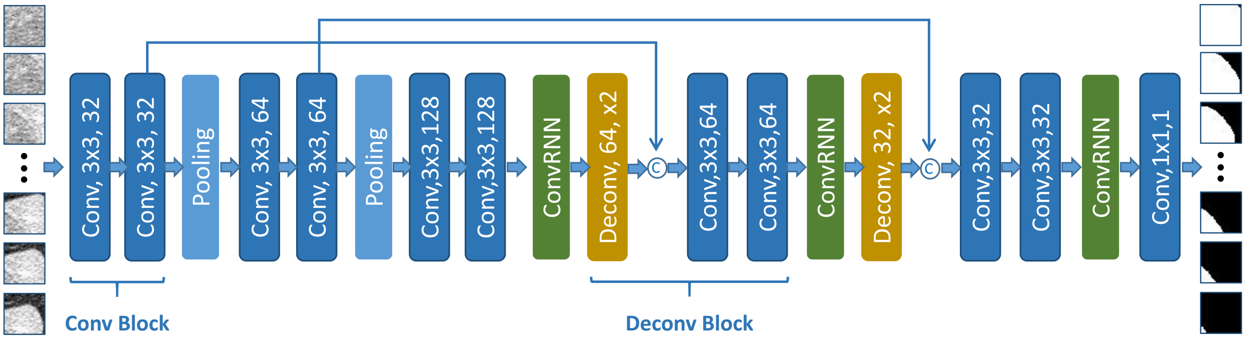

Fig. 4. Details of the model. The high resolution feature maps from contracting path are concatenated with the output of deconvolution operation.

tensor is hidden state $\mathbf{H}_t$ at t-th step, which encodes spatial relation among adjacent patches. The reset gate $\mathbf{r}_t$ measures the correlation between current input and hidden state $\mathbf{H}_{t-1}$, aiming to control that how much information from $\mathbf{H}_{t-1}$ is allowed to flow into current hidden state $\mathbf{H}_t$. If $\mathbf{r}_t$ is close to zero, the memory from previous hidden state $\mathbf{H}_{t-1}$ will be forgotten and meanwhile the new hidden state $\mathbf{H}_t$ will be determined only by current input $\mathbf{X}_t$ to handle the sudden change of intensity. The `relu` Nair and Hinton (2010) activation function is employed to generate the final hidden state instead of `tanh`, since `tanh` is easy to saturate and change the range of value in feature map. The hidden state is used directly as the new representation of $\mathbf{X}_t$.

Given an input sequence starting from the foreground (object) region and ending at the background, a single direction with a memory unit could only propagate the memory in a forward or backward view. However, it is still hard to determine the location of stopping propagation for predicting the boundary of the object. To improve the performance of the visual memory model, this paper innovatively employs bi-directional RNN for sequential patch learning to integrate the information from different two directions. As illustrated in Fig. 3, the first unit processes the patches of a sequence in a forward direction, and the second unit processes the patches of the same sequence in a reversed direction. Notably, the two units show different favors when propagating visual memory: the forward unit tends to classify pixels to be inside the object (classified as foreground), while the backward unit prefers to classify pixels to be outside the object (classified as background). Thus, when only one direction is used, more pixels are classified to be foreground and there is no chance to correct it. After applying the bi-directional RNN, the network can correct its prediction with the backward unit. Finally, A convolution layer with 1 × 1 kernel is employed as a memory fusion module to fuse the different memories.

4.3. Network Structure

Inspired by vanilla U-net Ronneberger et al. (2015), the network also contains a contracting path and an expanding path (see Fig. 4). The contracting path consists of three convolution blocks, in which each block contains two time-distributed convolution layers. The max-pooling operation between two contracting blocks aims to keep the notable features and increase the size of the receptive field of the network. The two deconvolution blocks in expanding paths are employed to increase the size of the feature map.

Compared with common ways of using ConvGRU Tokmakov et al. (2017) and ConvLSTM Xingjian et al. (2015), the method embeds the ConvRNN modules into the different levels of the model. The consideration includes that (1) the features from the deeper layers are more effective to measure the correlation between two patches in a general view because the size of the receptive field will increase when the convolution and pooling operations are employed; (2) the features from the shallower layers contain more details which are useful to delineate the boundary of an object precisely (3) from another perspective, the ConvRNN embedded in the top-level (where the feature maps share the same size with original patch) performs like learning pixel-wise non-linear affinity matrix which proves to be effective for segmentation refinement Liu et al. (2017); Maire et al. (2016). As illustrated in Fig. 4, the feature maps are fed into a gated memory propagation module aiming to learn the spatial relationship between adjacent patches. Then, the output of ConvRNN is upsampled and concatenated with the feature map from the contracting path. In other words, the detailed spatial relation at a certain level is captured by both memories generated in a general view and the feature map from the corresponding level in the contracting path. Since every aforementioned component is differentiable, the model can be trained in an end-to-end way.

The binary cross-entropy is employed as a loss function in the model. Adam Kingma and Ba (2014) is applied with $\beta_1 = 0.95, \beta_2 = 0.99$ to optimize the model. Besides, weight decay is employed to avoid overfitting, which is set to 1e-4. The learning rate, batch size, epoch are set to 1e-4, 4, and 30, respectively for all experiments. Considering the small-sized datasets widely exist in the real case, 30 epoch is enough for our model to fit the three datasets used in our work. According to our observation, the over-fitting will occur roughly at 10, 15, 10 epoch for Kidney, PROMISE12, and MR prostate, respectively. As a matter, too many epochs might not help the model to achieve much better performance but have a high risk of over-fitting to affect the generalization ability. At the same time, redundant computation will also be introduced. This paper combines the patch-level segmentation result to the whole

