Multi-step segmentation for prostate MR image based on reinforcement learning

Si, Xiangyu, Tian, Zhiqiang, Li, Xiaojian, Chen, Zhang, Li, Gen, et al.

Xiangyu Si, Zhiqiang Tian, Xiaojian Li, Zhang Chen, Gen Li, James D. Dormer, "Multi-step segmentation for prostate MR image based on reinforcement learning," Proc. SPIE 11315, Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling, 113152R (16 March 2020); doi: 10.1117/12.2550448

SPIE.

Event: SPIE Medical Imaging, 2020, Houston, Texas, United States

Multi-step Segmentation for Prostate MR Image based on Reinforcement Learning

Xiangyu Si[a], Zhiqiang Tian[a], Xiaojian Li[a], Zhang Chen[a], and Gen Li[a]
[a]School of Software Engineering, Xi'an Jiaotong University, Xi'an, China

ABSTRACT

Medical image segmentation is a complex and critical step in the field of medical image processing and analysis. Manual annotation of the medical image requires a lot of effort by professionals, which is a subjective task. In recent years, researchers have proposed a number of models for automatic medical image segmentation. In this paper, we formulate the medical image segmentation problem as a Markov Decision Process (MDP) and optimize it by reinforcement learning method. The proposed medical image segmentation method mimics a professional delineating the foreground of medical images in a multi-step manner. The proposed model get notable accuracy compared to popular methods on prostate MR data sets. Meanwhile, we adopted a deep reinforcement learning (DRL) algorithm called deep deterministic policy gradient (DDPG) to learn the segmentation model, which provides an insight on medical image segmentation problem.

Keywords: Prostate MR Image, Deep Deterministic Policy Gradient, Multi-step Segmentation

1. INTRODUCTION

Medical Image segmentation is a specific image processing technique that aims to identify the pixels of organs or lesions from medical images such as CT or MRI images [1]. This technique is crucial in health care, because many kinds of diagnostic procedures involve image processing. Since CNN came out, many researchers have proposed various automated segmentation structures such as Fully Convolution Network (FCN) [2], U-Net [3], V-Net [4] etc. These methods have been applied to the task of medical image segmentation.

With the development of deep learning, deep reinforcement learning methods are gradually applied to various tasks, such as games, autonomous driving, recommended systems. Deep Q-Network is one of the most widely known and widely used DRL algorithms. Previously, some researchers tried to use Deep Q-Network (DQN) [5] to solve image segmentation task. DeepOutline[11] and SeedNet[12] are two segmentation models they have proposed in recent years. DeepOutline copies a user holding a pen to draw the outline of objects in the image to achieve the semantic image segmentation task. SeedNet proposed an automatic seed generation system for the task of interactive image segmentation. However, DQN can only deal with discrete action space. This defect more or less affects the performance of both models.

In this paper, we propose a multi-step method for prostate MR image segmentation. In intermediate steps, the last segmentation mask (foreground and background) is treated as prior knowledge to the next step. We formulate semantic segmentation problem as a sequential decision-making problem and train a segmentation agent with deep reinforcement learning. Because DQN cannot handle continuous action space, DDPG algorithm [6] is adopted in this paper, which is a combination of neural network and Deterministic Policy Gradient (DPG) [7]. DDPG uses the Actor-Critic (AC) framework [6], which combines policy-based algorithm and value-based algorithm. In AC framework, the policy network is called actor, and the value network is called the critic. In the task of medical image segmentation, actor selects an action in action space based on current state, while critic makes an evaluation for the decision of actor. The goal of actor is to get a better rating, and the goal of critic is to be more accurate. The actor is typically used to reduce spatial dimension information. Meanwhile, in order to progressively restore target and spatial dimension information and directly output the segmentation mask, we train a segmentation executor by a neural network as segmentation action operator. The segmentation executor maps action parameters to a meaning region and overlays the meaning region on the foreground region of the input segmentation mask to further optimize the segmentation results.

A high-performance agent needs the appropriate reinforcement learning method for training. We choose DDPG method that has ability in handing continuous action space. Since the segmentation executor is trained based on a Quadratic Bezier Curve (QBC), we formulate a set of continuous parameters to precisely control the location, shape, and other information of the foreground in each step of segmentation operation.

Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling, edited by Baowei Fei, Cristian A. Linte, Proc. of SPIE Vol. 11315, 113152R · © 2020 SPIE
CCC code: 1605-7422/20/$21 · doi: 10.1117/12.2550448

2. METHOD

2.1 Overview

We propose an automatic multi-step segmentation model based on deep reinforcement learning. Given a prostate image, the ultimate goal of the proposed model is to generate an accurate segmentation mask under the limited steps based on current segmentation policy π. The proposed method divides the image segmentation task into multiple steps, which could yield more accurate segmentation results. The segmentation agent is trained in each step to get an optimized segmentation policy π based on the reward of the last step.

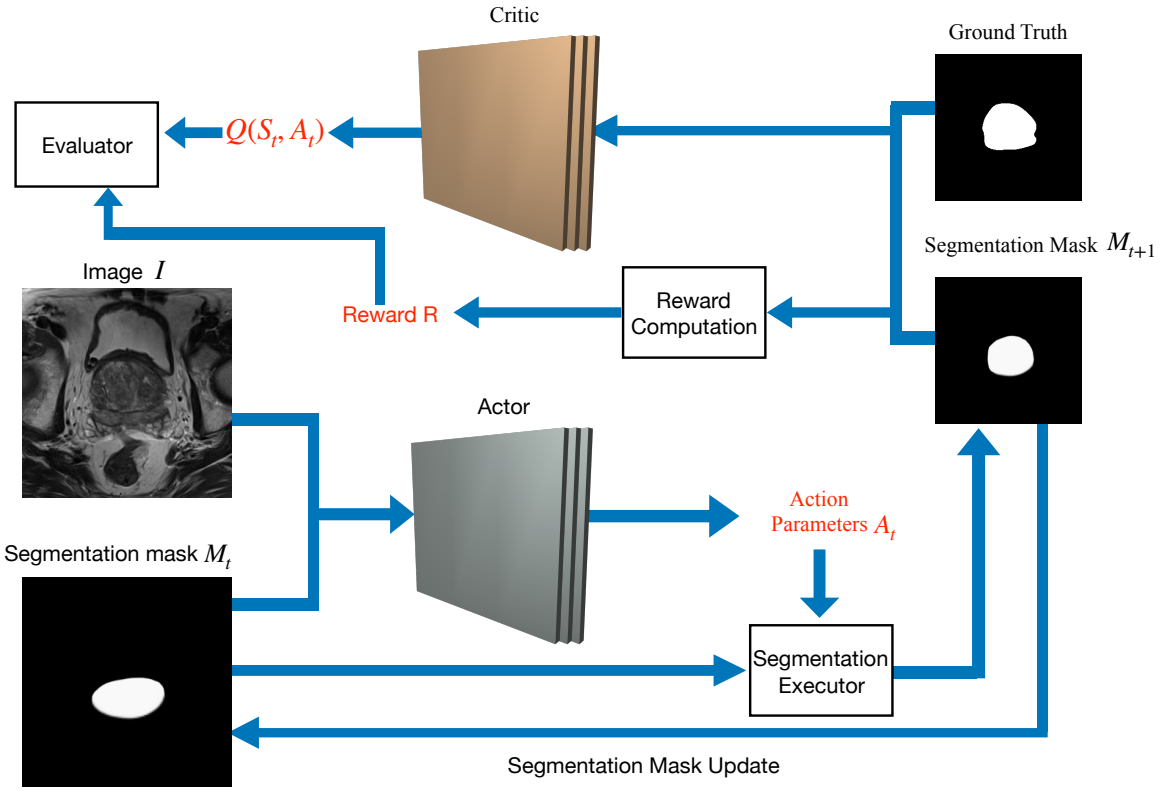

Figure 1. Overview of the proposed method. The actor gives a set of action parameters based on image and segmentation mask at each step. The segmentation model generates a revised segmentation mask M_{t+1} based on the action parameters and the segmentation mask M_t of the last step. The revised segmentation mask has three roles: 1. Calculate the reward value by comparing it with ground truth mask. 2. As the input of critic with ground truth mask to get long-term expected return $Q(S_t, A_t)$ for evaluation of the action A_t. 3. Update the previous segmentation mask M_t. The actor and critic are implemented by neural network. Q value and reward R are used to update the network parameters of actor and critic. Evaluator will further evaluate the current segmentation policy π based on the long-term expected return Q value and reward R.

