Proceedings of 2023 IEEE
International Conference on Mechatronics and Automation
August 6 - 9, Harbin, Heilongjiang, China

A DDPG-Based Method of Autonomous Catheter Navigation in Virtual Environment

Wei Tian[1], Jian Guo[1,2]*
[1]Tianjin Key Laboratory for Control Theory & Applications
In Complicated systems and Intelligent Robot Laboratory
Tianjin University of Technology
Binshui Xidao Extension 391, Tianjin, 300384,China
tianwei202102@163.com;
**Corresponding author: jianguo@tjut.edu.cn*

Shuxiang Guo[1,2,3]*
[2]Shenzhen Institute of Advanced Biomedical Robot Co.,Ltd.
No.12, Ganli Sixth Road, Jihua Street, Longgang District, Shenzhen,518100,China
**Corresponding author: guoshuxiang@hotmail.com*

and Qiang Fu[1]
[3]Key Laboratory of Convergence Medical Engineering System and Healthcare Technology,The Ministry of Industry and Information Technology,School of Life Science
Beijing Institute of Technology
No.5,Zhongguancun South Street,Beijing,100081,China
fuqiang6369@hotmail.com

***Abstract* - Vascular interventional surgery is the main method for treating cardiovascular diseases. But navigating endovascular catheters through the vascular tree is a highly challenging task even for highly trained specialists. Automation of this task can reduce the burden on surgeons and is expected to improve the surgical outcomes. Although there have been relevant studies utilizing reinforcement learning algorithms to realize autonomous navigation of catheter in the virtual environment Cathsim. However, the kinematics model of the catheter in Cathsim does not conform to the operating mode of the catheter in real vascular interventional surgery. Besides, there are problems such as low success rates of catheter autonomous navigation tasks. To address these issues, this paper modifies the kinematics model of the catheter in Cathsim and designs a catheter autonomous navigation model based on reinforcement learning DDPG (Deep Deterministic Policy Gradient) algorithm. The experimental results show that the agents trained through DDPG in this paper performs better than the agents trained through PPO (Proximal Policy Optimization) in other studies in terms of navigation task success rate, completion time, and contact force between the catheter and vascular wall during the navigation process.**

Index Terms - Vascular interventional surgery, Catheter navigation, Deep reinforcement learning, Autonomous surgery, Virtual environment

I. INTRODUCTION

Cardiovascular disease is one of the most causes of death in the world [1]. Vascular interventional surgery are minimally invasive surgery procedures frequently used to treat cardiovascular diseases [2]. One of the key technical component of vascular interventional surgery is a safe navigation of catheter (or guidewire) to a target place in a blood vessel [3]. This is a highly challenging task even for highly trained specialists. The dynamics of catheters are complex and non-linear and the patients' vasculature are often complex and diverse which increases the operative risk [4][5]. This will undoubtedly bring great mental pressure to surgeons, and the effectiveness of surgery often depends on the experience level and status of the doctor [6]. For these reasons, automation of cannulation can reduce the burden on surgeons and is expected to improve the average treatment result.

Several robotic platforms have been developed to assist surgeons in vascular intervention surgery [7]. These robotic platforms aim to reduce the risk for orthopedic injuries for surgeons [8], and to reduce their exposure to X-rays [9]. Several clinical trials and surgeons also report that the use of the robot increased stability and accuracy of catheter navigation and facilitates some complex procedures [10]. However, these robots are all used as teleoperation tools and offer a low level of robotic autonomy [11].

