Application of Deep Reinforcement Learning in Mobile Robot Path Planning

Jing Xin,Huan Zhao,Ding Liu
Shaanxi Key Laboratory of Complex System Control and Intelligent Information Processing
Xi'an University of Technology
Xi'an 710048, P.R.China
xinj@xaut.edu.cn

Minqi Li
Department of the Information and Communications
Xi'an Polytechnic University
Xi'an 710048, P.R.China

***Abstract*—In order to make the robot obtain the optimal action directly from the original visual perception without any hand-crafted features and features matching, a novel end-to-end path planning method--mobile robot path planning using deep reinforcement learning is proposed. Firstly, a deep Q-network (DQN) is designed and trained to approximate the mobile robot state-action value function. Then, the Q value corresponding to each possible mobile robot action (i.e., turn left, turn right, forward) is determined by the well trained DQN, here, the input of the DQN is the original RGB image (image pixels) captured from the environment without any hand-crafted features and features matching; Finally, the current optimal mobile robot action is selected by the action selection strategy. Mobile robot reach to the goal point while avoiding obstacles ultimately. 30 times path planning experiments are conducted in the seekavoid_arena_01 environment on DeepMind Lab platform. The experimental results show that our deep reinforcement learning based robot path planning method is an effective end-to-end mobile robot path planning method.**

Keywords—Mobile robot; End-to-end Path planning; Deep Reinforcement Learning; DQN

I. INTRODUCTION

Path planning problem can be described as the task of navigating a mobile robot around a space in which a number of obstacles that should be avoided. The task usually is under some optimization criteria, such as least working cost, shortest walking distance, minimal walking time, etc. It is an important and challenging topic in robotics [1]. In many applications of robots, the working environments are complex and unpredictable, which require the path planning methods to show self-study, adaptation and robust abilities. To overcome the weakness of these approaches, researchers explored variety of solutions. Reinforcement learning (RL) techniques can learn appropriate actions from the environment states, whose benefits are based on the concept of online learning, and rewards or punishments from the environments. Therefore, the agent is allowed to modify its policy from the rewards or punishments it receives. Presently, reinforcement learning algorithms have been well applied in mobile robot path planning problems and achieved important achievements [2].

However, traditional RL_based path planning methods heavily rely on hand-crafted features of the task representation, which are not powerful for raw high dimensional input images, so, learning a control strategy from a raw image directly is still an important challenging for RL[3,4]. Recent advances in deep learning have made it possible to exploit high-level feature from raw image data, leading to great breakthroughs in speech recognition, computer vision and other applications [5].

Specifically, Mnih et al. [5] from Google Deepmind team combined the convolutional neural network and Q-learning of the Traditional RL, then proposed a deep Q-network model to solve the high dimensional perception based decision problem. They utilize a deep Q-network (DQN) to evaluate the Q-function for Q-learning. Many followed papers tried to make improvements from the model of Mnih et al [5], as DQN is essentially state-of-the art and was the main catalyst for deep RL [5,6].

Based on this, in this paper, deep RL technology is applied in mobile robot path planning and achieve the end-to-end mobile robot path planning, that is, the proposed planning method can determine the optimal action to make the mobile robot reach to the goal point while avoiding obstacles only using the original visual perception without any hand-crafted features and features matching.

The paper is organized as follows: The general framework of the proposed path planning method and the basic implement principle of the each components are introduced in section 2; In section 3, some experiment results and analysis are presented. Finally, some conclusion and improvement issues for future research are described in section 4.

II. PROPOSED METHOD

A. The overall framework of the proposed planning method

The overall framework of the proposed method for mobile robot path planning using deep reinforcement learning is shown in Fig.1, which is composed of agent (also known as intelligent systems) and environment. The agent is a computer system in an environment that has the ability to act autonomously in this environment to achieve a design goal [7]. Reinforcement learning is the agent in the process of

This work is supported by the Shaanxi Provincial Natural Science Foundation of China under Grant No. 2016JM6006,the Shaanxi Provincial Department of Education Key Laboratory of scientific research program under Grant No.16JS071, the Shaanxi Provincial Modern equipment of green manufacturing Collaborative Innovation Center(Research on Key Technology of Intelligent Perception and Cooperative Control for Industry Robot), Grant No.304-210891702.