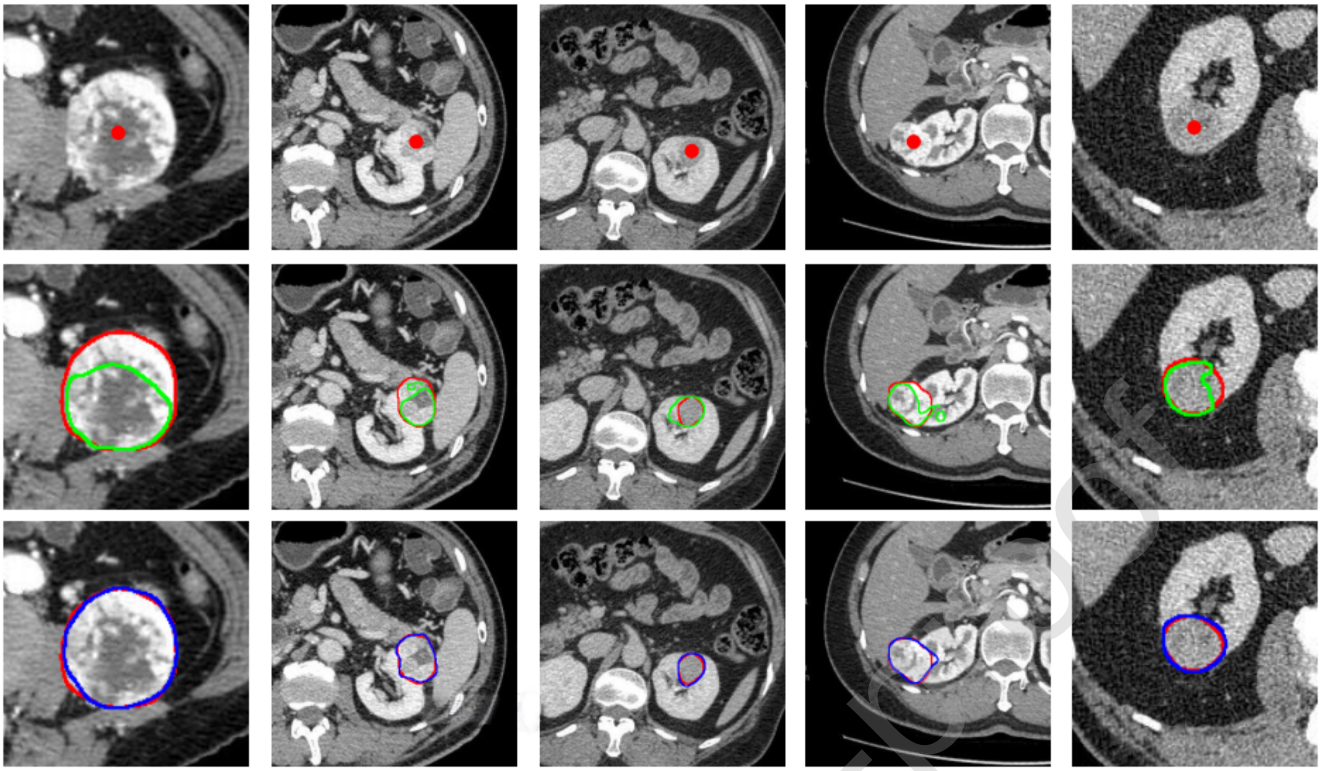

Fig. 5. Typical results of tumor segmentation. (First row) The input images. (Second row) Results of U-net. (Last row) Results of the method. The red dots given by physicians indicates the rough center area of tumor. The red, green and blue curves denote the ground truth, result of U-net and the method, respectively.

slice to generate the final segmentation result in refinement.

5. Results

This paper reports the qualitative and quantitative segmentation results of the method on three medical image segmentation datasets, including CT kidney tumor segmentation, MR prostate segmentation, and PROMISE12 challenge. For the evaluation metrics, the paper employ the Dice Ratio Score (DSC) (%) and Centroid Distance (CD) (mm) along 3 different directions (*i.e.*, lateral x-axis, anterior-posterior y-axis, and superior-inferior z-axis), which are widely used in previous literatures Shi et al. (2015, 2017). Also, for PROMISE12 challenge, several additional metrics (*e.g.*, ABD, 95HD) are reported according to the automatic calculation provided by the challenge organizers Litjens et al. (2014). We extract patches along 16 rays in all experiments. The experiments use one 1080Ti card to train the model and the CPU is Intel Xeon Processor E5-2620 v4 with a base frequency of 2.1 GHz. The memory is 128G. All experiments are implemented with Pytorch.

5.1. CT Kidney Tumor Segmentation

5.1.1. Setting

The CT kidney dataset consists of about 2500 CT slices, scanned from 60 different patients. The resolution of each CT image after image preprocessing is 296 × 296 × 40, with the in-plane voxel size as 1×1 mm^2 and the inter-slice thickness as 1 mm. The manual delineation results are available for all the images which could be considered as the ground truth for performance evaluation. For each input sequential, 15 patches with 4 strides are extracted. For segmenting a new coming testing image, the method first cuts the image into slices, and asks the physician to perform clicks for every slice to indicate the rough center of the tumor. Basically, the method performs 2-fold cross-validation on these 60 CT images where there are in total 1269 and 1260 images for training and testing set, respectively. We adopt the 2-fold cross-validation because we found that the 2-fold has no statistical difference with the 5-fold (the standard deviation for the kidney dataset is only 0.23%). The possible reason is that our method is based on sequential patches where a sufficient number of sequential patches could be generated to make the learning process relatively stable. For a fair comparison, we adopt the same folding for the different reference methods. Note that all slices of the testing patients have not been used for training.

