基于深度强化学习的车道保持策略研究

杨新光[1,2]　郭孔辉[1]

（1.福建理工大学新能源车辆运动控制研究院，福州 350118；2.福建理工大学电子电气与物理学院，福州 350118）

【摘要】针对传统车道保持方法依赖高精度模型且计算负荷大的局限性，提出一种基于深度确定性策略梯度（DDPG）算法的车道保持辅助策略，融合了车辆动力学模型与深度强化学习（DRL）框架，并设计了一种考虑任务完成度和横向误差的奖励函数以优化模型训练收敛速度和跟踪精度，从而在不同弯道场景下实现更为精准和高效的车道保持性能。仿真验证结果表明，相较于传统的模型预测控制（MPC）方法，所提出方法的横向误差绝对值均值较MPC降低了42%~72%。

关键词：车道保持　深度强化学习　深度确定性策略梯度算法

中图分类号：U463；TP273　　**文献标志码：**A　　**DOI：** 10.20104/j.cnki.1674-6546.20250253

Research on Lane Keeping Strategy Based on Deep Reinforcement Learning

Yang Xinguang[1,2], Guo Konghui[1]

(1. New Energy Vehicle Motion Control Research Institute, Fujian University of Technology, Fuzhou 350118; 2. School of Electronic, Electrical Engineering and Physics, Fujian University of Technology, Fuzhou 350118)

【Abstract】 To address the limitations of traditional lane-keeping methods that rely on high-fidelity models and significant computational burden, this paper proposes a lane-keeping assist strategy based on the Deep Deterministic Policy Gradient (DDPG) algorithm. The proposed system integrates vehicle dynamics models with a Deep Reinforcement Learning (DRL) framework. Furthermore, a specialized reward function considering both task completion and lateral error is designed to optimize training convergence speed and tracking accuracy, thereby achieving more precise and efficient performance across various curve scenarios. Simulation experiments demonstrate that the proposed strategy yields significantly smaller lateral errors compared to the traditional Model Predictive Control (MPC) method. Specifically, the Mean Absolute Lateral Error (MALE) is reduced by 42%~72% compared to MPC, highlighting the potential of DRL in lane-keeping tasks.

Key words: Lane keeping, Deep Reinforcement Learning (DRL), Deep Deterministic Policy Gradient (DDPG) algorithm

【引用格式】 杨新光, 郭孔辉. 基于深度强化学习的车道保持策略研究[J]. 汽车工程师, 2026(5): 26-33.

YANG X G, GUO K H. Research on Lane Keeping Strategy Based on Deep Reinforcement Learning[J]. Automotive Engineer, 2026(5): 26-33.

1　前言

车道保持系统（Lane Keeping System，LKS）通过实时感知车辆位姿与道路边界信息实现车辆行驶轨迹的自动纠偏与稳定性控制[1]。高效可靠的横向控制策略不仅能有效弥补驾驶员因疲劳、分心导致的感知与操作缺失，显著减轻驾驶负荷，更能从根本上规避因非预期车道偏离引发的交通事故[2]。

目前，车道保持的横向控制方法主要分为传统控制方法和基于学习的方法[3]。传统的控制方法主要包括基于规则的比例-积分-微分（Proportional-Integral-Derivative，PID）控制器[4-6]，以及基于模型的优化控制策略，如线性二次型调节器（Linear Quadratic Regulator，LQR）[7-8]、模型预测控制（Model

Predictive Control，MPC）等。文献[9]指出，相较于PID控制器，MPC控制器在执行车道保持时能够使车辆与车道中心间的横向误差更小、行驶更加稳定。文献[10]表明，MPC虽然能够有效处理约束和预测未来状态，但其在线优化计算量大，难以在计算资源受限的汽车电子控制单元上实时运行。文献[11]指出，基于模型的方法需要建立精确的车辆动力学模型，然而在实际工程应用中，难以兼顾模型精度与计算效率。

