智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

基于人工势场的防疫机器人改进近端策略优化算法

伍锡如, 沈可扬

引用本文:
伍锡如, 沈可扬. 基于人工势场的防疫机器人改进近端策略优化算法[J]. 智能系统学报, 2025, 20(3): 689-698.
WU Xiru, SHEN Keyang. Improved proximal policy optimization algorithm for epidemic prevention robots based on artificial potential fields[J]. *CAAI Transactions on Intelligent Systems*, 2025, 20(3): 689-698.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202407026

DOI: 10.11992/tis.202407026
网络出版地址: https://link.cnki.net/urlid/23.1538.TP.20250425.1808.006

基于人工势场的防疫机器人改进近端策略优化算法

伍锡如, 沈可扬
(桂林电子科技大学 电子工程与自动化学院, 广西 桂林 541004)

摘　要: 针对防疫机器人在复杂医疗环境中的路径规划与避障效果差、学习效率低的问题, 提出一种基于人工势场的改进近端策略优化 (proximal policy optimization, PPO) 路径规划算法。根据人工势场法 (artificial potential field, APF) 构建障碍物和目标节点的势场, 定义防疫机器人的动作空间与安全运动范围, 解决防疫机器人运作中避障效率低的问题。为解决传统 PPO 算法的奖励稀疏问题, 将人工势场因子引入 PPO 算法的奖励函数, 提升算法运行中的奖励反馈效率。改进 PPO 算法网络模型, 增加隐藏层和 Previous Actor 网络, 提高了防疫机器人的灵活性与学习感知能力。最后, 在静态和动态仿真环境中对算法进行对比实验, 结果表明本算法能更快到达奖励峰值, 减少冗余路径, 有效完成避障和路径规划决策。

关键词: PPO 算法; 人工势场; 路径规划; 防疫机器人; 深度强化学习; 动态环境; 安全性; 奖励函数
中图分类号: TP183; TP391.41　　**文献标志码:** A　　**文章编号:** 1673−4785(2025)03−0689−10

中文引用格式: 伍锡如, 沈可扬. 基于人工势场的防疫机器人改进近端策略优化算法 [J]. 智能系统学报, 2025, 20(3): 689–698.
英文引用格式: WU Xiru, SHEN Keyang. Improved proximal policy optimization algorithm for epidemic prevention robots based on artificial potential fields[J]. CAAI transactions on intelligent systems, 2025, 20(3): 689–698.

Improved proximal policy optimization algorithm for epidemic prevention robots based on artificial potential fields

WU Xiru, SHEN Keyang
(College of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin 541004, China)

Abstract: This paper presents an improved proximal policy optimization (PPO) path planning algorithm based on artificial potential fields (APFs) to address poor path planning, obstacle avoidance effectiveness, and low learning efficiency of epidemic prevention robots in complex medical environments. The potential fields of obstacles and target nodes are constructed using the APF method, defining the action space and safe motion range for epidemic prevention robots to resolve the low obstacle avoidance efficiency during operation. To tackle the sparse reward issue in traditional PPO algorithms, APF factors are incorporated into the reward function of the PPO algorithm to enhance the feedback efficiency of reward mechanisms during algorithm execution. The network model of the PPO algorithm is improved by adding hidden layers and a previous actor network, thereby enhancing the flexibility and learning perception capabilities of epidemic prevention robots. Finally, comparative experiments conducted in static and dynamic simulation environments demonstrate that the proposed algorithm achieves faster attainment of reward peaks, reduces redundant path segments, and effectively completes obstacle avoidance and path planning decisions.

Keywords: PPO algorithm; artificial potential field; path planning; epidemic prevention robot; deep reinforcement learning; dynamic environment; safety; reward function

近年来, COVID-19 病毒、埃博拉病毒等疫情的爆发使全球防疫形势愈发严峻, 疫情的发展受到人们的普遍关注。面对日益复杂的医疗环境, 人们需要多样化、智能化的防疫设备。防疫机器人能在极大程度上减少人们与病原体的直接接触[1], 并能有效地防止医患间交叉感染, 其在防疫工作中的地位越发凸显[2]。在实际应用中, 防疫机器人常在多障碍物环境下进行防疫工作[3], 如何快速规划出一条到达目标点的最短安全路径, 对防

收稿日期: 2024−07−24.　**网络出版日期:** 2025−04−27.
基金项目: 国家自然科学基金项目(62263005); 广西自然科学基金重点项目(2020GXNSFDA238029); 广西高校人工智能与信息处理重点实验室开放基金重点项目(2022GXZDSY004); 桂林电子科技大学研究生教育创新计划项目(2024YCXS119, 2024YCXS131).
通信作者: 伍锡如. E-mail: xiruwu520@163.com.

疫机器人应用的发展具有重要的研究意义。

在传统的路径规划算法中，人工势场法 (artifical potential field, APF) 结构简单[4-5]，实时性较好[6]。但传统 APF 算法在复杂的障碍物场景中容易停滞在局部最小值[7]，甚至丢失导航目标[8]。为提高 APF 算法的工作效率，学者们尝试了多种解决办法。Yang 等[9] 在传统 A*算法的代价函数中引入人工势场，确保智能体与障碍物保持安全距离。Yu 等[10] 根据具体障碍物设置人工势场，在平滑路径的同时减少算法的无效分支和规划时间。上述算法在处理静态环境时表现良好，但在存在动态障碍物的复杂医疗环境中依然存在搜索缓慢、导航停滞等问题。

近端策略优化算法 (proximal policy optimization, PPO) 在处理连续控制问题时展现出了卓越的性能[11-12]。当应用于连续的路径规划任务时，该算法能有效发挥算法的学习能力[13-14]。对于防疫机器人路径规划任务，算法可处理具有时序性的数据[15]，通过内部神经网络解决连续状态动作空间问题[16]。PPO 算法已经广泛应用于路径规划领域[17-19]：Guo 等[20] 引入优先级经验重播方法提高算法的学习效率，并集成深度强化学习算法实现更复杂的规划任务；Huang 等[21] 通过嵌入网络分别提取感知特征和状态特征以增强网络记忆能力，从而减少观测时的碰撞；Guan 等[22] 在 PPO 算法的损失函数中加入广义优势估计，使 PPO 算法中的基线能自我调整。但上述算法普遍存在稀疏奖励问题[23]，复杂环境对奖励的干扰直接影响防疫机器人的路径规划能力[24]，而以往对算法的研究主要集中于对目标节点的搜寻，对奖励函数的构造缺少针对性的改进。

本文从防疫机器人的工作需求出发，针对传统的 APF 和 PPO 算法进行改进和优化，提出了一种基于 APF 和 PPO 的改进路径规划与避障算法。利用人工势场对 PPO 算法的运动状态空间和奖励函数进行了改进，解决传统 PPO 算法的奖励稀疏问题，提升算法寻路效率，同时提高防疫机器人避障能力。改进 PPO 算法的网络结构，提升防疫机器人的灵活性和学习感知能力。最后，通过仿真实验和实物实验验证了改进算法的有效性。

