DOI: 10.19650/j.cnki.cjsi.J2514733

# 融合 MPC 参数化控制与改进 SAC 的全向移动机器人能耗优化与路径规划*

单泽彪[1,2],何立本[1],刘小松[1],郑海平[1],梁法辉[3]

(1. 长春理工大学电子信息工程学院 长春 130022; 2. 长春理工大学吉林省智能机器人高校协同创新中心 长春 130022; 3. 长春汽车职业技术大学电气工程学院 长春 130012)

**摘 要**:针对全向移动机器人在多障碍物不确定环境下的节能路径规划问题,提出一种考虑能耗的融合模型预测控制参数化与改进的软演员-评论家(SAC)框架的路径规划(E-QPSAC)算法。首先,将局部路径规划构造为二次规划问题,并采用改进原始-对偶混合梯度算法高效求解;其次,构建了 SAC 的 12 维状态空间(包括机器人位姿、目标信息及障碍物信息)、级联式网络架构及动态权重融合机制,有效平衡避障与目标趋向;同时考虑全向轮移动机器人因环境不确定性、系统非线性及运动特性导致的高能耗的特性,建立五维能耗模型,并依据能耗模型设计节能奖励函数。最后,在仿真平台中,与 SAC、深度确定性策略梯度(DDPG)及基于优先经验回放与长短期记忆网络的双延迟深度确定性策略梯度(PL-TD3)算法进行对比实验。实验结果表明:所提算法任务完成率达 98.7%(SAC 为 85.0%、DDPG 为 98.5%、PL-TD3 为 98.2%),总能耗低至 18.274 kJ(SAC 为 25.765 kJ、DDPG 为 22.418 kJ、PL-TD3 为 21.191 kJ),平均路径长度 6.45 m(SAC 为 6.56 m、DDPG 为 6.97 m、PL-TD3 为 7.13 m),所提算法在任务完成率、能耗、路径长度及鲁棒性上均更优。最终通过实测实验验证了所提方法的有效性。

**关键词**: 全向移动机器人;路径规划;能耗优化;强化学习

**中图分类号**: TH134 TP242 **文献标识码**: A **国家标准学科分类代码**: 460.40

# Energy consumption optimization and path planning of omnidirectional mobile robots integrating MPC parameterized control and improved SAC

Shan Zebiao[1,2], He Liben[1], Liu Xiaosong[1], Zheng Haiping[1], Liang Fahui[3]

(*1. School of Electronic and Information Engineering, Changchun University of Science and Technology, Changchun 130022, China; 2. Jilin Provincial Collaborative Innovation Center for Intelligent Robots, Changchun University of Science and Technology, Changchun 130022, China; 3. College of Electrical Engineering, Changchun Vocational University of Automobile Technology, Changchun 130012, China*)

**Abstract**: To address the energy-efficient path planning problem of omnidirectional mobile robots in uncertain environments with multiple obstacles, this paper proposes an energy-aware path planning algorithm named energy-aware quadratic programming parameterization soft actor-critic (E-QPSAC), which integrates model predictive control parameterization with an improved soft actor-critic (SAC) framework. First, local path planning is formulated as a Quadratic Programming problem and efficiently solved by an improved primal-dual hybrid gradient algorithm. Second, a 12-dimensional state space (including robot pose, target information, and obstacle information) for SAC is constructed, along with a cascaded network architecture and a dynamic weight fusion mechanism to effectively balance obstacle avoidance and goal orientation. Meanwhile, considering the high energy consumption characteristics of omnidirectional mobile robots caused by environmental uncertainty, system nonlinearity, and motion characteristics, a five-dimensional energy consumption model is established, and an energy-saving reward function is designed based on this model. Finally, comparative experiments are conducted on a simulation platform with SAC, deep deterministic policy gradient (DDPG), and the prioritized replay and LSTM-based twin delayed deep deterministic policy gradient (PL-TD3) algorithm. The experimental results show that the proposed algorithm achieves a task completion rate of 98.7% (85.0% for SAC, 98.5% for DDPG, and 98.2% for PL-TD3), a total energy consumption as low as

收稿日期:2025-11-27 Received Date: 2025-11-27

*基金项目:吉林省科技发展计划项目(YDZJ202503CGZH002)资助

18.274 kJ (25.765 kJ for SAC, 22.418 kJ for DDPG, and 21.191 kJ for PL-TD3), and an average path length of 6.45 m (6.56 m for SAC, 6.97 m for DDPG, and 7.13 m for PL-TD3). The proposed algorithm outperforms the others in task completion rate, energy consumption, path length, and robustness. The effectiveness of the proposed method is finally verified through practical experiments.

**Keywords**: omnidirectional mobile robot; path planning; energy consumption optimization; reinforcement learning

## 0 引 言

无人车路径规划与轨迹跟踪控制是自动驾驶系统的核心环节,也是保障车辆安全行驶、提升行驶平顺性与智能决策落地的关键技术支撑[1-3]。三轮全向移动机器人(three-wheeled omnidirectional mobile robot, TOMR)凭借其全向运动能力,在物流配送、医疗服务和灾难救援等复杂场景中展现出不可替代的应用价值。然而,在多障碍物不确定环境下的路径规划面临双重挑战:一方面,环境扰动与系统非线性特性导致传统规划方法产生路径跟踪误差;另一方面,TOMR 独特的运动机制引发严峻能耗问题——自转工况下电机扭矩相互抵消造成额外的功耗,减速阶段反接制动能量回收缺失使电池利用率降低,频繁转向更使典型工况续航时间缩短,这种能耗瓶颈严重制约机器人的持续作业能力。尽管传统方法如 A* 算法[4]、RRT* 算法[5]在静态环境中具有完备性保证,但其复杂环境适应性不足;模型预测控制(model predictive control, MPC)虽能处理约束,但计算复杂度随障碍物数量呈指数增长;深度强化学习(deep reinforcement learning, DRL)方法如 Chu 等[6]提出的双重深度 Q 网络(double deep q-network, DDQN)和 Xue 等[7]提出的多传感器融合虽提升环境自适应性,却因黑箱特性缺乏 Lyapunov 稳定性证明,在安全关键场景部署存在风险。在连续动作空间算法改进方面,Lin 等[8]提出的基于优先经验回放与长短期记忆网络的双延迟深度确定性策略梯度(prioritized replay and LSTM-based twin delayed deep deterministic policy gradient, PL-TD3)方法,有效提升了算法在障碍物环境下的样本利用效率与时序感知能力,然而该方法未将能耗建模纳入优化框架,在高能耗场景下存在明显局限。现有节能方案存在系统性缺陷:Hou 等[9-10]提出的启动阶段功耗模型仅覆盖加速工况,未包含匀速与制动阶段能量流;文献[11]的电池能耗模型忽略电机本体机电转换损耗;文献[12]未考虑反接制动能量回馈机制;Pal 等[13]提出的节能 A* 算法将地面摩擦简化为恒定参数;Li 等[14]提出的轨迹规划未耦合完整动力学模型;而王义娜等[15]提出的五维能耗模型虽创新性地量化了电机抵消效应,但其鲁棒补偿控制未解决多障碍物避障的实时规划问题,且未将能耗优化与 DRL 的探索能力结合,导致室外复杂场景的泛化能力受限。

