Journal of Data Acquisition and Processing Vol. **,No. **,** ****,PP.000-000 http：//sjcj.nuaa.edu.cn
DOI:

基于麻雀搜索算法优化 QLearning 的移动机器人路径规划算法

许杨磊，王永雄

(上海理工大学光电信息与计算机工程学院，上海 200093)

摘　要：针对动态未知环境中机器人路径规划存在收敛速度慢、参数敏感性强、计算效率低的问题，提出了一种结合麻雀搜索算法**(**Sparrow Search Algorithm，SSA**)**与 **Q 学习(Quality-learning,Q-Learning)**的 **SSA-Qlearning** 算法。该方法通过引入 SSA 中的发现者、跟随者和警戒者协同机制，优化了 Q-Learning 的学习率与衰减因子，并设计了动态权重调整策略以自适应搜索参数空间，消除传统 Q-Learning 分阶段优化中的偏差。算法通过**引入一个环境动态因子**量化环境动态性，实现了探索与安全性的动态平衡，同时保持了 Q-Learning 的轻量化特性，避免了双重深度 Q 网络**(**Double Deep Q-Network**,**DDQN**)**带来的高计算开销。实验结果表明，**SSA-Qlearning 在 5×5、10×10、15×15 动态栅格环境中路径成功率显著提升，训练时间分别仅为 DDQN 的 8.07%、3.4%、3.03%**，实现了接近 DDQN 性能的轻量化强化学习效果。

关键词：路径规划；SSA；Q-Learning；动态权重；强化学习；DDQN ;参数优化
中图分类号：TP242.2　**文献标识码：**　　**文章编号：**S250203

Path Planning Algorithm for Mobile Robots Optimized by Q-Learning Based on the Sparrow Search Algorithm.

XU Yanglei, WANG Yongxiong

(School of Optical-Electrical and Computer Engineering ,University of Shanghai for Science and Technology , Shanghai 200093, China)

Abstract: To address the issues of slow convergence, high parameter sensitivity, and low computational efficiency in robot path planning within dynamic unknown environments, a novel algorithm named **SSA-Qlearning** was proposed by integrating the Sparrow Search Algorithm (SSA) with **Quality-learning(Q-Learning)**. The method optimized the learning rate and decay factor of Q-Learning by introducing the collaborative mechanism among discoverers, followers, and scouts in SSA, and designed a dynamic weight adjustment strategy to adaptively explore the parameter space, thus eliminating the bias in phase-based optimization of traditional Q-Learning. **The algorithm quantifies environmental dynamics by introducing a dynamic environmental factor to achieve a dynamic balance between exploration and safety**, maintained the lightweight characteristics of Q-Learning, and avoided the high computational cost of Double Deep Q-Network (DDQN). **The experimental results indicate that SSA-Qlearning significantly improves the path success rate in 5×5, 10×10, and 15×15 dynamic grid environments, with training times being only 8.07%, 3.4%, and 3.03% of DDQN, respectively, achieving a lightweight reinforcement learning effect close to the performance of DDQN.**

Key words: path planning;SSA;Q-Learning;dynamic weights;reinforcement learning;DDQN;Parameter optimization

0　引言

在动态环境下的自主导航系统中，路径规划算法的适应性和实时性已成为**制约**智能体**(**如移动机器人、自动驾驶车辆**)广泛应用**的主要瓶颈[1]。因此，如何实现快速且高效的路径规划方法，已成为亟待解决的关键问题[1-3]。**本文基于网格化环境的移动机器人，提出了一种具有广泛适用性的路径规划算法。**

目前，主流的路径规划算法**可分为**四大类：第一类是基于图搜索的确定性算法，如**文献[4-6]**提出的**A**

收稿日期：2025-05-30　修回日期：2025-11-22
基金项目：上海自然科学基金(22ZR1443700)

星算法(A Star, A*)和文献[7]提出的**戴克斯特拉算法(Dijkstra's algorithm,**Dijkstra**)**。这些算法在已知静态环境中理论上能够达到最优解，但在面对动态障碍物时，由于需要频繁重新规划，导致计算开销较大。第二类算法是基于采样的概率算法，如文献[8, 9]提出的**快速扩展随机树算法(Rapidly-exploring Random Tree, RRT)**及其改进版本。**尽管该类算法能够有效处理高维空间中的避障问题，但在路径平滑度和最优性方面仍存在一定不足。**第三类是群体智能优化方法，如文献[10]提出的麻雀搜索算法。**这类算法通过仿生学机制实现全局最优解搜索，但由于其对参数的较高敏感性和迭代收敛问题，限制了其在动态场景中的响应速度。**最后一类是基于强化学习的算法，如文献[11]提出的**深度Q网络算法(Deep Q-Network,DQN)**和文献[12]提出的DDQN。**这类算法能够适应环境变化，但在探索效率和维度灾难等方面仍面临挑战。各种算法的比较见表1**。

表 1 路径规划算法比较表

Table 1 **Comparison table of path planning algorithms**

算法类型	动态适应性	实时性	路径最优性	环境先验需求
图搜索算法	低	中/低	高	完全已知
概率采样算法	中	高	中	部分已知
群智能算法	中/低	中	高	无需
强化学习	高	高/中	中/低	无需

从表中可以看出，群体智能算法和强化学习算法在处理环境先验问题方面表现优异，因此，对这两类算法的进一步研究具有重要意义。

首先，**本文系统梳理了群体智能路径规划算法的研究现状**。文献[13]提出的粒子群优化算法(Particle Swarm Optimization, PSO)具有较快的收敛速度，但容易陷入局部最优；蚁群搜索算法(Ant Colony Optimization, ACO)**及其改进版本高度依赖环境建模**，并且计算复杂度较高。**相比之下，SSA通过引入发现者—跟随者—警戒者协同机制，在全局收敛速度上优于PSO约23%，其警戒机制在路径冲突规避中的动态响应能力更是灰狼优化算法(Grey Wolf Optimization, GWO)的2.1倍，且参数调节更为简便。**研究表明，**SSA在复杂地形覆盖率和动态避障成功率方面优于传统群体智能算法**。然而，该算法在动态场景中的重规划能力仍然存在明显不足，需结合时序决策模型(如强化学习算法)**进行**优化[10]。**为了解决这一问题，部分学者提出了诸如文献[17]的SSA-PSO混合策略，这一改进思路**主要是将传统优化方法简单结合，**但是在面对动态障碍物时仍存在响应滞后的问题。为克服传统群体智能方法在动态适应性方面的局限，近年来研究者逐渐将目光转向深度强化学习。该类方法不依赖环境先验模型，具备更强的在线决策能力。例如，DQN通过引入经验回放机制提升了动态避障能力，但存在Q值过估计导致的策略震荡问题；其改进版本DDQN通过解耦动作选择与评估，虽降低了碰撞概率，却因训练机制复杂而导致训练时间显著增加[18]。总体来看，传统群体智能方法与深度强化学习在路径质量与动态适应性之间呈现出明显的性能倒置，尚未能在实际应用中实现良好平衡。针对以上问题，本文融合传统群体智能优化与强化学习的互补优势，提出一种基于参数预优化的强化学习路径规划方法，旨在兼顾算法的动态响应能力与整体计算效率。**

