Journal of System Simulation

Volume 35 | Issue 3 Article 13

3-22-2023

Obstacle Avoidance and Simulation of Carrier-Based Aircraft on the Deck of Aircraft Carrier

Junxiao Xue
1.School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou 450002, China;,
xuejx@zzu.edu.cn

Xiangyan Kong
1.School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou 450002, China;,
m15537229290@163.com

Bowei Dong
2.School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou 450001, China;

Hao Tao
3.China Ship Research and Design Center, Wuhan 430064, China;

See next page for additional authors

Obstacle Avoidance and Simulation of Carrier-Based Aircraft on the Deck of Aircraft Carrier

Abstract

Abstract: A predictive depth deterministic policy gradient (PDDPG) algorithm is proposed by combining the least squares method with deep deterministic policy gradient(DDPG) for the problems of strong randomness, poor real-time performance, and slow planning speed by obstacle avoidance on aircraft carrier deck. *The short-term trajectory of dynamic obstacles on the deck is predicted by the least square method. DDPG is used to provide agents with the ability to learn and make decisions in continuous space by the short-term trajectory of dynamic obstacles. The reward function is set based on the artificial potential field to improve the convergence speed and accuracy of the algorithm.* A high dynamic complex scene of aircraft carrier deck is constructed using unity 3D to simulate experiments of obstacle avoidance method. The experimental results show that the method can accurately realize the hybrid obstacle avoidance of carrier aircraft on the aircraft carrier deck, and the accuracy is improved by 7% ~ 30% compared with other methods. Compared with deep Q network (DQN), the path length and turning angle are reduced by 100 units and 400o~450o respectively.

Keywords

path planning, obstacle avoidance, least square method, artificial potential field, DDPG(deep deterministic policy gradient)

Authors

Junxiao Xue, Xiangyan Kong, Bowei Dong, Hao Tao, Haiyang Guan, Lei Shi, and Mingliang Xu

Recommended Citation

Junxiao Xue, Xiangyan Kong, Bowei Dong, Hao Tao, Haiyang Guan, Lei Shi, Mingliang Xu. Obstacle Avoidance and Simulation of Carrier-Based Aircraft on the Deck of Aircraft Carrier[J]. Journal of System Simulation, 2023, 35(3): 592-603.

第 35 卷第 3 期 系统仿真学报© Vol. 35 No. 3
2023 年 3 月 Journal of System Simulation Mar. 2023

航母甲板上舰载机的混合避障和仿真

薛均晓[1]，孔祥燕[1*]，董博威[2]，陶浩[3]，管海洋[2]，石磊[1]，徐明亮[2]
(1. 郑州大学 网络空间安全学院，河南 郑州 450002；2. 郑州大学 计算机与人工智能学院，河南 郑州 450001；
3. 中国船舰研究设计中心，湖北 武汉 430064)

摘要：针对航母甲板上舰载机混合避障随机性强、实时性差、规划速度慢等问题，结合最小二乘法与 DDPG(deep deterministic policy gradient)算法提出一种 PDDPG(predictive depth deterministic policy gradient)算法。*该方法利用最小二乘法预测航母甲板上动态障碍物的短期轨迹。DDPG 根据动态障碍物的短期轨迹为智能体提供在连续空间里学习和决策行为的能力。基于人工势场设置奖励函数，提高混合避障算法的收敛速度和准确率。*使用 Unity 3D 构建了航母甲板高动态复杂场景，进行舰载机混合避障仿真实验。实验结果表明，PDDPG 能较准确地实现航母甲板上舰载机的混合避障，与其他方法相比，在精度上提高了 7%~30%。与 DQN(deep Q network)相比，路径长度和转弯角度上分别减少了 100 个单位和 400º~450º。

关键词：路径规划；混合避障；最小二乘法；人工势场；DDPG
中图分类号：TP 391.9 文献标志码：A 文章编号：1004-731X(2023)03-0592-12
DOI: 10.16182/j.issn1004731x.joss.21-1145
引用格式: 薛均晓, 孔祥燕, 董博威, 等. 航母甲板上舰载机的混合避障和仿真[J]. 系统仿真学报, 2023, 35(3): 592-603.
Reference format: Xue Junxiao, Kong Xiangyan, Dong Bowei, et al. Obstacle Avoidance and Simulation of Carrier-Based Aircraft on the Deck of Aircraft Carrier[J]. Journal of System Simulation, 2023, 35(3): 592-603.

Obstacle Avoidance and Simulation of Carrier-Based Aircraft on the Deck of Aircraft Carrier

Xue Junxiao[1], *Kong Xiangyan*[1*], *Dong Bowei*[2], *Tao Hao*[3], *Guan Haiyang*[2], *Shi Lei*[1], *Xu Mingliang*[2]
(1. School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou 450002, China; 2. School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou 450001, China; 3. China Ship Research and Design Center, Wuhan 430064, China)

Abstract: A predictive depth deterministic policy gradient (PDDPG) algorithm is proposed by combining the least squares method with deep deterministic policy gradient(DDPG) for the problems of strong randomness, poor real-time performance, and slow planning speed by obstacle avoidance on aircraft carrier deck. *The short-term trajectory of dynamic obstacles on the deck is predicted by the least square method. DDPG is used to provide agents with the ability to learn and make decisions in continuous space by the short-term trajectory of dynamic obstacles. The reward function is set based on the artificial potential field to improve the convergence speed and accuracy of the algorithm.* A high dynamic complex scene of aircraft carrier deck is constructed using unity 3D to simulate experiments of obstacle avoidance method. The experimental results show that the method can accurately realize the hybrid obstacle avoidance of carrier aircraft on the aircraft carrier deck, and the accuracy is improved by 7% ~ 30% compared with other methods. Compared with deep Q network (DQN), the path length and turning angle are reduced by 100 units and 400º~450º respectively.

收稿日期：2021-11-09 修回日期：2022-01-18
基金项目：国家自然科学基金(62036010，61972362)；河南省自然科学基金(202300410378)；河南省高等学校青年骨干教师培养计划(22020GGJS014)
第一作者：薛均晓(1982-)，男，副教授，博士，研究方向为人工智能、虚拟现实。E-mail：xuejx@zzu.edu.cn
通讯作者：孔祥燕(1996-)，女，硕士生，研究方向为强化学习、路径规划。E-mail：m15537229290@163.com

Keywords: path planning; obstacle avoidance; least square method; artificial potential field; DDPG(deep deterministic policy gradient)

0 引言

路径规划是指在有障碍物的工作环境中，寻找一条从起点到终点、无碰撞地避开所有障碍物的平滑运动路径。航母甲板空间狭窄，作业危险系数高，障碍物密度大[1]。舰载机人工路径规划和调度方法速度慢，准确率低。一旦调度失误，将带来不可估量的损失和重大伤害。

舰载机是航空母舰搭载的核心战斗力量，其出动与回收的安全性和高效性是衡量航母作战和综合支援能力的重要技术指标。舰载机的加油、装弹、维保、布列、引导、起飞和着陆等各种任务都在航母甲板上完成。与陆上机场相比，航母甲板是一个高度异构动态的危险作业场景，在一块占地不到20 000 m^2的空间内挤满了舰载机、预警直升机、反潜直升机以及各种作业设备。航母甲板上同时存在多个机务保障点，且作业区域存在部分重合且没有固定的作业通道，因此会同时出动多架舰载机。要在高度受限的空间内、“人–机–物”混杂分布的环境、规程复杂繁巨且工序衔接紧凑的作业要求，保证航母甲板舰载机路径规划工作的高效和安全开展，需要对航母甲板的人、机、物等多个重要素进行统筹管理。与普通场景下的路径规划问题不同的是，航母甲板上舰载机的路径规划不仅需要考虑规划的精度和效率，还要考虑舰载机在高密度、高动态的障碍物存在情况下的混合避障问题。