针对上述问题,提出一种考虑能耗的融合 MPC 参数化与改进软演员-评论家(soft actor-critic, SAC)框架的路径规划(energy-aware quadratic programming parameterized soft actor-critic, E-QPSAC)算法,实现能耗与鲁棒性的协同优化。首先,构建类 MPC 结构的参数化控制器,将局部路径规划建模为带约束的二次规划(quadratic programming, QP)问题,并采用原始-对偶混合梯度算法进行求解。其次,构建 12 维状态空间,融合机器人位姿、目标信息及障碍物信息;设计由状态编码器、障碍物感知模块、QP 参数动态生成模块和目标导向模块构成的级联式 Actor 网络,并通过动态权重融合机制自适应平衡避障与目标追踪。然后,基于 TOMR 的电机特性与运动特性,设计了融合五维能耗模型的奖励函数,结合运动平滑性约束与目标朝向奖励,引导节能策略学习。最后,通过仿真分析以及实际场景路径规划实验验证了所提算法的有效性。

## 1 MPC 启发的参数化控制器

### 1.1 QP 问题的构建

基于 MPC 的有限时域优化控制问题通常采用的近似方法是将其截断到有限视界,并在每个时间步长中求解式(1)所示的问题,以 $\boldsymbol{x}_0$ 为当前状态,从最优解的第 1 个控制输入 $\boldsymbol{u}$ 以滚动时域的方式应用到系统中。

$$\begin{cases} \min\limits_{\boldsymbol{x}_1:N,\boldsymbol{u}_0:N-1} \sum\limits_{k=0}^{N-1} (\boldsymbol{x}_{k+1}-\boldsymbol{x}_r)^{\mathrm{T}}\boldsymbol{Q}(\boldsymbol{x}_{k+1}-\boldsymbol{x}_r) + \boldsymbol{u}_k^{\mathrm{T}}\boldsymbol{R}\boldsymbol{u}_k \\ \text{s.t. } \boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}\boldsymbol{u}_k,\ k=0,\cdots,N-1 \\ \boldsymbol{u}_{\min} \leqslant \boldsymbol{u}_k \leqslant \boldsymbol{u}_{\max} \\ \boldsymbol{x}_{\min} \leqslant \boldsymbol{x}_{k+1} \leqslant \boldsymbol{x}_{\max} \end{cases} \tag{1}$$

式中:$\boldsymbol{x}_k$ 为状态向量;$\boldsymbol{u}_k$ 为控制输入向量;$\boldsymbol{x}_r$ 为参考信号;$\boldsymbol{A}$ 和 $\boldsymbol{B}$ 为系统动态的状态转移和控制输入矩阵;$\boldsymbol{Q}$、$\boldsymbol{R}$ 均为系统的权重正定矩阵;$\boldsymbol{u}_{\min}$、$\boldsymbol{u}_{\max}$ 是控制输入约束;$\boldsymbol{x}_{\min}$、$\boldsymbol{x}_{\max}$ 是状态约束。

令 $\boldsymbol{y}=[\boldsymbol{u}_0^{\mathrm{T}},\cdots,\boldsymbol{u}_{N-1}^{\mathrm{T}}]$ 表示所有控制输入的决策序列并代入式(1),最后简化为 QP 问题的标准形式,标准形式如式(2)所示,产生 QP 问题的大小和参数如式(3)和(4)所示。

$$\min_{\boldsymbol{y}} \frac{1}{2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{y} + \boldsymbol{q}^{\mathrm{T}}\boldsymbol{y} \quad \text{s.t. } \boldsymbol{H}\boldsymbol{y} + \boldsymbol{b} \geqslant 0 \tag{2}$$

式中:$\boldsymbol{y} \in \mathbf{R}^{n_{\mathrm{qp}}}$;$\boldsymbol{P} \in \boldsymbol{S}_{++}^{n_{\mathrm{qp}}}$、$\boldsymbol{q} \in \mathbf{R}^{n_{\mathrm{qp}}}$ 由系统动态矩阵及权重参数确定,约束矩阵 $\boldsymbol{H} \in \mathbf{R}^{m_{\mathrm{qp}} \times n_{\mathrm{qp}}}$ 和向量 $\boldsymbol{b} \in \mathrm{R}^{m_{\mathrm{qp}}}$ 编码了

动作限制与障碍物约束。

$$\begin{cases} n_{\mathrm{qp}} = Nm_{\mathrm{sys}} \\ m_{\mathrm{qp}} = 2N(m_{\mathrm{sys}} + n_{\mathrm{sys}}) \end{cases} \tag{3}$$

式中:$n_{\mathrm{qp}}$ 为 QP 问题中决策变量的维数;$m_{\mathrm{qp}}$ 为约束条件的维数;$N$ 为预测时域。

$$\begin{cases} \boldsymbol{P} = \boldsymbol{B}'^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{B}' + \boldsymbol{R} \\ \boldsymbol{q} = 2\boldsymbol{B}'^{\mathrm{T}}\boldsymbol{Q}(\boldsymbol{A}'\boldsymbol{x}_0 - \boldsymbol{r}) \\ \boldsymbol{H} = -\boldsymbol{C}_x\boldsymbol{B}' - \boldsymbol{C}_u \\ \boldsymbol{b} = \boldsymbol{e}_x + \boldsymbol{e}_u - \boldsymbol{C}_x\boldsymbol{A}'\boldsymbol{x}_0 \end{cases} \tag{4}$$

式中:$\boldsymbol{I}$ 为单位矩阵;$\boldsymbol{C}_x$ 为状态约束系数矩阵;$\boldsymbol{C}_u$ 为控制约束系数矩阵;$\boldsymbol{e}_x$ 为状态约束常数项矩阵;$\boldsymbol{e}_u$ 为控制约束常数项向量;$\boldsymbol{r}$ 为参考轨迹向量;$\boldsymbol{B}' = \begin{bmatrix} \boldsymbol{B} & & & \\ \boldsymbol{AB} & \boldsymbol{B} & & \\ \vdots & & \ddots & \\ \boldsymbol{A}^{N-1}\boldsymbol{B} & \cdots & \cdots & \boldsymbol{B} \end{bmatrix}$;$\boldsymbol{A}' = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{A}^2 \\ \vdots \\ \boldsymbol{A}^N \end{bmatrix}$;$\boldsymbol{C}_x = \boldsymbol{I}_N \otimes \mathrm{diag}(\boldsymbol{I}_{n_{sys}}, -\boldsymbol{I}_{n_{sys}}, \boldsymbol{0}_{2m_{sys}\times 2n_{sys}})$;$\otimes$为Kronecker积;$\boldsymbol{C}_u = \boldsymbol{I}_N \otimes \mathrm{diag}(\boldsymbol{0}_{2n_{sys}\times 2n_{sys}}, \boldsymbol{I}_{m_{sys}}, -\boldsymbol{I}_{m_{sys}})$;$\boldsymbol{e}_x = \boldsymbol{I}_N \otimes [\boldsymbol{x}_{\max}^{\mathrm{T}} - \boldsymbol{x}_{\min}^{\mathrm{T}}]^{\mathrm{T}}$;$\boldsymbol{e}_u = \boldsymbol{I}_N \otimes [\boldsymbol{u}_{\max}^{\mathrm{T}} - \boldsymbol{u}_{\min}^{\mathrm{T}}]^{\mathrm{T}}$;$\boldsymbol{r} = \boldsymbol{I}_N \otimes \boldsymbol{x}_r$;$\boldsymbol{Q} = \boldsymbol{I}_N \otimes \boldsymbol{Q}$;$\boldsymbol{R} = \boldsymbol{I}_N \otimes \boldsymbol{R}$。

### 1.2　QP 问题的求解

针对传统 QP 求解器在动态环境中实时性不足的瓶颈,在 SAC 框架内设计了类 MPC 的嵌入式 QP 求解器。该求解器采用原始-对偶混合梯度(primal-dual hybrid gradient, PDHG)算法,通过交替更新原始变量和对偶变量来高效求解卡罗需-库恩-塔克条件(Karush-Kuhn-Tucker conditions, KKT)系统。