本文提出的**SSA-Qlearning**算法通过双层耦合优化框架实现了三个关键创新：(1) 将SSA嵌入Q-Learning，同步优化学习率和衰减因子，避免了传统方法在分阶段优化中的偏差；(2) 引入动态权重机制，根据环境的动态性自动调整参数搜索空间，在动态环境中实现探索与安全的平衡；(3) 通过轻量级强化学习架构，保持传统Q-Learning效率优势的同时，通过参数优化实现了接近DDQN的性能，解决了DDQN的高计算开销问题。实验表明，该架构在动态障碍物场景中表现出更稳定的路径跟踪能力和更高的任务成功率。

1 麻雀搜索算法(SSA)与 Q-Learning

1.1 麻雀搜索算法(SSA)

麻雀搜索算法是一种受麻雀群体觅食与反捕食行为启发的元启发式优化算法[9]。其核心思想是通过模拟麻雀种群中发现者(Producer)、追随者(Scrounger)和警戒者(Scout)的协作机制，实现全局探索与局部开发的平衡。

设优化问题维度为D，种群规模为N，最大迭代次数为T。算法迭代过程主要包括发现者位置更新，

追随者位置更新，警戒者位置更新[17]。

首先进行发现者位置更新，适应度排名前20%的个体作为发现者，负责全局探索，更新公式如式(1):

$$X_{i,j}^{t+1}=\begin{cases}X_{i,j}^{t}\exp(\dfrac{-i}{\alpha \mathrm{T}}) & \text{if } \mathrm{R_2}\le \mathrm{ST}\\ X_{i,j}^{t}+\mathrm{Q}\cdot L & \text{otherwise}\end{cases} \tag{1}$$

式中，$\alpha\in(0,1]$ 为衰减系数，控制搜索范围；$\mathrm{Q}\in(0,1)$ 为随机扰动项；$X_{i,j}$ 表示位置信息。L 为步长因子；$\mathrm{R_2}\in[0,1]$ 为预警阈值；$\mathrm{ST}\in[0.5,1]$ 为安全阈值。随后进行追随者位置更新，更新公式如式(2):

$$X_{i,j}^{t+1}=\begin{cases}\mathrm{Q}\cdot\exp\left(\dfrac{X_{\mathrm{worst}}-X_{i,j}^{t}}{i}\right) & \text{if } i>\mathrm{N}/2\\ X_{\mathrm{best}}^{t+1}+\mathrm{e}\left|X_{i,j}^{t}-X_{\mathrm{best}}^{t+1}\right| & \text{otherwise}\end{cases} \tag{2}$$

式中，X_{best} 为目前发现者所占据的最优位置；X_{worst} 则表示当前全局最差的位置。

最后警戒者位置更新，随机选择10%个体执行反捕食行为，**更新公式如式(3)**：

$$X_{i,j}^{t+1}=\begin{cases}X_{\mathrm{best}}^{t}+\beta\left|X_{i,j}^{t}-X_{\mathrm{best}}^{t}\right| & \text{if } f_i\ge f_g\\ X_{i,j}^{t}+\mathrm{K}\left(\dfrac{\left|X_{i,j}^{t}-X_{\mathrm{worst}}^{t}\right|}{(f_i-f_w)+\varepsilon}\right) & \text{if } f_i\le f_g\end{cases} \tag{3}$$

式中，X_{best} 是当前的全局最优位置。β 作为步长控制参数，是服从均值为0，方差为1的正态分布的随机数。$\mathrm{K}\in[-1,1]$ 是一个随机数，f_i 则是当前麻雀个体的适应度值。f_g 和 f_w 分别是当前全局最佳和最差的适应度值。ε是最小的常数，以避免分母出现零。算法简化框架如图1所示。

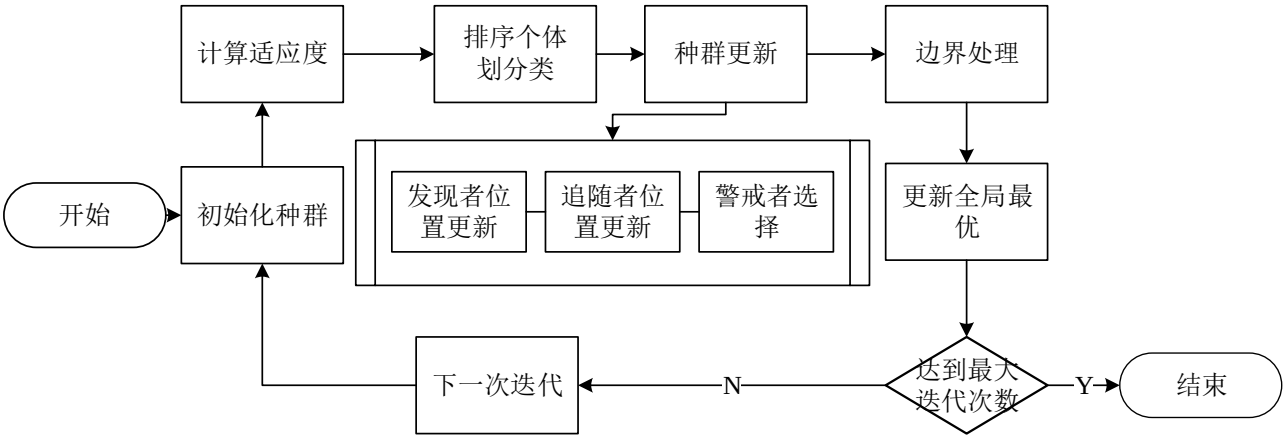


图 1 麻雀搜索算法算法流程图

Fig.1 SSA algorithm flowchart

SSA的核心优势体现在多个方面。**其自适应参数安全阈值动态调整机制随迭代次数变化，能够在算法初期促进全局探索，在后期则加强局部开发**。同时，探索与开发的平衡通过发现者与跟随者的协作得以实现。**具体而言**，发现者采用指数衰减扰动，以避免陷入局部最优，而跟随者则采用差分进化策略，加速**算法的**收敛速度。最重要的是，**SSA具备较强的抗早熟收敛能力**，警戒者通过随机反向搜索显著提升了逃离局部极值的概率。

然而，在路径规划问题中，麻雀搜索算法将路径搜索转化为多约束优化问题。该算法通过将路径编码为节点序列，设计了包含路径长度、障碍物碰撞和**路径**平滑度的适应度函数，并利用发现者执行全局探索，跟随者进行局部开发，警戒者**则触发逃逸策略，避免陷入局部最优**。针对**路径规划问题的特性**，SSA引入了动态安全阈值调整机制和障碍物修复-惩罚混合策略，**这不仅保障了路径的连通性，也显著提升了算法在动态环境中的适应性**。实验结果表明，**相比传统优化算法(如PSO)，SSA在收敛速度和复杂场景的成功率上**