1 相关工作

在这一部分，分别介绍传统路径规划算法、智能仿生路径规划算法和基于强化学习的路径规划算法。传统路径规划算法是最基本、最成熟的路径规划方法，其原理简单易实现，且应用广泛。智能仿生算法受启发于生物的行为规律，鲁棒性强。强化学习赋予智能体足够的智能不断与环境相互作用来获取未知环境的知识，然后实时的做出决策[2]。

1.1 传统路径规划算法

早期传统路径规划算法主要是基于图的路径规划[2-5]：例如C-space[2-3]，Dijkstra[4]，A*[5]等，此类方法无法解决动态障碍物的问题。人工势场[6]，快速探索随机树[7]和D*[8]可以用于解决动态障碍物的规避问题，但人工势场方法无法解决局部极小值的问题，而快速探索随机树方法虽然规划速度快，但是由于随机采样策略引入的随机性，导致每次规划的路径结果都不一样，而且路径较长，路径的平滑性也较差。在D*算法中，如果环境动态性强，环境的结构变化大，将造成局部重规划，过程异常繁琐且结果较差。因此，虽然传统路径规划算法具有参数少、易实现等优点，但此类算法无法处理动态环境下的避障问题，而且算法的实时性较差。

1.2 智能仿生算法

大部分智能仿生算法受启发于生物群体行为和特点，如蚁群算法[9]、粒子群算法[10]、遗传算法等[11]。蚁群算法是一种用来寻找优化路径的概率型算法，该方法受启发于蚁群留下信息素寻找路径的行为，用蚂蚁的行走路径表示待优化问题的可行解，将整个蚂蚁群体的路径定义为待优化问题的解空间。粒子群算法是一种通过模拟鸟群觅食行为的随机搜索算法，该方法首先初始化一组随机粒子(随机解)，然后通过迭代找到最优解，在每一次迭代中，粒子通过跟踪两个“极值”来更新自己。遗传算法是一种通过模拟自然进化过程搜索最优解的方法，该方法利用计算机仿真运算，将问题的求解过程转换成类似生物进化中的染色体基因的交叉、变异等过程。在求解较为复杂的

DOI: 10.16182/j.issn1004731x.joss.21-1145

组合优化问题时，相对一些常规的优化算法，通常能够较快地获得较好的优化结果。智能仿生算法简单易懂，鲁棒性强且具有较高的自组织能力，但此类算法容易陷入局部最优解，且收敛速度较慢[12]。

1.3 强化学习方法

强化学习[13]赋予智能体充足的智能不断与环境进行相互作用来获取未知环境的知识。Q-learning[14-15]可以被用来解决无人机的自主导航问题。文献[16]将Q-learning[17]用于含有动静态障碍物的路径规划中。虽然Q-learning适用于以上环境，但当状态空间和动作空间太大时，很难用表格的形式列举出所有状态动作对的Q值。2013年DeepMind提出了深度Q网络(deep Q network, DQN)，用深度学习拟合Q函数解决了维度灾难的问题，之后DQN被很多学者应用于路径规划。通过使用基于值的双DQN(double DQN, DDQN)[18]，文献[19]提出了一种分层深度Q网络和优先经验回放相结合的三维路径规划方法。文献[20]提出了一种深度Q网络和最小二乘法相结合的PDQN (parameterized DQN)算法。虽然深度Q网络成功的解决了的维度灾难的难题，但仍不能输出连续的动作。文献[21]提出了深度确定性策略梯度(deep deterministic policy gradient, DDPG)算法来实现连续动作的输出，但DDPG算法对于高动态、高密度场景的路径规划存在收敛速度慢，准确率低等问题。

为解决以上问题，本文提出了一种结合最小二乘法和确定性策略梯度算法的新算法。本文的主要贡献如下：

(1) 提出一种预测深度确定性策略梯度算法(predictive depth deterministic policy gradient, PDDPG)，利用人工势场(障碍和目标分别对舰载机施加排斥和吸引)的思想设计奖励函数，大大提高了舰载机混合避障和路径规划方法的效率。

(2) 根据目标的历史轨迹与环境信息，对目标的未来轨迹进行预测，降低了环境的动态性。

(3) 实验结果表明，PDDPG与DQN、A2C (advantage actor critic)和DDPG等算法相比在精度上高了7%~30%，在路径长度和转弯角度等评价指标上也有较大提升。

2 本文方法

2.1 场景建模

在甲板模型中，将障碍物分为静态和动态障碍物，将停靠位置、起飞位置等设置为作业点，作业时间随机生成。舰载机的任务是找到一条从起点到目标点最优或次优的无碰撞路径。

2.1.1 状态空间

状态空间是指舰载机在决策之前获得的环境信息。用于帮助舰载机评估环境情况，实时做出决策动作。状态分为静止环境状态和预测环境状态，可表示为$s(t)$: $\{s^t, u(t)\}$。

静止环境状态s^t是指当前环境中静止的障碍物和目标点与智能体舰载机的关系，表示为

$$s^t: \{(d^t_{e,x}, d^t_{e,y}), (d^t_{o1,x}, d^t_{o1,y}) \cdots (d^t_{on,x}, d^t_{on,y})\}$$

式中：$d^t_{e,x}$, $d^t_{e,y}$为舰载机当前位置与目标点在横、纵坐标上的距离差在环境中的占比，如式(1)，(2)所示；$d^t_{oi,x}$, $d^t_{oi,y}$表示舰载机当前位置与静态障碍物在横、纵坐标上的距离差在环境中的占比，如式(3)，(4)所示。

$$d^t_{e,x} = \frac{x_e - x^t}{W} \tag{1}$$

$$d^t_{e,y} = \frac{y_e - y^t}{H} \tag{2}$$

$$d^t_{oi,x} = \frac{x^t_{oi} - x^t}{W} \tag{3}$$

$$d^t_{oi,y} = \frac{y^t_{oi} - y^t}{H}, \quad i = 1, 2, \cdots, n \tag{4}$$

式中：(x_e, y_e)为目标点的坐标；(x^t, y^t)为舰载机当前位置坐标；(x^t_{oi}, y^t_{oi})为第i个静态障碍物的坐标，$i = 1, 2, \cdots, n$；W和H分别为环境的宽和高。

预测环境状态是指当前环境中动态障碍物预

测位置和目标点对智能体舰载机的影响。预测环境状态$u(t)$表示为

$$u(t):\{(d^t_{o'1,x},\ d^t_{o'1,y}),(d^t_{o'2,x},\ d^t_{o'2,y}),\cdots,(d^t_{o'm,x},d^t_{o'm,y})\}$$