1　基于人工势场的改进 PPO 算法

1.1　防疫机器人动力学模型

本文根据防疫机器人工作时的运行情况，设计防疫机器人动力学模型，并将状态数据输入算法网络并优化策略。设计算法的动作空间如图 1 所示，其中包含防疫机器人的运动方向和运动速度。结合分方向线速度v_1、v_2、v_3、v_4得到防疫机器人线速度v_r，运动角速度为ω_r。g为防疫机器人与终点的方位信息。动作空间 A 定义为

$$A=(\omega_r,v_r)$$

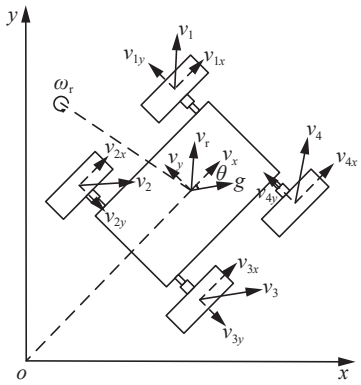


图 1　防疫机器人运动模型

Fig. 1　Epidemic prevention robots motion model

状态空间包含机器人的状态数据，机器人利用状态空间信息做出决策并评估动作的长期收益，有效提高算法的收敛性。防疫机器人运动状态空间 S 定义为

$$S=[x_r,y_r,A_r,R_r,\theta_r,D_g,\theta_g]$$

式中：(x_r,y_r) 为防疫机器人在二维平面上的位置，A_r、R_r、θ_r分别表示为防疫机器人所受引力势场、斥力势场和运动方向，D_g代表防疫机器人与目标节点的距离数据，θ_g代表防疫机器人运动方向与目标节点的夹角。

图 2 为描述防疫机器人安全工作的区域的模型图，表现为圆形，以防疫机器人中心点为圆心，半径为 d，说明了防疫机器人的最大探测距离。通过构建防疫机器人的安全运行模型，能在实验中定义防疫机器人的避障情况。

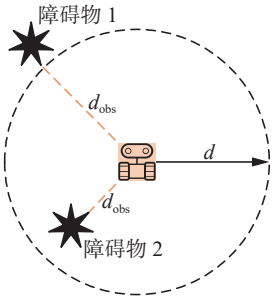


图 2　防疫机器人安全运行范围

Fig. 2　Epidemic prevention robots safe operating range

针对防疫机器人在复杂医疗环境中的实时避障要求，机器人从起始点出发沿初始最优路径前进；当防疫机器人检测到障碍物信息时，根据障碍势场做出决策，改变运动方向进行避障动作；当障碍物消失后，防疫机器人恢复最优路径行进至目标节点。本文改进算法的路径规划系统结构如图 3 所示。s_t 表示状态序列，a_t 表示动作信息，r_t 表示奖励信息。

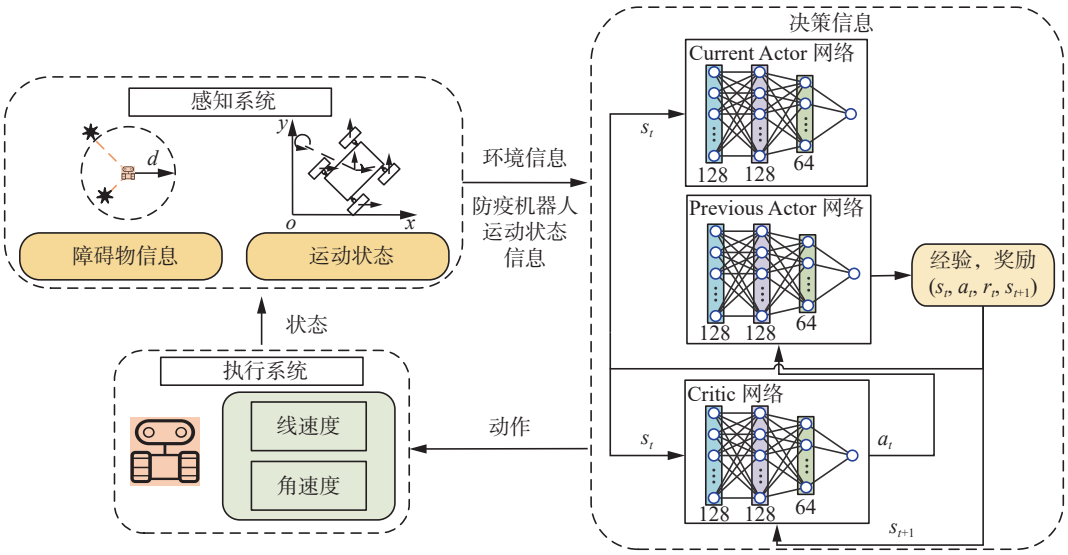


图 3　路径规划系统结构

Fig. 3　Path planning system architecture

1.2　APF 算法

避障能力对防疫机器人能否安全通过障碍区域起到关键作用。构建人工势场，在障碍物区域构建斥力场，在目标位置构建引力场[25]，防疫机器人根据所受合力规划路径。人工势场U由两种力场组成：目标位置形成的引力场U_{att}，吸引机器人向目标节点移动；障碍物位置形成的斥力场U_{rep}，排斥机器人使其避开障碍物[26]。组合公式为

$$U(q)=U_{\text{att}}(q)+U_{\text{rep}}(q) \tag{1}$$

引力场函数：

$$U_{\text{att}}(q)=\begin{cases}\dfrac{1}{2}\zeta d^2(q,q_{\text{goal}}), & d(q,q_{\text{goal}})\leqslant d^*_{\text{goal}}\\ d^*_{\text{goal}}\zeta d(q,q_{\text{goal}})-\dfrac{1}{2}\zeta(d^*_{\text{goal}})^2, & d(q,q_{\text{goal}})>d^*_{\text{goal}}\end{cases}$$

式中：d为防疫机器人与目标节点的欧氏距离。为防止不同节点的引力场互相干扰，设置吸引距离d^*_{goal}。

斥力场函数：

$$U_{\text{rep}}(q)=\begin{cases}\dfrac{1}{2}\eta\left(\dfrac{1}{D(q)}-\dfrac{1}{Q^*}\right)^2, & D(q)\leqslant Q^*\\ 0, & D(q)>Q^*\end{cases}$$

式中：Q^*是障碍物势场的作用范围，在该范围内障碍物的斥力才会对防疫机器人产生影响，超出此范围则不产生斥力影响；$D(q)$代表物体和障碍物之间的距离矢量。本文利用人工势场解决防疫机器人避障问题，通过添加斥力与引力因子改进 PPO 算法奖励函数，解决 PPO 算法的奖励稀疏问题。