continuous interaction with the external environment, repeatedly learn through the trial and error method, obtain the environmental information, constantly optimize itself action strategy. The goal of learning is to maximize the cumulative reward value that agent obtains from the environment [8]. In order to make the robot obtain the optimal action directly from the original visual perception through the end-to-end learning, a Deep Q Network is designed and trained to approximate the state-action value function of the mobile robot in this paper. RGB images captured by mobile robot system is directly viewed as the current state of the robot and the input of the trained Deep Q Network, the output of the network is the Q value corresponding to each possible action of the mobile robot. Finally, the mobile robot select the optimal action using the action selection strategy and move to target location while avoid the obstacles in the environment.

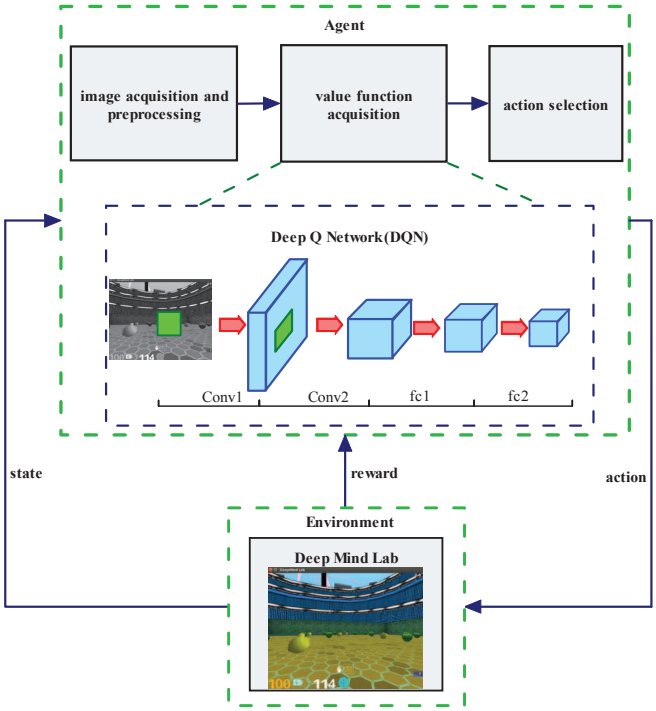


Fig. 1. General framework of mobile robot path planning using deep reinforcement learning

The agent in Fig.1 contains three modules: image acquisition and preprocessing, value function acquisition and action selection. The main function of each module is as follows:

(1) Image acquisition and preprocessing

The main function of this module is to reduce the image dimension and reduce the computational complexity through performing graying and dimension reduction operations on the original RGB images collected from the current environment.

(2) Value function acquisition

The main function of this module is to obtain the Q value of each possible action of the mobile robot. The Deep Q Network (DQN) is designed and trained to obtain the state-action value function Q. The input of the DQN is the last 4 frames of the preprocessed, and the output of the DQN is the Q value of each possible action of the robot.

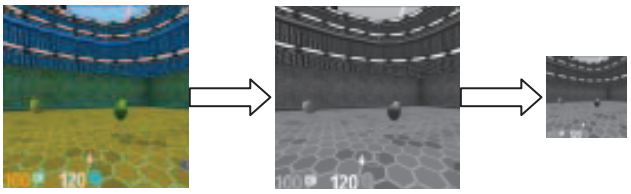
(3) Action selection

The main function of this module is to select the optimal action of the mobile robot according to the action selection strategy. During the DQN training, the ε-greedy strategy is used to select the action. During DQN computing, the action can be selected according to the optimal Q value directly.

The implementation principle of each module are described in detail as follows.

B. Image Acquisition and preprocessing

In order to reduce the computation load of the subsequent image operations, two kinds of operation, graying and resize can be used to reduce the dimension of the image. The whole process of the image preprocessing is shown in Fig.2. Fig.2 (a) is original RGB image captured by mobile robot system, Fig.2 (b) and Fig.2 (c) are images after graying and dimension reduction respectively.