式中：$d^t_{o'j,\,x}$, $d^t_{o'j,\,y}$为动态障碍物的预测位置与舰载机在横、纵坐标上的距离差在环境中的占比，如式(5)，(6)所示：

$$d^t_{o'j,x}=\frac{x^{t+1}_{o'j}-x^t}{W} \tag{5}$$

$$d^t_{o'j,y}=\frac{y^{t+1}_{o'j}-y^t}{H} \tag{6}$$

式中：$j\in 1,2,\cdots,m$；$(x^t_{o'j},\ y^t_{o'j})$为第j个静态障碍物的坐标。

2.1.2 动作空间

舰载机的动作空间表示舰载机根据状态空间决定要执行的动作。将动作空间设置为A:(X, Y)。X和Y分别为舰载机在x和y方向上移动的距离：

$$X=40x,\quad Y=40y,\quad x,\ y\in(-1,\ 1) \tag{7}$$

式中：x, y为DDPG中Actor网络的输出。

2.1.3 奖励函数

深度强化学习的奖励函数用于评估智能体采取的行为。奖励值设置的好坏决定智能体最终能否学到期望技能、影响算法的收敛速度和性能。其中最简单的设置方法是稀疏奖励，只有完成任务，智能体才能获得正回报。但是，此方法无法收集有用的经验数据以帮助智能体学习。因此，网络更新的收敛速度很慢，并且智能体学习到的策略很差。在本文中，将人工势场的思想(障碍物和目标分别对舰载机排斥和吸引)引入奖励函数的设计中。设置4类奖励函数：①目标引力奖励；②障碍斥力奖励；③碰撞奖励；④到达目标奖励。

目标引力的奖励函数是根据目标点对舰载机产生的引力势场设计的。为了奖励值设计合理，简化引力场计算方法。目标引力的奖励函数设置如式(8)~(10)所示：

$$\begin{cases}D^t_{i,e}=|p^t\,p^t_e|\\ D^{t+1}_{i,e}=|p^{t+1}\,p^{t+1}_e|\end{cases} \tag{8}$$

$$r_1=\begin{cases}L,\ \ L\leqslant D^t_{i,\mathrm{e}}-D^{t+1}_{i,\mathrm{e}}\\ D^t_{i,\mathrm{e}}-D^{t+1}_{i,\mathrm{e}},\ \ l<D^t_{i,\mathrm{e}}-D^{t+1}_{i,\mathrm{e}}<L\\ l,\ \ 0<D^t_{i,\mathrm{e}}-D^{t+1}_{i,\mathrm{e}}\leqslant l\end{cases} \tag{9}$$

$$r_1=\begin{cases}-l,\ \ -l\leqslant D^t_{i,\mathrm{e}}-D^{t+1}_{i,\mathrm{e}}<0\\ D^t_{i,\mathrm{e}}-D^{t+1}_{i,\mathrm{e}},\ \ -L<D^t_{i,\mathrm{e}}-D^{t+1}_{i,\mathrm{e}}<-l\\ -L,\ \ D^t_{i,\mathrm{e}}-D^{t+1}_{i,\mathrm{e}}\leqslant -L\end{cases} \tag{10}$$

式中：p^t，p^{t+1}分别为t和$t+1$时刻舰载机的位置坐标；p_e为目标点的位置坐标；$D^t_{i,e}$，$D^{t+1}_{i,e}$分别为t和t+1时刻舰载机与目标点之间的距离。当$D^t_{i,\mathrm{e}}-D^{t+1}_{i,\mathrm{e}}>0$时，奖励值$r_1$的计算如式(9)所示；当$D^t_{i,\mathrm{e}}-D^{t+1}_{i,\mathrm{e}}<0$时，奖励值$r_1$的计算如式(10)所示。$L$，$l$，$-l$，$-L$分别为奖励值的上限和下限。

障碍斥力的奖励函数是根据障碍物对舰载机产生的斥力势场设计的。同理，简化斥力场计算方法。障碍斥力的奖励函数设置如式(11)~(16)所示：

$$\begin{cases}D^t_{i,oj}=\{|p^t\,p^t_{oj}||j=1,\ 2,\cdots,n\}\\ D^{t+1}_{i,oj}=\{|p^{t+1}\,p^{t+1}_{oj}||j=1,2,\cdots,n\}\end{cases} \tag{11}$$

$$\begin{cases}D^t_{i,o'k}=\{|p^t\,p^t_{o'k}||k=1,\ 2,\cdots,m\}\\ D^{t+1}_{i,o'k}=\{|p^{t+1}\,p^{t+1}_{o'k}||k=1,\ 2,\cdots,m\}\end{cases} \tag{12}$$

$$r_2=\begin{cases}r_2+H,\ \ H\leqslant D^t_{i,oj}-D^{t+1}_{i,oj}\\ r_2+D^t_{i,oj}-D^{t+1}_{i,oj},\ \ h<D^t_{i,oj}-D^{t+1}_{i,oj}<H\\ r_2+h,\ \ 0<D^t_{i,oj}-D^{t+1}_{i,oj}\leqslant h\end{cases} \tag{13}$$

$$r_2=\begin{cases}r_2-h,\ \ -h\leqslant D^t_{i,oj}-D^{t+1}_{i,oj}<0\\ r_2+D^t_{i,oj}-D^{t+1}_{i,oj},\ \ -H<D^t_{i,oj}-D^{t+1}_{i,oj}<-h\\ r_2-H,\ \ D^t_{i,oj}-D^{t+1}_{i,oj}\leqslant -H\end{cases} \tag{14}$$

$$r_2=\begin{cases}r_2+H,\ \ H\leqslant D^t_{i,oj}-D^{t+1}_{i,o'k}\\ r_2+D^t_{i,o'k}-D^{t+1}_{i,o'k},\ \ h<D^t_{i,o'k}-D^{t+1}_{i,o'k}<H\\ r_2+h,\ \ 0<D^t_{i,o'k}-D^{t+1}_{i,o'k}\leqslant h\end{cases} \tag{15}$$

$$r_2=\begin{cases}r_2-h,\ \ -h\leqslant D^t_{i,o'k}-D^{t+1}_{i,o'k}<0\\ r_2+D^t_{i,o'k}-D^{t+1}_{i,o'k},\ \ -H<D^t_{i,o'k}-D^{t+1}_{i,o'k}<-h\\ r_2-H,\ \ D^t_{i,o'k}-D^{t+1}_{i,o'k}\leqslant -H\end{cases} \tag{16}$$

式中：p^t，p^{t+1}，p^t_{oj}，p^{t+1}_{oj}分别为舰载机、静态障碍物在t和t+1时刻的位置坐标；$p^t_{o'k}$，$p^{t+1}_{o'k}$分别为动态障碍物在t和t+1时刻的预测位置坐标；$D^t_{i,oj}$，$D^{t+1}_{i,oj}$分别为t和t+1时刻舰载机与静态障碍物之间

的距离；$D_{i,o'k}^{t}$，$D_{i,o'k}^{t+1}$分别为t和t+1时刻舰载机与动态障碍物预测位置之间的距离。由于静态障碍物和动态障碍物的数量为多个，因此利用累加奖励值作为障碍斥力奖励值。当$D_{i,oj}^{t}-D_{i,oj}^{t+1}>0$时，奖励值$r_2$的计算如式(13)所示；当$D_{i,oj}^{t}-D_{i,oj}^{t+1}<0$时，奖励值$r_2$的计算如式(14)所示。同理舰载机和动态障碍物的奖励值的计算如式(15)，(16)所示。H，h，$-h$，$-H$分别为奖励值的上限和下限。