The development of deep reinforcement learning has promoted task autonomy for robots. [12]. In recent years, in the field of vascular interventional surgery, many research institutions and universities have utilized deep reinforcement learning to realize autonomous navigation of catheters and guidewires in blood vessels. Karstensen et al. from the Fraunhofer IPA in Germany used DDPG with HER (Hindsight Experience Replay) to achieve autonomous navigation of catheter in complex two-dimensional rigid vascular model [13]. However they did not achieve autonomous navigation of the catheter in a 3D(three-dimensional) vascular model. In real vascular interventional surgery, the catheter moves in a 3D vascular environment. Yang et al. from Beijing University of Posts and Telecommunications used reinforcement learning algorithm SAC (Soft Actor Critic) to train robot agents on a real vascular interventional surgical robot platform, achieving simple point-to-point guidewire autonomous navigation tasks in a 3D vascular model [14]. However, the cost of training agents through a real robot platform is relatively high, as it will cause a certain degree of loss to the robot platform. In 2022, Tudor Jianu et al. from the University of Liverpool in the UK proposed an open-source 3D virtual environment, Cathsim, which provides a benchmark platform for the development and testing of algorithms for catheter autonomous navigation in vascular interventional surgery [15]. Their experimental results indicate that using the simulator Cathsim, reinforcement learning agents can be successfully trained to complete different autonomous cannulation tasks. However, the motion mode of the catheter in Cathsim does not match the motion mode in real vascular interventional surgery where the catheter can only be pushed, pulled, and rotated. And their success rate in completing autonomous cannulation tasks using PPO trained agents is not high.

In this paper, firstly, we modify the kinematics model of the catheter in Cathsim to approach the motion pattern of the catheter in real vascular interventional surgery. Secondly, we design a catheter autonomous navigation model based on reinforcement learning DDPG algorithm to improve the outcomes of autonomous cannulation tasks in Cathsim.

The structure of this paper is as follows: Section I introduces the background, significance, and research status of autonomous catheter navigation in vascular intervention surgery; Section II introduces the reinforcement learning method, training environment and DDPG algorithm for achieving endovascular catheter autonomous navigation; Section III introduces the experimental setup and analyzes the experimental results; Section IV presents the conclusion of this paper and future prospects.

II. Endovascular Catheter Autonomous Navigation

In this paper, we consider the navigation task of catheter in the aortic arch. The catheter was initially located within the ascending aorta and its task was to navigate to the brachiocephalic artery (BCA). The following will explain the methods of autonomous catheter navigation from three aspects: training environment, reinforcement learning algorithm and DDPG algorithm.

A. Training Environment

Here we choose Cathsim as our virtual training environment for reinforcement learning. Cathsim is an open-source simulation environment developed by Tudor Jianu et al. for the development of machine learning algorithms for autonomous endovascular catheter navigation. This environment is based on the MuJoCo physics engine, providing real-time force perception capabilities and a high-fidelity visualization interface for aortic , catheter and vascular intervention surgical robots, making it very suitable for using different machine learning methods to train autonomous catheter navigation agent.

In Cathsim, the catheter is modeled by a discretisation approach. It consists of 100 bodies joined together by two revolute joints. However, the manipulation of the catheter is controlled by the tip of the catheter. This is not in line with the fact that passive catheters can only perform push-pull and rotation movements in real vascular interventional surgery.

To mimick the motion pattern of the catheter in real vascular interventional surgery, we modify the kinematics model of the catheter in Cathsim. Firstly, we disable the motors that drive the revolute joints at the tip of the catheter so they can only move passively like other joints. Secondly, we add a rotating joint around the catheter axis at the proximal link of the catheter to mimick the rotational movement of the catheter.

B. Reinforcement Learning Methods

In this paper, we consider that the task of autonomous cannulation is an Partially Observable Markov Decision Process (POMDP) [16]. The agent, represented by a catheter, interacts with the environment represented by the aortic arch. The agent receives an observation, chooses to execute an action, receives a reward, and reaches a new state, and repeats it continuously. An episode is not terminated until the agent reaches the target position within the aorta. The schematic diagram of the reinforcement learning model [17] is shown in the Fig. 1.

Fig. 1 The Reinforcement learning model

In the aortic arch cannulation task, we have the following definitions:

1)State: The state is an image. The image is taken by a RGB camera of 128 × 128 × 3 resolution which is placed on the top of the aortic arch phantom. Then we map the image from RGB to grayscale and use it as input state. This state space simulates the clinical procedure of surgeons observing fluoroscopic images.

2)Action space: The action space of the catheter includes the rotation and push-pull actions of the catheter. This action space also simulate the operation of the catheter by surgeons.