(a) (b) (c)

Fig. 2. Schematic of the image preprocessing

(a)the original RGB image(80*80).(b)the image after graying(80*80). (c)the image after dimension reduction (40*40)

C. Value function acquisition

In this paper, a Deep Q Network is established to approximate the state-action value function Q of the mobile robot. The network input is the last 4 frames preprocessed image, and the output is the Q value of each possible action of the mobile robot.

Here, the state, action, and rewards can be defined as follows:

State: original RGB image collected from the environment.

Action: robot possible move, that is, turn left, turn right, forward.

Rewards:

$$r = \begin{cases} 1 & \text{reach apple} \\ -1 & \text{reach lemon} \end{cases}$$

The state-action value function Q is defined as an evaluation function after the action a_t is performed on the state s_t at time t, and its value can be updated using the Bellman equation shown in (1).

$$Q(s_t, a_t) = r_{t+1} + \gamma \max_{a \in A} Q(s_{t+1}, a) \tag{1}$$

Where r_{t+1} is the immediate reward obtained when action a_t is performed at the state s_t; a is all possible actions

at the state s_{t+1}, and the discount factor γ determines the importance of the future reward value.

Agent learning process contains many episodes (a run from the beginning to the end of a path planning), each episode repeat the following steps:

(1) Agent perceives the external environment state s_t at time t;

(2) Agent selects and executes an action a_t according to the action selection strategy;

(3) Action selected by agent is applied on the external environment, the state of the environment transfers from s_t to s_{t+1}. At the same time, agent receive an immediate reward r_{t+1};

(4) Update the target value of the Q according to the Bellman equation (1), and then update the DQN network parameters. $t \leftarrow t+1$, step into the next moment;

(5) If the new environmental state is the terminated one, then this episode is finished, else go back to step (1).

However, when the input is high-dimensional visual perception (image pixels), the state space becomes large and it is difficult to store the value function Q in the traditional table form. Therefore, the Convolutional Neural Network (CNN) can be used to approximate the Q function. This improved Q-learning algorithm is also called as Deep Q-Network (DQN) [5,6].

The DQN structure designed to approximate the Q function is shown in Fig.1. The network consists of two convolution layers (*conv1, conv2*) and two full connection layers (*fc1, fc2*). The first convolution layer *conv1* convolves the input image with 8 kernel of 3 * 3 with stride 2; The second convolution layer *conv2* is convoluted with 16 kernel of 3*3 with stride 2; The first full connection layer *fc1* has 128 nodes; The second full connection layer *fc2* has three nodes, and the value of the *fc2* output layer represent the Q value corresponding to each possible mobile robot action (i.e., turn left, turn right, forward) of a given input state (i.e., original RGB image collected from the environment). DQN training samples can be obtained using experience replay mechanism, and then the network parameters are updated using the general stochastic gradient descent and backpropagation algorithm. Here, the main idea of the experience replay mechanism is that during the DQN training, agent save the experience tuple e_t at each time point t, e_t = <current state, action, reward, next state>. The experience tuple is stored in the replay memory M of length N, $M = e_1, e_2, \cdots, e_N$. And then the random sampling experience e_t from the replay memory. The experience e_t would be used as training data to update the DQN weights.

D. Action selection

The main function of this module is to select the optimal mobile robot action. In other word, the input of the module is the Q values corresponding to three possible mobile robot action (i.e., turn left, turn right, forward), and the output of that is the optimal action to be executed by the mobile robot.

In the process of agent learning, there exists an exploration-exploitation trade-off problem. On the one hand, agent need to choose as many different actions as possible to find the optimal strategy, which can be called as exploration, on the other hand, agent would think about selecting the action with the largest Q value to obtain a big reward, which can be called as exploitation. Exploring is very important to learning. Optimal strategy can be determined only by exploring. However, too much exploration will reduce the performance of the mobile robot path planning system, and affect the learning speed. Therefore, a reasonable action selection strategy need to be designed to solve the above problem in the learning process, that is, exploration and exploitation need to be balanced.