碰撞奖励是指舰载机与障碍物发生碰撞所产生的奖励值。如图1所示，当$\exists D<d$时，$r_3=-50$。

$$D=\{\|p^t p_{ol}^t\| l=1, 2, \cdots, m, m+1, \cdots, m+n\}$$

式中：d为舰载机间的安全距离；p^t为舰载机t时刻的位置坐标；p_{ol}^t为任意一个障碍物t时刻的位置坐标。

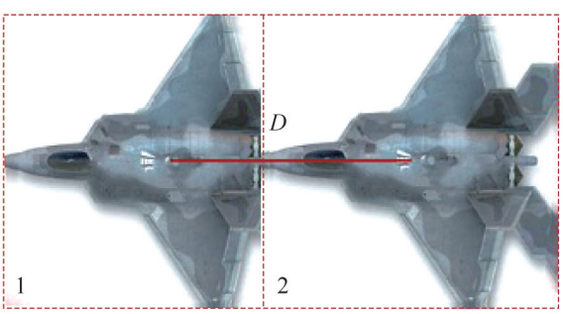


图1 碰撞状态
Fig. 1 Collision state

到达目标的奖励是指舰载机到达目标点所产生的奖励值。当$p^t=p_e^t$时，r_4=200。其中p^t表示舰载机当前位置坐标，p_e^t表示目标位置坐标。

综上所述，将所有可能产生的奖励值赋予其权重然后将累加的和作为总奖励值。总奖励值为

$$R=\lambda_1\cdot r_1+\lambda_2\cdot r_2+\lambda_3\cdot r_3+\lambda_4\cdot r_4 \tag{17}$$

式中：λ_1、λ_2、λ_3、λ_4为4种奖励值的权重。

2.2 算法框架图

2.2.1 轨迹预测

轨迹预测任务旨在根据目标当前或者历史轨迹与环境信息，对该目标未来的行驶轨迹进行预测。为了降低环境的动态性，使用最小二乘法来预测动态障碍物未来的坐标。如图2所示，最小二乘法的预测模型将静止环境状态s^t和动态障碍物的历史轨迹：$x_t^{t+N}=\{(p_1^t, p_1^{t+1}, \cdots, p_1^{t+N}), (p_2^t, p_2^{t+1}, \cdots, p_2^{t+N}) \cdots (p_n^t, p_n^{t+1}, \cdots, p_n^{t+N})\}$作为输入，通过最小二乘法的预测模型计算出动态障碍物的预测轨迹$y(t)=\{p_1^{t+N+1}, p_2^{t+N+1}, \cdots, p_n^{t+N+1}\}$。通过动态障碍物的预测轨迹$y(t)$和当前舰载机的位置计算预测环境状态$u(t)$。连接预测环境状态$u(t)$和静止环境状态$s^t$作为环境状态$s(t)$。

2.2.2 动作选择

强化学习决策的本质是动作选择。如图2所示，环境将历史轨迹x_t^{t+N}输入环境预测模块，预测模块通过计算将预测环境状态$u(t)$输入给环境，将$s(t)$输入DDPG主干网络的执行者，输出舰载机要执行动作a:(x, y)，其中x和y分别表示舰载机在x和y方向上移动的距离。舰载机执行动作后，静止环境状态由s^t: $\{(d_{e,x}^t, d_{e,y}^t), (d_{o1,x}^t, d_{o1,y}^t), \cdots, (d_{on,x}^t, d_{on,y}^t)\}$变为$s^{t+1}$: $\{(d_{e,x}^{t+1}, d_{e,y}^{t+1}), (d_{o1,x}^{t+1}, d_{o1,y}^{t+1}), \cdots, (d_{on,x}^{t+1}, d_{on,y}^{t+1})\}$动态障碍物的历史轨迹$x_t^{t+N}=\{(p_1^t, p_1^{t+1}, \cdots, p_1^{t+N}), (p_2^t, p_2^t, \cdots, p_2^{t+N}), \cdots, (p_n^t, p_n^{t+1}, \cdots, p_n^{t+N})\}$更新为$x_{t+1}^{t+N+1}=\{(p_1^{t+1}, p_1^{t+2}, \cdots, p_1^{t+N+1}), (p_2^{t+1}, p_2^{t+2}, \cdots, p_2^{t+N+1}), \cdots, (p_n^{t+1}, p_n^{t+2}, \cdots, p_2^{t+N+1})\}$。同时获得奖励值$r$。然后，将动态障碍物的历史轨迹$x_{t+1}^{t+N+1}$再输入给预测模块获得新的预测环境状态$u(t+1)$: $\{(d_{o'1,x}^{t+1}, d_{o'1,y}^{t+1}), (d_{o'2,x}^{t+1}, d_{o'2,y}^{t+1}), \cdots, (d_{o'm,x}^{t+1}, d_{o'm,y}^{t+1})\}$，之后连接静止环境状态$s^{t+1}$和预测环境状态$u(t+1)$得到新的环境状态$s(t+1)$: $\{s^{t+1}, u(t+1)\}$。最后环境将$(s(t), a_t, r_t, s(t+1))$存储在经验池中。

2.2.3 更新

DDPG是基于AC框架对DQN算法的扩展。它保留了DQN的双网络结构和经验重放。但与DQN的硬更新不同，DDPG采用软更新的方法来更新目标网络。更新Actor和Critic网络时，从经验池中采样N个小批量样本，然后根据当前的目标Q值，计算Critic网络的损失函数$L(\theta^Q)$，如式(18)，(19)所示。同时，通过策略梯度的方法更新Actor网络，如式(20)所示：

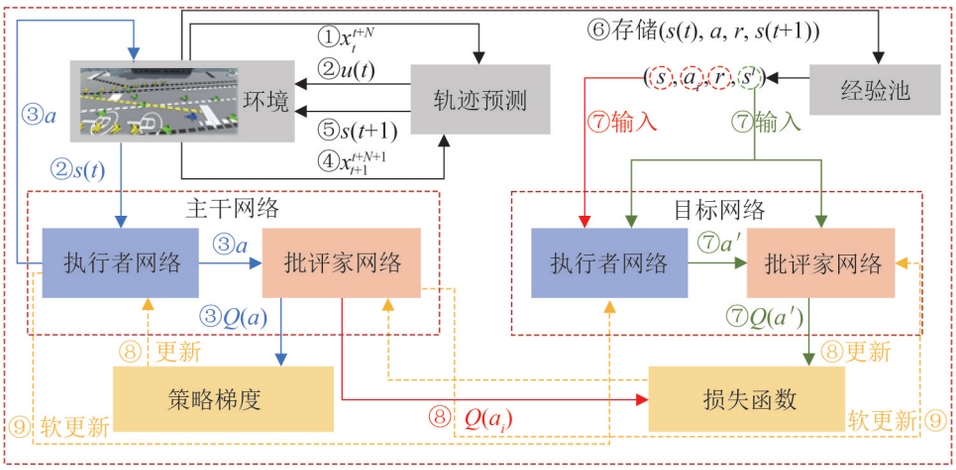


图2　算法框架图
Fig. 2　Framework of algorithm

$$y_t=\begin{cases}R_t\\ R_t+\gamma Q'(\phi(S'_t),\ \pi_{\theta'}(\phi(S'_t)),\omega)\end{cases} \tag{18}$$

$$L(\theta^Q)=\frac{1}{N}\sum_t^N(Y_t-Q(s_t,\ a_i|\theta^Q))^2 \tag{19}$$

$$\nabla_{\theta^\mu}J\approx N^{-1}\sum_t^N\nabla_{a_t}Q(s_t,\ a_t|\theta^Q)\nabla_{\theta^\mu}\mu(s_t|\theta^\mu) \tag{20}$$