基于强化学习的自动驾驶算法可以基于输入的状态空间变量直接输出决策，省去复杂的中间过程，实现端到端驾驶[12]，并通过车辆自主学习不断优化驾驶策略，提高对各种道路场景的适应能力，且不需要人工标注数据集。近年来，研究人员不断探索将强化学习应用于车道保持系统的横向控制，取得了显著成果。文献[13]指出，深度Q网络（Deep Q-Network，DQN）在车道保持任务中表现出色，误差较小且训练时间较短。文献[14]提出了一种结合深度Q网络与优先经验回放（Prioritized Experience Replay，PER）的安全驾驶策略，通过设计包含碰撞惩罚与安全距离约束的多维奖励函数有效解决了传统强化学习算法对高风险样本关注不足的问题。文献[15]研究了深度确定性策略梯度（Deep Deterministic Policy Gradient，DDPG）算法在车道保持任务中的可行性，展示了深度强化学习（Deep Reinforcement Learning，DRL）在自动驾驶车道保持辅助系统中的潜力，但仅针对单一道路场景进行了测试，模型在不同场景下的适应性尚未得到验证。

本文基于深度强化学习DDPG算法对不同场景下车道保持任务中的横向控制问题进行研究。首先，将车道保持过程建模为深度强化学习过程，从而通过智能体的自主学习适应并优化其在不同道路环境中的表现。然后，将任务完成度作为奖励函数的关键组成部分，以全面评估智能体在车道保持任务中的表现，进一步优化模型训练过程中的收敛速度和智能体在不同工况下的精准度和鲁棒性。最后，利用MATLAB/Simulink平台构建基于DDPG算法的车道保持仿真模型，并在不同弯道场景下验证模型的精准度和鲁棒性。

2　深度确定性策略梯度算法

深度强化学习[16]利用深度神经网络来近似强化学习中的价值函数或策略函数，从而在高维度状态和动作空间中进行决策。

强化学习的基本架构通常包括智能体、环境、状态、动作、奖励和策略[17]。在该框架中，智能体通过观测环境的状态，依据其内部策略选择动作并执行，从而使环境转移到新的状态，同时，环境向智能体反馈奖励信号，并对其动作效果进行即时评价。智能体的目标是学习一项策略，以在较长的时间跨度内实现累积奖励的最大化，其核心任务在于通过持续学习过程逐步优化从状态到动作的映射关系，以确保在连续的时间步长中能够高效地累积奖励。

在深度强化学习中，深度神经网络用于估计策略或值函数，解决了传统强化学习方法在处理高维连续状态空间时面临的维度灾难问题[18]。

DDPG算法是一种通过迭代优化来获取最优策略的方法，它适用于处理连续动作空间问题。连续动作空间意味着智能体需要在无穷多的可能动作中作出选择。DDPG算法通过演员（Actor）网络输出确定性的动作，并结合评论家（Critic）网络进行价值评估，从而有效地在连续动作空间中寻找最优策略。如图1所示，演员-评论家框架由Actor和Critic 2个神经网络组成，Actor网络根据环境提供的状态观测量S预测出决策动作A，环境据此动作生成奖励R并更新状态。Critic网络基于当前状态S和动作A估计奖励值Q，用于更新Actor网络的策略参数θ^{τ}，而Critic网络参数θ^{Q}则通过奖励函数进行调整。通过反复迭代实现Actor网络优化，以输出最佳动作。