3)Reward: To implement the navigation of the catheter to the target point, we calculate the distance between the catheter head h and the target point g : $d(h,g)=\|h-g\|$. When the distance between the catheter head and the target point is greater than δ =8 mm , the reward value for each step is $-d(h,g)$. When the distance between the catheter head and the target point is less than or equal to δ , the current eposide ends, the agent receives a reward value of 10. The reward function is shown in Equation(1).

$$r(h_t,g)=\begin{cases}10 & if\ d(h,g)\leq\delta \\ -d(h,g) & otherwise\end{cases} \quad (1)$$

C. Deep Deterministic Policy Gradient Algorithm

The DDPG algorithm is proposed based on the DPG(Deep Deterministic Policy) algorithm and belongs to the off policy algorithm of the actor critical method in model-free systems. It can be said that DDPG is an improvement on the

DQN algorithm bacause it use the skills of DQN: target network and experience replay for reference. However the problem with DQN is that it can only solve problems with discrete and low dimensional action spaces [18]. In general physical or control problems, the action space is continuous. According to II.B, the action space of the catheter is continuous, so it is appropriate to use the DDPG algorithm here. The process of DDPG algorithm[19] is shown in Table I.

TABLE I
PROCESS OF DDPG ALGORITHM

Algorithm 1 Deep Deterministic Policy Gradient algorithm

Randomly initialize critic network $Q(s,a|\theta^Q)$ and actor $\mu(s|\theta^\mu)$ with weights θ^Q and θ^μ.
Initialize target network Q' and μ' with weights $\theta^{Q'} \leftarrow \theta^Q, \theta^{\mu'} \leftarrow \theta^\mu$
Initialize replay buffer R
for episode = 1, M **do**
  Initialize a random process N for action exploration
  Receive initial observation state s_1
  for $t = 1, T$ **do**
    Select action $a_t = \mu(s_t|\theta^\mu) + N_t$ according to the current policy and exploration noise
    Execute action a_t and observe reward r_t and observe new state s_{t+1}
    Store transition (s_t, a_t, r_t, s_{t+1}) in R
    Sample a random minibatch of N transitions (s_t, a_t, r_t, s_{t+1}) from R
    Set $y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'})$
    Update critic by minimizing the loss:
$$L = \frac{1}{N}\sum_i (y_i - Q(s_i, a_i|\theta^Q))^2$$
    Update the actor policy using the sampled policy gradient:
$$\nabla_{\theta^\mu} J \approx \frac{1}{N}\sum_i \nabla_a Q(s,a|\theta^Q)|_{s=s_i, a=\mu(s_i)} \nabla_{\theta^\mu}\mu(s|\theta^\mu)|_{s_i}$$
    Update the target networks:
$$\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}$$
$$\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}$$
  end for
end for

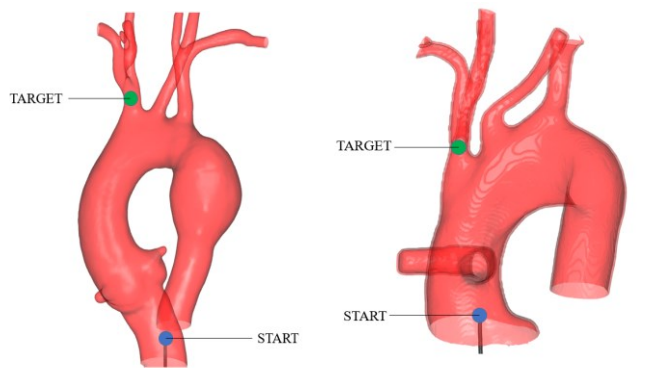
III. EXPERIMENTS AND RESULTS

A. *Experimental Setup*

We consider the autonomous navigation of the catheter in two different types of aortic arches. Within both setups, we place the catheter tip inside the ascending aorta as the starting point for navigation. When the catheter tip navigates within 8 mm of the target point (it is believed that the catheter has been inserted into the brachiocephalic artery), the task ends. The catheter navigation task experimental setups are shown in the Fig. 2. We train agents using DDPG algorithm and PPO algorithm based on the following settings which is shown in Table II. We use the parameters in [15] to train the DDPG agent and PPO agent, as shown in Table II. A total of 600000 time steps were trained. In each episode, when the catheter reaches the target point or the number of interaction steps with the environment reaches 2000, an episode ends. In section II. B, it was mentioned that the input state is an image, so a CNN strategy was used to extract features from the input image. CNN has three convolution layers, and the activation function is ReLU. The learning rate is 0.0003.