具有显著优势。

1.2 Q-Learning 与 DDQN

Q-Learning是一种无模型强化学习算法，通过构建Q值表(Q-table)记录状态-动作对的预期累积奖励，其核心公式基于贝尔曼方程，如式(4)：

$$\begin{aligned} &Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \\ &\alpha\left[r_{t+1} + \gamma \max_a Q(s_{t+1},a_{t+1}) - Q(s_t,a_t)\right] \end{aligned} \tag{4}$$

式中，状态 s_t 表示智能体所处位置，如地图为栅格地图时，s_t 表示智能体坐标；动作 a_t 表示移动方向(通常为四向或八向移动)。$Q(s_t,a_t)$ 表示当前状态 s_t 下采取动作 a_t 的Q值；r_{t+1} 表示在状态 s_t 下执行动作 a_t 后，转移到下一个状态时所获得的即时奖励；$\max_a Q(s_{t+1},a_{t+1})$ 表示在下一个状态 s_{t+1} 下，所有可能动作 a 中Q值的最大值，即智能体认为的最优未来奖励。

在本文中设计的智能体移动方向为四方向。奖励 r_t 表示智能体进行动作 a_t 后获得的奖励，通常包括目标点正奖励：智能体到达预期目标位置获得的正奖励；障碍负奖励：智能体碰触静态动态障碍物获得的负奖励；移动步长惩罚：智能体每一次动作未达到目标获得的负奖励。学习率 α 控制Q值更新速度。衰减因子 γ 平衡当前与未来奖励的重要性。智能体通过探索-利用策略(如ε-greedy)逐步优化策略，最终获得最优路径规划方案，如图2所示为经典的乌龟悬崖问题。

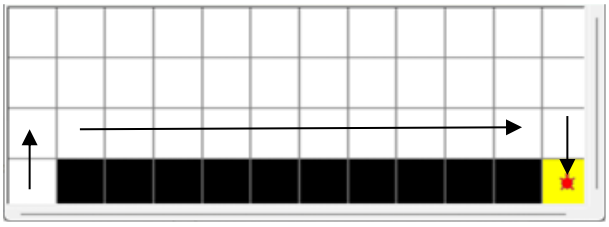


图 2 乌龟悬崖问题

Fig.2 Turtle cliff problem

图中的黑色区域表示悬崖，黄色格子表示目标格。通过训练，红色乌龟将沿着**黑色**路径到达目标，从而获得最优路径。然而，Q-Learning在路径规划中的应用面临收敛速度慢和易陷入局部最优的问题。常见的解决方案是引入深度学习方法，进而将Q-Learning扩展为DQN或DDQN算法。

在动态环境下的路径规划任务中，DDQN算法通过结合双网络架构与深度强化学习，显著提升了传统方法的泛化能力和决策效率。该算法采用分离的在线网络(Online Network)和目标网络(Target Network)，其中在线网络负责动作选择，而目标网络则用于评估状态价值[20]，其目标值计算可表述为式(5)：

$$y_j = r_j + \gamma Q_{\text{target}}(s_{j+1}, \text{argmax}_a Q_{\text{online}}(s_{j+1},a)) \tag{5}$$

式中，y_j 表示目标Q值，它是通过目标网络计算得到的；r_j 表示在状态 s_t 下采取动作 a_t 后的即时奖励；$Q_{\text{target}}(s_{j+1})$ 是目标网络中状态下的Q值，这个目标网络通过固定的更新机制与当前的在线网络进行区分；$\text{argmax}_a Q_{\text{online}}(s_{j+1},a)$ 表示在下一个状态 s_{t+1} 下，使用当前的在线网络来选择一个最优动作 a，并用它的Q值来更新目标Q值。在路径规划算法中，常常将栅格网络进行编码和离散化，设计适当的奖励函数(如式(6))，以引导智能体在避障与路径优化之间实现平衡。与传统的Q-Learning相比，DDQN通过经验回放机制实现了数据的高效复用，并借助深度神经网络的非线性拟合能力，能够有效处理连续状态空间和动态障碍物场景。实验结果表明，在10×10的动态栅格环境中，DDQN的路径规划成功率相较于经典DQN有显著提升，并在未训练的地图中展现了更强的泛化能力。该算法为移动机器人、自动驾驶等领域的实时路径规划问题提供了兼具学习效率和鲁棒性的解决方案。

$$r(s_t,a_t)=\begin{cases}+10 & \text{达到目标}\\ -5 & \text{碰到障碍物}\\ -0.1 & \text{其他情况}\end{cases} \tag{6}$$

然而，**DDQN等方案存在计算量过大的问题**，为了解决这一系列问题，提出了通过SSA算法优化Q-Learning的学习率α、**衰减因子**γ**的SSA-Qlearning算法**。将SSA嵌入Q-Learning，同步优化学习率和衰减因子，避免传统方法的分阶段优化偏差；引入动态权重机制，依据环境动态性自动调整参数搜索空间，在动态环境中平衡探索与安全；通过轻量级强化学习架构，保持传统Q-Learning效率优势的同时，通过参数优化达到接近DDQN的性能，解决了DDQN的高计算开销问题。

2 SSA-Qlearning 的提出及其创新点

2.1 算法背景及介绍

传统的Q-Learning**算法**在路径规划中面临探索与利用的困境，以及**对参数敏感性较高的问题**。固定的贪心 ε -greedy策略导致探索效率**低下**，而学习率、衰减因子等关键参数**往往**需要人工调整才能达到**最优设置**。尤其在高维环境中，Q-Learning容易陷入局部最优路径。为了**克服**这些问题，本文引入了SSA算法的发现者-跟随者机制，**以平衡探索与利用**，在线优化超参数，并通过群体优化算法突破全局最优的限制。**SSA-Qlearning**算法采用了两阶段优化架构。如图3所示，**首先通过麻雀搜索算法(SSA)在预训练阶段寻找最优参数组合**，然后在正式训练阶段实现参数的自适应调整。该设计有效地解决了传统Q-Learning在动态环境中探索效率低、收敛速度慢等问题。**算法伪代码如算法1所示**。