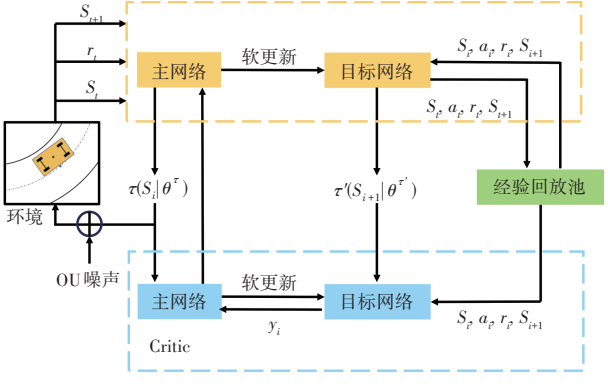


图1　DDPG算法结构

为提升训练稳定性，DDPG算法引入了目标网络与经验回放机制。目标网络作为原始网络的延迟复制，通过软更新策略逐步调整参数，有效缓解了因网络快速更新导致的训练振荡问题。经验回放机制通过存储历史经验并随机采样，降低了数据间的相关性，从而提高了训练效率。此外，算法采

用奥恩斯坦-乌伦贝克（Ornstein-Uhlenbeck，OU）噪声增强探索能力，将确定性动作决策转化为随机过程$\tau'(S)$，通过采样生成动作值。这种设计不仅提升了算法在复杂环境中的探索多样性，还避免了陷入局部最优解的风险。

结合DDPG算法在连续动作空间中处理、训练稳定性高的优势，本文基于DDPG算法研究车道保持任务中的横向控制问题，通过不断学习策略网络输出的前轮转角，使智能体能够实时调整车辆的行驶轨迹，以实现精准的车道保持。

3 车道保持控制模型

3.1 车辆动力学模型

本文主要研究车道保持任务中的横向问题，因此采用二自由度车辆动力学模型，同时假设在车道保持过程中车速不变，忽略悬架作用，且车辆为前轮驱动。

可以推导出以下二自由度车辆动力学方程：

$$\begin{bmatrix}\dot{v}_y\\ \ddot{\varphi}\end{bmatrix}=\begin{bmatrix}-\dfrac{C_f+C_r}{mv_x} & \dfrac{bC_r-aC_f}{mv_x}-v_x\\ \dfrac{bC_r-aC_f}{Iv_x} & -\dfrac{a^2C_f+b^2C_r}{Iv_x}\end{bmatrix}\begin{bmatrix}v_y\\ \dot{\varphi}\end{bmatrix}+\begin{bmatrix}\dfrac{C_f}{m}\\ \dfrac{aC_f}{I}\end{bmatrix}\delta \quad (1)$$

式中：v_x为纵向车速，v_y为横向车速，φ为车辆横摆角，I为车辆的转动惯量，a、b分别为质心到前、后轴的距离，C_f、C_r分别为前、后轮的侧偏刚度，m为车身质量，δ为车辆前轮转角。

3.2 车道保持模型

基于图2所示的车道保持模型，车辆外置传感设备实时获取左、右车道线的曲率ρ_l、ρ_r和车辆与左、右车道线的距离y_l、y_r，以及车辆当前航向与左、右车道线切线方向的夹角ε_l、ε_r。