1.3　PPO 算法

PPO 算法通过调整策略参数来最大化平均奖励，其通过与环境交互生成轨迹数据[27-28]，每个轨迹包含一系列状态、动作和奖励。PPO 算法通过近端比率裁剪损失 Clip 限制策略更新幅度[29-30]，从而保持训练稳定性。

比率裁剪损失$L_{\text{CLIP}}(\theta)$定义为

$$L_{\text{CLIP}}(\theta)=\hat{E}_t[\min(r_t(\theta)\hat{A}_t,\text{clip}(r_t(\theta),1-\varepsilon,1+\varepsilon)\hat{A}_t)]$$

如图 4 所示，优势函数$\hat{A}_t>0$时算法认为该动作较平均动作更优，$\hat{A}_t<0$时算法认为该动作较平均动作更差。裁剪函数为$\text{clip}(r_t(\theta),1-\varepsilon,1+\varepsilon)$。将概率比率$r_t(\theta)$限制在区间$(1-\varepsilon,1+\varepsilon)$，防止策略变化过大。

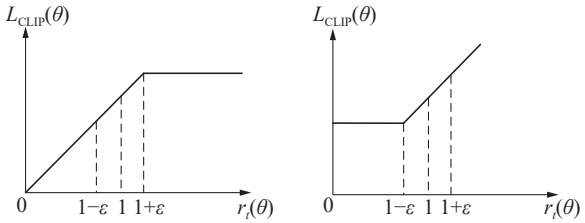


图 4　目标函数限定范围

Fig. 4　Path planning system architecture diagram

PPO 算法的总损失函数定义为

$$L(\theta)=L_{\text{CLIP}}(\theta)-c_1L_{vf}(\theta)+c_2S[\pi_\theta]$$

式中：c_1和c_2为超参数，用于平衡不同损失函数的权重；$L_{vf}(\theta)$为值函数优化；S为策略熵。计算出总损失函数L后反向传递数据，更新算法的策略和价值函数直至达到停止条件。本文结合人工势场构建奖励函数，改进 PPO 算法的网络结构，获取奖励优化算法策略，增强 PPO 算法的学习能力与反馈频率。

1.4　APF-PPO 算法设计

PPO 算法对采集的具有时间特性的数据进行训练，能较好地处理连续控制问题，但存在奖励稀疏、收敛性较差等问题。APF 算法能根据设定的势场受到斥力作用和引力作用的影响，生成避

障路径，但存在易陷入局部最优解的问题。本文融合 PPO 算法的策略学习能力与 APF 算法的路径规划优势，构造 APF-PPO 算法。APF-PPO 融合改进算法的训练流程如图 5 所示。

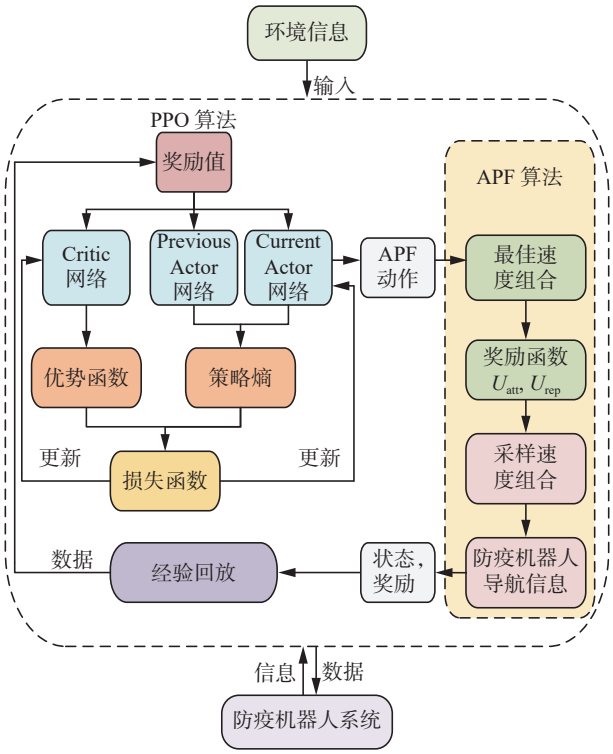


图 5　APF-PPO 算法训练流程

Fig. 5　APF-PPO algorithm training process

1.4.1　APF-PPO 算法奖励函数设计

使用传统 PPO 算法进行路径规划时，奖励函数局限于终点奖励、避障奖励和下一步移动奖励，在复杂医疗环境中会导致智能体在训练过程中很难获得有意义的反馈，延缓学习进度和收敛速度。本文基于防疫机器人的路径规划与避障的需求，结合 APF 算法中势场的分布特性，设计改进的奖励函数，能有效提高算法反馈效率。该函数结合防疫机器人的当前状态进行避障决策，同时确保算法收敛。

防疫机器人在每个时间步中获得的奖励 R 分为 3 部分：正常动作奖励 R_{n}、终点奖励 R_{goal}、碰撞奖励 R_{obs}。

$$R=\begin{cases}R_{\mathrm{n}}, & \text{正常动作}\\ R_{\mathrm{goal}}, & \text{终点}\\ R_{\mathrm{obs}}, & \text{碰撞}\end{cases}$$

正常动作奖励 R_{n} 定义为防疫机器人避开障碍物但未达到终点节点时，移动一个时间步获得的奖励。为了解决传统 PPO 算法中的奖励稀疏问题，根据人工势场对奖励函数进行了改进。改进后的正常动作奖励函数 R_{n} 定义为

$$R_{\mathrm{n}}=\alpha R_{\mathrm{rep}}+\beta R_{\mathrm{att}} \tag{2}$$

式中：R_{rep} 为斥力项奖励，R_{att} 为引力项奖励，α 和 β 为权重系数。

斥力项奖励：

$$R_{\mathrm{rep}}=\begin{cases}\dfrac{1}{2(d-d_{\mathrm{obs}})^2}\eta\left(\dfrac{Q^*-D(q)}{D(q)Q^*}\right)^2, & D(q)\leqslant Q^*\\ 0, & D(q)>Q^*\end{cases}$$

式中：d 为防疫机器人障碍物探测半径；d_{obs} 为障碍物与防疫机器人的距离，障碍物距离防疫机器人越近，奖励价值越小。

引力项奖励：

$$R_{\mathrm{att}}=U_{\mathrm{att}}V_{\mathrm{r}}\cos(\theta-\theta_{\mathrm{goal}})$$

式中：U_{att} 为引力势场量，V_{r} 为防疫机器人当前速度值，θ_{goal} 为终点目标方向角。此奖励值与防疫机器人运动方向和终点方位的角度差成反比，角度差越大奖励价值越低。