5.1.2. Comparison with Baselines

This paper compares the model with two baselines, *e.g.*, U-net (image-based method), and patch-based FCN. It is worth noting that, for a fair comparison, the U-net in the experiment is designed with enough receptive fields to cover common kidney tumors. For the implementation of patch-based FCN, 1000 patches are extracted for each slice as the following way: (1) densely sampling the patches that are located inside the tumor or centered close to the tumor boundary, and (2) sparsely sam-

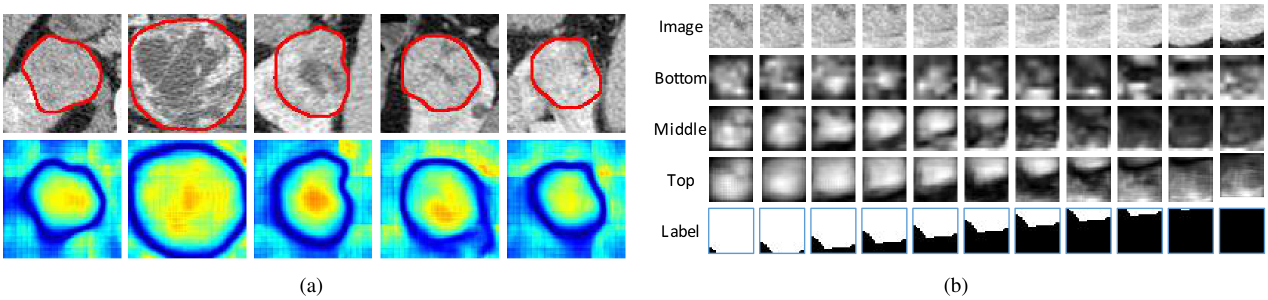

Fig. 6. The visualization of the method. As expected, the method shows the ability to distinguish the object from the background by a large margin. (a) The visualization of fused top-level feature maps. (b) The visualization of multi-level feature maps.

pling the patches that are far from the tumor boundary. There are about 64.5% (about 1,942,791) patches containing the foreground. Table 1 reports the quantitative result. It can be observed that the patch-based FCN performs worst, which is because that treating image patches individually as the direct input of the model might ignore the context information of the whole CT slice. For patch-based FCN, it is difficult to distinguish two patches (inside and outside the tumor) with similar appearance according to its setting. Moreover, U-net Ronneberger et al. (2015) performs better than patch-based FCN. As shown in Fig. 5, U-net can roughly locate the tumor by using the whole image context information. However, it still fails to delineate the tumor precisely when the texture of the tumor is similar to that of the renal parenchyma. Normally, the skip connection in U-net is applied to refine the rough segmentation generated by deeper layers. Unfortunately, there is a shortcoming of this approach: the receptive field of the shallower layers is small and limited (usually 7 × 7 or 15 × 15), thus the learned feature maps are not discriminative enough to determine the boundary in low contrast region. Different from U-net, all the layers in the network keep a large receptive field (even the shallower layers) and exploit crucial inside-out changing trends as what the physician does during delineation. It is worth noting that the model is much lighter (5.2MB) than U-net (135.9MB). Several typical segmentation examples are shown in Fig. 5, showing promising performance compared with U-net. The visualization of the feature maps of the method is also shown in Fig. 4.3. With the help of inside-out comparison, the method shows the ability to be aware of the blurry boundary.

Table 1. Comparison with other methods on CT kidney tumor dataset. The result of patch-based FCN is inferior, thus only the DSC is presented. the method outperforms other methods.

Method	Patch FCN	U-net	Ours
CD-x (mm)	-	0.71	**0.57**
CD-y (mm)	-	0.59	**0.33**
CD-z (mm)	-	0.37	**0.25**
DSC [%]	34.45	86.52	**93.24**

Table 2. Comparison with patch-based FCN, patch-based U-net, a traditional semi-automatic segmentation method DRLSE and a image-based method SegCaps. Two patch-based methods are both trained on patches extracting along the rays extending from the inside of object to the outside. And two image-based methods are also included for the comparison.

Method	DRLSE	Patch FCN	Patch U-net	SegCaps	2PG-CNN	Ours
CD-x (mm)	-	10.12	9.07	-	-	**0.57**
CD-y (mm)	-	9.97	15.19	-	-	**0.33**
CD-z (mm)	-	8.23	7.47	-	-	**0.25**
DSC [%]	60.68	69.73	75.55	87.50	88.20	**93.24**

Besides, the semi-automatic settings for patch-based FCN and patch-based U-net are also considered. Especially, the paper trains the patch-based FCN and patch-based U-net on patches which are extracted from the rays extending from the inside of the object to the outside. In other words, the patches are extracted in the same way as the method, while patch-based FCN and patch-based U-net still take a single patch as a training sample. Moreover, in the predictions stage, the point in the center of the object is given by the physician. Then the patches extracted along the rays are fed into patch-based FCN and patch-based U-net. Another popular-used traditional semi-automatic segmentation method DRLSE Li et al. (2010) is also added for comparison. Note that, the DRLSE does not any training. The alpha and lambda are all set to 3 and the number of the inner and outer loop to 5 and 100, respectively. Two recently proposed methods called SegCapsLaLonde and Bagci (2018) and 2PG-CNNRazzak et al. (2018) are also included for the comparison. Some quantitative results are shown in Table 2. As shown in Fig. 7, although the central point is given by the physician, the patch-based FCN and patch-based U-net still fail to locate the tumor and delineate the boundary of the tumor precisely.