最后，通过软更新来更新 2 个目标网络参数 $\theta^{\mu'}$ 和 $\theta^{Q'}$，如式(21)所示：

$$\begin{cases}\theta^{Q'}=\tau\theta^Q+(1-\tau)\theta^{Q'}\\ \theta^{\mu'}=\tau\theta^\mu+(1-\tau)\theta^{\mu'}\end{cases} \tag{21}$$

式中：τ 用于调节软更新因子。

3　仿真分析与验证

本文算法基于 CUDA 9.0 深度学习框架和 Tensorflow 1.14.0 利用 Unity 3D 进行仿真。运行环境为 Intel i7 870 0K 六核 3.7 GHZ 主频，GPU 为 GeForce RTX 2070 SUPER。

3.1　场景设置

在仿真系统中设计了 2 个不同的场景，如图 3 和 4 所示，在航母甲板上，蓝色舰载机要避开动静态障碍物(绿色和黄色舰载机分别代表动态障碍物和静态障碍物)，找到一条从起点到目标点(白板的位置)的无碰撞路径。在场景 1 中，设置 5 个动态和 15 个静态障碍物。为了验证算法在高动态环境中的有效性，在场景 2 中，设置了 9 个动态障碍物和 15 个静态障碍物。

图3　场景1
Fig. 3　Scene 1

图 4　场景2
Fig. 4　Scene 2

3.2　障碍物设置

障碍物分为动态障碍物和静态障碍物。甲板上的静态障碍物有弹射装置、舰岛、起降机及停放的舰载机。动态障碍物指在甲板上做任务的舰

第 35 卷第 3 期 系统仿真学报 Vol. 35 No. 3
2023 年 3 月 Journal of System Simulation Mar. 2023

载机，其状态在临时静态和临时动态之间转换。当舰载机在任务途中时其状态为临时动态，而在进行作业时其状态为临时静态。默认初始环境中所有动态舰载机都在任务途中。当其到达任务点时，在此位置上进行作业，完成后，动态舰载机随机生成新的速度去向新的任务点。

3.2 仿真分析

这一部分分别描述2个场景中智能体和动态障碍物的轨迹。如图5，6所示，智能体和动态障碍物分别由蓝色和绿色舰载机表示且其轨迹分别由蓝色和绿色虚线表示。黄色舰载机表示静态障碍物。图5，6中带方框的字母a表示各个场景中舰载机最终的落脚点。图5中带方框的字母b~f和图6中带方框的字母b~j分别表示各个场景中的动态障碍物在舰载机到达终点时的位置。图5，6中蓝色虚线上的圆点表示舰载机a每次决策后的位置，绿色虚线上的圆点表示障碍物每次移动后的位置。黑色数字表示对应动态障碍物在此位置停留的决策次数。算法决策时间在0.000 96~0.011 0 s之间。

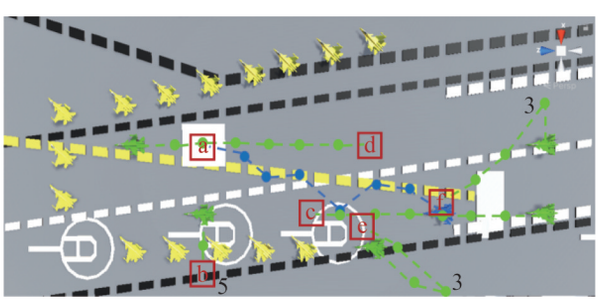


图 5 场景1轨迹图
Fig. 5 Trajectory diagram of Scene 1

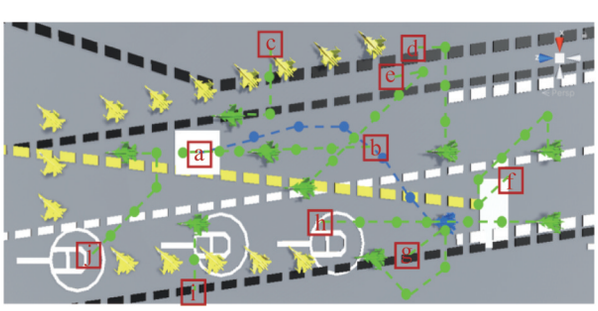


图 6 场景2轨迹图
Fig. 6 Trajectory diagram of Scene 2

舰载机每执行一个动作表示1步。如图5场景1的轨迹图所示，舰载机a未碰撞任何障碍物成功到达目标点。当舰载机a在第3步，向左下方走时，发现其离目标点更远了，然后迅速做出调整，向上移动；当舰载机a发现照此方向移动会与动态障碍物d发生碰撞时，舰载机a在第5步执行躲避动态障碍物d的动作。如图6场景2的轨迹图所示，当舰载机a发现障碍物b和e出现在舰载机a周围时，舰载机a向左下方移动以避免碰撞2个动态障碍物b，e，并确保舰载机a靠近目标。

3.4 评价指标和对比算法的设计

为了保证本文算法在大量动态障碍物中仍能保持较高的准确率，通过增加动态障碍物来设置多个实验场景，从准确率、路径长度、平均奖励值、平均转弯角度这4个方面和DDPG，DQN和A2C算法进行比较。

3.4.1 对比算法

DQN利用神经网络代替Q表生成每个动作的Q值，同时引入经验池且采用双网络结构延迟更新的方式。A2C作为AC算法的扩展，采用优势函数的方式代替批评家网络的原始回报来衡量动作的优劣，执行者网络根据优势函数的值调整网络参数。由于AC算法中批评家网络收敛速度较慢，执行者网络的更新取决于批评家网络传递的参数，所以整个算法存在难收敛的问题，DQN采用双网络结构，算法较为稳定，但无法处理连续动作。DeepMind在两者的基础上提出了DDPG算法，它保留了DQN的经验回放以及固定Q网络，采用双双网络结构，使用策略网络直接输出确定性动作。

为了实验公平起见，4种算法设置相同的实验参数、状态空间以及奖励函数。由于DQN只能处理离散动作，因此将DQN的动作空间设置为：$A=\{a_u, a_d, a_l, a_r\}$。其中，a_u，a_d，a_l，a_r分别表示舰载机向上、向下、向左、向右移动的决策。而其他2个算法的动作空间和PDDPG保留一致。

3.5 仿真结果

3.5.1 准确率

首先比较了2个场景中不同算法的准确率。如表1所示，随机抽样100组准确率计算平均值，在场景1中，PDDPG的平均准确率为0.98，DDPG和DQN算法分别为0.92、0.90。而A2C算法的准确率最低只有0.67。与DQN、A2C和DDPG相比，PDDPG的准确率分别增加了0.8、0.31和0.6。如图7场景1准确率所示，DDPG和PDDPG的准确率都可以达到100%，但PDDPG达到100%的次数最多且最稳定。而A2C和DQN算法的准确率较低，且A2C算法的稳定性最差。如表1和图8所示，场景2中，4种算法的准确率均有所下降，但PDDPG的准确率仍然最高，且可以达到0.91。