目标吸引奖励 R_{goal} 定义为基于防疫机器人当前位置与终点之间的距离给出的奖励。防疫机器人与终点的距离 $d_{\mathrm{goal}}\leqslant\dfrac{d}{2}$ 时，终点奖励值为 100；否则，目标吸引奖励值为 0。R_{goal} 定义为

$$R_{\mathrm{goal}}=\begin{cases}100, & d_{\mathrm{goal}}\leqslant\dfrac{d}{2}\\ 0, & d_{\mathrm{goal}}>\dfrac{d}{2}\end{cases} \tag{3}$$

避障奖励 R_{obs} 定义为防疫机器人在当前位置与最近静态或动态障碍物的距离所给予的奖励。当防疫机器人与最近静态或动态障碍物的距离 $R_{\mathrm{obs}}\leqslant\dfrac{d}{2}$ 时，终点奖励值为−100；否则，终点奖励值为 0。R_{obs} 定义为

$$R_{\mathrm{obs}}=\begin{cases}-100, & d_{\mathrm{obs}}\leqslant\dfrac{d}{2}\\ 0, & d_{\mathrm{obs}}>\dfrac{d}{2}\end{cases} \tag{4}$$

1.4.2　改进 PPO 网络模型设计

传统 Critic 网络的价值估计与网络的表达能力不足。如图 6 所示，将 Critic 网络改进为 3 个隐藏层处理价值估计，3 个隐藏层分别包含 128、128 和 64 个神经元，并采用 ReLU 激活函数。Critic 网络输入为防疫机器人状态信息，输出对应 Actor 网络选择动作的评分，保持 Actor 网络选择的动作在长期回报上最优。

为计算 PPO 算法的重要性采样比率，需要存储优化开始前的策略网络参数。同时，为减少更新过程中的方差和训练过程中出现的过拟合现象，在网络系统中引入三隐藏层 Previous Actor 网络，输出动作分布的参数，增强网络的表达能力以及对复杂环境和高维状态信息的拟合能力，使防疫机器人能更好地感知环境状态。改进后的神

经网络能够生成更优化的动作策略，提升了防疫机器人在复杂环境中的决策效率。

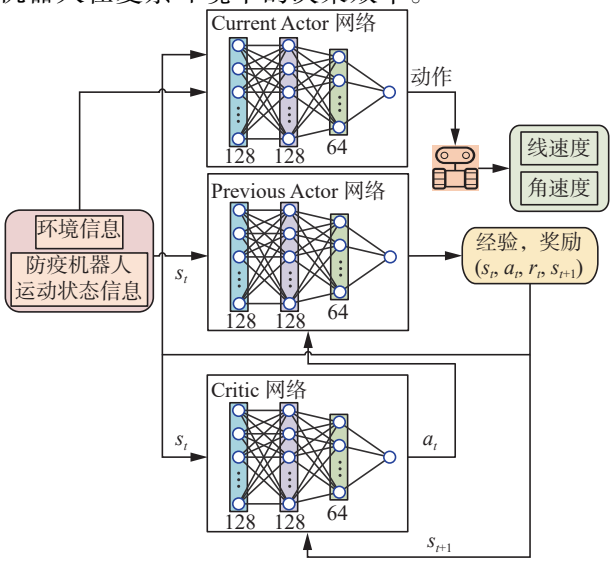


图 6　改进 PPO 算法神经网络结构

Fig. 6　Improvement of PPO algorithm neural network structure

2　仿真实验结果分析

2.1　实验模型设计

使用 PyTorch 框架构建算法模型和仿真环境，计算机配置：Intel 酷睿 i7-11900K，NVIDIA GTX3060，32 GB 内存，512 GB SSD 存储。

本文实验设计方法：首先，为模拟不同复杂医疗环境，实验构建静态障碍物以及包含动态障碍物的场景；其次，防疫机器人在构建路径规划环境中进行训练，确保其能完成寻路任务并有效避开障碍物。改进 PPO 算法的网络参数如表 1 所示。

表 1　算法网络参数设置

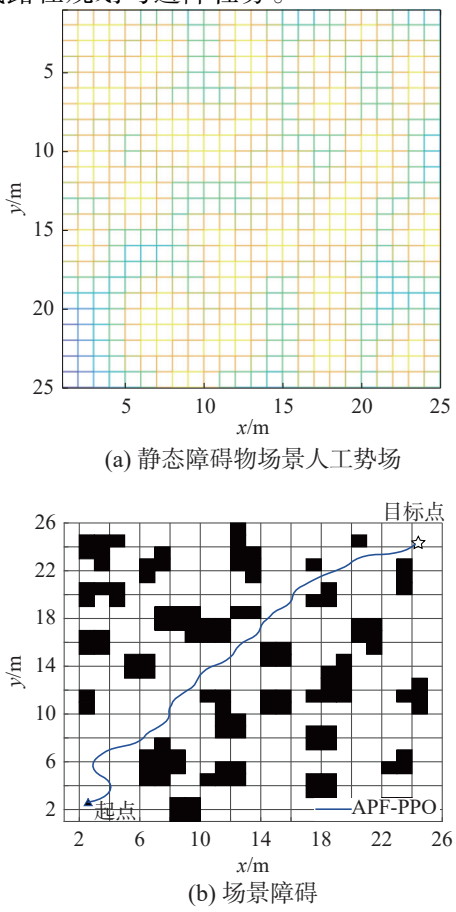
Table 1　Algorithm network parameter settings

超参数	数值
训练回合N	11 000
Actor网络学习率l_{r_Actor}	3×10^{-4}
Critic网络学习率l_{r_Critic}	1×10^{-4}
衰减因子γ	0.95
Clip参数ε	0.2

2.2　实验结果

2.2.1　静态场景

为了测试 PPO、深度 Q 网络 (deep Q-network, DQN)、深度确定性策略梯度 (deep deterministic policy gradient, DDPG)、柔性演员−评论家 (soft actor-critic, SAC) 算法与本文算法在静态障碍物场景中的路径规划与避障效果，首先构建如图 7 所示的静态仿真环境。场景人工势场设置如图 7(a) 所示，在障碍物周围设置斥力场，在终点设置引力场。在静态障碍物仿真实验中，障碍物设置如图 7(b)。实验结果表明本文提出的 APF-PPO 算法能完成路径规划与避障任务。



(a) 静态障碍物场景人工势场

(b) 场景障碍

图 7　静态障碍物场景人工势场与场景障碍

Fig. 7　Artificial potential field and scene obstacle maps for static obstacle scenes

如图 8 所示，实验结果表明，训练后的 5 种算法模型可以有效绕过障碍区域，在不碰到障碍物的同时规划出一条较短的路线。各算法的路径均有较大波动，其中 DQN、DDPG、PPO、SAC 算法均在接近终点的路径规划中产生了冗余路径。

