doi:10.6041/j.issn.1000-1298.2026.12.008

基于改进双 Q 学习线性自抗扰的水下机器人运动控制研究

周焕银[1]　刘凯伦[1]　刘金生[2]　张子民[1]　葛远香[1]

(1.东华理工大学电子与电气工程学院,南昌 330013;2.东华理工大学水资源与环境工程学院,南昌 330013)

摘要:在小型遥控水下机器人(Remotely operated vehicle, ROV)运动控制中存在模型较难描述、抗扰动能力差等问题,线性自抗扰控制器(Linear active disturbance rejection controller,LADRC)因对模型依赖程度低,对扰动有较好的观测补偿能力等,常被用于解决以上问题。而 LADRC 控制器在复杂环境中如参数固定会导致抗扰能力下降,因此提出一种基于改进双 Q 学习的 LADRC 参数自适应控制算法。采用一种新型分段式奖励函数指导参数调节方向,以加快调节速度,提高算法控制精度;引入双 Q 算法降低训练中单 Q 学习的过估计问题;根据系统运动状态实时调节 LADRC 参数,增强系统对外界扰动的观测补偿能力。Matlab/Simulink 转艏、定深试验结果表明:双 Q-LADRC 相较传统 LADRC 与 PID 控制器,阶跃扰动下最大超调量降低 1.54% 和 39.78%,稳定时间降低 38.2% 和 48.4%,在镇定控制上具有更快的收敛速度与更低的超调量,更小的跟踪误差。通过水池试验证明了双 Q-LADRC 算法的有效性,能较好跟踪预期信号。

关键词:水下机器人;镇定控制;双 Q 学习;线性自抗扰控制;参数自适应

中图分类号:TP242　　**文献标识码**:A　　**文章编号**:1000-1298(2026)12-0090-09　　**OSID**:

Motion Control of Underwater Robot Based on Improved Double Q-Learning and Linear Active Disturbance Rejection Control

ZHOU Huanyin[1]　LIU Kailun[1]　LIU Jinsheng[2]　ZHANG Zimin[1]　GE Yuanxiang[1]

(1. *School of Electronic and Electrical Engineering*, *East China University of Technology*, *Nanchang* 330013, *China*
2. *School of Water Resources and Environmental Engineering*, *East China University of Technology*, *Nanchang* 330013, *China*)

Abstract: Small remotely operated vehicles (ROVs) often suffer from difficulties in accurate model description and weak disturbance rejection in motion control. Although the linear active disturbance rejection controller (LADRC) has been widely used because of its low model dependence and strong disturbance estimation and compensation capability, its control performance may degrade in complex underwater environments when controller parameters remain fixed. To address this problem, an adaptive LADRC parameter tuning method based on an improved double Q-learning algorithm was proposed. A novel piecewise reward function was designed to guide the parameter adjustment process, thereby accelerating convergence and improving control accuracy. Meanwhile, the double Q-learning strategy was introduced to alleviate the overestimation problem in conventional Q-learning training. The proposed method enabled online adjustment of LADRC parameters according to the system state, which enhanced the disturbance observation and compensation capability of the controller. Matlab/Simulink simulations for heading and depth control showed that, compared with conventional LADRC and PID controllers, the proposed double Q-LADRC reduced the maximum overshoot under step disturbances by 1.54% and 39.78%, respectively, and shortened the settling time by 38.2% and 48.4%, respectively. In addition, it achieved faste convergence, low overshoot, and small tracking errors in stabilization control. Pool experiments further verified the effectiveness of the proposed method, demonstrating that it can track the desired signals well and maintain satisfactory control performance.

Key words: underwater robot; stabilization motion; double Q-learning; linear active disturbance rejection control; parameter adaptation

收稿日期:2025-12-28　修回日期:2026-02-08

基金项目:江西省科技厅重点基金项目(20224ACB204022)、国家自然科学基金项目(62063001)和江西省研究生省级创新基金项目(YC2024-S505)

作者简介:周焕银(1975—),女,教授,博士,主要从事水下机器人控制及应用研究,E-mail:hyzou@ecut.edu.cn

0　引言

传统水下机器人可分为水下遥控运载体(Remotely operated vehicle, ROV)与水下自主式无人运载体(Autonomous underwater vehicle, AUV)两类[1],ROV 通过电缆实现控制、数据接收和能源供应,具有易于携带作业相关传感器等优点,使 ROV 在水下信息采集、渔业养殖和浅海区生物资源探查等作业任务中得到广泛应用[2]。作业过程需要 ROV 保持如定深悬停,定向转角,定速巡航等动作,涉及水下机器人运动控制相关研究。水下机器人运动控制可分为轨迹跟踪控制与镇定控制两类[3],后者主要研究机器人的定点定深控制、姿态控制等。因此设计响应迅速,抗扰动能力更强的控制器对小型水下机器人的镇定运动控制具有现实意义和实用价值。

国内外研究者在水下机器人控制领域提供许多可行的方向。章盼梅[4]将模糊控制与传统 PID 控制相结合,极大提高了定深与转艏控制的速度稳定性并减少超调量,但模糊规则的建立需要大量专家经验及调试。郭东生等[5]设计了一种基于超前相位补偿的 PID 参数整定方法用于水下机器人定深控制,试验证明相较传统 PID 收敛速度更快,但仍然存在一定的超调量。刘学敏等[6]提出的 S 面控制算法应用于水下机器人运动控制,并考虑水下机器人作业环境及水声信号噪声问题,从模糊原理和 PID 控制器基础上推导出形式更简单鲁棒性较好的控制算法,但仍需要对参数进行复杂整定。JIANG 等[7]设计一种超扭滑模容错轨迹控制器,在超扭滑模控制基础上采用事件触发机制,降低推进器更新频率,提高系统整体控制精度,抗扰动效果强但存在抖振,且复杂控制率对小型低成本水下机器人的运算能力造成较大负担。黄健等[8]将韩京清提出的非线性自抗扰控制(Nonlinear active disturbance rejection control, NLADRC)[9]应用在欠驱动 AUV 控制中并设计非线性自抗扰控制器、线性自抗扰控制器[10]级联控制 3 种控制器对控制性能进行分析,针对一阶高频扰动抑制效果及估计效果做出不同评判,证明了算法的可行性,但 NLADRC 控制器调试参数过多,LADRC 在工程上虽易于实现,但仍需要对观测器及控制器带宽进行平衡兼顾抗扰、抗噪和跟踪精度。熊保星等[11]将模糊控制与线性自抗扰控制结合实现参数自适应整定,相较原算法在定深控制中有较好效果,但观测器带宽相对固定;WANG 等[12]利用优化算法实现线性自抗扰控制参数最优化。综合上述文献,本文将引用易于实现,抗扰动较强的 LADRC 控制,期望在水下机器人运用中对参数难改变造成的控制能力不足问题做出改善。