1)问题转化与拉格朗日函数构建

首先,将标准 QP 问题转化为易于求解的鞍点形式。为处理约束条件,引入拉格朗日乘子 $\boldsymbol{\lambda} \in \mathbf{R}_+^{m_{\mathrm{qp}}}$,构建拉格朗日函数如式(5)所示。

$$\mathcal{L}(\boldsymbol{y},\boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{y} + \boldsymbol{q}^{\mathrm{T}}\boldsymbol{y} - \boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{y} + \boldsymbol{b}) \tag{5}$$

根据 KKT 最优性条件,最优解 $(\boldsymbol{y}^*,\boldsymbol{\lambda}^*)$ 必须满足如式(6)所示条件。

$$\begin{cases} \nabla_y\mathcal{L} = \boldsymbol{P}\boldsymbol{y}^* + \boldsymbol{q} + \boldsymbol{H}^{\mathrm{T}}\boldsymbol{\lambda}^* = 0 & \text{(梯度条件)} \\ \boldsymbol{H}\boldsymbol{y}^* + \boldsymbol{b} \geqslant 0 & \text{(原始可行性)} \\ \boldsymbol{\lambda}^* \geqslant 0 & \text{(对偶可行性)} \\ \boldsymbol{\lambda}_i^*(\boldsymbol{H}\boldsymbol{y}^* + \boldsymbol{b})_i = 0, \quad \forall i & \text{(互补松弛条件)} \end{cases} \tag{6}$$

该条件为最优解提供了充要条件,同时也是 PDHG 算法的迭代基础。

2)原始-对偶梯度下降算法推导

为高效求解 KKT 系统,采用 PDHG 算法。其核心思想是通过交替更新原始变量 $\boldsymbol{y}$ 和对偶变量 $\boldsymbol{\lambda}$ 逼近鞍点,其迭代形式如式(7)、(8)所示。

(1)使用投影梯度上升对对偶变量进行更新。

$$\boldsymbol{\lambda}^{(k+1)} = \Pi_{\mathrm{R}^+}[\boldsymbol{\lambda}^{(k)} + \eta(\boldsymbol{H}\boldsymbol{y}^{(k)} + \boldsymbol{b})] \tag{7}$$

其中,$\Pi_{\mathrm{R}^+}[\cdot]$ 为投影算子,确保 $\boldsymbol{\lambda} \geqslant 0$。

(2)使用梯度下降对原始变量进行更新。

$$\boldsymbol{y}^{(k+1)} = \boldsymbol{y}^{(k)} - \eta(\boldsymbol{P}\boldsymbol{y}^{(k)} + \boldsymbol{q} + \boldsymbol{H}^{\mathrm{T}}\boldsymbol{\lambda}^{(k+1)}) \tag{8}$$

经过有限次迭代后,原始变量将收敛至满足约束且最小化目标函数的最优解,而对偶变量则收敛至最优拉格朗日乘子。

## 2　E-QPSAC 算法设计

### 2.1　SAC 算法

SAC 算法结合最大熵强化学习,适用于连续动作空间,其核心目标是通过最大化策略的随机性和累计奖励的加权和来实现探索与利用的平衡[16]。SAC 算法主要由 Actor 策略网络,Critic 价值网络,最大熵目标,经验回放等组成。策略网络的优化目标 $\mathcal{L}_{policy}$ 为最大化奖励和熵的组合,如式(9)所示。

$$\mathcal{L}_{policy} = \mathrm{E}_{s\sim D,\alpha\sim\pi_\theta}[\alpha\log\pi_\theta(\alpha\mid s) - \min_{j=1,2}Q_{\phi_j}(s,\alpha)] \tag{9}$$

式中:$\alpha$ 为温度系数;$\log\pi_\theta(\alpha\mid s)$ 为熵正则化项;$Q(s,\alpha)$ 为当前动作的价值估计。

Actor 网络根据当前策略生成动作 $u$,执行动作 $u$ 后,环境返回奖励 $r$ 与下一时刻状态 $s'$。当经验积累了足够数量后,从经验回放池中采样小批量数据,使用目标 $V$ 网络计算目标 Q 值 $y_Q$,如式(10)所示。

$$y_Q = r + \gamma V_{\mathrm{next}}(s') \tag{10}$$

式中:$V_{\mathrm{next}}(s')$ 为下一状态的目标值;$\gamma$ 为折扣因子。

通过最小化 Q 网络的损失函数进行 Critic 网络更新,如式(11)所示。

$$L_Q = \mathrm{E}_{(s,\alpha,r,s')\sim D}[(Q(s,\alpha) - y_Q)^2] \tag{11}$$

### 2.2　算法整体框架

E-QPSAC 算法整体架构如图 1 所示。下面将依次阐述状态空间设计、级联式 Actor 与 Critic 网络结构、融合 QP 残差正则化的损失函数,以及基于五维能耗模型的奖励设计。

### 2.3　状态空间设计

TOMR 的运动状态如图 2 所示,其中 $XOY$ 为大地坐标系,$X'O'Y'$ 为 TOMR 坐标系。

针对三轮全向轮机器人在多障碍物环境中的运动特性,设计 12 维状态空间,全面表征机器人的运动状态,其中包括 4 维机器人位姿、4 维目标信息、4 维障碍物信息,具体如式(12)所示。

$$\boldsymbol{s}_t = [x,y,\theta,v] \oplus [d_{\mathrm{g}},\phi_{\mathrm{g}},\Delta x_{\mathrm{g}},\Delta y_{\mathrm{g}}] \oplus [d_{\mathrm{o1}},\phi_{\mathrm{o1}},d_{\mathrm{o2}},\phi_{\mathrm{o2}}] \tag{12}$$

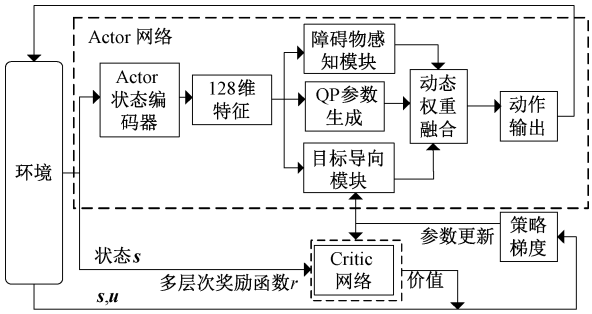


图 1 E-QPSAC 算法结构

Fig. 1 Structure diagram of E-QPSAC algorithm

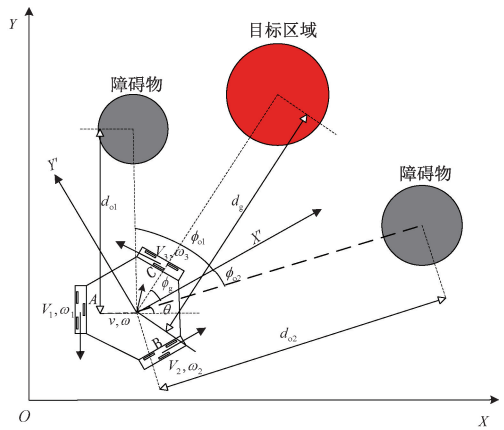


图 2 TOMR 运动状态

Fig. 2 Motion state of TOMR

式中：$x$ 和 $y$ 为位置坐标；$\theta$ 为航向角；$v$ 为线速度；$d_g$ 为目标距离；$\phi_g$ 为目标方位角；$\Delta x_g$ 和 $\Delta y_g$ 为目标相对位置；$d_{o1}$ 和 $\phi_{o1}$ 为最近障碍物的距离与方位角；$d_{o2}$ 和 $\phi_{o2}$ 为次近障碍物的距离与方位角。