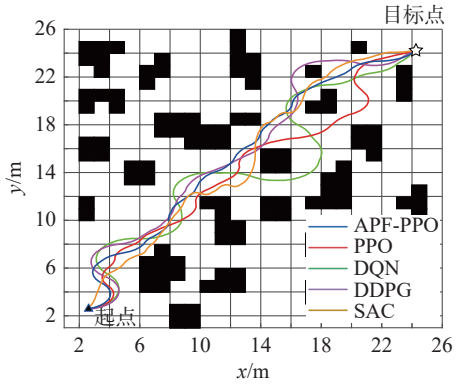


图 8　静态障碍物场景算法路径对比

Fig. 8　Comparison of paths for static obstacle scenarios

如表 2 所示，除 APF-PPO 算法外，4 种算法规划路径长度增大至 55 m 以上。实验结果表明 APF-PPO 算法能减少冗余路径生成，算法相较

于其他算法生成路径长度更短，减少冗余路径的能力更强，对最优路径的生成能力有较大提升。

表 2　算法数据对比

Table 2　Algorithm data comparison　　m

算法	路径长度
APF-PPO	48.683
PPO	64.371
DQN	72.542
DDPG	55.194
SAC	60.417

如图 9 所示，PPO 和 DQN 算法的奖励收敛速度相对较快，在 1 100 次左右时达到奖励峰值，但随后发生较剧烈波动。DDPG 和 SAC 算法奖励曲线较平稳，在后期数值低于 APF-PPO 算法。APF-PPO 算法借助改进的网络结构，相对于其他算法波动性较小且防疫机器人能更快到达目标节点，表明算法泛化能力较强，且收敛性较其他算法更好。

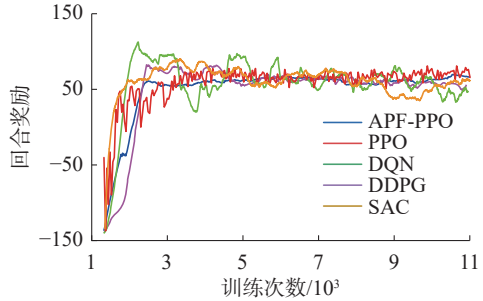
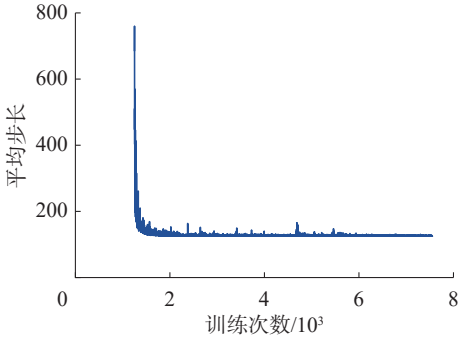


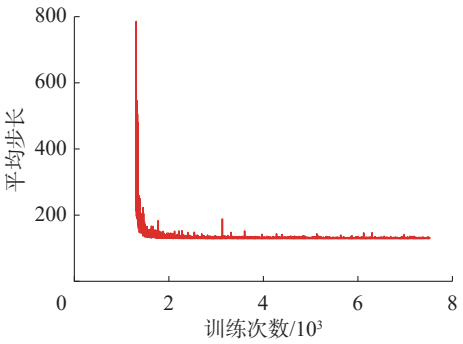
图 9　静态障碍物场景算法奖励

Fig. 9　Algorithmic rewards for static obstacle scenes

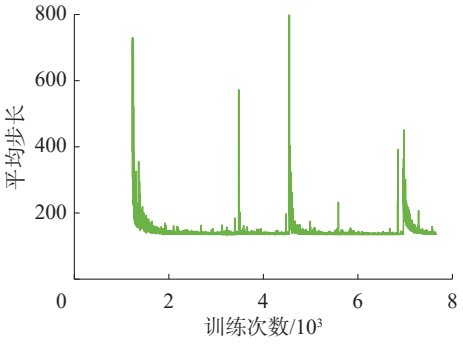
平均步数代表在每一个完整的环境交互周期各个次数中，防疫机器人需要执行多少步操作来完成任务。在训练过程中，平均步数随着时间的推移逐渐减少，表明防疫机器人对路径规划任务的执行效率不断提高。如图 10 所示，5 种算法在训练中的平均步长都下降迅速但均有波动：DDPG 和 DQN 算法所消耗时间步数量总体下降，但在 5 000 ~ 6 000 次的平均步长有跳变；PPO 和 SAC 算法在实验初期跳变较大。APF-PPO 算法的平均步长数据总体趋于平稳，大部分情况在 60 步上下波动，表明 APF-PPO 算法的收敛状态更好。



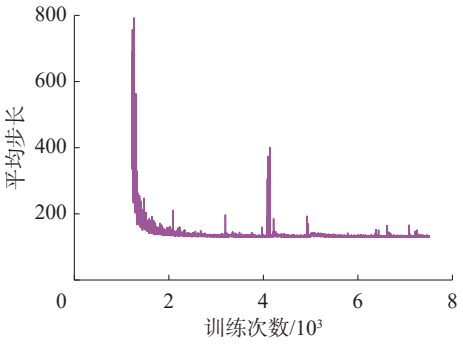
(a) APF-PPO



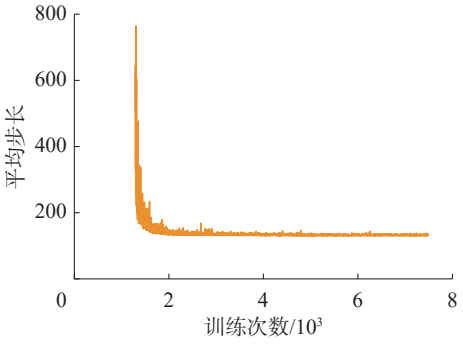
(b) PPO



(c) DQN



(d) DDPG



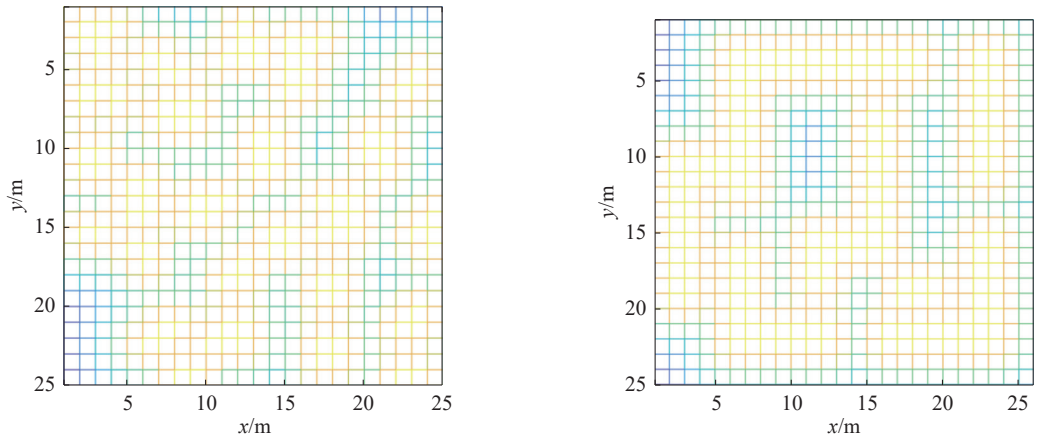
(e) SAC

图 10　静态障碍物场景平均步长

Fig. 10　Average step length for static obstacle scenarios

2.2.2　混合场景

为测试 5 种算法在动态障碍物与静态障碍物混合场景中的路径规划与避障效果，构建如图 11 所示的混合障碍物环境。人工势场如图 11(a)、(b) 所示。如图 11(c)、(d) 所示，实验加入两个动态障碍物，分别从第 5 s 和第 80 s 开始运动。实验结果表明，提出的 APF-PPO 算法能完成路径规划与避障。