The overall architecture of our method is shown in Figure 1. The method includes a pre-trained segmentation executor implemented using neural networks to perform segmentation operation. When the prostate image and initial segmentation mask are input to actor, the model aims to find an action sequence (A_0, …, A_t, A_{t+1}, …, A_T) for segmentation task. In step t, the segmentation executor adopts A_t to generate a foreground region which is overlaid on the foreground region of segmentation mask M_t and then output segmentation mask M_{t+1}. These operations are repeated throughout the segmentation process. Finally, we get the final segmentation mask by optimizing a Markov decision process.

The Markov Decision Process has three key parts, which are state space S, action A, and reward function R(S, A). The definitions of the three parts are shown as follows.

State: The state space contains all the information that agent can observe in the environment. In this work, a state contains the current segmentation mask M_t, prostate image I, and step number t. Therefore, $S_t = (M_t, I, t)$. The step number t is used to indicate every step of the segmentation process. The range of values for t is from 1 to the maximum number of step that needs to be set before training. When the step reaches the maximum number of steps, the agent enters the terminal state and the model output the final segmentation result. Current segmentation mask M_t represents current segmentation result, which is an image that pixel value is 0 or 255. The pixels in the background area are 0, and 255 in the foreground. When step t is 1, all pixel values of the initial segmentation mask are 0.

Action: The action space contains all the actions that agent can perform. Given a state, the agent selects a suitable action in the action space. We define the action as a set of parameters, which control the position and shape of the segmentation region. QBC is adopted to maximize the ability of the segmentation executor to fit the shape of the meaning region. The action parameters of QBC are defined as follows:

$$A_t = (x_0, y_0, x_1, y_1, x_2, y_2, r_0, r_1),$$

where $(x_0, y_0, x_1, y_1, x_2, y_2)$ are the coordinates of the three control points (P_0, P_1, P_2) of the QBC. The parameters (r_0, r_1) control the thickness of the two endpoints (P_0, P_2) of QBC curve. The formula of QBC is:

$$QBC(\alpha) = (1-\alpha)^2 P_0 + 2(1-\alpha)\alpha P_1 + \alpha^2 P_2, 0 \le \alpha \le 1,$$

Reward Function: The reward function acts to evaluate the result for the action of agent. A suitable reward function has always been the core key to whether the agent can accomplish the target task. We can get a segmentation mask M in each step during training stage. Thus, the accuracy of the mask can be calculated by comparing with the ground truth (GT) mask G. We adopt the mean square error (L2 loss) as the metric of evaluation. The reward function with L2 loss is described as R_{l2}:

$$R_{l2} = L2(M, G).$$

In order to make reward better represents the effect of each step, we need another reward function to obtain the change of R_{l2}. L2 loss measures the similarity between images. If the images are more similar, the L2 loss is closer to 0. Thus, the L2 loss negatively correlated with the similarity between the images. The reward function is described as R_{diff},

$$R_{diff} = L2(M_{prev}, G) - L2(M_{curr}, G),$$

where M_{prev} denotes the segmentation mask of the previous step and M_{curr} donates the current mask. The reward function gives a positive signal if the L2 loss is decreased and a negative signal if it is increased. The ultimate goal of the agent is to maximize the sum of the rewards after completing a segmentation task. Once Reward is obtained, it will be used to calculate the target $Q(S_t,A_t)$ based on Bellman equation:

$$Q(S_t, A_t) = R(S_t, A_t) + \gamma Q(S_{t+1}, \pi(S_{t+1})).$$

where $Q(S_t, A_t)$ represents the Q value function. $R(S_t, A_t)$ represents a reward when performing action A_t based on S_t. γ indicates the importance of the future returns $Q(S_{t+1},\pi(S_{t+1}))$ compared with the immediate reward $R(S_t, A_t)$. When γ is 0, it equivalent to only considering immediate reward without considering long-term returns. When γ is 1, long-term returns and immediate reward are equally important. π is the segmentation policy that agent learns. The critic estimates the long-term return Q for the agent decision, which is learned by using Bellman equation.