表1 准确率
Table 1 Accuracy

算法	场景1	场景2
DQN	0.90	0.90
A2C	0.67	0.64
DDPG	0.92	0.82
PDDPG	0.98	0.91

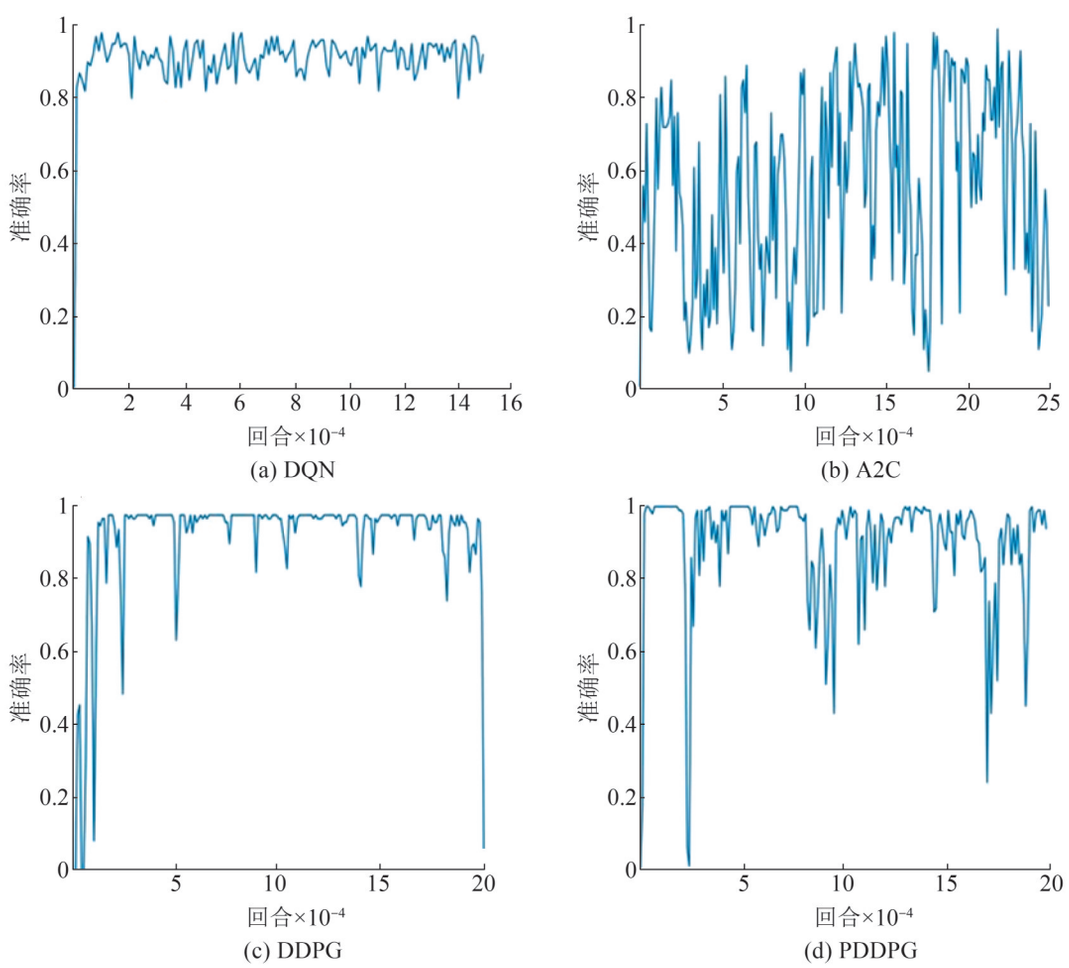


图7 场景1准确率
Fig. 7 Accuracy of Scene 1

3.5.2 平均奖励值

利用平均奖励值比较4种算法的学习效果。如图9场景1中4种算法的平均奖励值所示，由于A2C算法难收敛，其平均奖励值只能稳定在100左右，而其他算法都能达到150。如图9~10所示，相比于PDDPG，DDPG的平均奖励值更不稳

DOI: 10.16182/j.issn1004731x.joss.21-1145

第 35 卷第 3 期 系统仿真学报 Vol. 35 No. 3
2023 年 3 月 Journal of System Simulation Mar. 2023

定，即 DDPG 在没有预测的情况下学习效果较差。DQN 和 PDDPG 的平均奖励值都稳定在 150 左右，而 PDDPG 具有更高的稳定性。对比各个算法，PDDPG 的平均奖励值最高且最稳定，因此具有最佳的学习效果。

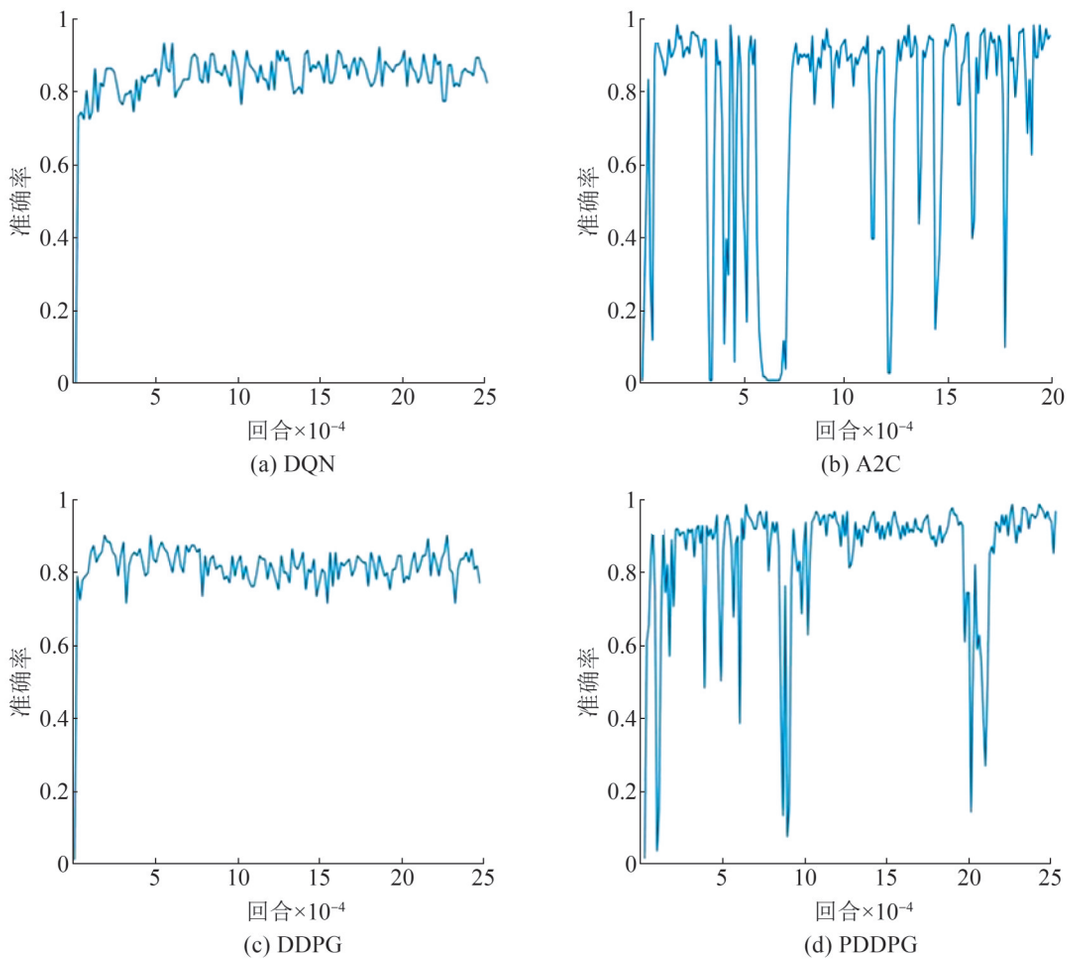
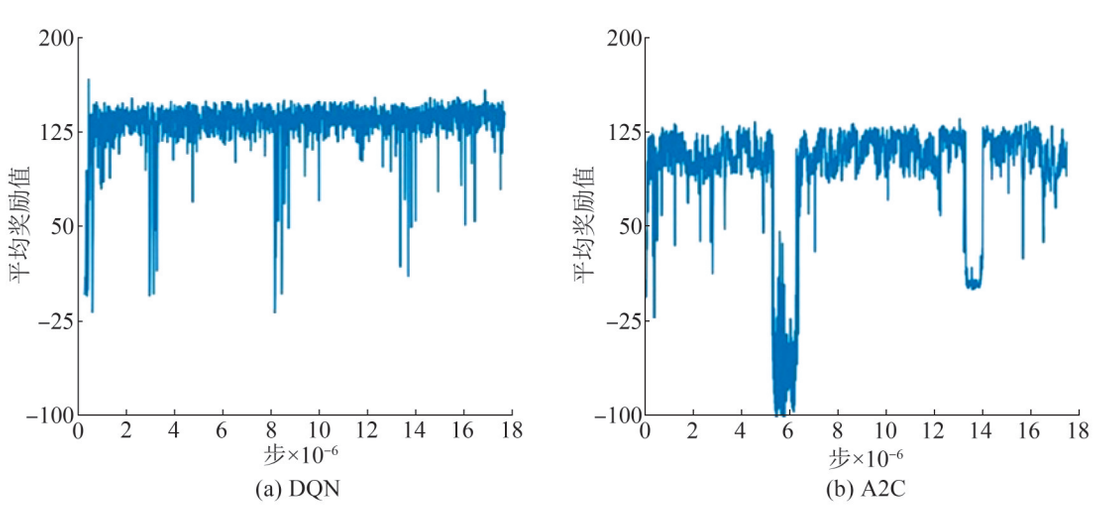