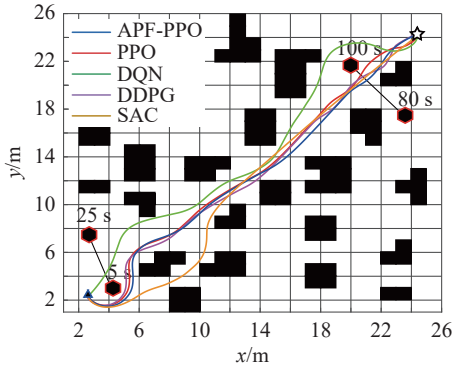
(a) 混合场景 1 人工势场　(b) 混合场景 2 人工势场

(c) 混合场景 1 场景障碍　(d) 混合场景 2 场景障碍

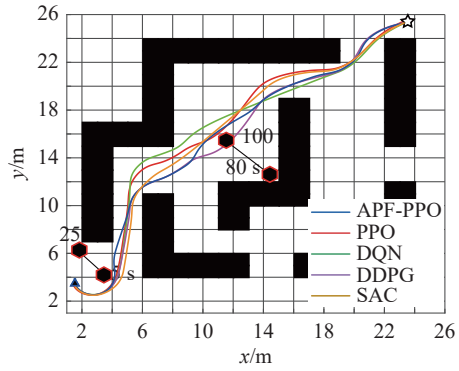
图 11　混合障碍物场景人工势场与障碍

Fig. 11　Mixed obstacle scene artificial potential field and obstacles

图 12 为包含动态障碍物场景的 5 种算法的结果图，所有算法均到达目标节点。根据图 12 和表 3 数据，在场景 1 中，SAC 算法在避障中选择了较多冗余的路段，DQN 算法到达第 2 个动态障碍物时选择绕行冗余路段，增加了防疫机器人的运行路程。在场景 1 和场景 2 中，PPO、SAC 和 DDPG 算法路径与障碍物的距离较近，在实际应用中易发生碰撞问题。APF-PPO 算法路径相较于静态和动态障碍物更远，具备更高的安全性和工作效率。



(a) 混合障碍物场景 1



(b) 混合障碍物场景 2

图 12　混合障碍物场景算法路径对比

Fig. 12　Comparison of algorithmic paths for mixed obstacle scenarios

图 13 和图 14 分别为混合场景 1 中算法奖励和平均步长。DQN 和 DDPG 算法在运行过程中累计奖励增长较快，约在 1 000 次训练时能得到正奖励，但较难收敛，在 8 000 次训练后仍有震荡。SAC 算法奖励增长较快，在中期发生较剧烈跳变。APF-PPO 算法相较于其他 3 种算法波动较小，表明 APF-PPO 算法具有更好的模型鲁棒性。如图 14 所示，在算法的学习过程中，PPO、DDPG、

DQN 和 SAC 算法需要更多时间步来完成路径规划任务，在 6000 ~ 8000 次训练中仍保持较高的平均步长，表明 APF-PPO 算法的经验学习能力更强。

表 3　混合障碍物场景算法结果对比

Table 3　Comparison of algorithm results for mixed obstacle scenarios　　m

算法	路径长度1	路径长度2
APF-PPO	33.099	38.327
PPO	34.127	42.645
DQN	51.646	40.691
DDPG	34.428	39.822
SAC	37.251	39.412

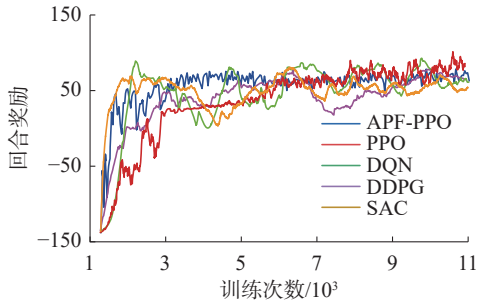
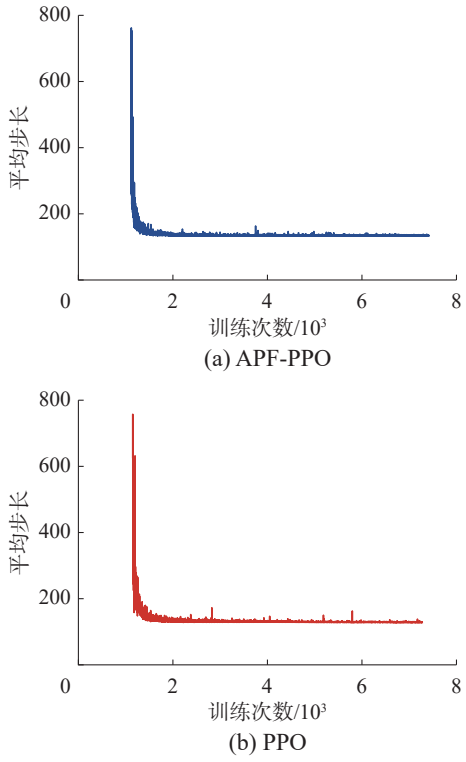