近年来针对以上传统控制器参数难以整定,或参数固定等问题的优化算法和智能算法逐渐被引入到机器人运动控制领域中[13]。CARLUCHO 等[14]提出一种增量式 Q 学习自适应调节 PID 参数,引入记忆算子有选择地定义状态变量,实现了陆地机器人智能巡航功能;在此基础上引入双 Q 学习机制进行 PID 参数调节,有效消除奖励高估问题,减轻局部最优带来的整体系统稳定性问题[15]。高强等[16]将 Q 学习与模糊控制结合设计一种复合控制器,对 LADRC 控制器参数进行自适应整定,在无人机姿态控制中精确度与准确度都有较好的改进。GU 等[17]将 Q 学习与 LADRC 结合应用于温度调控,控制效果得到显著提升。以上文献证明了 Q 学习对经典控制器参数自适应整定的可行性。

因此,本文设计一种结合改进的双 Q 学习算法与模糊控制器算法实现对线性自抗扰控制器参数自适应更新,增强抗干扰能力与控制效果。基于 Fluent 仿真得到水动力参数,并进行 LADRC 建模降低扩张状态观测器观测负担。基于改进的双 Q 学习算法,设计新型动作选择模型降低存量过程无效参数规模,设计新型分段式奖励函数,引导训练过程向更快更稳方向运行。经离线训练后得到综合 Q 值表用于参数更新,以期实现不同状态下 LADRC 参数的自动调整。最后基于 Matlab/Simulink 软件进行仿真试验,与传统 LADRC、PID 控制器进行对比分析。

1　水下机器人模型建立

1.1　水下机器人结构

水下机器人为自组装的六推进器水下机器人,主要由地面站,脐带缆线和 ROV 本体构成,ROV 本体如图 1 所示。

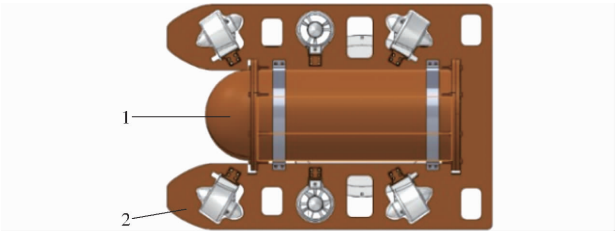

图 1　ROV 本体结构

Fig. 1　ROV body structure

1. 中间舱体　2. 外部推进器

水下机器人本体由中间舱体和外部推进器构成。中间舱体包含主控制器、深度计、陀螺仪和高清水下摄像头等。外部推进器包含推进器、照明灯以

及自行设计的采样器等。ROV 包含 4 个水平方向推进器与 2 个垂直方向推进器。推进器可由电缆供电也可由舱体内置的锂电池供电,可实现转速和转向控制。推进器分布如图 2 所示。

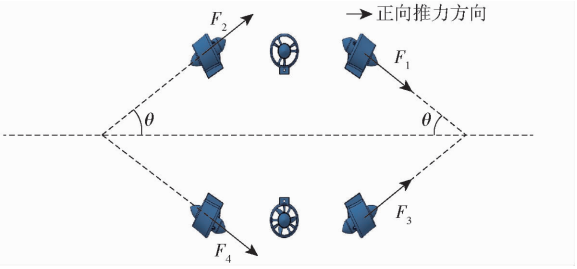

图 2 ROV 推进器分布

Fig. 2 ROV thruster layout

ROV 推进器呈环状包围主舱体,水平方向推进器与中心线呈 45°,6 个推进器配合实现纵荡、升沉、转艏动作。推进器作为系统的主要动力源,其精准模型是实现精准控制的重要依据。低成本小型机器人一般是低速运行,因此推进器推力计算式为

$$T = K_T \rho n^2 D^4 \tag{1}$$

式中 T——推进器产生的推力,N

K_T——螺旋桨推力系数,取 0.24

ρ——工作环境水体密度,kg/m^3

n——推进器螺旋桨转速,最高电压时为 140 r/s

D——螺旋桨直径,取 0.06 m

1.2 水下机器人模型

根据国际水池会议(ITTC)和造船与轮机工程学会(SNAME)推荐的坐标系,如图 3 所示建立 ROV 坐标系,惯性坐标系为 $E\xi\eta\zeta$,坐标原点可采用空间任意点,载体运动坐标系为 $OXYZ$,固定坐标原点取机器人运动起点,载体运动坐标原点与机器人重心点重合。

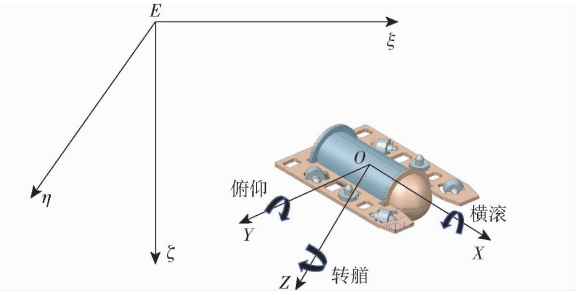

图 3 ROV 惯性坐标系和运动坐标系

Fig. 3 Inertial coordinate system and motion coordinate system of ROV

机器人运动相关参数符号及含义如表 1 所示。

ROV 数学模型为

$$\begin{cases} \boldsymbol{M}\dot{\boldsymbol{v}} + \boldsymbol{C}(\boldsymbol{v})\boldsymbol{v} + \boldsymbol{D}(\boldsymbol{v})\boldsymbol{v} + \boldsymbol{g}(\boldsymbol{\eta}) = \boldsymbol{\tau}_c + \boldsymbol{\tau}_d \\ \dot{\boldsymbol{\eta}} = \boldsymbol{J}(\boldsymbol{\eta})\boldsymbol{v} \end{cases} \tag{2}$$

表 1 运动学符号及定义

Tab. 1 Kinematic symbols and definitions

参数	含义	参数	含义
x/m	纵向位移	φ/(°)	横滚角
y/m	横向位移	θ/(°)	纵倾角
z/m	垂向位移	ψ/(°)	转艏角
u/(m·s^{-1})	纵向速度	p/((°)·s^{-1})	横滚角速度
v/(m·s^{-1})	横向速度	q/((°)·s^{-1})	俯仰角速度
w/(m·s^{-1})	垂向速度	r/((°)·s^{-1})	转艏角速度

式中 $\boldsymbol{v}$——速度向量

$\boldsymbol{M}$——包含附加质量的惯性矩阵

$\boldsymbol{C}(\boldsymbol{v})$——科氏及向心力矩阵

$\boldsymbol{D}(\boldsymbol{v})$——流体阻力矩阵

$\boldsymbol{g}(\boldsymbol{\eta})$——重力与浮力产生的恢复力向量

$\boldsymbol{\tau}_c$——推进器产生的推力矩阵

$\boldsymbol{\tau}_d$——外界干扰矩阵

$\boldsymbol{\eta}$——固定坐标系下位置以及姿态向量

$\boldsymbol{J}(\boldsymbol{\eta})$——载体运动坐标到惯性坐标系转换矩阵

1.3 水下机器人特点

自建 ROV 主要通过控制前进、横移、升沉、转艏 4 个自由度,即可控制机器人在整个空间运动[18]。机器人结构对称,通过配置浮力材料等可近似令浮心与重心重合。后文主要对系统升沉、转艏 2 个自由度进行仿真,因单一自由度运动过程耦合性强,镇定控制往往是对单一自由度进行控制,在对某一项进行控制时忽略弱耦合项以简化问题。因水平面与垂直面之间的耦合性较弱,将 ROV 运动解耦为水平面与垂直面 2 组实现对转艏运动和升沉运动控制。解耦后水平面运动模型为

$$\begin{cases} (m - X_{\dot u})\dot u = (m - Y_{\dot v})vr - Y_{\dot v} r^2 + \\ \qquad (X_u + X_{u|u|}|u|)u + \tau_X \\ (m - Y_{\dot v})\dot v = Y_{\dot r}\dot r - (m - X_{\dot u})ur + \\ \qquad (Y_v + Y_{v|v|}|v|)v + \tau_Y \\ (I_z - N_{\dot r})\dot r = N_{\dot v}\dot v - (X_{\dot u} - Y_{\dot v})uv + \\ \qquad (N_r + N_{r|r|}|r|)r + \tau_N \end{cases} \tag{3}$$