## 2.4 E-QPSAC 网络结构设计

提出的改进 SAC 框架中 Actor 采用“环境感知-优化求解-动作修正”级联架构，替代单一策略网络直接输出动作的方式。该架构嵌入了一个 QP 求解器，用于实现结构化决策，并通过动态权重融合机制输出最终动作。Critic 网络负责评估由 Actor 的 QP 求解器生成的动作，并提供 Q 值反馈。

1) Actor 和 Critic 网络

Critic 网络采用 SAC 的标准双 Q 网络减少估计偏差，每个网络通过独立分支处理状态和动作。每个 Q 网络都是 3 层全连接+LayerNorm 归一化层的结构，输入为 12 维状态向量，输出为状态-动作值函数 $Q(s,a)$。Actor 则采用级联架构，实现结构化决策，其结构为：

(1)状态编码器

采用 3 层 MLP 结构处理 12 维状态输入，将原始状态映射为 128 维特征向量，增强环境表征能力。

(2)障碍物感知模块

设计 1 个输入层、2 个全连接隐含层和 1 个输出层作为障碍物感知子网络，具体如式(13)和(14)所示。

$$u_{\text{obs}} = f_{\text{obs}}(s) \cdot \boldsymbol{v}_o \tag{13}$$

$$v_o = -\sum_{i=1}^{2} \frac{\nabla d_{\text{oi}}}{d_{\text{oi}}^2} \cdot \mathbb{I}(d_{\text{oi}} < d_{\text{safe}}) \tag{14}$$

式中：$f_{\text{obs}}(s)$ 是障碍物感知子网络；$\boldsymbol{v}_o$ 是排斥力向量，当障碍物距离小于阈值时输出排斥力向量；$d_{\text{oi}}$ 是最近障碍物距离；$d_{\text{safe}}$ 是安全距离阈值；$\mathbb{I}$ 为指示函数，仅当障碍物在警戒范围内激活。

(3)目标导向模块

设计 1 个输入层、1 个全连接隐含层和 1 个输出层作为目标导向子网络，如式(15)所示。

$$u_{\text{goal}} = f_{\text{goal}}(\Delta x_g, \Delta y_g, d_{\text{goal}}) \tag{15}$$

式中：$\Delta x_g$ 和 $\Delta y_g$ 是目标相对位置；$d_{\text{goal}}$ 是目标距离。

(4)QP 参数动态生成模块

通过式(7)和(8)最优解的迭代过程可以看出，其原始对偶变量依赖于 QP 问题参数 $(\boldsymbol{P},\boldsymbol{q},\boldsymbol{H},\boldsymbol{b})$，因此利用神经网络动态生成 QP 问题参数，然后通过对偶梯度求解得出迭代的最优解，其问题参数求解如下：

状态无关矩阵 $\boldsymbol{P}$、$\boldsymbol{H}$：由式(4)可知，对于 MPC 控制器，矩阵 $\boldsymbol{P}$、$\boldsymbol{H}$ 不受初始和参考状态 $(\boldsymbol{x}_0,\boldsymbol{x}_r)$ 的影响。因此可以独立建立 $\boldsymbol{P}$、$\boldsymbol{H}$ 矩阵参数。对于矩阵 $\boldsymbol{P}$，通过设计两层全连接层与 Tanh 激活函数生成了 $\boldsymbol{P}$ 矩阵生成网络。对于约束参数矩阵 $\boldsymbol{H}$，通过设计单层全连接网络与 Tanh 激活函数来动态构建障碍物约束的几何关系。

线性项 $\boldsymbol{q}$、$\boldsymbol{b}$：由式(4)可知，向量 $\boldsymbol{q}$ 与当前状态 $\boldsymbol{x}_0$ 和参考状态 $\boldsymbol{x}_r$ 存在线性映射，向量 $\boldsymbol{b}$ 可以看成 $\boldsymbol{x}_0$ 的仿射。通过单层全连接网络与 Tanh 激活函数生成了 $\boldsymbol{q}$ 和 $\boldsymbol{b}$ 向量生成网络，以此学习更复杂的状态-参数映射关系，适应非线性系统。

(5)动态权重融合机制

在复杂环境下的机器人控制任务，传统固定权重无法满足多种需求，会出现权重冲突、响应滞后等问题，因而设计了动态权重融合机制，保留基础的 QP 解并根据环境状态自适应调整策略，距离目标近时专注目标，有障碍物时优先避让。障碍物权重、目标权重动作计算公式如式(16)~(18)所示。

$$w_{\text{obs}} = 1 - 0.5 \cdot w_{\text{goal}} \tag{16}$$

$$w_{\text{goal}} = \sigma(10(0.3 - d_{\text{goal}})) \tag{17}$$

$$u_{\text{final}} = \tanh(u_{\text{QP}} + u_{\text{obs}} \cdot w_{\text{obs}} + u_{\text{goal}} \cdot w_{\text{goal}}) \cdot u_{\max} \tag{18}$$

式中：$w_{\text{obs}}$ 为障碍物避让权重；$w_{\text{goal}}$ 为目标趋向权重，通过 Sigmoid 函数实现距离相关的动态权重分配；$u_{\text{QP}}$ 为 QP 求解器输出的基础动作；$u_{\max}$ 为最大控制幅值；$u_{\text{final}}$ 为最终的动作决策输出。

### 2.5　损失函数设计

提出的 E-QPSAC 算法在标准 SAC 损失函数的基础上,引入基于 MPC 启发的 QP 残差正则化项,形成多目标优化的损失函数体系。

Actor 网络的损失函数综合了策略改进与 QP 求解质量两方面的优化目标。其中,QP 残差正则化损失 $\mathcal{L}_{\mathrm{QP}}$ 用以保证嵌入网络中的 QP 求解器能够输出物理合理且数值稳定的解。具体如式(19)所示。

$$\mathcal{L}_{\pi}(\theta)=\mathcal{L}_{\mathrm{policy}}+\beta\cdot\mathcal{L}_{\mathrm{QP}} \tag{19}$$

式中:$\mathcal{L}_{\mathrm{policy}}=E_{s\sim D,\alpha\sim\pi_\theta}[\alpha\log\pi_\theta(\alpha\mid s)-\min_{j=1,2}Q_{\phi_j}(s,\alpha)]$;$\mathcal{L}_{\mathrm{QP}}=\mathrm{E}_{s\sim D}[\parallel\nabla_{\mathrm{y}}\mathcal{L}(\boldsymbol{y}^{*},\boldsymbol{\lambda}^{*})\parallel^{2}+\parallel\min(\boldsymbol{\lambda}^{*},\boldsymbol{H}\boldsymbol{y}^{*}-\boldsymbol{b})\parallel^{2}]$。其中,$\alpha\log\pi_\theta(\alpha\mid s)$ 鼓励策略随机性以增强探索,$Q_{\phi_j}(s,\alpha)$ 驱使策略选择高价值动作,两者通过温度参数 $\alpha$ 实现平衡;$(\boldsymbol{y}^{*},\boldsymbol{\lambda}^{*})$ 是 QP 问题的原始-对偶最优解,$\mathcal{L}$是拉格朗日函数;正则化系数 $\beta$ 通过实验调优确定,用于平衡策略优化与 QP 求解质量。

### 2.6　五维能耗模型融合奖励函数设计

针对节能路径规划问题,将五维能耗模型[15]融入奖励函数,通过实时计算各能耗分量并施加差异化惩罚,驱动策略在保证路径跟踪精度同时降低能量消耗。五维能耗模型 $E_{\mathrm{total}}$ 包括电机电枢电阻消耗($E_{\mathrm{Ra}}$)、垂直方向力抵消损耗($E_{\mathrm{F1}}$)、平行方向力抵消损耗($E_{\mathrm{F2}}$)、动能变化($E_{\mathrm{k}}$)和机械损耗($E_{\mathrm{e}}$),如式(20)所示。

$$E_{\mathrm{total}}=E_{\mathrm{Ra}}+E_{\mathrm{F1}}+E_{\mathrm{F2}}+E_{\mathrm{k}}+E_{\mathrm{e}} \tag{20}$$