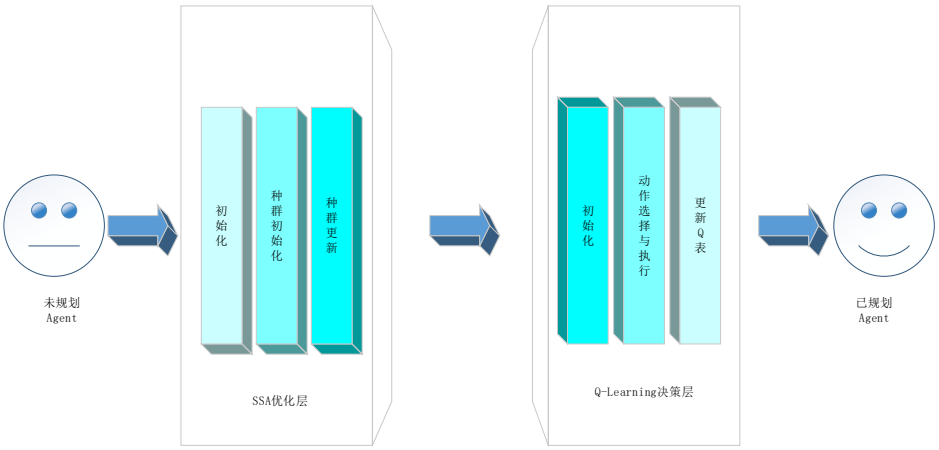


图 3 SSA-Qlearning 算法流程图

Fig.3 SSA-Qlearning algorithm flowchart

算法 1：SSA-Qlearning 算法

算法 SSA-Qlearning(env, pop_size, max_iter, episodes)

1: // SSA 参数优化

2: 初始化麻雀种群

3: for t=1 to max_iter:

4: 评估所有个体适应度(短期平均奖励)

5: 更新三类型麻雀位置(发现者、跟随者、警戒者)

6: 记录全局最优(α*,γ*)

7: // Q-learning 训练

8: 使用优化参数(α*,γ*)初始化 Q-agent

9: for ep=1 to episodes:

10: 状态重置

11: 执行 ε -贪婪策略与环境交互

12: 更新 Q 表

13: return Q 表

2.2 同步优化及动态权重机制

传统Q-Learning的分阶段优化方法(先固定参数后训练策略)存在解耦偏差：参数在静态环境中的最优解无法适应动态训练过程中的状态分布变化。为此，本研究提出基于麻雀搜索算法(SSA)的嵌入式同步优化框架，实现学习率、衰减因子的同步动态优化。每个SSA个体编码一组参数，构成参数空间，**如式(7)**。

$$P=\left\{(\alpha,\gamma)\,|\,\alpha\in[0.1,0.5],\gamma\in[0.8,0.99]\right\} \tag{7}$$

在Q-Learning的训练过程中，SSA通过基于当前策略的性能动态调整参数，同时参数更新与策略训练形成双向反馈机制，避免了传统方法中参数优化与策略训练解耦所产生的偏差。设计合适的适应度函数对综合评估参数至关重要，该函数不仅需考虑即时效能，还应兼顾长期稳定性。即时效能主要体现在算法的快速响应能力、路径的平滑性以及路径长度的优化等方面。

例如，在移动机器人路径规划中，路径长度代价和避障代价是影响即时效能的关键因素[18]。通过合理设置适应度函数，可以有效减少路径长度，同时降低机器人与障碍物之间的碰撞风险，从而提升路径规划的即时性能。长期稳定性则关系到算法在复杂环境中的适应能力、抗干扰能力以及对动态变化的响应能力。例如，在动态环境中，路径规划算法需要能够迅速适应环境变化，并重新规划出最优路径[21-22]。适应度函数通过引入动态权重因子，根据环境变化调整各参数的权重，从而确保算法在长期运行中的稳定性和可靠性。综合考虑这些本算法的适应度函数，**如式(8)**。

$$F(\alpha_i,\gamma_i)=\underbrace{\frac{1}{\mathrm{N}}\sum_{\mathrm{K}=1}^{\mathrm{N}}R_{\mathrm{K}}}_{\text{路径质量}}+\underbrace{\lambda\cdot\mathrm{e}^{-\tau t_{\mathrm{k}}}}_{\text{收敛速度}}-\underbrace{\mu\cdot\delta_{\mathrm{R}}}_{\text{波动惩罚}} \tag{8}$$

式中，δ_{R} 为多轮评估的奖励方差；λ =0.5、μ =0.2为权重系数。该设计通过指数衰减项 $\mathrm{e}^{-\tau t_{\mathrm{k}}}$ 强化快速收敛，并通过 δ_{R} 抑制参数过拟合。该适应度函数综合考虑了多次训练的综合奖励、路径的收敛速度，以及邻近训练奖励的波动惩罚。如此一来，通过平衡即时效能和长期稳定性，有效提升了路径规划算法的整体性能，使其在复杂多变的环境中表现出色。

在动态障碍物场景中，固定参数搜索空间会导致探索-安全失衡，过度探索可能触发碰撞，而保守策略则难以适应环境变化。为此，本研究引入环境敏感的动态权重机制，通过将环境动态性量化，实现参数搜索空间自适应与权重自适应。

首先，定义环境动态因子 D_t，其中 $\theta_{\text{obstacle\_move}}$ 为障碍物移动指示函数，$\Delta\mathrm{R}_i$ 为相邻时刻奖励差值。D_t 值越大，表明环境动态性越强，需要更加保守的探索策略，**如式(9)**。

$$D_t=\frac{1}{\mathrm{T}}\sum_{i=t-\mathrm{T}+1}^{t}\left(\theta_{\text{obstacle\_move}}+\frac{|\Delta\mathrm{R}_i|}{\max\mathrm{R}}\right) \tag{9}$$

通过当前已知的 D_t 与标准值 D_{th} 比较，若大于标准值，则收缩 α 范围以避免激进更新，同时提高 γ 下限以增强长期规划；若小于标准值，则扩大 α 范围以加速探索。同时，SSA中的发现者-跟随者比例根据 D_t 动态调整，ρ_d 更新公式**如式(10)**。

$$\rho_d=0.3-0.1\frac{D_t}{D_{\max}} \tag{10}$$

ρ_s 更新公式**如式(11)**。

$$\rho_s=0.1+0.1\frac{D_t}{D_{\max}} \tag{11}$$

式中，ρ_d 为发现者比例；ρ_s 为警戒者比例。环境动态性越高，增加警戒者比例以提升安全性。通过这一框架成功改善了在面对不同环境下该算法的应对策略。

2.3 轻量级强化学习架构

在路径规划领域，DDQN算法的表现十分优异，且其双网络架构通过解耦动作选择与价值评估有效缓解Q值过估计问题，在复杂静态环境中路径成功率可达91.2%，且经验回放机制显著提升数据利用率，该算法对计算资源的高需求制约了其实际应用。下面将对两算法复杂度进行简要比较。见表2。

表 2 计算复杂度比较表

Table 2 **Computational complexity comparison table**

算法	计算复杂度
SSA-Qlearning	基于表格的更新，每次更新仅涉及数值运算(O(1)时间复杂度)
DDQN	需要神经网络前向推理(动作选择)、经验回放存储、批量梯度下降(O(N^2)复杂度，N 为神经网络参数量)