5.1.3. Ablation Study

To systematically study the efficacy of each component in the method, several ablation experiments are conducted. Table 3 reports the corresponding results. Typically, the performance of Vanilla U-net without ConvRNN is inferior to that of the method, with only 86.52% on DSC. Also, the method first embeds a ConvRNN block into the bottom level of the U-net

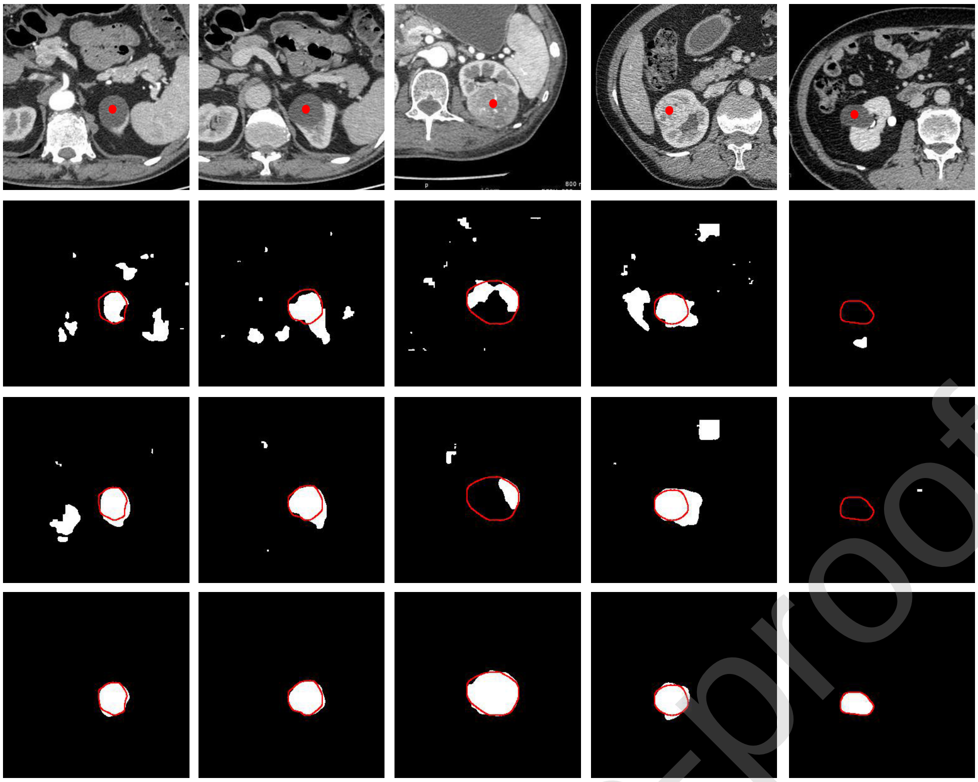

Fig. 7. Typical results on kidney tumor dataset. (First row) The input images. (Second row) Results of patch-based FCN. (Third row) Results of patch-based U-net. (Last row) Results of the method. The red dots are the central points given by the physician. The red curves denote the ground truth delineated by the physician.

Table 3. Ablation study results. Seq-FCN denotes the seq-based FCN. U-net-B, U-net-BM denote the U-net embedded with ConvRNN in the bottom level, both the bottom middle level, respectively.

Method	U-net	Seq-FCN	U-net-B	U-net-LSTM	Single direction	U-net-BM	Ours
DSC [%]	86.52	90.42	91.28	91.51	92.31	92.34	**93.24**

and then feed sequential patches, leading to a 4.76% improvement. The U-net with ConvRNN embedded into both the bottom and middle-level further advances the DSC to 92.34%. As observed, the ConvRNN could successfully capture the spatial relation, incorporate the context to improve the performance. Moreover, this paper modifies the patch-based FCN to seq-based FCN by embedding ConvRNN into its original structure. Compared to patch-based FCN, seq-based FCN achieves substantial improvement on DSC, which reveals the efficacy of sequential patches in this method. Furthermore, this paper evaluates the single direction design of ConvRNN. As shown in Fig. 8, without the spatial relation from the bi-direction and the memory fusion, this architecture leads to a slight drop in DSC. Finally, the paper replaces ConvRNN with ConvGRU Tokmakov et al. (2017) or ConvLSTM Xingjian et al. (2015) to evaluate the influence. Unfortunately, the network with ConvGRU Tokmakov et al. (2017) is hard to train and fails to get a good performance. The variant with ConvLSTM Xingjian et al. (2015) performs a little worse than the method especially on these tumors with blurry boundaries. One possible reason for this problem is that the long time memory encoded in the cell state is not sensitive to slight intensity changing, thus the ConvLSTM Xingjian et al. (2015) unit may get inferior result around the blurry boundary area.

5.1.4. Influence of Initial Click Location

As an interactive segmentation method, the initial click location is very important. To evaluate how different initial click locations (different coordinates) affect the segmentation result, some points are selected near the central point and test their dice values, since it is a rare case for an experienced physician to click a very incorrect central point of the object. The relative Dice Ratio Score to the current setting is shown to emphasize its effect.

As Table 5 conveys, the coordinates of the different initial points really influence the result. Specifically, (0,0) indicates the ground truth of the central point, and different numerical values in x and y-axis mean different offsets according to the central point. The farther away from the real central point of the tumor, the worse the result will be. Fig. 9 shows an example

Table 4. Influence of hyper-parameters and noisy label. (size, number, stride) denotes patch size, sequence number and patch stride respectively. * indicates the network is trained with noisy label.

(size, number, stride)	(16, 15, 4)	(32, 15, 4)	(48, 15, 4)	(32, 5, 4)	(32, 10, 4)	(32, 15, 8)	(32, 15, 12)	(32, 15, 4)*
Relative DSC [%]	-3.87	0.00	+0.68	-0.37	-0.13	+0.51	+0.19	-0.02

Table 5. Influence of Initial Click Location. (x, y) is the relative offset to the ground truth of the central point.