In this paper, S_t and ground truth are fed into critic rather than S_t and A_t for the evaluator to get a more accurate Q value, which draws on the idea of supervised learning. The modified value function $V(S_t, G)$ is learned by using the following equation,

$$V(S_t, G) = R(S_t, A_t) + \gamma V(S_{t+1}, G).$$

The actor learns a policy π that maps a state S_t to A_t. The critic estimates the expert return for the agent taking action A_t at state S_t, which is learned using Bellman equation.

Finally, the final segmentation can be obtained by optimizing the MDP based on DDPG.

2.2 Action bundle

Inspired by frame skip [13], action bundle [14] strategy is adopted to further improve the accuracy of our method. Frame skip is a hyper-parameter, which plays an important role in many reinforcement learning algorithms. In some reinforcement

learning tasks, agent does not need to take an action for each state, such as game. Frame skip determines how often agent takes an action during interaction with the environment. When the value of frame skip is K, the agent only performs a selected action at K frames. This strategy can explore the connection of similar states and save computing resources. And the action bundle strategy is used to explore the connection between different actions. Action bundle strategy makes actor can better explore the action space. When the value of action bundle is K, actor picks out K actions from the action space based on current state to form an action bundle. Then, the segmentation executor performs the K actions of the action bundle, which can further improve the accuracy of the segmentation result. We set K = 5, which means that one bundle contains 5 actions.

3. RESULTS

Dataset. The prostate MRI dataset contains 172 T2 transverse subjects. 142 subjects are used for training, which are from PROMISE12, ISBI2013, and in-house data sets. 30 subjects from PROMISE12 test data set are used for testing. All these images are fully labeled by the radiologists.

Experiment Details. The model structure of policy network (actor) and value network (critic) are similar to ResNet-18 [15]. Batch normalization [16] was adopted for actor and weight normalization [17] with Translated ReLU (TReLU) [18] was used for the critic. The segmentation executor network consists of fully connected layers and convolution layers. Sub-pixel [19] strategy was used to increase the performance segmentation executor.

Our experiments were performed on a single NVIDIA GeForce RTX 2080Ti with 11G memory. The prostate images were resized to 128×128 during training and testing. We use Adam [20] to optimize the segmentation strategy π. The mini-batch size was 64. The learning rate range of actor is [3e-4, 1e-4] and critic is [1e-3, 3e-4], which both decay every 800 training episodes. The reward discount factor γ is set as 0.955. We set the action bundle K equal to 5 and step number t equal to 3.

Evaluation Metrics. The proposed method was evaluated based on the manually labeled ground truth. Two quantitative metrics are used for segmentation evaluation, which are Dice similarity coefficient (DSC) and Hausdorff distance (HD) [8]. The DSC formula is:

$$\mathrm{DSC} = \frac{2|S_{gt}+S_m|}{|S_{gt}|+|S_m|},$$

where $|S_{gt}|$ is the number of pixels of the prostate from the manually segmentation ground truth. $|S_m|$ is the number of pixels of the prostate from the proposed method.

A distance from a pixel x to a surface Y is defined as $\mathrm{d}(\mathrm{x},\mathrm{Y}) = min_{y\in Y}||x-y||$. The HD between two surfaces X and Y is calculated as:

$$\mathrm{HD}(\mathrm{X},\mathrm{Y}) = \max\left[\max_{x\in X} d(x,Y), \max_{y\in Y} d(y,X)\right].$$

Quantitative Comparison. We choose five segmentation methods to evaluate our method, which are PSPNet [9], FCN, U-Net, V-Net, and DeepLabV3+ [10]. The comparison results are shown in Table 1.

Table 1. Quantitative comparison between the proposed method and five methods.