电机电枢电阻消耗($E_{\mathrm{Ra}}$),如式(21)所示。

$$E_{\mathrm{Ra}}=\int_0^T\left(\frac{V_{\mathrm{s}}^2}{R_{\mathrm{a}}}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}-\frac{2k_{\mathrm{e}}V_{\mathrm{s}}}{LR_{\mathrm{a}}}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}+\frac{n^2k_{\mathrm{e}}^2}{L^2R_{\mathrm{a}}}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}\right)\mathrm{d}t \tag{21}$$

式中:$k_{\mathrm{e}}$ 是反电动势常数;$R_a$ 是电枢电阻;$L$ 是全向轮半径;$\boldsymbol{u}=[u_1,u_2,u_3]^{\mathrm{T}}$ 是电机的控制输入;$n$ 是传动比;$V_s$ 是电源电压。

力抵消损耗($E_{\mathrm{F1}}$ 和 $E_{\mathrm{F2}}$),$E_{\mathrm{F1}}$ 是电机驱动力垂直于速度方向抵消的能量,$E_{\mathrm{F2}}$ 是电机驱动力平行于速度方向抵消的能量,如式(22)、(23)所示。

$$E_{\mathrm{F1}}=\int_0^T\left[\sum_{i=1}^{3}\left|F_i\sin\sigma_i\times\boldsymbol{v}_i\sin\sigma_i\right|-\left|\sum_{i=1}^{3}F_i\sin\sigma_i\times\boldsymbol{v}_i\sin\sigma_i\right|\right]\mathrm{d}t \tag{22}$$

$$E_{\mathrm{F2}}=\int_0^T\left[\begin{array}{l}\left|F_1\cos\sigma_1\times\boldsymbol{v}_1\cos\sigma_1\right|+\\\left|F_2\cos\sigma_2\times\boldsymbol{v}_2\cos\sigma_2\right|+\\\left|F_3\cos\sigma_3\times\boldsymbol{v}_3\cos\sigma_3\right|-\\\left|F_1\cos\sigma_1\times\boldsymbol{v}_1\cos\sigma_1+\right.\\\left.F_2\cos\sigma_2\times\boldsymbol{v}_2\cos\sigma_2+\right.\\\left.F_3\cos\sigma_3\times\boldsymbol{v}_3\cos\sigma_3\right|\end{array}\right]\mathrm{d}t \tag{23}$$

动能变化($E_{\mathrm{k}}$)如式(24)所示。

$$E_{\mathrm{k}}=\frac{1}{2}M\Delta v^2 \tag{24}$$

式中:$E_{\mathrm{k}}$ 反映机器人动能变化,加速时消耗能量,减速时回收部分能量;$M$ 为电机本体质量。

机械损耗($E_e$)如式(25)所示。

$$E_{\mathrm{e}}=\int_0^T h\parallel\boldsymbol{\omega}\parallel^{2}\mathrm{d}t \tag{25}$$

式中:$E_{\mathrm{e}}$ 由电机齿轮摩擦引起;$h$ 为粘性摩擦系数;$\boldsymbol{\omega}=[\omega_1,\omega_2,\omega_3]^{\mathrm{T}}$ 是电机转动角速度。

所设计方法的奖励函数 $r_{\mathrm{total}}$ 如式(26)所示。

$$r_{\mathrm{total}}=r_{\mathrm{goal}}+r_{\mathrm{obstacle}}+r_{\mathrm{heading}}+r_{\mathrm{energy}}+r_{\mathrm{smooth}} \tag{26}$$

式中:$r_{\mathrm{goal}}$ 表示目标奖励;$r_{\mathrm{obstacle}}$ 表示障碍物奖励;$r_{\mathrm{heading}}$ 表示目标朝向奖励;$r_{\mathrm{energy}}$ 表示五维能耗惩罚 $r_{\mathrm{smooth}}$ 为运动约束奖励。

(1) 目标奖励函数如式(27)所示,当机器人到达目标点区域时触发高额奖励。

$$r_{\mathrm{goal}}=\omega_{\mathrm{goal}}\cdot 2\,000 \tag{27}$$

式中:$\omega_{\mathrm{goal}}$ 为目标权重系数。

(2) 障碍物避障与碰撞奖励函数如式(28)所示,在决策评估中,机器人特别重视避障行为的判断和奖惩机制。

$$r_{\mathrm{obstacle}}=\begin{cases}-1\,000, & \text{碰撞}\\ -100\cdot\left(\dfrac{20-d_{\mathrm{obs}}}{20}\right)^{2}, & d_{\mathrm{obs}}<20\\ -20\cdot\dfrac{40-d_{\mathrm{obs}}}{30}, & 20\leqslant d_{\mathrm{obs}}<40\\ 1.0, & d_{\mathrm{obs}}\geqslant 40\end{cases} \tag{28}$$

式中:$d_{\mathrm{obs}}$ 为障碍物距离。当 $d_{\mathrm{obs}}<20$ 时惩罚急剧增加,当 $20\leqslant d_{\mathrm{obs}}<40$ 时惩罚线性增加。

(3) 朝向目标奖励函数如式(29)所示。该项奖励鼓励 TOMR 朝向目标方向运动,目的是提高对目标的定位精度,减少不必要的转向动作,从而使行驶轨迹更加平滑,奖励函数为:

$$r_{\mathrm{heading}}=w_{\mathrm{heading}}\times\left(1-\frac{\Delta\theta}{\pi/6}\right) \tag{29}$$

式中:$\Delta\theta$ 是机器人动作方向与目标方向的最小夹角,当 $\Delta\theta<30$ 朝向目标时给予奖励,如果 $\Delta\theta>30$ 背向目标,则给予惩罚;$w_{\mathrm{heading}}$ 是目标朝向权重系数。

(4) 运动约束奖励函数,如式(30)所示。该项奖励抑制剧烈的控制动作,促进运动的连续性与平滑性。定义动作幅值 $\parallel\boldsymbol{u}\parallel=\sqrt{\boldsymbol{u}_x^2+\boldsymbol{u}_y^2}$,当其超过阈值 0.8 时,施加惩罚。

$$r_{\text{smooth}} = \begin{cases} -2.0(\|\boldsymbol{u}\| - 0.8), & \|\boldsymbol{u}\| > 0.8 \\ 0, & \text{其他} \end{cases} \tag{30}$$

式中：$\|\boldsymbol{u}\|$ 为控制输入幅值，当 $\|\boldsymbol{u}\| - 0.8$ 时，惩罚剧烈控制行为。

（5）五维能耗惩罚函数

五维能耗惩罚项将前述构建的能耗模型嵌入奖励函数，实现能耗感知的策略学习。根据各能耗分量的增量，构造加权惩罚如式(31)所示。

$$R_{\text{energy}} = -\omega_{\text{energy}} \cdot (\alpha_1 \cdot E_{\text{Ra}} + \alpha_2 \cdot (E_{\text{F1}} + E_{\text{F2}}) + \alpha_3 \cdot E_{\text{e}} + \alpha_4 \cdot E_{\text{k}}) \tag{31}$$

式中：$\omega_{\text{energy}}$ 和 $\alpha_{1\sim4}$ 是用于平衡各能耗分量的权重系数。

## 3 仿真与实验验证

### 3.1 仿真条件与内容

实验采用 TOMR 模型的基本参数如表 1 所示。TOMR 的路径规划任务定义为：从固定起点出发，避开所有障碍物到达目标点。成功到达视为任务成功；若碰撞障碍物或超过最大步数，则任务失败并结束当前回合。