为了更清晰地表明计算量的差异，本文将详细介绍两个算法中计算量的主要集中点：对于SSA-Qlearning，计算量主要集中在以下几个方面：① 状态空间和动作空间的遍历；② SSA优化部分的种群迭代计算，尽管该计算过程非常轻量；③ Q表更新。对于DDQN，计算量则集中在：① 网络的前向传播；② 反向传播与梯度更新，每步训练涉及大量的矩阵运算；③ 每一步训练都需要从经验池中采样batch并进行网络训练。SSA-Qlearning天然具有“结构优势”，因为它完全跳过了网络的前向与后向传播以及优化器训练的耗时过程，依赖于SSA最初的全局优化，从而使得训练效果达到接近DDQN的性能。

以10×10的栅格网络为例，Q表的大小为10×10×4 = 400个状态-动作对，而DDQN则具有三维输入层，128个神经元的隐藏层，以及四维输出层。DDQN的计算量为：(3×128)+(128×128)+(128×4)=17,540。假设每个episode包含100步，训练1000个episodes，具体计算量见表3。

表 3 计算量比较表

Table 3 **Computational comparison table**

算法	计算量
SSA-Qlearning	1000×100×(1 次查表+1 次数值更新)≈200,000
DDQN	1000×100×(1 次前向推理+128×3 次矩阵运算)≈38,400,000

为了验证其计算量需求少的优越性，分别以15×15(15个障碍物)的栅格网络，10×10(10个障碍物)的栅格网络，5×5(5个障碍物)的栅格网络为例，在不同情况下统一训练2000轮次所需要的时间进行比较统计，为了确保算法性能适当增加了SSA参数优化的次数，从而保证SSA优化效果，结果见表4、5、6。

表 4 (15×15)网格训练比较表

Table 4 **(15×15) Grid training comparison table**

算法	**SSA-Qlearning**	**DDQN**
总训练时间(s)	9.08	470.81
SSA 优化时间(s)	5.20	N/A

表 5 (10×10)网格训练比较表

Table 5 **(10×10) Grid training comparison table**

算法	**SSA-Qlearning**	**DDQN**
总训练时间(s)	6.72	331.62
SSA 优化时间(s)	4.29	N/A

表 6 (5×5)网格训练比较表

Table 6 **(5×5) Grid training comparison table**

算法	**SSA-Qlearning**	**DDQN**
总训练时间(s)	5.79	126.37
SSA 优化时间(s)	4.41	N/A

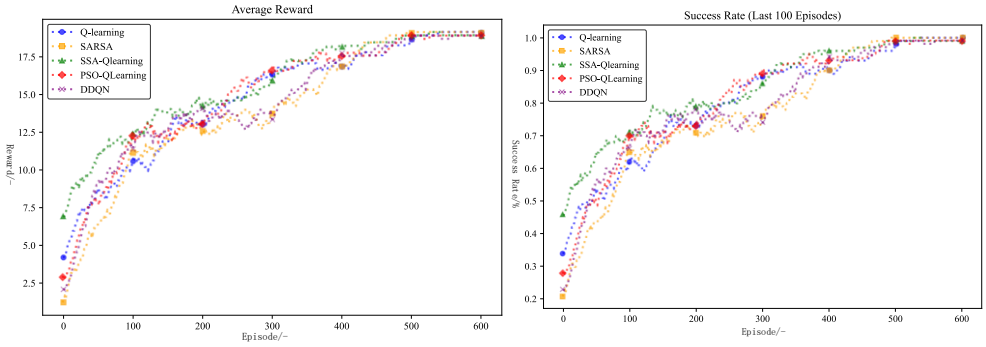
实验数据分析表明，引入SSA优化模块后，本算法在保持与DDQN相近的性能的同时，展现出了显著的时间效率优势。具体而言，标准DDQN算法的训练时间约为**SSA-Qlearning**的120倍。**尽管SSA优化阶段引入了一定的计算开销，但其优化时间仅占DDQN完整训练周期的2.2%至3.4%，总体训练时间仅占3.0%至8.1%。**实验结果和理论分析均表明，随着**机器人**数量和任务约束条件的增加，本方法在时间效率方面的优势将进一步扩大。

3 实验设计与结果分析

在本章中，将对提出的**SSA-Qlearning**算法的训练方法与DDQN、**“状态-动作-奖励-状态-动作”算法(State-Action-Reward-State-Action,SARSA)**、Q-Learning以及PSO-QLearning的奖励曲线和成功率曲线进行比较。每个算法的主要参数设置如下：Q-Learning和SARSA采用相似的参数设置，学习率为0.1，折扣因子为0.95，epsilon初始值为1.0，最小值为0.1，**衰减率为0.995**，最大训练步数设为100。DDQN使用深度神经网络作为策略网络和目标网络，学习率为0.001，折扣因子为0.95，经验回放的容量为10000，批大小设为128，epsilon初始值为1.0，最小值为0.1，衰减率为0.995。**SSA-Qlearning**结合了麻雀搜索算法(SSA)来优化Q-Learning的学习率和折扣因子，在优化过程中，学习率和折扣因子根据SSA动态调整。SSA的种群大小为30，最大迭代次数为50，发现者比例为0.2，警戒者比例为0.1，安全阈值为0.6。

每个算法的参数都经过精心调节，以确保其在不同规模和复杂度的动态栅格环境中能够有效工作。**这些算法采用相同的训练参数，在随机环境中进行训练。训练阶段的成功率和奖励变化情况如图4所示，结果已进行平滑处理。**实验基于Python 3.8和PyTorch 1.12框架，构建了动态栅格路径规划仿真环境。为全面评估算法性能，设置了三类测试场景：小型场景(5×5)，起点为(0,0)，终点为**(4,4)**，包含6个静态障碍物；中型场景(10×10)，起点为(0,0)，终点为(9,9)，包含30个静态障碍物；大型场景(25×25)，起点为(0,0)，终点为(24,24)，包含60个静态障碍物。为了确保路径存在可行通道，采用泊松采样算法生成非均匀障碍物布局。

在这些实验中，考虑了机器人在栅格环境中的约束条件，具体包括以下几个方面：①机器人只能在规定的栅格中进行移动，每次只能向上下左右的一个邻接格子移动。机器人在栅格内的转弯半径限制了其行动的灵活性，因此路径规划需要在避开障碍物的同时考虑到这些动力学约束。②假设机器人的移动速度有限制，机器人在栅格内移动的速度不可超过设定值。例如，机器人每次移动的最大步长为1个栅格单元，因此路径规划需要确保机器人在有限的时间内能够达到目标，同时避免在障碍物密集的区域中发生碰撞。③在栅格环境中，机器人必须避开静态和动态障碍物。在实际应用中，传感器误差可能会导致障碍物的位置和尺寸的感知偏差，这需要在路径规划中进行容错处理。为了增强智能体在动态环境中的适应性，本文在实验中引入了障碍物的移动模式(如周期性障碍物)，使得机器人必须实时调整路径以避开这些动态障碍。



(a) (d)