	PSPNet [9]	FCN [2]	U-Net [3]	V-Net [4]	DeepLabV3+ [10]	Ours
DSC (%)	75.49	82.37	84.71	85.29	86.45	**88.89**
Std. of DSC	9.41	5.56	6.52	6.82	5.09	**3.12**
HD (mm)	24.58	19.64	15.92	16.78	23.08	**13.44**
Std. of HD	15.27	19.79	6.85	6.60	19.07	**6.15**

Our method achieved the highest volume DSC and the lowest HD in the quantitative comparison. Meanwhile, the proposed method has the lowest standard deviation of both DSC and HD, which means that our method is robust to the different prostate MR volumes.

Action Bundle Setting. In this paper, we adopt a strategy called action bundle to further improve the accuracy of our method. In order to get the most appropriate action bundle value for this medical image task, we set the K value to 1 (without action bundle), 3, 5, 7, respectively. The comparison results are shown in Table 2.

Table 2. The comparison results of different action bundle setting.

	K=1	K=3	K=5	K=7
DSC (%)	86.21	87.74	88.89	88.87

From the table, our method can get best result when K is set as 5.

Qualitative Evaluation Results. The performance of the proposed method was evaluated qualitatively by visualizing contours of the proposed method and the manually segmented ground truth. Figure 2 shows the qualitative comparison results.

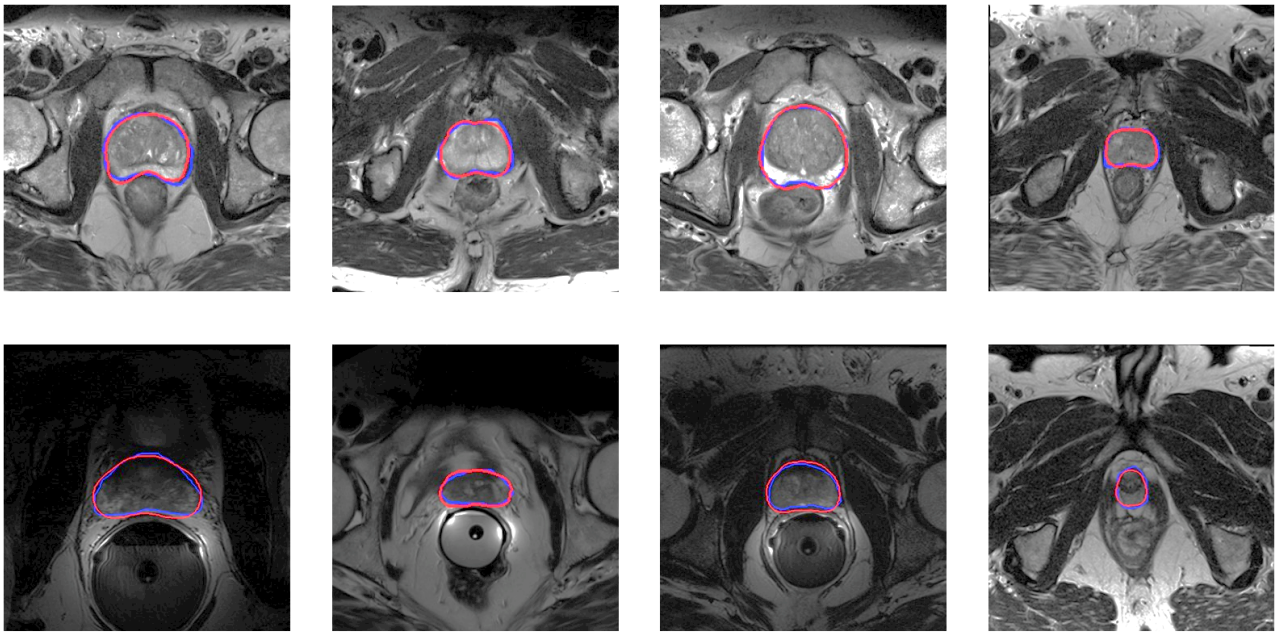

Figure 2. The qualitative results of the proposed method. The red curves represent the prostate contours obtained by the proposed method, and the blue curves represent the contours obtained from manual segmentation by the experienced radiologists.

4. NEW OR BREAKTHROUGH WORK TO BE PRESENTED

To the best of our knowledge, this is the first study to formulate prostate MR image segmentation problem as a MDP, and optimize it by DDPG. The proposed method can automatically segment prostate, which could alleviate the workload of the radiologists.