图 8　场景 2 准确率

Fig. 8　Accuracy of Scene 2



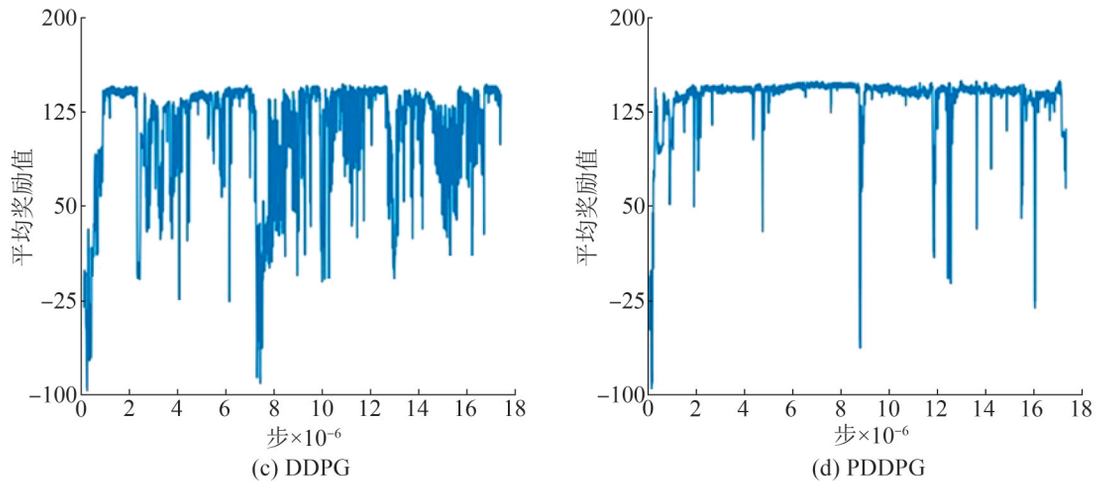


图9　场景1平均奖励值

Fig. 9　Average reward value of Scene 1

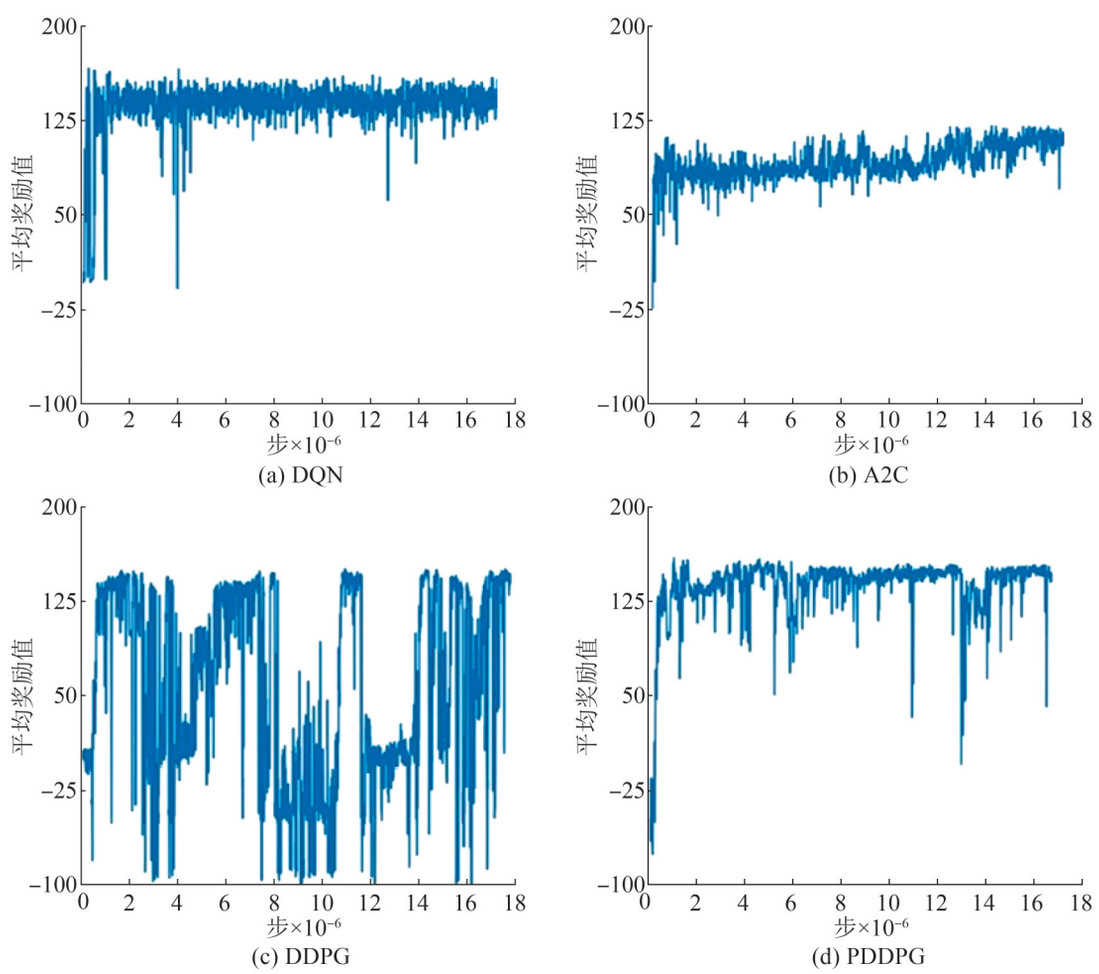


图10　场景2平均奖励值

Fig. 10　Average reward value of Scene 2

DOI: 10.16182/j.issn1004731x.joss.21-1145

3.5.3 平均路径长度和平均转弯角度

在航母甲板上中，找到一条平滑的最优或次优的路径不仅可以减少资源消耗，缩短规划时间而且还可以提高任务调度的效率。因此比较各个算法的平均路径长度和平均转弯角度。如表 2 和表 3 所示，场景 1 和场景 2 中 4 种算法的平均路径长度和平均转弯角度差距不大。对比 4 种算法，A2C、DDPG 和 PDDPG 规划的路径长度和平均转弯角度基本相似。由于 DQN 只能处理离散动作，DQN 所规划的路径的平均路径长度最长且平均转弯角度最大。

表 2　平均路径长度
Table 2　Average path length

算法	场景 1	场景 2
DQN	438.22	465.45
A2C	352.60	324.16
DDPG	328.97	320.61
PDDPG	328.95	315.68

表 3　平均转弯角度
Table 3　Average turning angle (º)

算法	场景 1	场景 2
DQN	586.73	522.62
A2C	142.03	158.00
DDPG	139.96	134.28
PDDPG	139.80	132.58

4　结论

本文设计了一种基于 DDPG 和最小二乘法的路径规划新算法 PDDPG，该算法结合了深度强化学习的感知和决策能力以及最小二乘法的预测能力，应用于高度不确定的场景。最小二乘法实现了动态障碍物的轨迹预测，大大降低了环境的不确定性，有效地解决了传统算法在动态环境中面临的收敛速度慢，泛化能力差等问题。DDPG 适用于解决连续状态空间的问题，该问题更符合真实场景。为了验证算法的效率，选择 Unity3D 来对甲板上舰载机的路径规划进仿真，并对 2 个场景中智能体和动态障碍物的轨迹进行了分析。其结果表明，与 DDPG，DQN 和 A2C 等算法相比，PDDPG 在精度上提高了 7%~30%；与 DQN 相比，PDDPG 在路径长度和转弯角度方面分别减少了 100 个单位和 400º~450º。

参考文献：

[1] 薛均晓, 徐明亮, 李亚飞, 等. 面向航空母舰电子显灵板的多智能体建模技术研究进展[J]. 计算机辅助设计与图形学学报, 2021, 33(10): 1475-1485.
Xue Junxiao, Xu Mingliang, Li Yafei, et al. Research Progress of Multi-agent Technology for Aircraft Carrier Electronic Display Panel[J]. Journal of Computer Aided Design and Graphics, 2021, 33(10): 1475-1485.

[2] Wang J, Meng M Q H. Optimal Path Planning Using Generalized Voronoi Graph and Multiple Potential Functions[J]. IEEE Transactions on Industrial Electronics (S0278-0046), 2020, 67(12): 10621-10630.

[3] Dang T, Mascarich F, Khattak S, et al. Graph-Based Path Planning for Autonomous Robotic Exploration in Subterranean Environments[C]// 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). China: IEEE, 2019: 3105-3112.

[4] DIJKSTRA E W. A Note on Two Problems in Connexion with Graphs[J]. Numerische Mathematik (S0299-599X), 1959, 1(1): 2569-271.

[5] Hart P E, Nilsson N J, Raphael B. A Formal Basis for The Heuristic Determination of Minimum Cost Paths[J]. Acm Sigart Bulletin (S0029-599X), 1972, 4(2): 28-29.

[6] Hwang Y K, Ahuja N. A Potential Field Approach to Path Planning[J]. IEEE Transactions on Robotics and Automation (S1070-9932) 1992, 8(1): 23-32.

[7] Bruce J, Veloso M M. Real-Time Randomized Path Planning for Robot Navigation[C]// Robot Soccer World Cup. Lausanne, Switzerland: Springer; 2002: 288-295.