TABLE II
TRAINNING PARAMETERS

Policy	CNN
Training steps	600000
Learning rate	0.0003
Max episode steps	2000



(a) Aortic arch-I (b) Aortic arch-II

Fig. 2 The experimental setups of the catheter navigation task

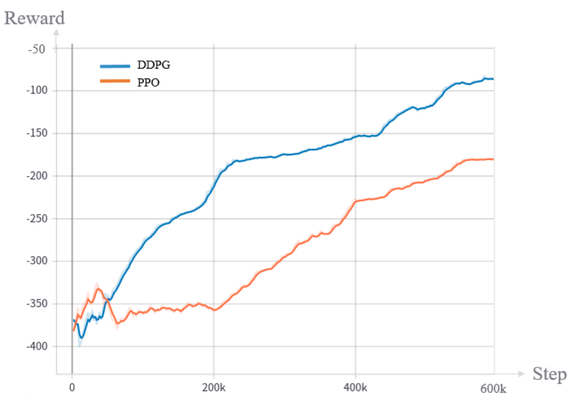
B. *Experimental Results*

The trend of rewards when using different reinforcement learning algorithms DDPG and PPO to train autonomous catheter navigation agents in aortic arch-I and aortic arch-II are as shown in Fig. 3(a) and Fig. 3(b), respectively. From the trend of rewards during the training process, it can be seen that agent trained with DDPG can obtain higher rewards in a shorter time compared to PPO in both aortic arches. This indicates that the DDPG algorithm performs better than the PPO algorithm in the given catheter autonomous navigation task.

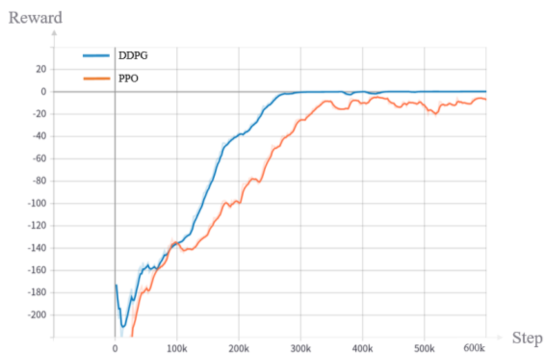
When the training is over, we evaluated the agents trained with DDPG and PPO algorithms in both aortic arches for 10 episodes, extracting force interaction and reward information for each episode, as well as the success rate of navigation tasks and the time required to complete them. The average results are shown in the Table III.

The experimental results show that in both aortic arches, compared to PPO, the agents trained with DDPG have a higher success rate, shorter time to complete catheter autonomous navigation tasks, and less interaction with blood vessel walls. Overall, the DDPG algorithm performs better than the PPO algorithm in catheter autonomous navigation tasks.

We qualitatively evaluate the cannulation performance of the agent which has trained with DDPG by observing the successful and failed autonomous catheter navigation process in aortic arch-I because the cannulation in aortic arch-I is more difficult. The successful catheter navigation process is shown in the Fig. 4.



(a) Aortic arch-I



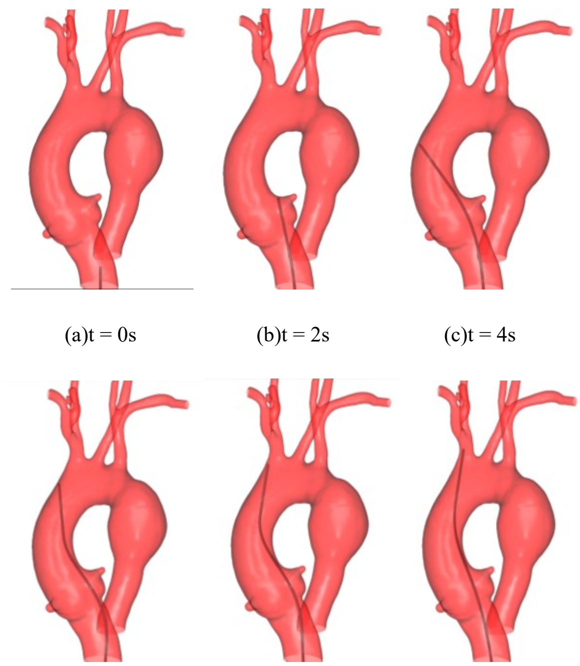
(b) Aortic arch-II

Fig. 3 The rewards when trraining autonomous catheter navigation agents using DDPG and PPO in the aortic arch-I and aortic arch-II