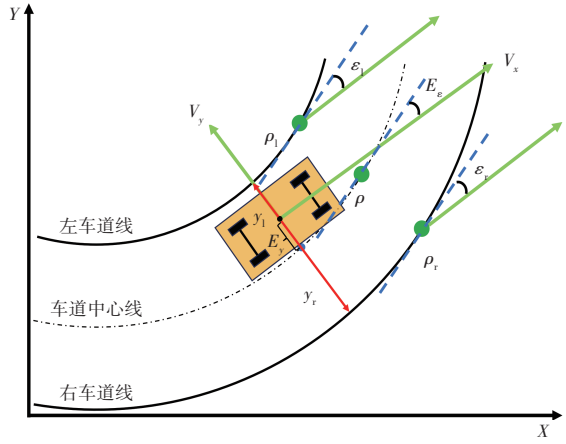


图2 车道保持模型

基于上述信息进一步计算出车道中心线的曲率ρ、车辆与车道中心线的横向误差E_y和航向角误差E_ε：

$$\rho=(\rho_l+\rho_r)/2 \quad (2)$$

$$E_y=(y_l-y_r)/2 \quad (3)$$

$$E_\varepsilon=(\varepsilon_l+\varepsilon_r)/2 \quad (4)$$

在车道保持控制系统中，E_y和E_ε是衡量跟踪性能的核心状态变量：E_y表征车辆质心相对于车道中心线的侧向偏离程度，直接反映了当前的路径跟踪精度；E_ε表征车辆行驶方向与道路切线方向的夹角，它不仅反映了车辆偏离车道的趋势，还可能影响车辆的横向加速度和乘坐舒适性。故需综合考虑E_y和E_ε，以确保车辆能够稳定地保持在车道内。

3.3 马尔可夫决策过程建模

3.3.1 状态空间

在控制车辆进行车道保持时，不仅需要考虑本车的行驶状态，还需要考虑车道线与车身的相对状态，为了使模型易于收敛，状态量的选取不宜过多，本文选取的状态空间为：

$$S=\{E_y,E_\varepsilon,\delta,\rho\} \quad (5)$$

式中：δ为车辆的前轮转角。

3.3.2 动作空间

本文主要研究一般行驶速度下车辆的横向运动，主要通过控制前轮转角实现车道保持功能，因此动作空间A_t定义为车辆的前轮转角δ：

$$A_t=\{\delta\},\ \delta\in[-0.1,0.1]\ \text{rad} \quad (6)$$

3.3.3 奖励函数

奖励函数旨在引导智能体通过与环境交互，逐步逼近并收敛至最优策略。为实现车辆在车道保持任务中的精准控制与平稳驾驶，本文构建了包含精准性奖励、稳定性奖励及任务完成度奖励的综合奖励函数。

针对车道保持精准性，分别设计横向误差奖励R_l和航向误差奖励R_ε。其中R_l的表达式为：

$$R_l=-\omega_l E_y^2+H_l \quad (7)$$

式中：$\omega_l=2$为横向误差奖励权重，H_l为分段式离散奖励项。

分段式离散奖励项用于进一步引导车辆向车道中心线收敛并保持高精度跟踪：

$$H_l=\begin{cases}5, |E_y|\le 0.02\ \text{m}\\ 3, 0.02\ \text{m}<|E_y|<0.05\ \text{m}\\ -1, |E_y|>0.3\ \text{m}\end{cases} \quad (8)$$

航向误差奖励R_ε的表达式为：

$$R_\varepsilon=-\omega_2 E_\varepsilon^2+H_\varepsilon \tag{9}$$

式中：ω_2=14为航向误差奖励权重，H_ε为分段奖励。

分段奖励用于鼓励车辆保持行驶方向与车道方向一致：

$$H_\varepsilon=\begin{cases}5, |E_\varepsilon|\leqslant 0.02\ \text{rad}\\ 3, 0.02\ \text{rad}<|E_\varepsilon|<0.05\ \text{rad}\\ -1, |E_\varepsilon|>0.3\ \text{rad}\end{cases} \tag{10}$$

考虑到车道保持控制的稳定性，应确保前轮转角尽可能小且变化平稳，因此基于前轮转角设计舒适性奖励函数R_c：

$$R_c=-\omega_3\delta^2+H_c \tag{11}$$

式中：ω_3=10为稳定奖励权重，H_c为正值奖励。

正值奖励用于限制控制器的输出幅度，抑制高频振荡，引导智能体学习平滑的控制策略：

$$H_c=10, |\delta|<0.04\ \text{rad} \tag{12}$$

在仿真训练过程中，为确保车辆行驶的安全性和准确性，本文采用基于横向误差与航向角误差的终止条件。横向误差能够直接反映车辆相对于车道中心线的偏移程度，航向角误差则反映行驶方向与车道线切线方向的偏差。当E_y>0.35 m或E_ε>1 rad时，系统判定当前状态为“失败”并触发终止条件。此时，当前训练回合立即结束，环境状态重置，随即进入下一回合的训练。