[8] Stentz A. Intelligent Unmanned Ground Vehicles[M]. Germany: Springer, 1997: 203-220.

[9] Dorigo M, Gambardella L M. Ant Colony System: A Cooperative Learning Approach to the Traveling Salesman Problem[J]. IEEE Trans on Evolutionary Computation (S1089-778X), 1997, 1(1): 53-66.

[10] Li G, Chou W. Path Planning for Mobile Robot Using Self-adaptive Learning Particle Swarm Optimization[J]. Science China Information Sciences (S1674-733X), 2018, 61(5): 052204.

[11] Katoch S, Chauhan S S, Kumar V. A Review on Genetic Algorithm: Past, Present, and Future[J]. Multimedia Tools and Applications (S1380-7501), 2021, 80(5): 8091-8126.

[12] Zhang H, Lin W, Chen A. Path Planning for the Mobile

Robot: A Review[J]. Symmetry (S2073-8994), 2018, 10 (10): 450.

[13] Sutton R S, Barto A G. Reinforcement Learning: An Introduction[M]. England: MIT press, 2018.

[14] Yan C, Xiang X. A Path Planning Algorithm for UAV Based on Improved Q-learning[C]// 2nd International Conference on Robotics and Automation Sciences (ICRAS).Wuhan, China: IEEE, 2018: 1-5.

[15] Bouhamed O, Ghazzai H, Besbes H, et al. Q-learning Based Routing Scheduling for a Multi-task Autonomous Agent[C]// 2019 IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS). USA: IEEE, 2019: 634-637.

[16] Sichkar V N. Reinforcement Learning Algorithms in Global Path Planning for Mobile Robot[C]// 2019 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM). Sochi, Russia: IEEE, 2019: 1-5.

[17] Watkins C J, Dayan P. Q-learning[J]. Machine Learning (S0885-6125), 1992 (3/4): 279-292.

[18] Mohit Sewak. Deep Reinforcement Learning[M]. Germany: Springer, 2019: 95-108.

[19] Sun Yushan, Ran Xiangrui, Zhang Guocheng, et al. AUV 3D Path Planning Based on the Improved Hierarchical Deep Q Network[J]. Journal of Marine Science and Engineering (S2077-1312), 2020, 8(2): 145.

[20] 薛均晓, 孔祥燕, 郭毅博, 等. 基于深度强化学习的舰载机动态避障方法[J]. 计算机辅助设计与图形学学报 (S1003-9775), 2021, 33(7): 1102-1112.
Xue Junxiao, Kong Xiangyan, Guo Yibo, et al. Dynamic Obstacle Avoidance Method for Carrier Aircraft Based on Deep Reinforcement Learning[J]. Journal of Computer Aided Design and Graphics (S1003-9775), 2021, 33(7): 1102-1112.

[21] Zhang Xi, Liu Youbo, Duan Jiajun, et al. DDPG-based Multi-agent Framework for SVC Tuning in Urban Power Grid with Renewable Energy Resources[J]. IEEE Transactions on Power Systems (S1558-0679), 2021, 36 (6): 5465-5475.

DOI: 10.16182/j.issn1004731x.joss.21-1145