Click Offset	(-5, -5)	(0, -5)	(5, -5)	(-5, 0)	(0, 0)	(5, 0)	(-5, 5)	(0, 5)	(5, 5)
Relative DSC [%]	+0.43	+0.50	−0.16	−0.02	0.00	−0.32	−0.36	−0.33	−1.48

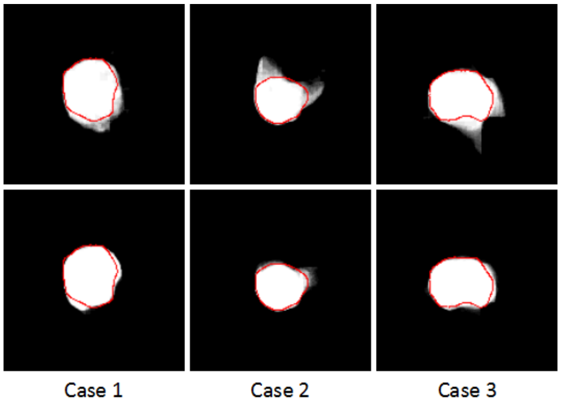

Fig. 8. Tumor-likelihood maps generated by single directional architecture and bi-directional architecture. (First row) Results under single direction setting. (Second row) Results under bi-directional setting.

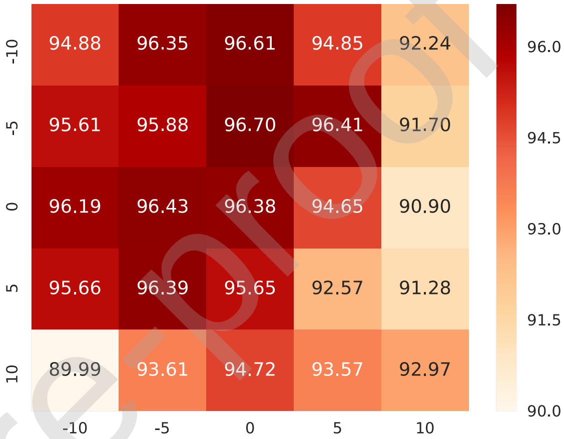

Fig. 9. Illustration of DSC with different initial coordinates.

of a dice range on a single slice. However, it is noteworthy that the method still can obtain robust results and also outperform vanilla U-net.

5.1.5. Influence of hyper-parameters

There are three important hyper-parameters in the method. The paper tests the influence of different patch sizes, sequence numbers, and patch strides. The combination of them is denoted as (patch size, sequence number, patch stride). The relative Dice result to the current setting is shown in Table 4. It can be observed that the larger patch size, longer sequence size, and appropriate patch stride, usually help produce a better segmentation result (*i.e.*, (48, 15, 4) achieves the best performance). However, this combination usually increases the computational burden. In future work, the author will explore the adaptive method to trade-off performance and complexity.

5.1.6. Influence of noisy labels

Sometimes the ground truth provided by the expert might not be perfect due to the factor from inter- and intra- observation, which may significantly degenerate the performance. Therefore, this paper evaluates the method in a simulated noisy label situation. For the original ground truth segmentation maps of the training images provided by the expert, the noisy label case is simulated by using random dilation and erosion operators. Specifically, for a training image, its ground truth map is with the probability of 0.5 to be preprocessed by the dilation operator, and 0.5 to be preprocessed by the erosion operator. The dilation and erosion operator are all performed 3 times. Note that, for all the testing images, they still use the ground truth as the evaluation for a fair comparison. The result is shown in Table4 denoted by $*$. It shows that the method is robust to the relative imperfect labels.

Table 6. Comparison with other methods on MR prostate dataset. the method achieves best performance.

Method	Patch FCN	APSL	U-net	Ours
CD-x (mm)	1.91	0.74	0.76	**0.47**
CD-y (mm)	1.31	0.49	0.56	**0.28**
CD-z (mm)	0.57	0.27	0.19	**0.12**
DSC [%]	71.35	84.00	84.69	**90.91**

5.2. MR Prostate Segmentation

5.2.1. Setting

The method is also validated on an MR prostate segmentation dataset. This dataset is collected by scanning 22 different patients. The ground truth of the prostate region is manually delineated by the physician for comparison. The resolution of MR images after preprocessing is 193 × 153 × 60. 11 patients

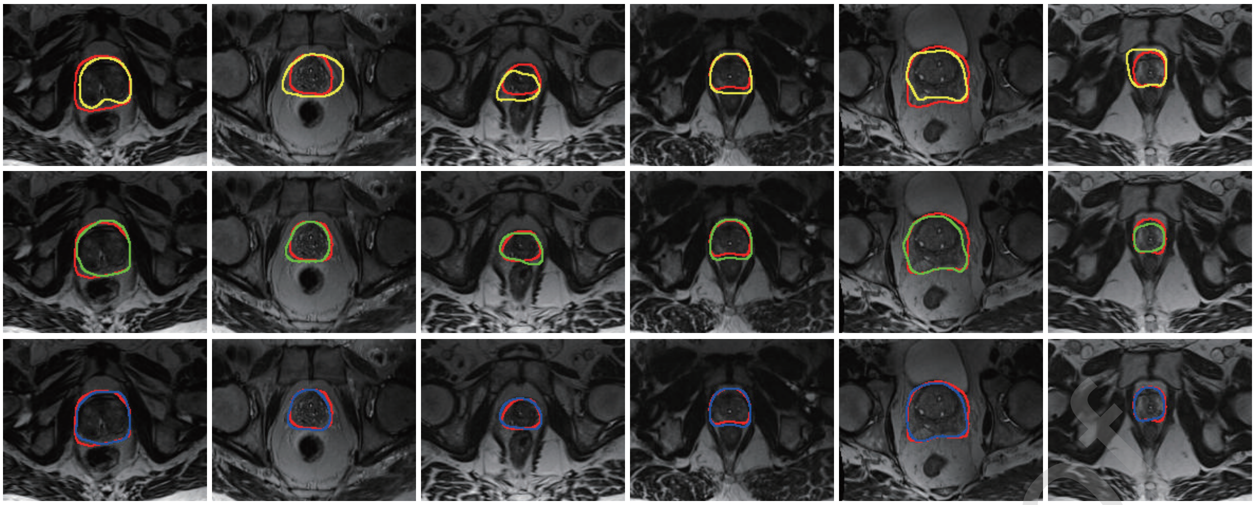

Fig. 10. Typical results. (First row) Results of patch-based FCN. (Second row) Results of U-net. (Last row) Results of the method. The red curve denotes the ground truth. The yellow, green, blue curves denote the results of patch-based FCN, U-net and the method, respectively.