式中 m——机器人质量,kg

τ_i——推进器输出推力,N

X_i、Y_i、N_i——水动力系数

I_z——Z 轴转动惯量,kg·m^2

垂直面运动模型为

$$\begin{cases} (m - X_{\dot u})\dot u = -(m - Z_{\dot w})wq + Z_{\dot q} q^2 + \\ \qquad (X_u + X_{u|u|}|u|)u + \tau_X \\ (m - Z_{\dot w})\dot w = Z_{\dot q}\dot q + (m - X_{\dot u})uq + \\ \qquad (Z_w + Z_{w|w|}|w|)w + \tau_Z \\ (I_y - M_{\dot q})\dot q = M_{\dot w}\dot w + (X_{\dot u} - Z_{\dot w})uw + \\ \qquad (M_q + M_{q|q|}|q|)q + \tau_M \end{cases} \tag{4}$$

式中　Z_i、M_i——水动力系数

I_y——Y 轴转动惯量，kg·m^2

2　基于改进双 Q 学习的 LADRC 参数自适应算法

针对 LADRC 参数固定问题设计一种将双 Q 学习算法和模糊控制结合实现 LADRC 参数的自适调整。根据已有信息构建水下机器人 LADRC 控制器，再对双 Q 学习进行改进，划分状态目标，设计分段式奖励函数，优化动作选取策略，加强学习收敛性，实现对 LADRC 参数的增量式优化，算法结构如图 4 所示。

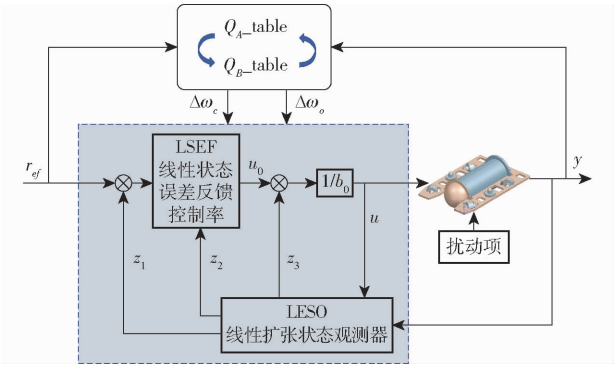

图 4　改进算法结构

Fig. 4　Improved algorithm structure

2.1　水下机器人线性自抗扰控制器

非线性自抗扰控制器源于韩京清研究员对传统 PID 控制算法的思考[9]，最大优势在于对精确模型的依赖性低，将内外部扰动定义为总扰动，通过主动误差补偿方式极大提升控制鲁棒性，同时规避 PID 中积分环节的饱和副作用。然而因其包含大量非线性函数，导致需整定的参数过多，难以在工程中实现[19]。GAO[10] 将带宽引入该控制器，不仅降低了参数整定难度，物理意义也更加清晰。依据其结构构建水下机器人 LADRC 控制器，结构如图 5 所示。

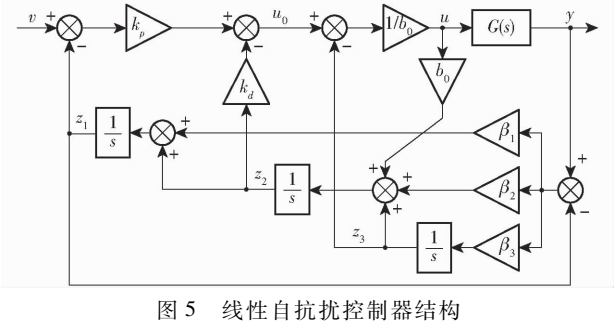

图 5　线性自抗扰控制器结构

Fig. 5　Structure of linear active disturbance rejection controller

LADRC 主要由信号微分跟踪（Tracking differentiator，TD），线性扩张状态观测器（Linear extended state observer，LESO）以及线性状态误差反馈控制率（Linear state error feedback，LSEF）3 部分构成。

首先构造线性扩张状态观测器，是算法的核心部分，对总体扰动做出观测并实时补偿。以深度控制为例，由式（4）可知，升沉运动动力学模型为二阶系统。根据已知部分信息构建深度控制简易模型，降低估计干扰的难度，计算式为

$$\ddot{z} = bu_z - Z_w\dot{z} - f(z,\dot{z},t) \tag{5}$$

式中　b——控制输出增益

u_z——控制输出

$f(z,\dot{z},t)$——包括外部扰动在内总扰动

构建线性状态观测器为

$$\begin{cases} \dot{\hat{x}}_1 = \hat{x}_2 + \beta_1(y - \hat{x}_1) \\ \dot{\hat{x}}_2 = \hat{x}_3 + b_0(u_z - Z_w\hat{x}_2) + \beta_2(y - \hat{x}_1) \\ \dot{\hat{x}}_3 = \beta_3(y - \hat{x}_1) \end{cases} \tag{6}$$

式中　b_0——升沉运动方向总质量

y——系统状态值

$\hat{x}_i$——状态观测值

β_i——观测器待整定参数

待整定参数合适的取值可以使状态观测器快速精准预估扰动[19]。根据式（6）得到观测器状态误差特征多项式 $D(s)$ 为

$$D(s) = s^3 + (\beta_1 + b_0Z_w)s^2 + (\beta_2 + \beta_1 b_0 Z_w)s + \beta_3 \tag{7}$$

要使观测器稳定，需将极点配置在坐标系左半平面，最终得到待整定参数计算式为

$$\begin{cases} \beta_1 = 3\omega_o - b_0Z_w \\ \beta_2 = 3\omega_o^2 - 3\omega_o b_0 Z_w + (b_0Z_w)^2 \\ \beta_3 = \omega_o^3 \end{cases} \tag{8}$$

式中　ω_o——观测器带宽

经补偿后，令

$$\ddot{z} = u_0 \tag{9}$$

式中　u_0——中间控制量

线性反馈控制率设计，类似 PD 控制器，计算式为

$$u_0 = k_p(r_{ef} - \hat{x}_1) - k_d\hat{x}_2 \tag{10}$$

式中　k_p——比例增益系数

k_d——微分增益系数

r_{ef}——系统预期值

为保持系统稳定，配置系统模型极点在坐标系左平面可得

$$\begin{cases} k_p = \omega_c^2 \\ k_d = 2\omega_c \end{cases} \tag{11}$$

式中　ω_c——控制器带宽

系统实际输出 u 为

$$u=\frac{1}{b_0}(u_0-\hat{x}_3) \tag{12}$$

由式(8)~(12)可知,系统主要需要调整的参数是控制器带宽 ω_c 和观测器带宽 ω_o。因 LADRC 定义扰动为可导且变化缓慢,因此当出现高频率扰动时,固定观测器带宽使控制器效果受限,因此采用双 Q 学习算法对 2 个参数进行增量式调整,以提升控制效果。