Through the observation of the successful autonomous catheter navigation process, we can see that the DDPG agent has successfully navigated the catheter to the target point at t =12s as shown in Fig. 4(f). At t=4s, as shown in Fig. 4(c) , the agent successfully passes the catheter through the curved part of the blood vessel by rotating and pushing it with the help of the contact between the catheter and the blood vessel wall. Also with the help of the contact between the catheter and the blood vessel wall, the agent successfully inserted the catheter into the target branch, as shown in Fig. 4(d) - Fig. 4(f).

The skill of agent's cannulation is similar to that of surgeons in real vascular interventional surgery. During the surgery, the catheter used by the surgeons is a passive catheter, which cannot be steerable. The catheter can only be either pushed, pulled or rotated along its longitudinal axis. Therefore, navigation of catheters in blood vessels relies heavily on the contact between the catheter and the vessel wall, especially when passing through the curved and bifurcated parts of the vessel. As a consequence, from the successful catheter navigation process, we believe that the agent has successfully learned the skills of cannulation in blood vessels through the the reinforcement learning algorithm DDPG.

The failed situation is that the agent inserted the catheter into the wrong blood vessel branch, as shown in Fig. 5. One possible reason for the failed cannulation is that the setting of the reward function which is shown in Equation(1). Whilst this reward function assists in agent convergence, it is also prone to local minima. As a consequence, the catheter is inserted into a bifurcation of blood vessels that are closer to the target bifurcation. In the future, the reward function will be optimized, such as increasing the value of δ (the distance between catheter tip and target) to improve the success rate of cannulation.



(a)t = 0s (b)t = 2s (c)t = 4s

(d)t = 8s (e)t = 10s (f)t = 12s

Fig. 4 The successful catheter navigation proces

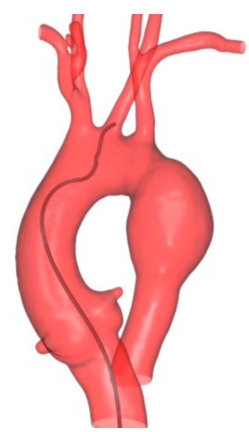


Fig. 5 The failed catheter navigation situation

TABLE III

EVALUATION RESULTS

Aorta	Algorithm	Reward	Mean Force（N）	Max Force（N）	Success %	Time（s）
Aortic arch-I	DDPG	-64 ± 50	0.002 ± 0.0001	0.079 ± 0.027	80	12 ± 1.2
	PPO	-151 ± 84	0.004 ± 0.0015	0.193 ± 0.175	60	20 ± 2.6
Aortic arch-II	DDPG	-10 ± 1	0.002 ± 0.0002	0.017 ± 0.002	100	4 ± 0.8
	PPO	-50 ± 12	0.008 ± 0.0021	0.125 ± 0.004	80	7 ± 1.1

IV. CONCLUSIONS

In this paper, we modified the kinematics model of the catheter in Cathsim to approach the motion pattern of the catheter in real vascular interventional surgery. Besides, we designed a catheter autonomous navigation model based on reinforcement learning DDPG. The comparative experimental results with the reinforcement learning algorithm PPO showed that in catheter navigation task in different environments, the success rate of the agent trained with DDPG is higher, the average and maximum contact force with the blood vessel wall is smaller, and the time required to complete the task is shorter. The overall performance of DDPG was better than the PPO algorithm.

Although we had implemented autonomous navigation of catheters in a virtual environment, there is still a long way to go to transplant the models trained in the virtual environment to laboratory environments and ultimately apply them to clinical applications. Firstly, it is necessary to establish a highly realistic vascular environment and catheter guidewire model; Secondly, it is difficult to establish a one-to-one correspondence between various quantities in the virtual environment and the real environment. How to reduce the impact of registration errors is a problem that needs to be solved.