**表 1 机器人基本参数**

**Table 1 Basic parameters of the robot**

| 物理意义 | 符号/单位 | 值 |
|---|---|---|
| 机器人质量 | kg | 20.000 |
| 转动惯量 | kg·m$^2$ | 5.000 |
| 全向轮半径 | m | 0.100 |
| 电池电压 | V | 24.000 |
| 电枢电阻 | Ω | 0.560 |
| 转矩常数 | N·m/A | 0.028 |
| 反电动势常数 | V·s/rad | 0.028 |
| 传动比 | | 1.000 |
| 电机齿轮间粘性摩擦系数 | N·m·s/rad | 0.054 |

### 3.2 实验结果分析

1）学习过程分析

为验证所提算法相对 PL-TD3、DDPG、SAC 算法的优越性，在同一环境下，对比各算法的训练曲线（路径长度、能耗、平均奖励）以评估学习效率与稳定性。图 3～5 分别展示了在复杂环境下，各算法训练过程中每个回合的路径长度、能耗及平均奖励值的变化曲线。

从图 3 可以看出，E-QPSAC 的路径长度曲线在训练过程中，经前期短暂波动后迅速收敛至较低数值区间，整体路径长度显著低于对比算法 SAC、DDPG 和 PL-TD3。这表明 E-QPSAC 能更高效地探索并规划出更短且稳定的路径，路径的最优性与稳定性更突出。

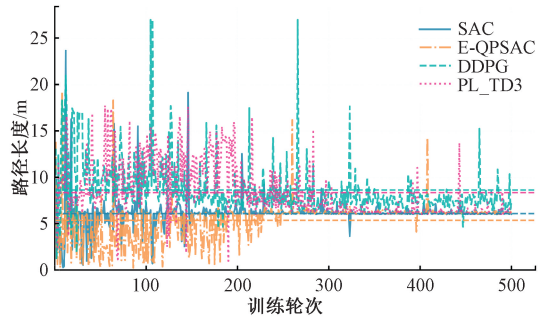


图 3 路径长度曲线

Fig. 3 Path length curve

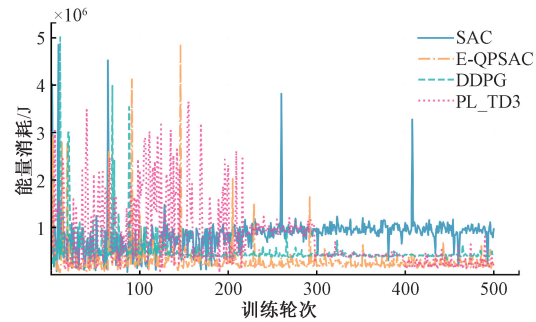


图 4 能耗曲线

Fig. 4 Energy consumption curve

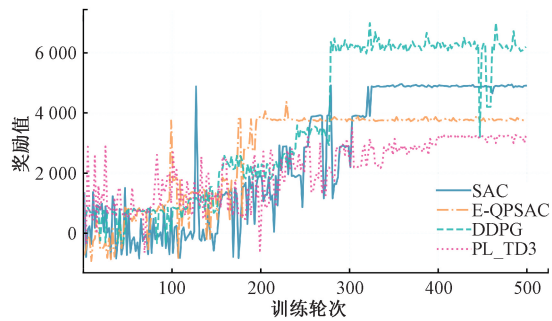


图 5 平均奖励值曲线

Fig. 5 Average reward value curve

从图 4 可以看出，E-QPSAC 的能耗虽然在训练前期存在小幅波动，但全程平均能耗远低于对比算法，且收敛后能耗波动极小。结果表明 E-QPSAC 算法能够有效控制三轮全向移动机器人的能量消耗，在路径规划中实现了较为显著的节能效果。

从图 5 可以看出，E-QPSAC 所设计的奖励机制更为合理，约在 200 轮后趋于稳定；SAC 算法的奖励机制直至约 300 轮才逐渐收敛，且在训练后期仍偶发奖励值骤降，表明其策略稳定性不足；DDPG 算法的收敛速度仍慢于 E-QPSAC，同时因对环境噪声较为敏感，在训练后期出现阶段性不稳定，影响了学习效果的持续性。PL-TD3 算法

整体奖励值偏低,始终维持在较低水平且波动持续存在,表明其在当前任务场景下样本利用效率有限,策略优化效果相对较弱。

综上所述,所提算法的收敛速度与学习稳定性均优于对比算法,路径长度与能耗曲线对比验证了其在复杂任务中的综合优势。

2)不同复杂度静态环境测试分析

为验证算法的基础性能和泛化能力,分别验证不同算法在不同障碍物环境下的路径规划能力。在简单障碍物环境中,各算法的规划路径对比如图 6 所示,定量性能指标对比详见表 2。

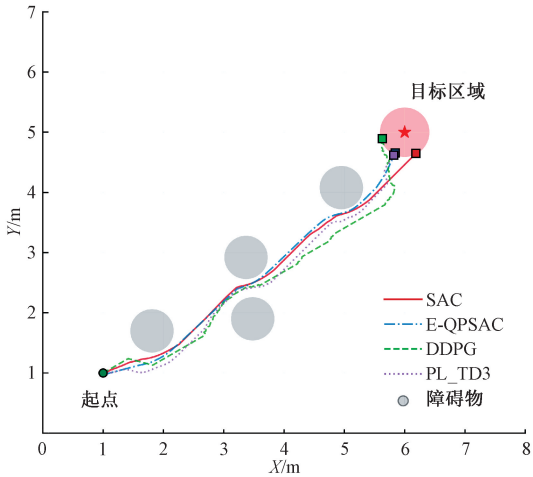


图 6　简单障碍物条件下规划路径

Fig. 6　Planned paths under simple obstacle conditions

**表 2　性能指标对比**

**Table 2　Comparison of performance indices**

| 方法 | 任务完成率/% | 能量损耗/kJ | 平均运动时间/s | 平均路径长度/m |
|---|---|---|---|---|
| SAC | 85.0 | 25.765 | 12.56 | 6.56 |
| DDPG | 98.5 | 22.418 | 10.81 | 6.97 |
| PL-TD3 | 98.2 | 21.191 | 12.69 | 7.13 |
| E-QPSAC | 98.7 | 18.274 | 12.32 | 6.45 |

分析图 6 和表 2 可知,各算法均能规划出从起点到目标区域的可行路径,但路径质量存在明显差异。SAC 算法存在明显的曲折与震荡,导致其任务完成率最低(85.0%),且能量损耗最高。PL-TD3 算法虽能到达目标区域,但其路径呈现出较大幅度的迂回绕行,在障碍物稀疏的简单障碍物环境中路径偏离最优轨迹较为明显,表明其路径规划效率有待提升。DDPG 算法在任务完成率(98.5%)和平均运动时间(10.81 s)上表现较好,但能量损耗(22.418 kJ)和平均路径长度(6.97 m)均高于 E-QPSAC。本文提出的 E-QPSAC 算法表现最为优异,在简单障碍物环境中取得了最高的任务完成率(98.7%),其规划的路径更贴近最优轨迹。

值得注意的是,E-QPSAC 的平均运动时间(12.32 s)略高于 DDPG(10.81 s),这主要源于算法中引入的五维能耗惩罚机制主动抑制高速急加速与反接制动,使策略习得中低速平稳行驶模式,以降低动能突变和力抵消损耗。从综合效益来看,E-QPSAC 的路径长度(6.45 m)短于 DDPG(6.97 m),单位路程能耗(2.83 kJ/m)低于 DDPG(3.22 kJ/m),以约 1.5 s 的时间代价换取了 18.49% 的能耗节约与更高的任务可靠性。综上,E-QPSAC 在能量损耗、路径长度及任务完成率上均为最优,综合性能最佳。