5. CONCLUSIONS

In this paper, we propose an automatic medical image segmentation method based on deep reinforcement learning. An agent is trained by DRL, which could segment prostate from MR image by a multi-step manner. Experimental results show that, the proposed method could yield satisfactory segmentation of prostate MR images. In the future, we will try to use this method for multi-organ semantic segmentation and other modalities image.

ACKNOWLEDGEMENT

This work was supported in part by NSFC under grant No. 61876148. This work was also supported in part by the Fundamental Research Funds for the Central Universities No. XJJ2018254, and China Postdoctoral Science Foundation No. 2018M631164.

REFERENCES

[1] Hesamian M H, Jia W, He X, et al. "Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges." In Journal of digital imaging, 1-15(2019).

[2] Long J, Shelhamer E, Darrell T. "Fully convolutional networks for semantic segmentation." In Proceedings of the IEEE conference on computer vision and pattern recognition, 3431-3440(2015).

[3] Ronneberger O, Fischer P, Brox T. "U-net: Convolutional networks for biomedical image segmentation." In International Conference on Medical image computing and computer-assisted intervention, 234-241(2015).

[4] Milletari F, Navab N, Ahmadi S A. "V-net: Fully convolutional neural networks for volumetric medical image segmentation." In 2016 Fourth International Conference on 3D Vision (3DV), 565-571(2016).

[5] Mnih V, Kavukcuoglu K, Silver D, et al. "Human-level control through deep reinforcement learning." In Nature, 518(7540): 529(2015).

[6] Lillicrap T P, Hunt J J, Pritzel A, et al. "Continuous control with deep reinforcement learning." In arXiv preprint arXiv:1509.02971(2015).

[7] Silver D, Lever G, Heess N, et al. "Deterministic policy gradient algorithms." In ICML, 2014.

[8] Tian Z, Liu L, Zhang Z, Fei B. "Superpixel-based segmentation for 3D prostate MR images." In IEEE transactions on medical imaging, 35(3): 791-801(2015).

[9] Zhao H, Shi J, Qi X, et al. "Pyramid scene parsing network." In Proceedings of the IEEE conference on computer vision and pattern recognition, 2881-2890(2017).

[10] Chen L C, Zhu Y, Papandreou G, et al. "Encoder-decoder with atrous separable convolution for semantic image segmentation." In Proceedings of the European conference on computer vision (ECCV), 801-818(2018).

[11] Zhenxin Wang, Sayan Sarcar, et al. "Outline Objects using Deep Reinforcement Learning." In Computer Vision and Pattern Recognition, 2018.

[12] Song, Gwangmo, Heesoo Myeong, and Kyoung Mu Lee. "Seednet: Automatic seed generation with deep reinforcement learning for robust interactive segmentation." In Computer Vision and Pattern Recognition, 2018.

[13] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M. "Playing atari with deep reinforcement learning." In arXiv preprint arXiv:1312.5602(2013).

[14] Huang, Z., Heng, W., and Zhou, S. "Stroke-based artistic rendering agent with deep reinforcement learning." In arXiv preprint arXiv:1903.04411(2019).

[15] He, K., Zhang, X., Ren, S., and Sun, J. "Deep residual learning for image recognition. " In Proceedings of the IEEE conference on computer vision and pattern recognition, 770–778(2016).

[16] Ioffe, S. and Szegedy, C. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." In arXiv preprint arXiv:1502.03167(2015).

[17] Salimans, T. and Kingma, D. P. "Weight normalization: A simple reparameterization to accelerate training of deep neural networks." In Advances in Neural Information Processing Systems, 901–909(2016).

[18] Xiang, S. and Li, H. "On the effects of batch and weight normalization in generative adversarial networks." In arXiv preprint arXiv:1704.03971(2017).

[19] Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., and Wang, Z. "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network." In Proceedings of the IEEE conference on computer vision and pattern recognition, 1874–1883(2016).

[20] Kingma, D. P. and Ba, J. "Adam: A method for stochastic optimization." In arXiv preprint arXiv:1412.6980(2014).