In the future, we will investigate how to apply models trained with reinforcement learning algorithm DDPG in virtual environments to our laboratory environments.

ACKNOWLEDGMENT

This research is supported by National Natural Science Foundation of China (61703305), Key Research Program of the Natural Science Foundation of Tianjin (18JCZDJC38500) and Innovative Cooperation Project of Tianjin Scientific and Technological (18PTZWHZ00090).

REFERENCES

[1] S.-X. Guo, et al, “Machine learning-based operation skills assessment with vascular difficulty index for vascular intervention surgery,” *Medical & Biological Engineering & Computing*, vol. 58, no. 8, pp. 1707- 1721, 2020.

[2] J. F. Granada, et al, “First-in-human evaluation of a novel robotic-assisted coronary angioplasty system,” *JACC Cardiovascular Interventions*, vol. 4, no. 4, pp. 460-465, 2011.

[3] J. Guo, H. Zhao and S.-X. Guo, “Design of a novel virtual sensor for the vascular interventional robotic system,” *IEEE International Conference on Mechatronics and Automation*, pp. 1134-1139, 2020

[4] X.-Q. Bao, S.-X. Guo, N. Xiao and L.-W. Shi, “Design and evaluation of sensorized robot for minimally vascular interventional surgery,” *Microsystem Technologies*, vol. 25, no. 7, pp. 2759-2766, 2019.

[5] X.-Y. Zhou, Y. Guo, M. Shen and G.-Z. Yang, “Application of artificial intelligence in surgery,” *Frontiers of Medicine*, vol. 14, no. 4, pp. 417-430, 2020.

[6] A. K. Ghamraoui and J. J. Ricotta, “Current and future perspectives in robotic endovascular surgery,” *Current Surgery Reports*, vol. 6, no. 21, pp. 1-7, 2018.

[7] G.-Z. Yang, et al, “Medical robotics regulatory, ethical, and legal considerations for increasing levels of autonomy,” *Science Robotics*, vol. 2, no. 8, pp.8638, 2017.

[8] C. V. Riga, et al, “Robot-assisted fenestrated endovascular aneurysm repair (FEVAR) using the Magellan system,” *Journal of Vascular and Interventional Radiology*, vol. 24, no. 2, pp. 191-196, 2013.

[9] H.-J. Cha, B.-J. Yi and J.-Y. Won, “An assembly-type master–slave catheter and guidewire driving system for vascular intervention,” *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 231, no. 1, pp. 69-79, 2017.

[10] H. You, et al, “Automatic control of cardiac ablation catheter with deep reinforcement learning method,” *Journal of Mechanical Science and Technology*, vol. 33, no. 6, pp. 5415-5423, 2019.

[11] Y. Zhao, et al, “A CNN-based prototype method of unstructured surgical state perception and navigation for an endovascular surgery robot,” *Medical & Biological Engineering & Computing*, vol. 57, pp. 1875-1887, 2019.

[12] D. Silver, et al, “Mastering the game of go without human knowledge,” *Nature*, vol. 550, no. 7676, pp. 354-359, 2017.

[13] L. Karstensen, et al, “Autonomous guidewire navigation in a two dimensional vascular phantom,” *Current Directions in Biomedical Engineering*, vol. 6, no. 1, 2020.

[14] D. Yang, J. Song and Y. Hu, “Guidewire feeding method based on deep reinforcement learning for vascular intervention robot,” *IEEE International Conference on Mechatronics and Automation*, pp. 1287-1293, 2022.

[15] T. Jianu, et al, “CathSim: An open-source simulator for autonomous cannulation,” unpublished.

[16] V. Mnih, et al, “Playing atari with deep reinforcement learning,” unpublished.

[17] T. Behr, et al, “Deep reinforcement learning for the navigation of neurovascular catheters,” *Current Directions in Biomedical Engineering*, vol. 5, no. 1, pp. 5-8, 2019.

[18] L. Karstensen, et al, “Learning-based autonomous vascular guidewire navigation without human demonstration in the venous system of a porcine liver,” *International Journal of Computer Assisted Radiology and Surgery*, vol. 17, no. 11, pp. 2033-2040, 2022.

[19] T. P. Lillicrap, et al, “Continuous control with deep reinforcement learning,” unpublished.