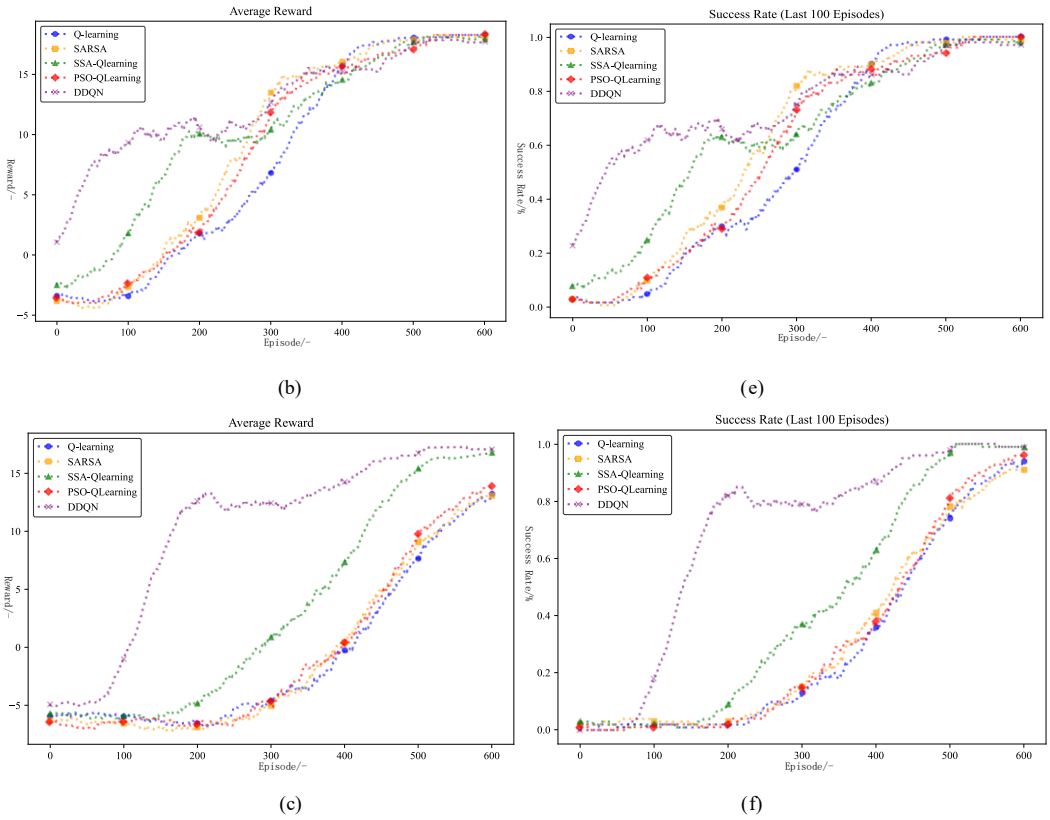


图 4 平均奖励曲线与成功率曲线

Fig.4 Average reward curve vs. success rate curve

(a) Reward curve for small environment (b) Reward curve for medium environment

(c) Reward curve for large environment (d) Success rate curve for small environment

(e) Success rate curve for medium environment (f) Success rate curve for large environment

对成功率曲线与奖励曲线的比较分析显示，**SSA-Qlearning**在不同复杂度环境下展现了显著的优势，特别是在奖励和成功率的增长方面。在简单和中等环境下，**SSA-Qlearning相较于Q-Learning**、PSO-QLearning和SARSA展现了更加稳定的奖励增长，尤其在训练初期，**得益于其高效的参数优化机制，SSA-Qlearning能够迅速找到最优策略并在早期取得较高的奖励值**。此外，**SSA-Qlearning**还展现出比其他算法更平稳的奖励曲线，避免了**Q-Learning**和SARSA中出现的奖励波动。

在成功率方面，**SSA-Qlearning**在简单和中等环境下也表现出色，特别是在训练的初期，探索策略优化得益于其快速提升成功率。相比之下，**Q-Learning**、PSO-QLearning和SARSA在这一过程中进展较慢。即使在复杂环境下，**SSA-Qlearning**的成功率依旧保持较高水平，尽管某些时刻DDQN的表现更强劲，**SSA-Qlearning仍能保持稳定，避免了Q-Learning**和SARSA在复杂环境中出现的性能波动。

与DDQN相比，**SSA-Qlearning**在复杂环境下的最终表现相似，但其训练时间明显较短。DDQN依赖于深度神经网络的复杂结构，尽管其后期能够展示出较为优异的性能，但训练过程相对较慢，特别是在复杂环境下，DDQN需要更多训练周期才能稳定表现。相比之下，SSA-Qlearning通过有效的策略优化，在较短的训练时间内迅速找到最优解，确保其在简单和中等环境中能够快速收敛。

具体来说，在简单和中等环境下，SSA-Qlearning的奖励和成功率增长速度较DDQN更快，并且相较于Q-learning、PSO-QLearning和SARSA，SSA-Qlearning能够在训练早期取得较好的效果。尽管在复杂环境中，DDQN表现略好，但SSA-Qlearning的差距不大，并且其训练速度远远领先于DDQN，显著减少了计算资源消耗和训练时间。

为了更好的比较算法的优，将平均奖励(MR)、最终成功率(FSR)、奖励标准差(R STD)、最大/最小值(M/m)进行对比，**如表7所示**。

表 7 结果对比表

Table 7 **Result comparison table**

Algorithm	MR	FSR	R STD	M/m
Q-Learning(小)	14.11	0.98	8.94	19.3/-5.4
SARSA(小)	13.40	1	9.46	19.3/-6.9
SSA-Q(小)	15.08	1	8.21	19.3/-4.6
PSO-Q(小)	14.28	1	8.86	19.3/-6.69
DDQN(小)	13.74	0.99	9.26	19.3/-8.89
Q-Learning(中)	7.64	0.99	10.95	18.3/-11.4
SARSA(中)	8.89	0.99	10.96	18.3/-9.99
SSA-Q(中)	9.98	1	10.41	18.3/-11.4
PSO-Q(中)	8.41	1	10.99	18.3/-10.5
DDQN(中)	12.03	0.97	9.5	18.3/-9.99
Q-Learning(大)	-0.33	0.92	9.98	17.3/-11.7
SARSA(大)	-0.29	0.96	10.43	17.1/-11.8
SSA-Q(大)	3.43	0.99	11.16	17.3/-11.8
PSO-Q(大)	-0.02	0.89	10.29	17.2/-11.9
DDQN(大)	9.61	0.99	10.14	17.3/-11.6

在本实验中，SSA-Qlearning显示出显著的优势，尤其是在较大规模和复杂的动态环境中。相比于其他对比算法，SSA-Qlearning在多个环境下均保持了较高的最终成功率和较高的平均奖励，特别是在中型和大型环境中，其表现优于**Q-Learning**和PSO-Qlearning。此外，SSA-Qlearning在成功率上表现稳定，且在每个环境中均能达到较高的成功率，即使在复杂场景中也能有效地规避障碍物并成功到达目标。