为全面评估车辆在单次训练回合中的表现，设计任务完成度奖励G，并在触发终止条件时进行结算。该奖励值与车辆的行驶时长呈正相关，车辆在车道内行驶的时间越长，表征其任务完成度越高。这一机制旨在防止智能体因过度追求单步高奖励而导致控制策略激进，引导其关注长周期的任务存续性，从而在保证控制精度的同时提升整体任务质量。

任务完成度奖励G的表达式为：

$$G=-1\ 000/(1+t) \tag{13}$$

式中：t为车辆的运动时间。

因此，综合奖励函数表达式为：

$$R=R_1+R_\varepsilon+R_c+G \tag{14}$$

3.4 网络设计和超参数选取

Actor网络如图3所示，由多个层级组成，先使用输入层接收观测信息，并连接3层全连接层（Fully Connected Layer，FC Layer），每层全连接层都连接线性修正单元（Rectified Linear Unit，ReLU）激活函数以提取特征。输出层后连接Tanh激活函数，将输出限制在[−1,1]范围内，最后通过缩放层调整动作输出的比例。

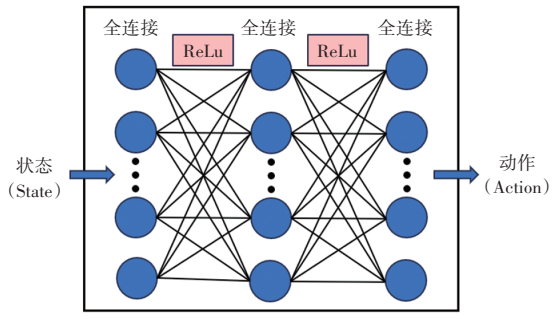


图3 Actor网络

如图4所示，Critic网络的主路径开始于一个特征输入层，用于接收观测信息，并使用2层全连接层和ReLU激活函数逐步提取特征，通过加法层整合其他输入信息，最后经过全连接层输出动作价值Q。

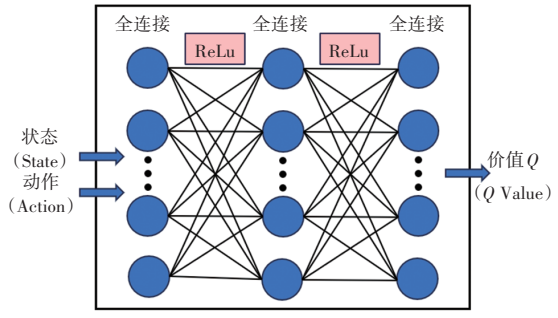


图4 Critic网络

网络中的超参数设置如表1所示，部分超参数的设置参考文献[19]~文献[21]，并多次对比试验确定。

表1 训练超参数

参数	取值	参数	取值
目标网络折扣因子	0.99	经验池大小	1 000 000
Actor学习率	0.000 1	最大训练回合数	5 000
Critic学习率	0.001	噪声方差	0.05
批量大小	64	噪声衰减率	0.000 1

4 验证与分析

为了验证所提出的基于深度强化学习的车道保持决策控制模型的有效性，本文利用MATLAB/Simulink仿真平台对智能车辆车道保持控制过程进行仿真分析。

4.1 仿真环境设置

强化学习仿真训练时采用的车辆模型参数如表2所示，根据车道保持场景参数定义搭建仿真场景，在训练过程中被控车辆的初始位置固定为(0 m, 0 m)，设定每个训练回合的最大仿真时长T_f=25 s、采样周期T_s=0.1 s。

表2　智能车参数

参数	取值
质心离前轴的距离 a/m	1.2
质心离后轴的距离 b/m	1.6
整车质量 m/kg	1 575
绕 z 轴转动惯量 I_z/kg·m^2	2 875
前轮胎侧偏刚度 C_{af}/N·rad^{-1}	19 000
后轮胎侧偏刚度 C_{ar}/N·rad^{-1}	33 000