为检验算法在更加复杂场景下的鲁棒性,进一步在复杂障碍物环境中进行测试,结果如图 7 所示。

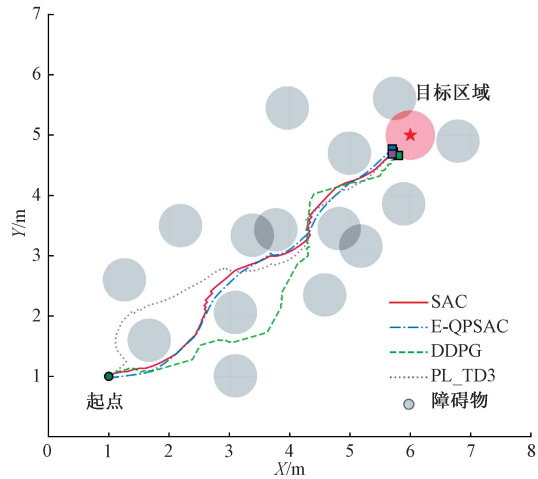


图 7　复杂障碍物条件下规划路径

Fig. 7　Planned paths under complex obstacle conditions

从图 7 可见,在障碍物显著增加的复杂环境下,SAC 算法规划的路径出现更明显的曲折,导致路径平滑性不足;DDPG 和 PL-TD3 算法虽能完成规划任务,但其路径在密集障碍物间也表现出较多的绕行与转折,导致路径冗余度增高。相比之下,E-QPSAC 算法所规划的路径在高密度障碍物环境中均保持整体平滑与连贯,展现出良好的环境适应性与决策鲁棒性。

3)运动特性分析

为深入对比各算法在规划路径过程中的底层控制性能,基于在复杂障碍物环境下的路径测试结果,选取训练成熟的策略,分析 TOMR 在典型回合中速度及航向角的变化情况,以评估算法在运动平滑性与控制稳定性方面的表现。图 8 和 9 分别对比了 E-QPSAC、SAC、DDPG、PL-TD3 算法控制 TOMR 运动时的速度变化与航向角变化。

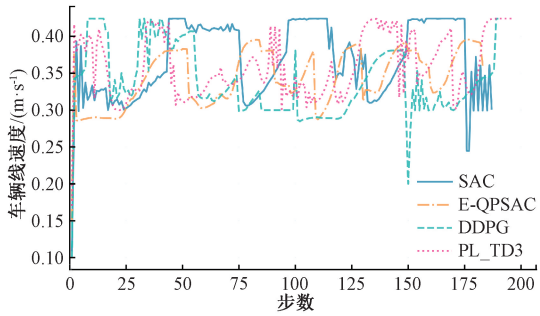


图 8 车辆行驶速度

Fig. 8 Vehicle driving speed

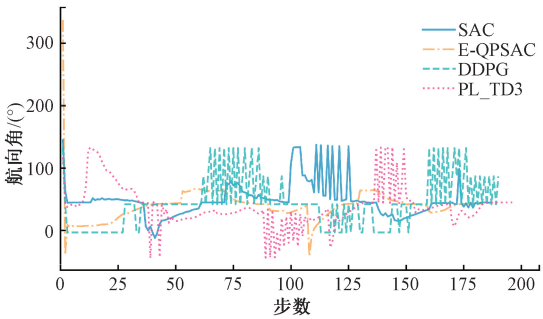


图 9 车辆行驶航向角

Fig. 9 Vehicle heading angle

由图 8 所示的线速度变化对比可知,在运动控制过程中,E-QPSAC 算法控制下的速度始终维持在合理区间内,整体曲线平稳;SAC 算法的速度曲线存在频繁波动,多次出现瞬时起伏,反映出其控制策略的不稳定性;DDPG 和 PL-TD3 算法在速度控制中虽整体趋势可控,但仍存在一定范围的周期性波动,平滑性不及 E-QPSAC,表明其速度控制策略不够连贯,存在较为明显的控制抖振问题。

由图 9 可见,在航向角控制方面,E-QPSAC 算法整体在较小角度区间内平缓波动,变化过程连续且波动范围小,偶发的角度突变幅度也明显小于其他算法;SAC 算法在整个过程中出现较多的剧烈跳变,平稳性较差;DDPG 和 PL-TD3 算法的航向角变化幅度较大,方向调整频繁,多次出现超过 100°的大幅波动,表明其航向角控制连续性最差,转向调整代价较高。

### 3.3 实车测试实验

为验证所提算法在实际环境中的有效性,实验所用三轮全向移动机器人如图 10 所示。该车辆由微控制器、驱动电路、惯性测量单元、带编码器的电机、电源模块、深度相机和 UWB 定位模块组成。其中 3 个车轮均为全向轮,以 120°夹角分布在车体周围。

将训练好的改进 E-QPSAC 网络模型部署到移动机器人,在实际环境中进行路径规划,规划路径如图 11 所示。由图 11 可知,TOMR 能在复杂环境中完成避障导航并成功到达预设终点,表明所提算法在实际环境中具有良好的适用性。

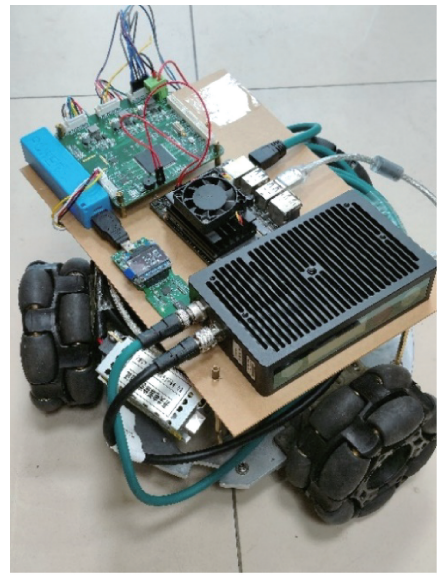

图 10 TOMR 实验装置

Fig. 10 TOMR experimental setup

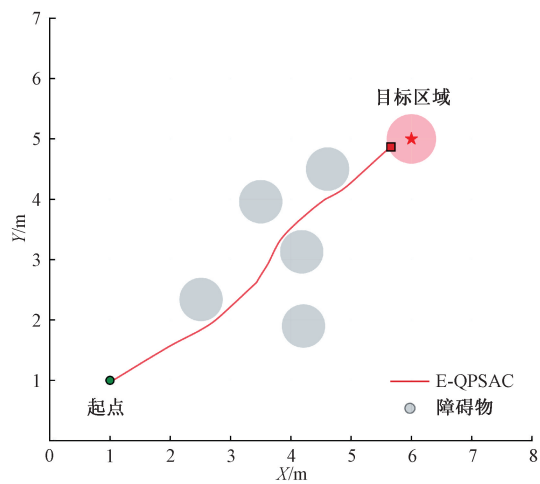


图 11 实际规划路径

Fig. 11 Actual planned path

进一步分析图 11 所示的规划轨迹可知,E-QPSAC 算法规划的路径整体平滑连贯,转向动作集中于障碍物附近,直行段占比高。从路径轨迹来看,TOMR 在绕避各障碍物过程中仅产生局部曲率,未出现大幅迂回或多余转折,体现出级联式 Actor 网络中障碍物感知模块与动态权重融合机制的协同作用;完成绕避后,目标导向模块权重迅速回升,策略及时收敛至最佳趋目方向,保证了路径整体的高效性。