虽然SSA-Qlearning的奖励标准差(R STD)略高，但这也反映了其在探索过程中对环境的适应性。相较于其他算法，SSA-Qlearning 能够在更多的情境中进行充分的探索，从而在复杂动态环境中做出更加灵活的决策，保持较高的表现。

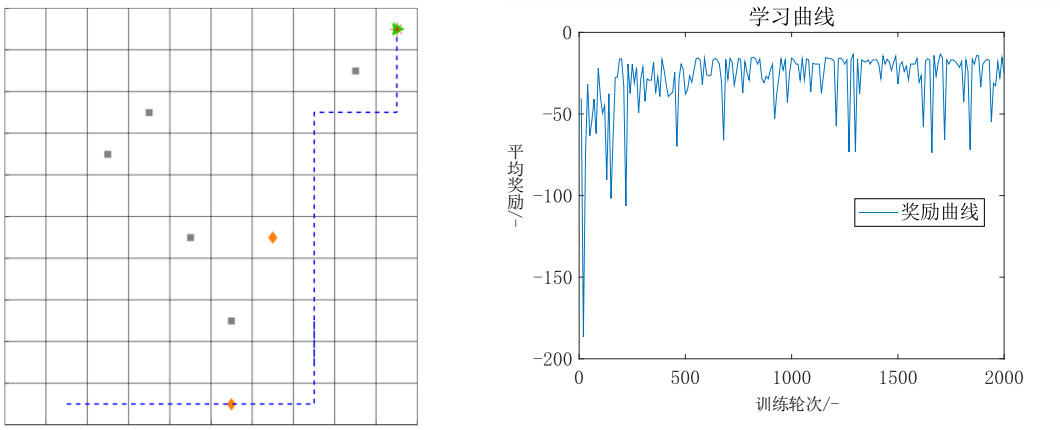
在动态环境下，SSA-Qlearning展现出了更强的稳定性和较快的适应能力，尤其是在大型环境中，其表现优于**Q-Learning**和PSO-Qlearning，训练时间也更短，计算开销显著减少。总的来说，SSA-Qlearning在各个对比算法中脱颖而出，尤其适用于需要较高动态适应性的任务。

因此，在时间和计算资源有限的情况下，SSA-Qlearning是一个非常理想的选择。它不仅能够在较短时间内达到与DDQN相似的效果，还能在更长时间的训练过程中展现出更高的效率和更稳定的收敛性。SSA-Qlearning的稳定性和高效性使其在多种环境下都能够提供较高的性能。

4 仿真实验

为了进一步验证本算法在动态环境下的优异性，本章将通过Matlab 2022a进行仿真实验。网格世界规格为10×10的动态栅格环境，设置三种不同复杂度的环境：(1)5个不可移动静态障碍物，2个动态周期性障碍物；(2)8个不可移动静态障碍物，4个动态周期性障碍物静态障碍物在图中显示为灰色实心方块；(3)17个不可移动静态障碍物，4个动态周期性障碍物。静态障碍物在图中为灰色实心方块，动态周期性障碍物在图中为橙色菱形块。

首先在环境(1)进行分析，仿真的路径图和奖励曲线图如图5所示。



(a) (b)

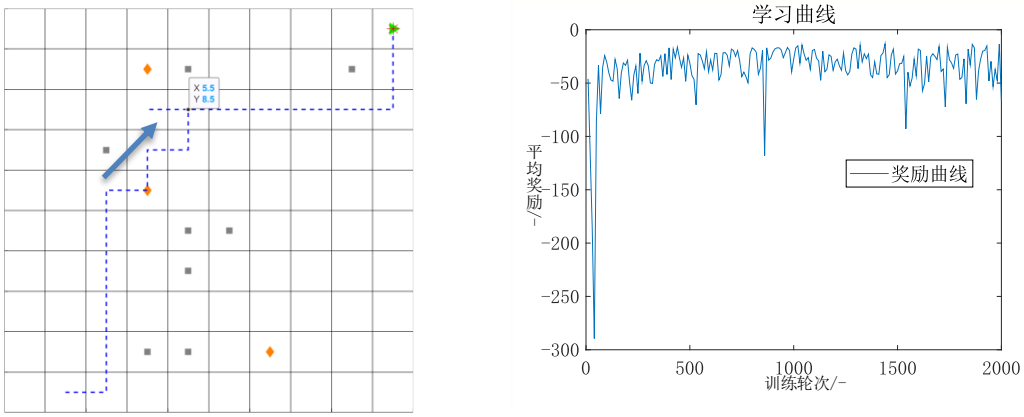
图 5 环境(1)结果图

Fig.5 Environmental (1) Result Chart

(a) Path simulation diagram (b) Reward curve diagram

在简单环境下的实验表明，该算法展现出显著的初期收敛特性，仅需150次迭代即实现奖励值稳定于-30波动区间且**最终奖励为-16.50**，验证了其在低复杂度场景中的快速策略优化能力。

然后基于环境(2)进行仿真，仿真的路径图和奖励曲线图如图6所示。



(a) (b)

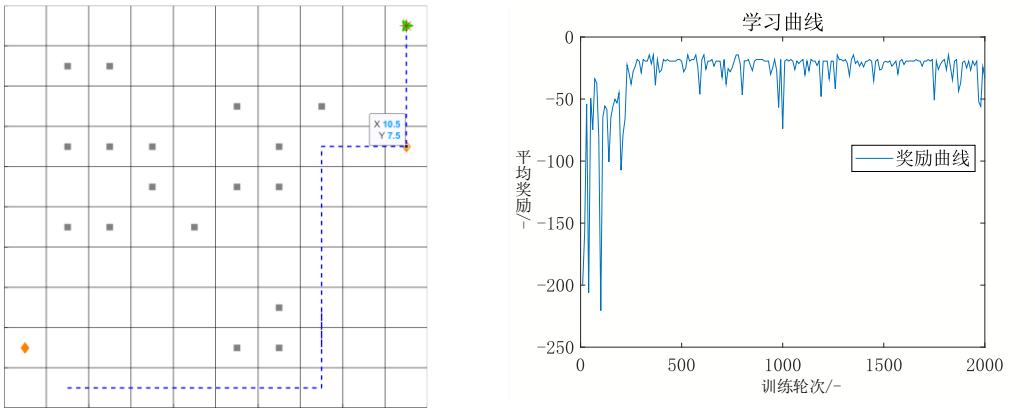
图 6 环境(2)结果图

Fig.6 Environmental (2) Result Chart

(a) Path simulation diagram (b) Reward curve diagram

根据路径仿真实验的轨迹分析，智能体在(5.5,8.5)呈现出阶段性回溯行为，以规避动态障碍物(4.5,9.5)的潜在碰撞风险。轨迹数据显示，智能体通过回溯成功规避碰撞风险，对应的奖励曲线在此阶段产生短期波动最终仍稳定收敛至-30波动区间，**最终奖励为-32.41**。该过程验证了算法在动态环境下的自主避障决策能力。