ε-greedy strategy can make the system to avoid falling into a local optimal state, so ε-greedy strategy is used to complete the mobile robot action selection in this paper. In ε-greedy strategy, a certain probability of random changes is added in the process of the action selection. Agent in the current state, will select action randomly with the probability ε to ensure that all the state space may be explored, and select the action $a_{\max}$ with the largest current Q value with the probability $1-\varepsilon$ to make use of the learned knowledge as much as possible [9]. So the probability $P(a_{\max}, s)$ of action $a_{\max}$ being selected can be calculated by the (2) [10]

$$P(a_{\max}, s) = 1 - \varepsilon + \frac{\varepsilon}{N(A)} \tag{2}$$

Which $N(A)$ is the number of mobile robot actions in action set A.

III. EXPERIMENTAL RESULTS

To illustrate the performance of the proposed deep RL based mobile robot path planning method, we evaluate our algorithm in seekavoid_arena_01 environment on DeepMind Lab platform. DeepMind Lab is a first-person 3D game platform designed for research and development of general artificial intelligence and machine learning systems [11]. It provides a suite of challenging 3D navigation and puzzle-solving tasks for learning agents. The main purpose is to act as a testbed for research in artificial intelligence, particularly, deep reinforcement learning.

In the seekavoid_arena_01 map with 15 apples and 8 lemons, the task is to collect apples (positive reward +1) while avoiding lemons (negative reward -1). The highest score is 15, and lowest score is -8. Seekavoid_arena_01 consists tasks of navigation and path planning from the first-person perspective of the agent, which is suitable of the real world robotic applications. Therefore, seekavoid_arena_01 is a good testing platform for mobile robot path planning algorithms. In the path planning experiment, apples and lemons can be considered as goal points and obstacles respectively. A good path planning algorithms should make the mobile robot reach to the as many goal points (i.e., apples) as possible while avoiding obstacles (i.e., lemons).

Using the raw RGB images obtained from the seekavoid_arena_01 and the experiences saved in the replay memory, we train the DQN shown in Fig. 1. The related parameters in our experiment are set as Table 1.

TABLE I. IN THE EXPERIMENT

Parameter	Value
γ	0.99
Initial_ $\mathcal{E}$	1.0
End_ $\mathcal{E}$	0.1
Explore	150,000
Replay_memory_size	50,000
Batch_size	32
Step	500,000

Score change curve in the training process is shown in Fig.3. In Fig. 3, the y-axis shows the average score of each 5,000 steps. It can be seen that the scores are increasing obviously in the first 200,000 steps during the training process, which means the performance of the path planning is improving, more and more goal points are reachable, and more obstacles are avoided. After 200,000 steps of training, the leaning net would basically converge.

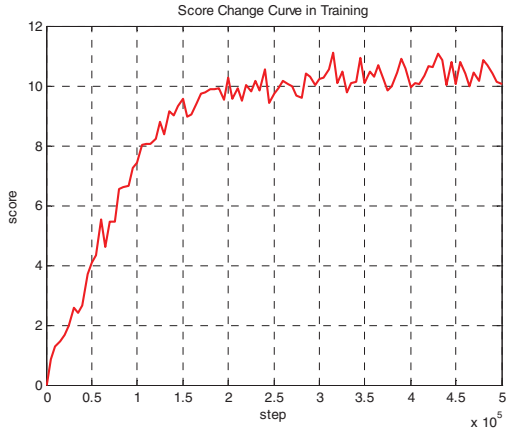


Fig. 3. Score change curve in the training process

We run 30 times path planning tests. The results are shown in Table II. The first column shows the trial number. The second column represents the score obtained using the agent proposed in this paper, where, the mean of the score is 10.03, the standard deviation of the score is 1.30, the maximum score is 12.0, and the minimum score is 7.0.

It can be seen from the above experimental results that the proposed path planning method is able to reach more goal points while avoiding obstacles and have better path planning performance.

IV. CONCLUSION

In order to make the robot obtain the optimal action directly from the original visual perception without hand-crafted features and features matching, a novel end-to-end path planning method--mobile robot path planning using deep reinforcement learning is proposed in this paper. The effectiveness of the proposed method is verified in seekavoid_arena_01 environment on DeepMind Lab platform. In the future, we will implement the proposed planning method in the real mobile robot combined with our previous work [12,13].