Table 7. Comparison with methods proposed by other competitors.

Method	ABD [mm]			95HD [mm]			DSC [%]			aRVD [%]			Score
	Whole	Base	Apex	Whole	Base	Apex	Whole	Base	Apex	Whole	Base	Apex	
CAMP-TUM2	2.23	2.46	2.03	5.71	5.84	4.62	86.91	84.31	85.40	14.98	20.84	21.21	82.39
OncoB	2.20	2.20	2.47	5.85	**5.15**	5.65	86.99	86.09	79.24	13.35	14.74	28.65	82.72
UdeM 2D	2.17	2.39	2.07	6.12	6.44	4.71	87.42	84.93	84.16	12.37	18.37	21.90	83.02
ScrAutoProstate	2.13	2.23	2.18	5.58	5.60	4.93	87.45	86.30	83.47	13.56	14.46	23.78	83.49
Emory	2.14	2.65	2.41	5.94	5.45	4.73	87.99	86.06	84.53	8.64	15.70	20.32	83.66
Imorphics	2.10	2.18	1.96	5.94	5.45	4.73	87.99	86.06	84.53	11.65	13.33	20.75	84.36
methinks	2.06	**2.11**	2.01	5.53	5.45	4.62	87.91	86.79	84.58	8.71	**10.84**	21.21	85.41
CREATIS	1.93	2.14	1.74	5.59	5.62	**4.22**	89.33	86.60	86.77	9.20	14.65	16.64	85.74
CUMED	1.95	2.13	1.74	5.54	5.41	4.29	89.43	86.42	**86.81**	**6.95**	11.04	**15.18**	86.65
MedicalVision	**1.79**	**2.11**	2.29	**5.35**	6.20	5.86	**89.81**	**87.49**	81.74	8.24	11.52	19.40	85.33

are randomly selected as the training set and use the remaining patients as the testing set, which results in 390 and 387 images respectively. All the parameters in the model keep the same as that in segmenting CT kidney tumors. Compared to the tumor, the size of the prostate is relatively smaller, thus, for each input sequential patches, 8 patches are extracted with 4 strides. There are about 57.2% (about 389,981) patches containing the foreground for the training of patch-based FCN.

5.2.2. *Results*

Fig. 10 illustrated several typical results of the method on MR prostate segmentation. As observed, the method can precisely segment the prostate boundary compared with two baselines, even for these difficult cases (*i.e.*, top and bottom slices). The numerical results of the method, patch-based FCN, a state-of-the-art interactive method denoted as APSL Sun et al. (2017) and U-net are also shown in Table 6, showing the best performance achieved by the method on both DSC and CD.

5.3. *PROMISE12 Prostate Segmentation Challenge*

5.3.1. *Setting*

The method is also evaluated on MICCAI Prostate MR Image Segmentation (PROMISE12) challenge dataset, which is usually considered as the most famous challenge in MR prostate segmentation. In PROMISE12, the training dataset consists of 50 T2-weighted MR images of the prostate with the corresponding ground truth. The organizers also provide additional 30 MR images as testing without the ground truth. Thus, after segmentation, the results of the testing images are submitted to the organizers, and the organizers calculated the quantitative results. Specifically, beyond DSC, the percentage of the absolute difference between the volumes (ABD), the average over the shortest distances between the boundary points of the volumes (aRVD), and the 95% Haussdorf distance (95HD) are employed as the evaluation metrics. Since these MR images are collected from different medical institutions, there is a large variance of voxel spacing among different MR images.

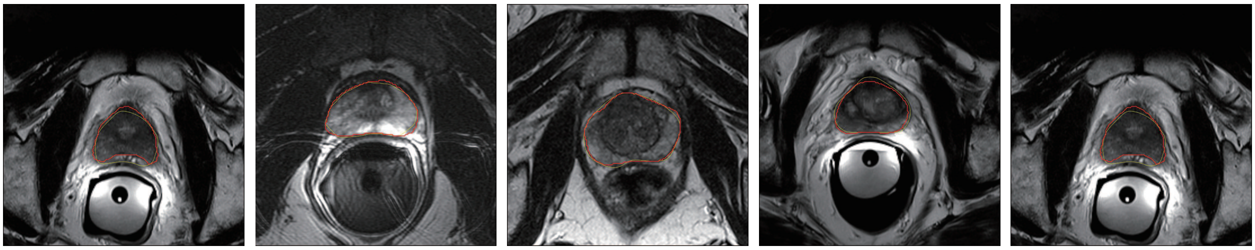

Fig. 11. Several typical results of the method on testing dataset. The yellow and red curve denote the ground truth and the results, respectively.

To reduce the influence of spacing, the spacing in the z-axis is kept unchanged, and resize these images into 0.625 × 0.625 mm in the x-axis and y-axis. As Fig. 1 shows, the sequential patches are extracted from 16 different rays extending from the inside of the prostate to the outside with the stride as 6. Also, each sequence consists of 12 patches with a size of 32 × 32.

5.3.2. Results

Fig. 11 shows several typical results of the method on the testing dataset. As observed, the method can localize the prostate and delineate the boundary accurately.

10 teams are listed in PROMISE12 in Table 7. Seven out of ten teams utilized deep learning methods (ScrAutoProstate, Emory, and Imorphics are traditional methods). They all performed segmentation on 3D volumes, except for UdeM2D Drozdzal et al. (2017) and ours. the method achieves competitive results compared to the state-of-the-art methods: CUMED Yu et al. (2017). Also, the method obtains the promising performance on several metrics (*i.e.*, Whole ABD, Base ABD, Whole 95HD, Whole DSC, and Base DSC), which shows the effectiveness of the method.