2.2 改进双 Q 学习算法

Q 学习(Q-Learning)作为一种无模型、基于值函数的经典强化学习算法,通过更新状态动作价值网络或者表格以达到学习最优策略的目的[20],该算法以马尔科夫决策过程作为数学系框架,如图 6 所示。

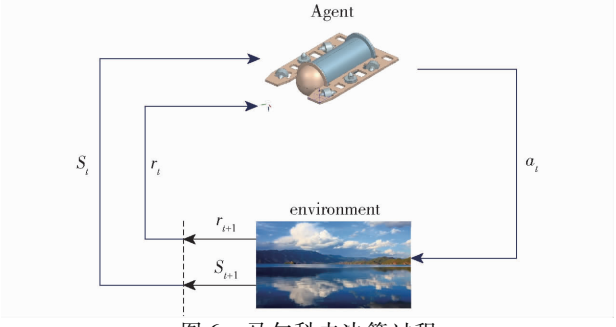

图 6 马尔科夫决策过程

Fig. 6 Markov policy process

该过程可概述为智能体当前所处状态为 S_t,根据当前策略 $\pi(a|s)$ 在动作集中获取随机动作,通过状态转移函数得到下一时刻状态 S_{t+1},伴随状态转移通过奖励函数获取当前动作的即时奖励,通过大量交互学习到最优策略 $\pi^*(a|s)$,并实现最大化累计折扣奖励,计算式为

$$Q^*(s,a)=\sum_{s'}P(s'|s,a)\cdot(R(s,a,s')+\gamma\max_{a'}Q^*(s',a')) \tag{13}$$

式中 s——当前状态

a——选择的动作

s'——下一状态

a'——下一状态的候选动作

$Q^*(s,a)$——最优动作价值函数

$\sum_{s'}P(s'|s,a)$——状态转移函数

$R(s,a,s')$——状态转移获取的奖励

γ——折扣率

$\max_{a'}Q^*(s',a')$——最优未来回报

γ 通常取 0~1,衡量即时奖励和长期回报的重要性。式(13)同时也解释了马尔科夫性质:下一状态 s' 和当前奖励 R 仅依赖于当前状态 s 和动作 a,与历史状态无关。因此单 Q 学习迭代过程为

$$Q(s,a)\leftarrow Q(s,a)+\alpha[r_{ew}+\gamma\max_{a'}Q(s',a')-Q(s,a)] \tag{14}$$

式中 r_{ew}——即时奖励 α——学习率

双 Q 学习算法与前者不同的地方在于引入 2 个独立的 Q 函数 Q_A 和 Q_B,每次随机更新其中 1 个 Q 表时该表作为动作选择,而另 1 个 Q 表则未改动作提供下一状态的动作价值[21]。

假设此时选择更新 Q_A 表,则选取下一状态动作价值最大动作对应为 a^*,对此时状态动作价值更新为

$$Q_A(s,a)\leftarrow Q_A(s,a)+\alpha[r_{ew}+\gamma Q_B(s',a_A^*)-Q_A(s,a)] \tag{15}$$

其中

$$a_A^*=\arg\max_{a'}Q_A(s',a') \tag{16}$$

式中 $\arg\max_{a'}Q_A(s',a')$——最大值所对应的动作选择算子

同理可得 Q_B 表更新式为

$$Q_B(s,a)\leftarrow Q_B(s,a)+\alpha[r_{ew}+\gamma Q_A(s',a_B^*)-Q_B(s,a)] \tag{17}$$

其中

$$a_B^*=\arg\max_{a'}Q_B(s',a') \tag{18}$$

双 Q 算法相较单 Q 算法,可避免使用同 1 个 Q 函数进行动作选择和奖励值计算而导致的状态动作价值被系统过高估计的问题,更适合水下这种高噪声或者环境变化较为随机的场景,可有效规避水下环境中复杂扰动对动作的影响。

2.3 探索策略

传统 Q 学习中动作策略常以 ε-greedy 策略代替贪婪策略,ε 值与探索概率成正比,动作选择策略通过调节 ε 值维持探索与利用的平衡,但 ε-greedy 策略非最优即随机的特点,对 Q 值较小的动作重复选择会产生大量无效数据,因此本文结合动作空间设计一种 ε-greedy 与玻尔兹曼混合动作选择策略,计算式为

$$P_{\text{total}}(a|s)=\alpha_0P_{\varepsilon\text{-greedy}}(a|s)+(1-\alpha_0)P_{\text{Boltzmann}}(a|s) \tag{19}$$

其中

$$P_{\varepsilon\text{-greedy}}(a|s)=\begin{cases}1-\varepsilon+\dfrac{\varepsilon}{|A(s)|} & \left(a=\arg\max\limits_{a'\in A(s)}Q(s,a')\right)\\ \dfrac{\varepsilon}{|A(s)|} & \left(a\neq\arg\max\limits_{a'\in A(s)}Q(s,a')\right)\end{cases} \tag{20}$$

$$P_{\text{Boltzmann}}(a|s)=\frac{e^{\frac{Q(s,a)}{\tau}}}{\sum\limits_{a'}e^{\frac{Q(s,a)}{\tau}}} \tag{21}$$

式中 ε——随机探索概率

$A(s)$——状态 s 下动作集合

α_0——权重

τ——调节参数

式(20)为 ε-greedy 策略；式(21)为玻尔兹曼动作选择策略，该策略基于 softmax 函数特性，可根据 Q 值赋予不同动作被抽取概率，τ 决定概率分布的平缓度，值越小 Q 值高的动作被抽取概率越大。

2.4　状态量与动作空间定义

强化学习状态量定义为输出误差 e 及导数 $\dot{e}$ 的组合，引用模糊控制论域定义，将 e 及 $\dot{e}$ 分为{负大(NB)，负中(NM)，负小(NS)，零(ZO)，正小(PS)，正中(PM)，正大(PB)}7 个等级，因此将状态划分 49 个，相较文献[14]极大减少了状态量数量，可加快收敛速度，状态量定义如表 2 所示。

表 2　状态量定义

Tab.2　Definition of state variables

$\dot{e}$	e						
	NB	NM	NS	ZO	PS	PM	PB
NB	1	8	15	22	29	36	43
NM	2	9	16	23	30	37	44
NS	3	10	17	24	31	38	45
ZO	4	11	18	25	32	39	46
PS	5	12	19	26	33	40	47
PM	6	13	20	27	34	41	48
PB	7	14	21	28	35	42	49

本文动作定义为控制器带宽 ω_c 和观测器带宽 ω_o 增量组合。动作选取以 2 种动作的增量形式，这样能保持稳定性防止参数变化过大导致输出突变问题。9 个动作如表 3 所示。