TABLE II. AGENT SCORES OF 30 TRIALS

Trial number	Score
1	10.0
2	8.0
3	11.0
4	11.0
5	11.0
6	10.0
7	10.0
8	7.0
9	11.0
10	11.0
11	10.0
12	10.0
13	8.0
14	10.0
15	12.0
16	9.0
17	10.0
18	12.0
19	11.0
20	10.0
21	9.0
22	11.0
23	10.0
24	10.0
25	10.0
26	10.0
27	12.0
28	9.0
29	7.0
30	11.0

REFERENCES

[1] X. Bu, H. Su, W. Zou, P. Wang, H. Zhou, " Ant Colony Path Planning Based on Non-uniform Modeling of Complex Environment," ROBOT,vol.38,pp.276-284,May 2016.
卜新苹, 苏虎, 邹伟, 王鹏,周海, "基于复杂环境非均匀建模的蚁群路径规划," 机器人,vol.38,pp.276-284,May 2016.

[2] J. Shreyas, J.Sandeep, "Modern Machine Learning Approaches for Robotic Path Planning," International Journal of Computer Science and Information Technologies, Vol. 8, pp.256-259, 2017.

[3] Q. Liu, J.W. Zhang, Z. C. Zhang, S. Zhong, Q. Zhou, P. Zhang, et.al, "A Survey on Deep Reinforcement Learning," Chinese Journal of Computers, Vol. 8, pp.256-259, 2017.
刘 全, 翟建伟,章宗长,钟珊,周倩,章鹏等,"深度强化学习综述," 计算机学报,Vol. 40, pp.1-28, 2017.

[4] L.W. Qiu, "Application of Deep Reinforcement Learning In Video Game Playing," Master Thesis of South China University of Technology,2015.
邱立威, "深度强化学习在视频游戏中的应用," 华南理工大学硕士论文,2015.

[5] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing Atari with Deep Reinforcement Learning," In Deep Learning, Neural Information Processing Systems Workshop, 2013.

[6] V. Mnih, K. Kavukcuoglu, D. Silver et al,“Human-level control through deep reinforcement learning,” Nature, vol. 518,pp. 529-553,February 2015.

[7] M.J. Wooldridge, and N.R. Jennings, “Intelligent Agent: Theory and Practice,” Knowledge Engineering Review, vol.10, pp.115-152, 1994.

[8] Y. Gao, S.F. Chen, X. Lu, “Research on Reinforcement Learning Technology: A Review,” ACTA Automatica Sinica, vol.30, pp. 86-100, January 2004.

高阳,陈世福,陆鑫, “强化学习综述,” 自动化学报, vol.30, pp. 86-100, January 2004.

[9] R.S. Sutton ,A.G. Barto,“Reinforcement Learning: an Introduction,” Cambridge, MA：MIT Press, 1998.

[10] J. G. Ren, “Navigation Control for Autonomous Mobile Robots Based on Reinforcement Learning,” Master Thesis of Harbin Institute of Technology, 2010.

任建功, “基于强化学习的自主式移动机器人导航控制,” 哈尔滨工业大学硕士论文,2010.

[11] C. Beattie, J.Z. Leibo, D. Teplyashin, et al, “Deepmind lab,” arXiv:1612.03801, https://arxiv.org/abs/1612.03801v1, 2016.

[12] J. Xin, X.L. Jiao, Y. Yang, D. Liu, “Visual Navigation for Mobile Robot with Kinect Camera in Dynamic Environment,” Proceedings of the 35th Chinese Control Conference, Chengdu, China, pp. 4757-4764,July 2016.

[13] J. Xin, J. Gou, X.M. Ma, K. Huang, D. Liu, Y. Zhang, “A Large Viewing Angle 3-Dimensional V-SLAM Algorithm with a Kinect-based Mobile Robot System” ROBOT,vol.36, pp.560-568,May 2014.

辛菁,苟蛟龙,马晓敏,黄凯,刘丁,张友民, “基于 Kinect 的移动机器人大视角三维 V-SLAM,” 机器人, vol.36, pp.560-568,May 2014.