Compared to the 3D methods, the method can be directly borrowed to deal with 2D medical image segmentation tasks. Also, the method has its potential to extend to its 3D version by fusing the segmentation results from different directions. Moreover, please note that, for the point-based interaction, the user clicks the central point according to their own experience. This will bring the noise especially for the top and bottom slices (Whole DSC and Base DSC of the proposed method are the best, while Apex DSC is worse than other methods).

6. Discussion

Several issues are discussed in this section. First, in this paper, the method is illustrated in its 2D case rather than the 3D case with the following considerations.

1. The way of point-based interaction: the method requires an initial clicked point roughly located at the center of the object. As known, to localize and click the rough center point on a 2D (MR or CT) slice is easier than on a 3D volume. Thus, the center point-based interaction in the 2D image is more feasible and reliable than that in the 3D image. Also, the central points in different slices might have very different locations. It is reasonable to assign each slice with its own central point.
2. The calculation of different rays: Compared with the 2D case, the calculation of rays in the 3D case requires more parameters to be pre-defined before training the segmentation model.
3. The generalization of 2D and 3D cases: Several medical image segmentation tasks are 2D-based, (*e.g.*, histopathology image segmentation). the method keeps the generalization ability to traditional 2D-based segmentation. Also, the method can be extended to its 3D case by fusing the 2D results from different directions.

Second, there is an interesting idea that uses smaller patches closer to the selected point and larger patches away from the point. However, it is not implemented in this work due to the following consideration:

1. The sizes of the receptive field in the same ray are different which might make the inside-out comparison on the same scale is difficult to realize.
2. Using different sizes of the patch for inside-out modeling might introduce more parameters to set before training.

Third, in the experiment, the organs (*e.g.*, prostate, kidney tumor) belong to the circular shape. In this case, the inside-out comparison between different projection lines is consistent since their distances from the clicked point to the boundary are roughly the same. However, for the other object (*e.g.*, spine or pancreas), the complex and irregular shapes might pose a challenge to our method. Actually, in our previous studies, we have indeed investigated our method on various complex shapes (e.g., vessel, brain), where the results greatly depend on the specific segmentation tasks. We would like to clarify: 1) In our method, the central point was required to be clicked first before segmentation. However, in some irregular shapes (e.g., tree-shaped), the central point was very hard to determine. 2) For very irregular shape, we tried to click more than one points before segmentation and the results could be improved by using more and more initial points but this strategy is case by case. Also, this setting increases the additional computational burden and goes against our original goal, *i.e.*, a simple click to enhance the performance. 3) Compared with automatic segmentation methods that usually build an end-to-end learning model, semi-automatic (interactive) segmentation methods are able to incorporate the user's hint (e.g., point, bounding box, scribbles) to improve the results or make the segmentation model pay more attention to hard regions. Thus, it is very difficult to design a general interactive model for all complex structures.

Finally, to improve the robustness of the segmentation methods to the imperfect label, the error tolerance-based modeling will be investigated to obtain a robust segmentation. For example, a module or block could be added before the final output layer to impose the error tolerance-based constraint. An alternative is to generate several fake ground truths to enhance the ability to label noise. They will be explored in future work.

7. Conclusion

In this paper, a novel method is presented for medical image segmentation using sequential patches, by imitating inside-out comparison as what the physician does during manual delineation. During the segmentation, the method first asks the physician to take a few seconds to click on the rough central point of the object. Then, according to this point, the method extracts different sequential patches and hence train a sequential patch learning model for prediction, by designing a specific gated memory propagation unit. Finally, as U-net, a multi-layer architecture is performed. Besides, the model is light (5.2MB), easy to train, and fast to test. This paper evaluates the method of CT kidney tumor segmentation, MR prostate segmentation, and PROMISE12 challenge. The results are promising.

References

Bearman, A., Russakovsky, O., Ferrari, V., Fei-Fei, L., 2016. What's the point: Semantic segmentation with point supervision, in: European Conference on Computer Vision, Springer. pp. 549–565.

Cai, J., Lu, L., Xie, Y., Xing, F., Yang, L., 2017. Improving deep pancreas segmentation in ct and mri images via recurrent neural contextual learning and direct loss function. arXiv preprint arXiv:1707.04912 .

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2016. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. arXiv preprint arXiv:1606.00915 .

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3d u-net: learning dense volumetric segmentation from sparse annotation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 424–432.

Ciresan, D., Giusti, A., Gambardella, L.M., Schmidhuber, J., 2012. Deep neural networks segment neuronal membranes in electron microscopy images, in: Advances in neural information processing systems, pp. 2843–2851.

Clark, T., Wong, A., Haider, M.A., 2017. Fully deep convolutional neural networks for segmentation of the prostate gland in diffusion-weighted mr images, in: Machine Learning for Medical Image Computing, 14th Int. Conf. on Image Analysis and Recognition (ICIAR), Springer. pp. 1–8.

Drozdzal, M., Chartrand, G., Vorontsov, E., Di Jorio, L., Tang, A., Romero, A., Bengio, Y., Pal, C., Kadoury, S., 2017. Learning normalized inputs for iterative estimation in medical image segmentation. arXiv preprint arXiv:1702.05174 .

Gao, Y., Shao, Y., Lian, J., Wang, A.Z., Chen, R.C., Shen, D., 2016. Accurate segmentation of ct male pelvic organs via regression-based deformable models and multi-task random forests. IEEE transactions on medical imaging 35, 1532–1543.

Grady, L., 2008. Minimal surfaces extend shortest path segmentation methods to 3d. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 321–334.

Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.M., Larochelle, H., 2017. Brain tumor segmentation with deep neural networks. Medical image analysis 35, 18–31.

Kingma, D., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 .

LaLonde, R., Bagci, U., 2018. Capsules for object segmentation. arXiv preprint arXiv:1804.04241 .