表 3　动作定义

Tab.3　Action definition

序号	ω_c 变化趋势	ω_o 变化趋势
1	减少	大幅减少
2	小幅减少	小幅减少
3	保持不变	保持不变
4	小幅增加	小幅增加
5	增加	大幅增加
6	减少	小幅减少
7	大幅减少	增加
8	大幅增加	减少
9	增加	小幅减少

2.5　奖励函数

在水下机器人运动控制中，奖励函数作为状态对动作的及时评判，直接决定了 LADRC 动作的选择效率、整体控制合理性以及收敛性。本文设计一种分段式奖励，将水下机器人运动过程分为上升阶段和稳定阶段 2 个部分，奖励函数 r_t 计算式为

$$r_t=\begin{cases}r_{\text{up}} & (e_t>\lambda)\\ r_{\text{stable}} & (e_t\leqslant\lambda)\end{cases} \tag{22}$$

式中　λ——阶段切换阈值

e_t——当前误差

r_{up}——上升阶段奖励函数

r_{stable}——稳定阶段奖励函数

取预期状态值的 5%，高于该阈值时注重快速性与安全性，低于阈值注重减少超调量与振荡。

2.5.1　上升阶段奖励函数

上升阶段奖励指导智能体快速接近系统稳态区间，注重系统响应快速性，同时对系统输出做出一定限制，防止推进器因长时间满负荷推进或快速反向转动造成机械损伤，因此该阶段奖励函数由 a_1、a_2、a_3 3 部分组成，计算式为

$$r_{\text{up}}=a_1\left[-\left(1-\mathrm{e}^{-\frac{e_t^2}{2\sigma_1^2}}\right)\right]+a_2(-|u_t|^2)+a_3(-\text{Ⅱ}(u_tu_{t-1}<0)) \tag{23}$$

式中　u_t——当前控制输出

u_{t-1}——上一时刻控制输出

Ⅱ(·)——指示函数，前后输出方向相反条件成立

a_1、a_2、a_3——权重系数

σ_1——上升阶段正态分布标准差

通过调节权重系数 a_1 改变误差所占奖励的权重，指导智能体更快到达预期，同理 a_2、a_3 分别表示控制量奖励权重与反向转动惩罚权重。σ_1 表明误差灵敏度越小惩罚值增长越快。a_1 为正态分布误差惩罚，当误差项趋近于 0 时惩罚项最小，惩罚值趋近于零，当误差增大时惩罚项逐渐趋近于 1，最终惩罚值为 $-a_1$；a_2 为输出惩罚值，抑制推进器在水中长时间满负荷运动，避免系统过载；a_3 惩罚控制量反向行为，如电机正转快速到反转将触发惩罚，可减少机器人机械损耗和系统振荡。

2.5.2　稳定阶段奖励函数

稳定阶段奖励引导智能体到达预期状态同时减少超调量与系统抖振问题，因此在调整误差灵敏度 σ_2 的同时，引入平滑性奖励函数，可有效限制过高超调量与变化量，计算式为

$$r_{\text{stable}}=b_1\left[-\left(1-\mathrm{e}^{-\frac{e_t^2}{2\sigma_2^2}}\right)\right]+b_2(-|e_t-e_{t-1}|) \tag{24}$$

式中　b_1、b_2——权重系数

此阶段误差远小于上升段误差，因此需调节参数使误差惩罚增长速度适应稳定阶段，b_2 为平滑性奖励项的权重系数，当前状态与上一状态误差值相差为零时该项惩罚最小，当变化量逐渐增大，对波动的惩罚值逐渐增大。

本文算法输入为系统误差及变化率，系统通过相关传感器采集位置与角度等相关信息，用作双 Q 表的训练。实际控制过程中，每次获取的信息与预期信号差值 e 及变化率通过模糊化后组合为双 Q 学习的状态，训练过程通过奖励函数指导动作选择，训练的双 Q 模型进行输出选择，输出动作为 LADRC 算法的系数变化值。

3 试验与结果分析

3.1 仿真试验

双 Q－LADRC 算法使用 Matlab/Simulink 进行仿真测试。建立水下机器人运动学与动力学模型，对输出作出限制，水平方向最大推力为 80 N，深度控制最大推力为 50 N。以转艏控制试验为例，验证 LADRC 参数对控制效果影响，确定调节参数范围。然后进行转艏连续阶跃信号跟踪仿真，为验证双 Q－LADRC算法抗扰动能力，在连续阶跃信号中加入白噪声和脉冲信号模拟水流扰动；为验证深度控制信号跟踪能力进行方波信号跟踪试验，仿真时加入高斯白噪声模拟水流扰动。为验证双 QLADRC 算法的控制效果，与 PID、参数固定式 LADRC 进行抗扰动能力、信号响应性能对比。

3.1.1 转艏角度控制仿真试验

以控制定艏 50°为例，为验证控制器带宽对控制效果影响，对转艏运动阶跃信号仿真，在第 2 秒输入幅值 50°阶跃信号；为平衡观测器带宽变化带来的误差补偿能力和干扰放大缺点，仿真时在第 8 秒加入幅值为 8 N 阶跃扰动信号，仿真结果如图 7 所示。

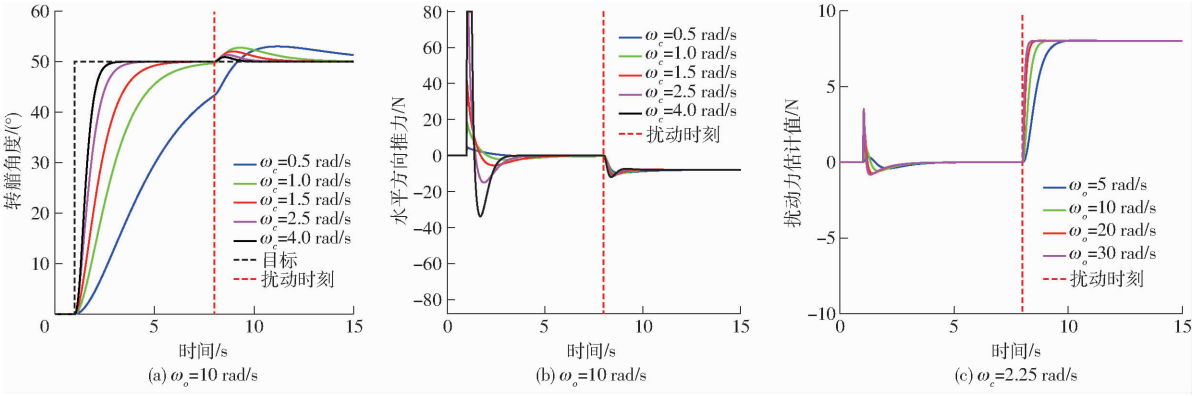

图 7 ROV 转艏控制 LARC 仿真结果

Fig. 7 Simulation results of LARC control for ROV turning bow

由图 7a 可知，当观测器带宽固定为 10 rad/s，控制器带宽越大系统响应越快，当控制器带宽低于 0.5 rad/s 响应过慢；超过 4 rad/s 时，由图 7b 可知，系统输出的水平推力持续，低于 0.5 rad/s 时响应速度过慢；由图 7c 可知，控制器带宽为 2.25 rad/s，观测器带宽一般取其 3 ~ 5 倍，当观测器带宽超过 20 rad/s 后，对扰动过于敏感，综上 ω_c 取 0.5 ~ 4 rad/s，观测器带宽取 2 ~ 20 rad/s。