图 13　混合障碍物场景算法奖励

Fig. 13　Mixed obstacle scene algorithm bonus



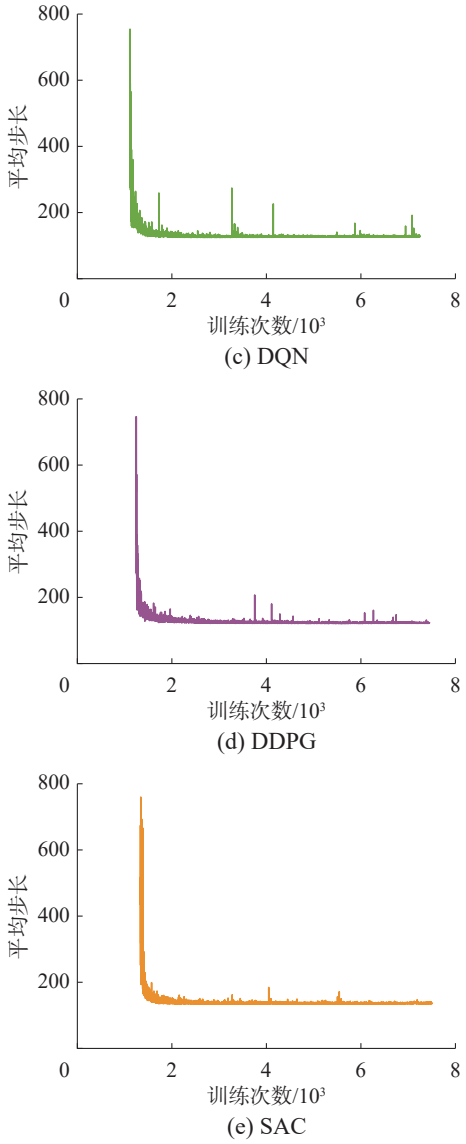


图 14　混合障碍物场景平均步长

Fig. 14　Average step length for mixed obstacle scenarios

3　实物实验验证

实验使用如图 15 所示的一个防疫机器人和一个移动机器人及若干静态障碍物作为测试对象。其中移动机器人模拟动态障碍物，防疫机器人需避开障碍物并完成路径规划。实验场景为 2.0 m×3.0 m 的矩形环境。防疫机器人车身尺寸为 0.5 m×0.6 m，移动机器人车身尺寸为 0.35 m×0.4 m。两个机器人由顶部摄像头观察环境信息并与主机通信。防疫机器人从左下角出发，规划前往右上角叉型终点的路径；移动机器人保持直线向下运动。

实验结果如图 16 所示，在考虑实物机器人具有碰撞体积的情况下，防疫机器人能根据本文算法实时规避障碍物，同时完成路径规划任务要求，规划的路径情况较优。这表明 APF-PPO 算法对机器人路径规划具有实用价值。

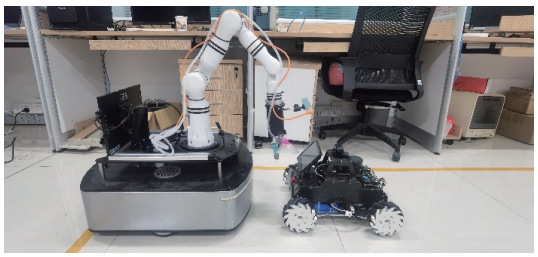


图 15　实验设备
Fig. 15　Experiment apparatus

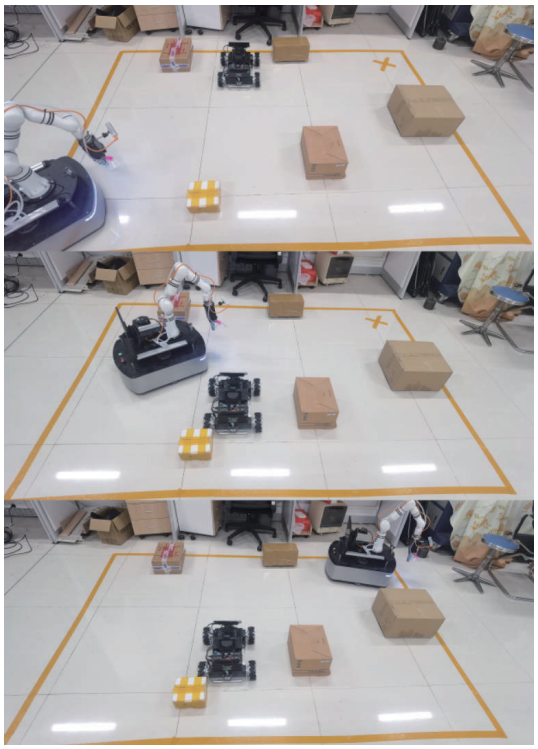


图 16　算法实验情况
Fig. 16　Experimental situation of algorithm

4　结束语

为提升复杂医疗环境下防疫机器人路径规划及避障效率, 本文提出了一种基于人工势场的改进 PPO 算法。该算法结合人工势场改进 PPO 算法的运动状态空间与奖励函数, 在新的奖励函数中添加引力场与斥力场影响因子, 弥补传统 PPO 算法奖励稀疏的缺陷, 在避障的同时规划出最优路径。改进传统 PPO 算法的网络结构, 添加隐藏层和 Actor 网络, 增强了网络的拟合能力与价值估计能力, 优化了防疫机器人对复杂医疗环境的感知能力。通过对比实验, 证明了本文算法的有效性。在后续研究中, 将结合机器视觉技术进一步探索多任务环境下的防疫机器人路径规划与避障。

参考文献:

[1] 张帆, 谭跃刚. 生成式预训练模型机器人及其潜力与挑战[J]. 中国机械工程, 2024, 35(7): 1241−1252.
ZHANG Fan, TAN Yuegang. Generative pre-trained model robot: potential and challenges[J]. China mechanical engineering, 2024, 35(7): 1241−1252.

[2] 鞠庆, 刘飞飞, 李光昌, 等. 室内环境自主消毒防疫机器人系统设计[J]. 传感器与微系统, 2023, 42(12): 103−106.
JU Qing, LIU Feifei, LI Guangchang, et al. Design of autonomous disinfection and prevention robot system for indoor environment[J]. Sensors and microsystems, 2023, 42(12): 103−106.

[3] JULIAN R E, BERNARDINO C T, STEFANO D G, et al. An algorithm for dynamic obstacle avoidance applied to UAVs[J]. Robotics and autonomous systems, 2025, 186: 104907.

[4] 黄郑, 谢彧颖, 张欣, 等. 基于运动预测与改进 APF 的无人机路径规划方法[J]. 电子测量技术, 2023, 46(24): 103−111.
HUANG Zhen, XIE Yuying, ZHANG Xin, et al. UAV path planning method based on motion prediction and improved APF[J]. Electronic measurement technology, 2023, 46(24): 103−111.

[5] 鲜斌, 宋宁. 基于模型预测控制与改进人工势场法的多无人机路径规划[J]. 控制与决策, 2024, 39(7): 2133−2141.
XIAN Bin, SONG Ning. Path planning of multi-UAV based on model predictive control and improved artificial potential field method[J]. Control and decision, 2024, 39(7): 2133−2141.

[6] 邓冬冬, 许建民, 孟寒, 等. 基于蚁群算法与人工势场法融合的移动机器人路径规划[J]. 仪器仪表学报, 2025, 3(1): 1−16.
DENG Dongdong, XU Jianmin, MENG Han, et al. Path planning of mobile robot based on fusion of ant colony algorithm and artificial potential field method[J]. Chinese journal of scientific instrument, 2025, 3(1): 1−16.