Li, C., Xu, C., Gui, C., Fox, M.D., 2010. Distance regularized level set evolution and its application to image segmentation. IEEE transactions on image processing 19, 3243–3254.

Li, W., Jia, F., Hu, Q., 2015. Automatic segmentation of liver tumor in ct images with deep convolutional neural networks. Journal of Computer and Communications 3, 146.

Liang, X., Shen, X., Feng, J., Lin, L., Yan, S., 2016. Semantic object parsing with graph lstm, in: European Conference on Computer Vision, Springer. pp. 125–143.

Liao, S., Shen, D., 2012. A feature-based learning framework for accurate prostate localization in ct images. IEEE transactions on image processing 21, 3546–3559.

Litjens, G., Toth, R., van de Ven, W., Hoeks, C., Kerkstra, S., van Ginneken, B., Vincent, G., Guillard, G., Birbeck, N., Zhang, J., et al., 2014. Evaluation of prostate segmentation algorithms for mri: the promise12 challenge. Medical image analysis 18, 359–373.

Liu, S., De Mello, S., Gu, J., Zhong, G., Yang, M.H., Kautz, J., 2017. Learning affinity via spatial propagation networks, in: Advances in Neural Information Processing Systems, pp. 1519–1529.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431–3440.

Ma, X., Hovy, E., 2016. End-to-end sequence labeling via bi-directional lstm-cnns-crf. arXiv preprint arXiv:1603.01354 .

Maire, M., Narihira, T., Yu, S.X., 2016. Affinity cnn: Learning pixel-centric pairwise relations for figure/ground embedding, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 174–182.

Mesejo, P., Ibáñez, O., Cordón, O., Cagnoni, S., 2016. A survey on image segmentation using metaheuristic-based deformable models: state of the art and critical analysis. Applied Soft Computing 44, 1–29.

Milletari, F., Navab, N., Ahmadi, S.A., 2016. V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: 3D Vision (3DV), 2016 Fourth International Conference on, IEEE. pp. 565–571.

Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines, in: Proceedings of the 27th international conference on machine learning (ICML-10), pp. 807–814.

Nguyen, N., Guo, Y., 2007. Comparisons of sequence labeling algorithms and extensions, in: Proceedings of the 24th international conference on Machine learning, ACM. pp. 681–688.

Poon, M., Hamarneh, G., Abugharbieh, R., 2008. Efficient interactive 3d livewire segmentation of complex objects with arbitrary topology. Computerized Medical Imaging and Graphics 32, 639–650.

Razzak, M.I., Imran, M., Xu, G., 2018. Efficient brain tumor segmentation with multiscale two-pathway-group conventional neural networks. IEEE journal of biomedical and health informatics 23, 1911–1919.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 234–241.

Shi, Y., Gao, Y., Liao, S., Zhang, D., Gao, Y., Shen, D., 2015. Semi-automatic segmentation of prostate in ct images via coupled feature representation and spatial-constrained transductive lasso. IEEE transactions on pattern analysis and machine intelligence 37, 2286–2303.

Shi, Y., Yang, W., Gao, Y., Shen, D., 2017. Does manual delineation only provide the side information in ct prostate segmentation?, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 692–700.

Sun, J., Shi, Y., Gao, Y., Shen, D., 2017. A point says a lot: An interactive segmentation method for mr prostate via one-point labeling, in: International Workshop on Machine Learning in Medical Imaging, Springer. pp. 220–228.

Tokmakov, P., Alahari, K., Schmid, C., 2017. Learning video object segmentation with visual memory. arXiv preprint arXiv:1704.05737 .

Wang, Z., Liu, C., Cheng, D., Wang, L., Yang, X., Cheng, K.T., 2018. Automated detection of clinically significant prostate cancer in mp-mri images based on an end-to-end deep neural network. IEEE transactions on medical imaging 37, 1127–1139.

Xingjian, S., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.c., 2015. Convolutional lstm network: A machine learning approach for precipitation nowcasting, in: Advances in neural information processing systems, pp. 802–810.

Yan, P., Xu, S., Turkbey, B., Kruecker, J., 2010. Discrete deformable model guided by partial active shape model for trus image segmentation. IEEE

Transactions on Biomedical Engineering 57, 1158–1166.

Yang, X., Liu, C., Wang, Z., Yang, J., Le Min, H., Wang, L., Cheng, K.T.T., 2017. Co-trained convolutional neural networks for automated detection of prostate cancer in multi-parametric mri. Medical image analysis 42, 212–227.

Yang, Y., Van Reeth, E., Poh, C.L., Tan, C.H., Tham, I.W., 2014. A spatiotemporal-based scheme for efficient registration-based segmentation of thoracic 4-d mri. IEEE journal of biomedical and health informatics 18, 969–977.

Yu, L., Yang, X., Chen, H., Qin, J., Heng, P.A., 2017. Volumetric convnets with mixed residual connections for automated prostate segmentation from 3d mr images., in: AAAI, pp. 66–72.

Zewei, Z., Tianyue, W., Li, G., Tingting, W., Lu, X., 2014. An interactive method based on the live wire for segmentation of the breast in mammography images. Computational and mathematical methods in medicine 2014.

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network, in: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 2881–2890.

Zhao, L., Jia, K., 2016. Multiscale cnns for brain tumor segmentation and diagnosis. Computational and mathematical methods in medicine 2016.

Zhu, Q., Du, B., Turkbey, B., Choyke, P., Yan, P., 2018. Exploiting interslice correlation for mri prostate image segmentation, from recursive neural networks aspect. Complexity 2018.

Zhuang, X., Rhode, K.S., Razavi, R.S., Hawkes, D.J., Ourselin, S., 2010. A registration-based propagation framework for automatic whole heart segmentation of cardiac mri. IEEE transactions on medical imaging 29, 1612–1625.