控制参数取值范围确定后进行转艏控制仿真，并将系统控制目标设置为幅值 50°连续阶跃信号。为模拟实际水流环境，试验过程加入方差为 0.01 白噪声；为模拟水下真实突发情况，验证系统抗扰动能力，在第 10 秒时加入幅值为 10 N、时间为 2 s 脉冲扰动信号。图 8a 为系统转艏角度仿真结果，图 8b 为系统实际输出力仿真结果，仿真结果如表 4 所示。

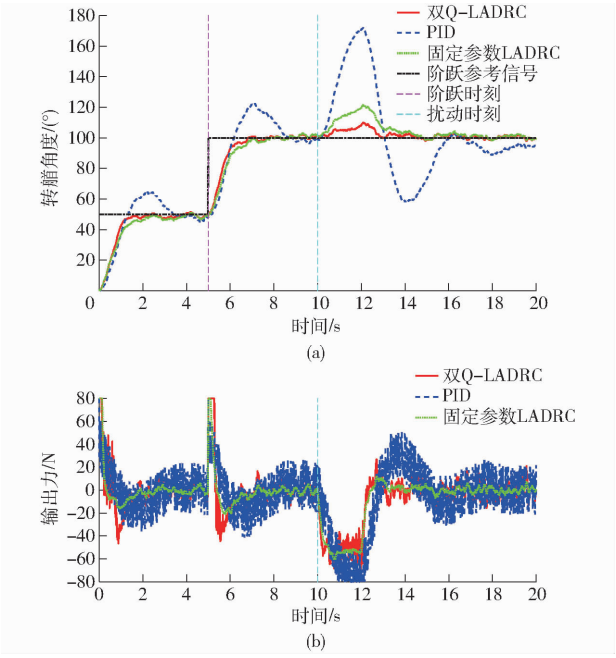

图 8 转艏角度控制及输出力变化曲线

Fig. 8 Yaw angle control and output force variation curves

由图 8a 可知，双 Q－LADRC 在超调量、稳定时间、扰动恢复速度上最优，抗扰性最强，固定 LADRC 稳定时间长，自适应能力不足，扰动恢复速度一般。由表 4 可知，双 Q－LADRC 相较传统 LADRC 与 PID 控制器，阶跃扰动下最大超调量降低 1.54% 和 39.78%，稳定时间降低 38.2% 和 48.4%。由图 8b 可知，双 Q－LADRC 控制器在整体控制过程对干扰

反应更加灵敏，输入转艏角度信号变化时输出变化更快，结合图 8a 第 5 秒时上升阶段以及表 4 可知，系统输出力相较其他控制器更高，稳定阶段对干扰抑制效果更好，但输出相较固定参数 LADRC 存在一定波动。

表 4　仿真结果

Tab. 4　Simulation results

控制器	最大超调量/%	稳定时间/s	平均稳态误差/(°)
双 Q－LADRC	5.05	2.84	2.245 4
LADRC	6.59	4.60	5.210 0
PID	44.83	5.21	21.677 5

3.1.2　镇定深度控制仿真试验

镇定控制是对深度的控制，为仿真 ROV 作业中对目标跟踪效果，系统输入深度为 2 m、周期为 10 s 方波信号。为模仿水下真实干扰，加入方差为 0.02 白噪声干扰信号，仿真试验结果如图 9 所示。

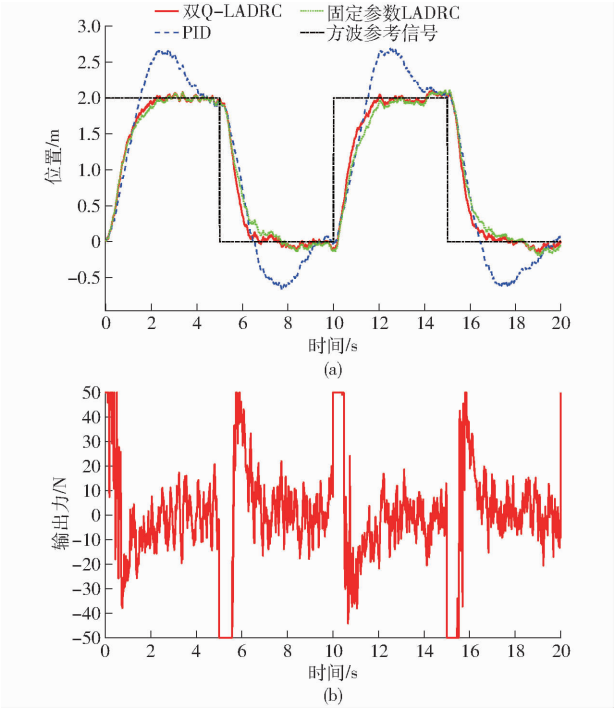

图 9　深度控制及输出力变化曲线

Fig. 9　Depth control and output force variation curves

由图 9a 可知，双 Q－LADRC 平均稳态误差为 0.333 4 m，相较 PID 控制器平均稳态误差 0.557 3 m，固定参数 LADRC 平均稳态误差 0.401 5 m，分别降低 40% 和 23%。由图 9b 可知，双 Q－LADRC 针对白噪声扰动时，反应更快，能更好地对误差进行补偿。综上双 Q－LADRC 在深度轨迹跟踪中稳态误差方差最优，跟踪延迟最低。

3.2　实际试验

试验环境为室内水池，深度约 2 m。测试平台为小型六自由度 ROV，包括地面站、电缆绳与 ROV 本体，ROV 本体控制器树莓派 4B 采用 Broadcom BCM2711 四核 ARM Cortex－A72 架构处理器，主频 1.5 GHz，运行内存 8 GB，采样时间 10 ms 可满足实时性要求。离线训练为 Q 表的迭代更新过程，核心操作为状态离散化、动作选择、Q 表迭代，计算复杂度为 $O(s \times a \times t)$。本方法中 $s=49$、$a=9$，为极小的状态/动作空间，即训练步数 $T=10^4$，总迭代次数仅为 $49 \times 9 \times 10^4 = 4.41 \times 10^6$，在普通上位机（Intel i5 及以上）中训练耗时小于 5 min，满足工程化离线训练需求。试验平台及试验环境如图 10 所示。