如图5所示，训练道路设计为直线道路和曲线道路组合的转弯路径，其中包含直道、左弯道、右弯道，两端直道长度分别为100 m，弯道长度为450 m，宽度为3.6 m，道路总长550 m。

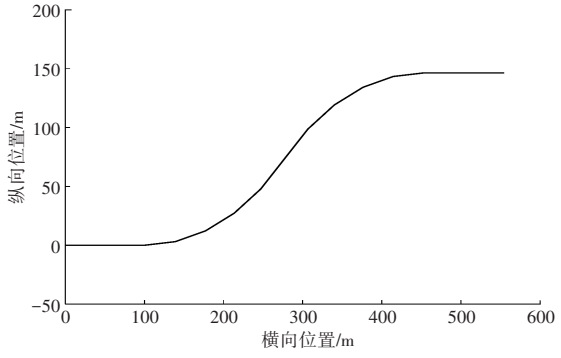


图5　训练道路

4.2　模型训练结果

在DDPG算法的训练过程中，平均奖励值是衡量智能体学习进展与策略探索成效的关键指标。图6展示了整个训练周期内平均奖励值的动态变化过程。

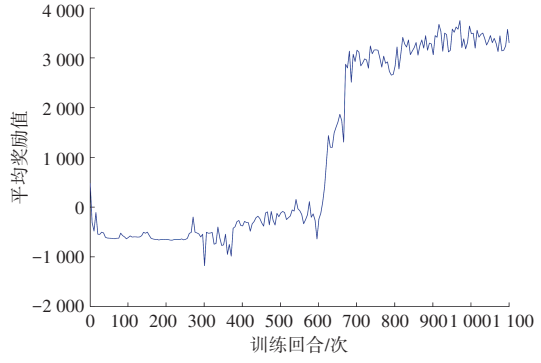


图6　平均奖励值

在训练的前300回合，奖励值维持在−500左右。这一阶段，智能体处于广泛的探索阶段，由于频繁触发终止条件，导致每回合奖励值较低且波动幅度有限。在第300回合和第590回合左右，奖励值出现了明显的波动。这一现象归因于DDPG算法中引入的随机噪声策略，该策略旨在帮助智能体跳出局部最优解，进一步拓宽探索范围。在第600回合左右，奖励值显著上升，智能体逐步发现更为有效的动作策略。这一转变体现了智能体通过不断试错与策略调整，探索到了性能更优的解决方案。最终在约第700回合时达到了逐步收敛状态，此时，智能体已探索到一套相对优越的策略，并在此基础上继续深化探索，使得最终奖励值稳定在3 000左右。

选取训练过程中奖励值较高的智能体策略作为车道保持控制策略，利用MATLAB/Simulink搭建控制模型仿真系统，建立对应的车道保持场景，分别采用DDPG算法和MPC算法对车道保持控制模型进行仿真验证。

车道保持的精准性通过观察车辆与车道中心线的横向误差来评估。图7所示为2种算法的横向误差对比情况。可以看出，在整体表现上，DDPG算法的横向误差较MPC算法小，这表明在车道保持的精准性方面，DDPG算法更具优势。