最后，对环境(3)进行仿真，仿真的路径图和奖励曲线图如图7所示。



(a) (b)

图 7 环境(3)结果图

Fig.7 Environmental (3) Result Chart

(a) Path simulation diagram (b) Reward curve diagram

根据路径仿真实验的轨迹分析，智能体呈现出良好的避障能力。**虽然这个动态障碍物最终落点在路径上，但实际运行中算法成功避开了它，没发生碰撞。**此过程导致奖励曲线出现短期波动，然而，曲线最终稳定收敛至-30的波动区间，**最终奖励为-18.15**。由此可见，在较为复杂的环境中，所提出的算法有效地实现了障碍物避让，并且在寻路任务中展现出较为优越的性能。

仿真实验验证了SSA-Qlearning算法在多种环境下取得了较好的成功率和动态避障能力。**通过协同优化框架的引入**，SSA-Qlearning显著提升了**动态路径规划任务中的学习效率、实时性和环境适应性**，表现出色的性能。**相比其他传统方法**，该算法不仅提高了路径规划的稳定性和准确性，还能够**高效应对动态障碍物的变化**，具有**广泛的应用前景**，特别是在移动机器人、无人机等实时导航系统中，**SSA-Qlearning表现出较强的适应能力和较高的可靠性，能够支持复杂的动态环境中的自主导航。**

5 结束语

针对动态环境下路径规划任务中，传统强化学习算法存在的探索-利用失衡、参数敏感性及实时性不足等问题，提出了一种基于麻雀搜索算法(SSA)与**Q-Learning**融合的协同优化框架(SSA-Qlearning)。通过理论分析、仿真与对比实验，验证了以下结论：

动态参数协同优化机制的有效性，通过SSA算法对学习率与衰减因子的在线联合优化，解决了传统**Q-Learning**中分阶段参数调优导致的策略滞后问题。麻雀种群的自适应搜索策略能够根据环境复杂度动态调整参数。

双决策融合策略的优越性，提出的动态权重机制通过量化环境动态性等级，实现了SSA全局规划与**Q-Learning**局部避障的自适应融合。

资源-性能的均衡性突破，相较于深度强化学习算法(如DDQN)，SSA-Qlearning在保持相近规划精度的前提下，大幅降低计算开销，训练时间大约为DDQN的百分之三。

总体而言，提出的SSA-Qlearning方法在动态路径规划任务中实现了性能与效率的突破性平衡，既为资源受限的场景提供了切实可行的解决方案，也为群体智能与强化学习交叉领域的研究开辟了新的路径。

参考文献

[1] Ait Saadi A, Soukane A, Meraihi Y, et al. UAV path planning using optimization approaches: A survey[J]. Archives of Computational Methods in Engineering, 2022, 29(6): 4233–4284.

[2] 于振中,李强,樊启高.智能仿生算法在移动机器人路径规划优化中的应用综述[J].计算机应用研究,2019,36(11):3210-3219.

[3] Zhang Hanye,Lin Weiming,Chen Aixia.Path planning for the mobile robot: A review [J]. Symmetry, 2018, 10(10): 450-466.

[4] Deng Zhiliang,Wang Dong.Research on Parking Path Planing Based on A-Star Algorithm [J]. Journal of New Media, 2023, 1(5):55-67.

[5] 齐款款,李二超,毛玉燕.改进 A*算法融合自适应 DWA 的移动机器人动态路径规划[J].数据采集与处理,2023,38(2):451-467.DOI:10.16337/j.1004-9037.2023.02.019.

[6] 陈靖辉,崔岩,刘兴林,等.基于改进 A*算法的移动机器人路径规划方法[J].计算机应用研究,2020,37(S1):118-119.

[7] Zhou Xiwei,Yan Jingwen,Yan Mei, et al.Path planning of rail-mounted logistics robots based on the improved Dijkstra algorithm [J]. Applied Sciences, 2023, 17(13): 9955-9972.

[8] 杨北辰,余粟.改进蚁群算法在路径规划中的应用[J].计算机应用研究,2022,39(11):3292-3297+3314.

[9] Li Binghui, Chen Badong. An adaptive rapidly-exploring random tree [J]. IEEE/CAA Journal of Automatica Sinica, 2021, 9(2): 283-294.

[10] 薛建凯.一种新型的群智能优化技术的研究与应用[D].上海:东华大学,2020:7-20.

[11] Li Zhaolun, Luo Xiaonan. Autonomous underwater vehicles (AUVs) path planning based on Deep Reinforcement Learning [C]// Proceedings of the 2022 9th International Conference on Digital Home (ICDH). IEEE, 2022: 257–262.

[12] 张荣霞,武长旭,孙同超,等.深度强化学习及在路径规划中的研究进展[J].计算机工程与应用,2021,57(19):44-56.

[13] Yu Zhenhua, Si Zhijie, Li Xiaobo, et al. A novel hybrid particle swarm optimization algorithm for path planning of UAVs [J]. IEEE Internet of Things Journal, 2022, 9(22): 22547-22558.

[14] 王宇,王文浩,徐凡,等.基于改进蚁群算法的植保无人机路径规划方法[J].农业机械学报,2020,51(11):103-112+92.

[15] 王艳春,郭永峰,夏颖,等.基于改进蚁群算法的机器人全局路径规划[J].电子科技,2024,37(05):88-94.

[16] 赵增旭,刘向阳,任彬.基于方向指引的蚁群算法机器人路径规划[J].计算机应用研究,2023,40(03):786-788+793.

[17] 张志文,刘伯威,张继园,等.麻雀搜索算法-粒子群算法与快速扩展随机树算法协同优化的智能车辆路径规划[J].中国机械工程,2024,35(06):993-999+1009.

[18] Zhang Xin, Shi Xiaoxu, Zhang Zuqiong, et al. A DDQN path planning algorithm based on experience classification and multi steps for mobile robots [J]. Electronics, 2022, 11(14): 2120-2139.

[19] 牟远明,卓然,高飞.基于混合改进麻雀搜索算法的农用移动机器人路径规划[J].中国农机化学报, 2024, 45(09): 234-243.

[20] 闫皎洁,张锲石,胡希平.基于强化学习的路径规划技术综述[J].计算机工程,2021,47(10):16-25.

[21] 刘光印,钱东海,王志国,等.基于改进 Q 学习的复杂环境下 AGV 路径规划研究[J].计量与测试技术,2025,51(03):84-88+94.

[22] 褚晶,邓旭辉,岳颀.基于 Q-learning 的搜救机器人自主路径规划[J].南京航空航天大学学报,2024,56(02):364-374.

作者简介：

许杨磊(2002-)，男，硕士研究生，研究方向：具身智能与路径规划, E-mail: 1944389915@qq.com。

王永雄(1970-)，**通信作者**，男，教授，博士，研究方向：机器视觉和智能机器人, E-mail: wyxiong@usst.edu.cn。