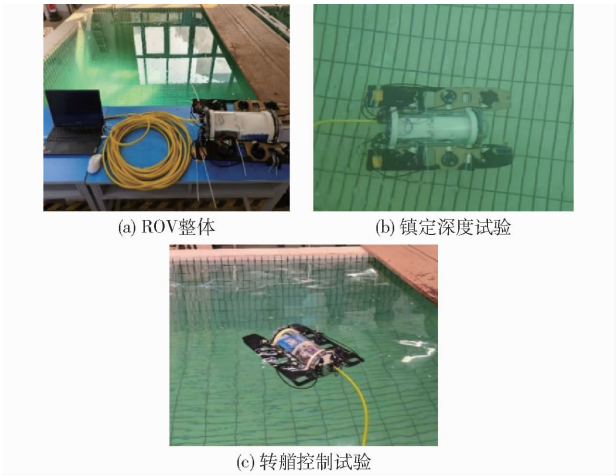

图 10　试验环境及试验现场

Fig. 10　Test environment and test site

为验证双 Q－LADRC 算法对镇定深度控制的可行性，将试验分为 2 组。第 1 组试验首先将机器人定深在水下 0.2 m 位置并保持悬停一段时间，随后上升至水平面并保持。第 2 组试验定深至 1.6 m 并保持深度稳定（图 10b）。试验结果如图 11 所示。

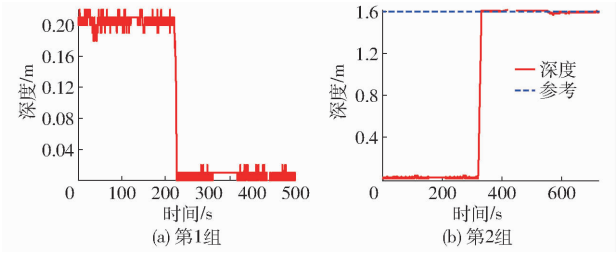

图 11　镇定深度试验结果

Fig. 11　Experimental results of pool depth

由图 11a 可知，在水下 0.2 m 悬停时 ROV 可保持良好的稳定，由于传感器噪声影响深度在 0.19 ~ 0.21 m 波动，在 270 s 时发送上升阶跃信号，3 s 后到达水平面并保持悬停状态，验证该算法可保持系统稳定。由图 11b 可知，水平面保持 5 min 后定深至 1.6 m 保持悬停，发送信号 12 s 后 ROV 到达指定位置悬停，其间深度变化绝对值小于 0.05 m，能够以较小超调量，较快速度到达稳定状态，证明双 Q－

LADRC 算法具有可行性,以较快速度,较小超调量完成定深任务。

为验证双 Q - LADRC 算法对转艏控制的可行性,ROV 按照一定转艏轨迹运动(图 10c)。试验结果如图 12 所示,初始状态由于水池较小,以及传感器存在噪声缘故,转艏角度在一定范围波动,绝对值小于 2°,ROV 水中静止 10 s 逆时针转动 160°,并保持 5 s,该过程平滑完成并未出现超调量,系统 3 s 完成调节并保持稳定;再顺时针回转试验过程发现缆绳本身材质以及缆绳在机体分布位置对回正过程造成一定影响,回正过程存在 18°左右超调量,但 5 s 后系统仍然实现稳定。

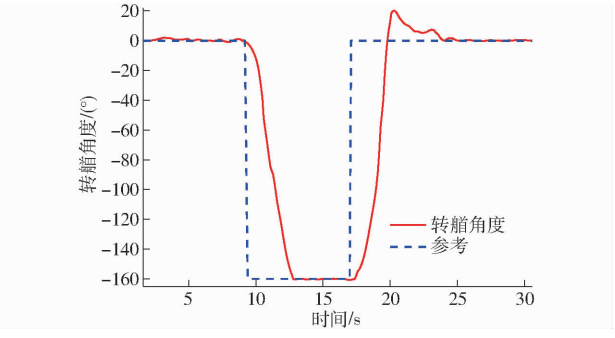

图 12 转艏试验结果

Fig. 12 Yawing experiment results of water tank

通过镇定深度及转艏角度控制试验,证明双 Q - LADRC 在实际任务中,能够有效完成目标任务,对扰动有较好的反应能力。

4 结论

(1)针对 ROV 镇定运动抗扰动能力差问题设计一种抗扰动性更强,控制效果更精准的控制器。将改进双 Q 算法用于调节 LADRC 参数信息,通过设计 Q 学习动作选择策略减少训练过程无效数据,加快收敛速;设计新型奖励函数对系统进行分段指导,误差大加快速度,误差小调节控制精度,减少超调,实现更快稳定、更小超调、更小误差的控制效果。

(2)通过 Matlab 仿真试验证明了双 Q - LADRC 控制器相较传统 PID 和 LADRC 控制器在无扰动下跟踪响应更快、超调量更小;加入随机扰动下,抗干扰方面效果更优,跟踪精度更高,验证了本文对 Q 学习算法设计的奖励函数与动作策略的可行性;对参数难以整定和控制器参数难改变带来的控制能力下降问题提供了一种解决方法。实际定深和转艏试验验证了本文算法的稳定性,超调量小,较快速度完成任务并保持稳定。

参 考 文 献

[1] 李硕,吴园涛,李琛,等. 水下机器人应用及展望[J]. 中国科学院院刊,2022,37(7):910 - 920.
LI Shuo, WU Yuantao, LI Chen, et al. Application and prospect of unmanned underwater vehicle[J]. Bulletin of Chinese Academy of Sciences, 2022, 37(7): 910 - 920. (in Chinese)

[2] AHN J, YOKOTA H, YASUKAWA S. Development of small-size jellyfish removal ROV and its evaluation of removal motion performance in tank[J]. Journal of Robotics Networking and Artificial Life, 2022, 9(1): 66 - 71.

[3] 边宇枢,高志慧,负超. 6 自由度水下机器人动力学分析与运动控制[J]. 机械工程学报,2007,43(7):87 - 92.
BIAN Yushu, GAO Zhihui, YUN Chao. Dynamic analysis and motion control of 6-dof underwater robot[J]. Chinese Journal of Mechanical Engineering, 2007, 43(7): 87 - 92. (in Chinese)

[4] 章盼梅. 模糊 PID 算法的水下机器人控制系统优化[J]. 机床与液压,2024,52(3):72 - 77.
ZHANG Panmei. Optimization of underwater robot control system based on fuzzy PID algorithm[J]. Machine Tool & Hydraulics, 2024, 52(3): 72 - 77. (in Chinese)

[5] 郭东生, 沈洋林, 张曦元,等. 一种基于改进 PID 的水下机器人悬停控制方法[J]. 西北工业大学学报, 2025, 43(4): 685 - 693.
GUO Dongsheng, SHEN Yanglin, ZHANG Xiyuan, et al. An underwater robot hovering control method based on improved PID algorithm[J]. Journal of Northwestern Polytechnical University, 2025, 43(4): 685 - 693. (in Chinese)

[6] 刘学敏,徐玉如. 水下机器人运动的 S 面控制方法[J]. 海洋工程,2001,19(3):81 - 84.
LIU Xuemin, XU Yuru. S control of automatic underwater vehicles[J]. The Ocean Engineering, 2001, 19(3): 81 - 84. (in Chinese)

[7] JIANG Y, ZHANG Q, HU Y. Fault-tolerant trajectory tracking control of underwater salvage robots based on super-twisting sliding mode[J]. Ocean Engineering, 2024, 312(P3): 119340.

[8] 黄健,舒晓葯. 基于线性 ADRC 的欠驱动 UUV 深度控制器设计[J]. 控制工程,2019,26(5):927 - 934.
HUANG Jian, SHU Xiaodi. Under-actuated UUV depth controller design based on linear ADRC technology[J]. Control Engineering of China, 2019, 26(5): 927 - 934. (in Chinese)

[9] 韩京清. 自抗扰控制器及其应用[J]. 控制与决策,1998(1):19 - 23.