[7] 孙传禹, 张雷, 辛山, 等. 结合 APF 和改进 DDQN 的动态环境机器人路径规划方法[J]. 小型微型计算机系统, 2023, 44(9): 1940−1946.
SUN Chuanyu, ZHANG Lei, XIN shan, et al. Combining APF and improved DDQN for robot path planning in dynamic environments[J]. Journal of Chinese computer systems, 2023, 44(9): 1940−1946.

[8] 张扬, 彭鹏菲, 曹杰. 基于改进 APF 算法的水面无人艇局部路径规划[J]. 兵器装备工程学报, 2023, 44(9): 42−48.
ZHANG Yang, PENG Pengfei, CAO Jie. Local path planning for surface unmanned craft based on improved APF algorithm[J]. Journal of ordnance equipment engineering, 2023, 44(9): 42−48.

[9] YANG Chaopeng, PAN Jiacai, WEI Kai, et al. A novel unmanned surface vehicle path-planning algorithm based on A* and artificial potential field in ocean currents[J]. Journal of marine science and engineering, 2024, 12(2): 285−310.

[10] YU Jiabin, WU Jiguang, XU Jiping, et al. A novel planning and tracking approach for mobile robotic arm in

obstacle environment[J]. Machines, 2023, 12(1): 19−35.
[11] 朱少凯, 孟庆浩, 金晟, 等. 基于深度强化学习的室内视觉局部路径规划[J]. 智能系统学报, 2022, 17(5): 908−918.
ZHU Shaokai, MENG Qinghao, JIN Sheng, et al. Indoor visual local path planning based on deep reinforcement learning[J]. CAAI transactions on intelligent systems, 2022, 17(5): 908−918.
[12] 赵玉新, 杜登辉, 成小会, 等. 基于强化学习的海洋移动观测网络观测路径规划方法[J]. 智能系统学报, 2022, 17(1): 192−200.
ZHAO Yuxin, DU Denghui, CHENG Xiaohui, et al. Reinforcement learning based observation path planning method for marine mobile observation networks[J]. CAAI transactions on intelligent systems, 2022, 17(1): 192−200.
[13] SUN Aijing, SUN Chi, DU Jianbo, et al. Optimizing energy efficiency in UAV-Assisted wireless sensor networks with reinforcement learning PPO2 algorithm[J]. IEEE sensors journal, 2023, 23(23): 29705−29721.
[14] CAI Peide, WANG Heng, HUANG Huaiyang, et al. Vision-based autonomous car racing using deep imitative reinforcement learning[J]. IEEE robot, 2021, 6(4): 7262−7269.
[15] GU Zhixin, JIA Keyi, XU Kaihong. Three-dimensional path planning method of agent based on fluid disturbance algorithm and PPO[J]. IAENG international journal of computer science, 2025, 52(2): 365−373.
[16] ZHU Zeyu, ZHAO Huijing. A survey of deep RL and IL for autonomous driving policy learning[J]. IEEE transactions on intelligent transportation systems, 2022, 23(9): 14043−14065.
[17] 沈骁, 赵彤洲. 基于 DDQN 的无人机区域覆盖路径规划策略[J]. 电子测量技术, 2023, 46(14): 30−36.
SHEN Xiao, ZHAO Tongzhou. DDQN-based path planning strategy for UAV area coverage[J]. Electronic measurement technology, 2023, 46(14): 30−36.
[18] XING Bowen, WANG Xiao, LIU Zhenchong. The wide-area coverage path planning strategy for deep-sea mining vehicle cluster based on deep reinforcement learning[J]. Journal of marine science and engineering, 2024, 12(2): 316−332.
[19] GUAN Yang, REN Yangang, SUN Qi, et al. Centralized cooperation for connected and automated vehicles at intersections by proximal policy optimization[J]. 2020, IEEE transactions on vehicular technology, 69(11), 12597–12608.
[20] GUO Hongda, XU Youchun, MA Yulin, et al. Pursuit path planning for multiple unmanned ground vehicles based on deep reinforcement learning[J]. Electronics, 2023, 12(23): 4759−4778.
[21] HUANG Xiangxiang, WANG Wei, JI Zhaokang, et al. Representation enhancement-based proximal policy optimization for UAV path planning and obstacle avoidance[J]. International journal of aerospace engineering, 2023, 2023: 1−15.
[22] GUAN Wei, CUI Zhewen, ZHANG Xianku. Intelligent smart marine autonomous surface ship decision system based on improved PPO algorithm[J]. Sensors, 2022, 22(15): 5732−5765.
[23] LIU Jinyuan FU Minglei, LIU Andong, et al. A homotopy invariant based on convex dissection topology and a distance optimal path planning algorithm[J]. IEEE robotics and automation letters, 2023, 8(11): 7695−7702.
[24] 邓修朋, 崔建明, 李敏, 等. 深度强化学习在机器人路径规划中的应用[J]. 电子测量技术, 2023, 46(6): 1−8.
DENG Xiuming, CUI Jianming, LI Min, et al. Deep reinforcement learning in robot path planning[J]. Electronic measurement technology, 2023, 46(6): 1−8.
[25] WU Haixiao, ZHANG Yong, HUANG Linxiong, et al. Research on vehicle obstacle avoidance path planning based on APF-PSO[J]. Proceedings of the institution of mechanical engineers, 2023, 237(6): 1391−1405.
[26] YAN Xun, JIANG Dapeng, MIAO Runlong, et al. Formation control and obstacle avoidance algorithm of a multi-USV system based on virtual structure and artificial potential field[J]. Journal of marine science and engineering, 2021, 9(2): 1−17.
[27] XU Haotian, YAN Zheng, XUAN Junyu, et al. Improving proximal policy optimization with alpha divergence [J]. Neuro computing, 2023, 534(C): 94−105.
[28] QIN Yunhui, ZHANG Zhongshan, LI Xulong, et al. Deep reinforcement learning based resource allocation and trajectory planning in integrated sensing and communications UAV network[J]. IEEE transactions on wireless communications, 2023, 22(11): 8158−8169.
[29] AN Haonan, WANG Lin. Robust topology generation of internet of things based on PPO algorithm using discrete action space[J]. IEEE transactions on industrial informatics., 2023, 20(4): 5406−5414.
[30] XU Yahao, WEI Yiran, WANG Di, et al. Multi-UAV path planning in GPS and communication denial environment[J]. Sensors (Basel, Switzerland), 2023, 23(6): 2997−3012.

作者简介：

伍锡如，教授，博士生导师，主要研究方向为深度学习、复杂网络、路径规划。主持国家自然科学基金项目 1 项、广西壮族自治区自然科学基金项目 1 项、广西高校人工智能与信息处理重点实验室开放基金重点项目 1 项。获发明专利授权 6 项，发表学术论文 50 余篇，出版专著 1 部。E-mail：xiruwu520@163.com。

沈可扬，硕士研究生，主要研究方向为路径规划。E-mail：1341391239@qq.com。