## 4 结　论

本文针对三轮全向移动机器人在多障碍物不确定环境中的节能路径规划问题,提出了一种考虑能耗的融合模型预测控制参数化方法与改进软演员-评论家框架的 E-QPSAC 算法。该算法通过构建局部路径规划的二次规划问题,并采用原始-对偶混合梯度算法高效求解;设计了包含 12 维状态信息的状态空间与级联式网络架构,引入动态权重融合机制,有效平衡了避障与目标趋向之间的冲突;将五维能耗模型集成至奖励函数,从电机电阻损耗、力抵消、动能变化等多维度实现能耗的精细化建模与优化。最后与 SAC、DDPG、PL-TD3 算法进行对比实验,所提算法任务完成率达 98.7%(SAC 为 85.0%、DDPG 为 98.5%、PL-TD3 为 98.2%),总能耗低至 18.274 kJ(SAC 为 25.765 kJ、DDPG 为 22.418 kJ、PL-TD3 为 21.191 kJ),平均路径长度 6.45 m(SAC 为 6.56 m、DDPG 为 6.97 m、PL-TD3 为 7.13 m),在各项指标上均显著优于对比算法。在此基础上,进一步开展了实车部署实验,在实验室环境下验证了算法的路径规划能力,表明所提算法具备从仿真到实际的良好泛化能力与工程可部署性。下一步可开展在更加复杂环境或更多障碍物尤其是动态障碍物条件下的路径规划及能耗优化研究。

## 参考文献


[1] 单泽彪,魏昌斌,刘小松,等. 基于轮式里程计的无人车分层控制策略[J]. 控制理论与应用, 2025, 42(10): 2066-2074.
SHAN Z B, WEI CH B, LIU X S, et al. heeled odometer based hierarchical control strategy for unmanned vehicle[J]. Control Theory & Applications, 2025, 42(10): 2066-2074.

[2] 刘小松,朱焕海,单泽彪,等. 基于双曲正切视线制导的轮式移动机器人目标跟踪自抗扰控制[J]. 仪器仪表学报, 2024,45(12):201-209.
LIU X S, ZHU H H, SHAN Z B, et al. Active disturbance rejection control for target tracking of wheeled mobile robots based on hyperbolic tangent line-of-sight guidance[J]. Chinese Journal of Scientific Instrument, 2024, 45(12): 201-209.

[3] 刘小松,魏昌斌,单泽涛,等. 基于扩张状态观测器的里程计定位补偿无人车轨迹跟踪控制[J]. 仪器仪表学报, 2024,45(7):313-320.
LIU X S, WEI CH B, SHAN Z T, et al. Trajectory tracking control of unmanned vehicles with odometer positioning compensation based on extended state observer[J]. Chinese Journal of Scientific Instrument, 2024, 45(7): 313-320.

[4] 金智新,罗实,李永安,等. 基于改进 A* 的模糊 PID 煤矿巡检机器人路径规划[J]. 电子测量技术, 2026, 49(5): 40-51.
JIN ZH X, LUO SH, LI Y AN, et al. Improved A* path planning with fuzzy PID control for coal mine inspection robots[J]. Electronic Measurement Technology, 2026, 49(5): 40-51.

[5] 刘小松,康磊,单泽彪,等. 基于双向目标偏置 APF-informed-RRT* 算法的机械臂路径规划[J]. 电子测量与仪器学报, 2024, 38(6): 75-83.
LIU X S, KANG L, SHAN Z B, et al. Path planning of robot arm based on APF-informed-RRT* algorithm with bidirectional target bias[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(6): 75-83.

[6] CHU ZH ZH, WANG F L, LEI T J, et al. Path planning based on deep reinforcement learning for autonomous underwater vehicles under ocean current disturbance[J]. IEEE Transactions on Intelligent Vehicles, 2023, 8(1): 108-120.

[7] XUE B X, ZHOU F Y, WANG CH Q, et al. Robot mapless navigation in VUCA environments via deep reinforcement learning[J]. IEEE Transactions on Industrial Electronics, 2025,72(1): 639-649.

[8] LIN Y H, ZHANG ZH J, TAN Y J, et al. Efficient TD3 based path planning of mobile robot in dynamic environments using prioritized experience replay and LSTM[J]. Scientific Reports,2025,15(1):18331.

[9] HOU L F, ZHANG L, KIM J. Energy modeling and power measurement for mobile robots[J]. Energies, 2018, 12(1): 27-42.

[10] HOU L F, ZHANG L, KIM J W. Energy modeling and power measurement for three-wheeled omnidirectional mobile robots for path planning[J]. Electronics 2019, 8(8): 843.

[11] WANG J B, YANG Y M, LI J. A minimum-energy consumption control algorithm for omni-directional mobile robots[J]. Applied Mechanics and Materials, 2013, 291:2408-2411.

[12] KIM H J, KIM B K. Online minimum-energy trajectory planning and control on a straight-line path for three-wheeled omnidirectional mobile robots[J]. IEEE Transactions on Industrial Electronics, 2013, 61(9):

4771-4779.

[13] PAL A, TIWARI R, SHUKLA A. Multi-robot exploration in wireless environments [J]. Cognitive Computation, 2012, 4(4):526-542.

[14] LIU J, LI ZH J, HE L P, et al. Energy efficient path planning for indoor wheeled mobile robots [C]. Conference on Global Reliability and Prognostics and Health Management, 2020: 1-7.

[15] 王义娜, 郑依伦, 杨俊友, 等. 基于节能考虑的全向移动机器人鲁棒补偿轨迹跟踪控制[J]. 控制与决策, 2022,37(11):3065-3072.
WANG Y N, ZHENG, Y L, YANG J Y, et al. Robust compensated trajectory tracking control of omni-directional mobile robot based on energy saving [J]. Control and Decision,2022,37(11):3065-3072.

[16] HAARNOJA T, ZHOU A, ABBEEL P, et al. Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor [C]. International Conference on Machine Learning, 2018: 1861-1870.

## 作者简介

**单泽彪**,2016 年于吉林大学获得博士学位,现为长春理工大学副教授、硕士生导师。主要研究方向为无人系统自主控制,智能感知与机器人控制技术。

E-mail:zbshan@126.com

**Shan Zebiao** received his Ph. D. degree from Jilin University in 2016. He is currently an associate professor and graduate supervisor at Changchun University of Science and Technology. His main research interests include autonomous control of unmanned systems, intelligent perception and robot control technology.

**何立本**,2022 年于长春理工大学获得学士学位,现为长春理工大学硕士研究生。主要研究方向为无人系统自主控制、机器人控制与路径规划。

E-mail:1045530226@qq.com

**He Liben** received his B. Sc degree from Changchun University of Science and Technology in 2022. He is currently a master's student at Changchun University of Science and Technology. His main research interests include autonomous control of unmanned systems, robot control and path planning.

**刘小松**(通信作者),2016 年于吉林大学获得博士学位,现为长春理工大学副教授、硕士生导师。主要研究方向为智能感知与先进控制技术,复杂系统建模、仿真与控制。

E-mail:liuxs@cust.edu.cn

**Liu Xiaosong** (Corresponding author) received her Ph. D. degree from Jilin University in 2016. She is currently an associate professor and graduate supervisor at Changchun University of Science and Technology. Her main research interests include intelligent perception and advanced control technology, modeling, simulation and control of complex systems.

**梁法辉**,2016 年于吉林大学获得硕士学位,现为长春汽车职业技术大学副教授。主要研究方向为智能装备系统集成、机器人智能控制技术。

E-mail:caii_lfh@qq.com

**Liang Fahui** received his M. Sc. degree from Jilin University in 2016. He is currently an associate professor at Changchun Technical University of Automobile. His main research interests include intelligent equipment system integration and robotic intelligent control technology.