[10] GAO Z Q. Scaling and bandwidth-parameterization based controller tuning[C]// Denver, CO, USA: IEEE, 2003: 4989 - 4996.

(下转第 110 页)

parameter-pre-tuned super-twisting sliding mode control[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025,56(2):136 - 144. (in Chinese)
[23] 胡子谦. 基于北斗定位的联合收获机自主导航技术研究[D]. 武汉:华中农业大学,2022.
HU Ziqian. Research on autonomous navigation technology of combine harvester based on Beidou positioning[D]. Wuhan: Huazhong Agricultural University,2022. (in Chinese)
[24] 张月新. 智能收获机的非线性运动控制方法研究[D]. 南京:东南大学,2021.
ZHANG Yuexin. Research on nonlinear motion control method for intelligent harvester[D]. Nanjing: Southeastern University, 2021. (in Chinese)
[25] 贾伟健, 刘西侠, 王可, 等. 基于 MPC 的无人履带车辆轨迹跟踪方法研究[J]. 计算机仿真,2024,41(5):429 - 435.
JIA Weijian, LIU Xixia, WANG Ke, et al. Research on trajectory tracking method of unmanned tracked vehicle based on MPC [J]. Computer Simulation,2024,41(5):429 - 435. (in Chinese)
[26] 吴洋,王聪,董国新,等. 考虑滑移的无人履带车自适应路径跟踪控制策略[J]. 机械工程学报,2024,60(24):211 - 225.
WU Yang, WANG Cong, DONG Guoxin, et al. Slip-aware adaptive trajectory tracking control strategy for autonomous tracked vehicle[J]. Journal of Mechanical Engineering, 2024,60(24):211 - 225. (in Chinese)
[27] WANG X, WANG Y, SUN Q, et al. Adaptive robust control of unmanned tracked vehicles for trajectory tracking based on constraint modeling and analysis[J]. Nonlinear Dynamics,2024,112: 9117 - 9135.
[28] 王法安, 杨全合, 张兆国, 等. 基于 ICR 的履带车辆路径跟踪与转向控制算法研究[J]. 农业机械学报,2024,55(1): 386 - 395,425.
WANG Faan, YANG Quanhe, ZHANG Zhaoguo, et al. Path tracking and turning control algorithm of tracked vehicle based on ICR[J]. Transactions of the Chinese Society for Agricultural Machinery,2024,55(1):386 - 395,425. (in Chinese)
[29] 李睿, 项昌乐, 王超, 等. 自动驾驶履带车辆鲁棒自适应轨迹跟踪控制方法[J]. 兵工学报,2021,42(6):1128 - 1137.
LI Rui, XIANG Changle, WANG Chao, et al. Robust adaptive trajectory tracking control approach for autonomous tracked vehicles[J]. Acta Armamentarii,2021,42(6):1128 - 1137. (in Chinese)
[30] 刘辉, 王振辉, 陈洪月, 等. 基于虚拟前轮反馈控制算法的履带车轨迹跟踪方案研究[J]. 机械设计,2024,41(1): 117 - 124.
LIU Hui, WANG Zhenhui, CHEN Hongyue, et al. Research on trajectory-tracking scheme for tracked vehicles based on virtual front-wheel feedback control algorithm[J]. Journal of Machine Design,2024,41(1):117 - 124. (in Chinese)
[31] 王辉, 王桂民, 罗锡文, 等. 基于预瞄追踪模型的拖拉机导航路径跟踪控制方法[J]. 农业工程学报,2019,35(4):11 - 19.
WANG Hui, WANG Guimin, LUO Xiwen, et al. Path tracking control method of agricultural machine navigation based on aiming pursuit model[J]. Transactions of the CSAE,2019,35(4):11 - 19. (in Chinese)

(上接第 98 页)

[11] 熊保星,甘文洋,陈铭治,等. 基于模糊线性自抗扰的水下机器人定深控制[J]. 控制工程,2026,33(1):161 - 169.
XIONG Baoxing, GAN Wenyang, CHEN Mingzhi, et al. Depth control of underwater vehicles based on fuzzy linear active disturbance rejection[J]. Control Engineering of China, 2026, 33(1): 161 - 169. (in Chinese)
[12] WANG T, NIE Y, WANG S, et al. Depth control of ROV using the improved LADRC based on nutcracker optimization algorithm[J]. Ocean Engineering,2024, 309(1):118370.
[13] BENAZOUZ R, CHEMORI A, CREUZE V. A novel intelligent RISE feedback control of autonomous tethered underwater vehicles: design & real-time experiments[J]. Ocean Engineering,2024,300(2):117470.
[14] CARLUCHO I,PAULA D M,VILLAR S, et al. Incremental Q-learning strategy for adaptive PID control of mobile robots[J]. Expert Systems With Applications,2017,80:183 - 199.
[15] CARLUCHO I, PAULA D M, ACOSTA G G. Double Q - PID algorithm for mobile robot control[J]. Expert Systems With Applications,2019,137:292 - 307.
[16] 高强,李旭,吉月辉,等. 基于 Q 学习的高超声速飞行器自抗扰控制研究[J]. 控制工程,2024,31(4):577 - 582.
GAO Qiang, LI Xu, JI Yuehui, et al. Research on active disturbance rejection control of hypersonic vehicle based on Q-learning[J]. Control Engineering of China, 2024, 31(4): 577 - 582. (in Chinese)
[17] GU Yanping, ZHANG Hao, XU Tao, et al. Temperature regulation of optical machine frame based onreinforcement learning active disturbance rejection control[J/OL]. Journal of Southeast University(English Edition),1 - 13[2025 - 12 - 08]. https://link.cnki.net/urlid/32.1325.N.20251112.1452.002.
[18] HASIB A S, GULZAR M M, OISHY R S, et al. An investigation of innovative strategies in underwater soft robotics[J]. Engineering Science and Technology, an International Journal,2025,70:102123.
[19] 袁东,马晓军,曾庆含,等. 二阶系统线性自抗扰控制器频带特性与参数配置研究[J]. 控制理论与应用,2013,30(12): 1630 - 1640.
YUAN Dong, MA Xiaojun, ZENG Qinghan, et al. Research on frequency-band characteristics and parameters configuration of linear active disturbance rejection control for second-order systems[J]. Control Theory & Applications,2013,30(12):1630 - 1640. (in Chinese)
[20] GANDHI J, NARMAWALA Z. Q-Learning and double Q-Learning based routing in opportunistic networks[J]. International Journal of Networked and Distributed Computing,2026,14(1):13.
[21] 周奇,柳丹,席磊,等. 基于超松弛双 Q 学习的源荷储协同频率稳定算法研究[J]. 中国电机工程学报,2024,44(4):1419 - 1429.
ZHOU Qi, LIU Dan, XI Lei, et al. Research on source load storage cooperative frequency stabilization algorithm based on super relaxed double Q learning[J]. Proceedings of the CSEE,2024, 44(4): 1419 - 1429. (in Chinese)