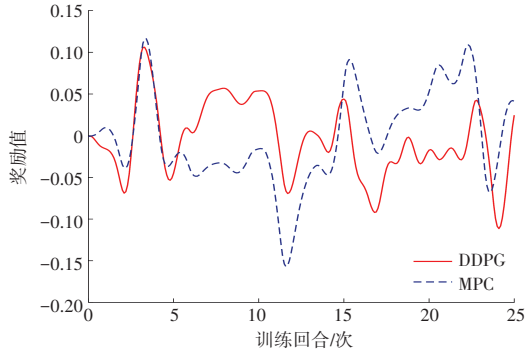
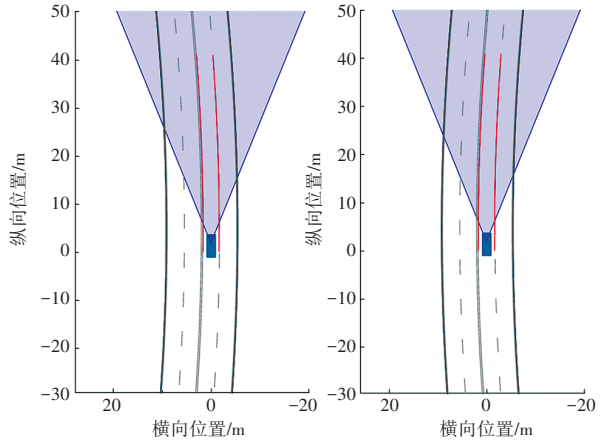


图7　训练道路横向误差

4.3　结果验证

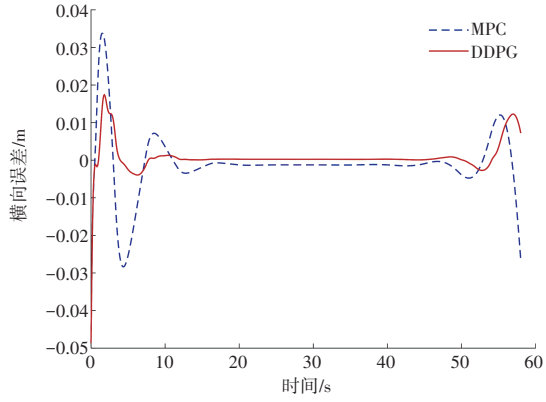
为了深入评估并验证本文所提出的车道保持控制模型在不同曲率弯道条件下的适应性和控制效能，设计了2组非固定曲率的测试场景，如图8所示。



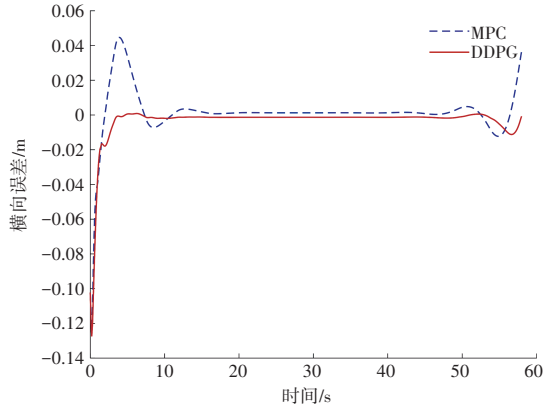
(a)左向弯道场景　(b)右向弯道场景

图8　非固定曲率测试场景

采用已训练成熟的车道保持控制策略在上述2个测试弯道上进行了车道保持精准性评估，不同车速下本文模型与MPC算法的路径跟踪精度如图9~图11所示，DDPG算法展现出优于MPC算法的误差收敛特性。在30 km/h低速工况下，DDPG的横向误差曲线呈现出更快速的收敛特性且波动峰值更小，这种差异在右弯道场景中更为显著，DDPG在5 s内就将横向误差稳定在±0.002 m，而MPC需要15 s完成收敛；表3数据显示，DDPG的横向误差绝对值均值相较于MPC减小了62%和42%，验证了其快速收敛特性与稳态精度的双重优势。当车速提升至60 km/h时，二者的横向误差明显增大，但DDPG的误差波动较小，且收敛速度依然较快，尽管与30 km/h时相比横向误差有所增加，但其横向误差绝对值均值依然比MPC减小了72%和68%。在90 km/h高速工况下，两种算法的横向误差进一步增大，但DDPG依然展现出更强的稳定性。表3数据显示，DDPG的横向误差绝对值均值较MPC降低了66.72%和63.27%。相比之下，MPC在高速场景下表现出明显的局限性：误差波动较大、收敛速度较慢，且在稳态阶段的误差波动范围显著大于DDPG。

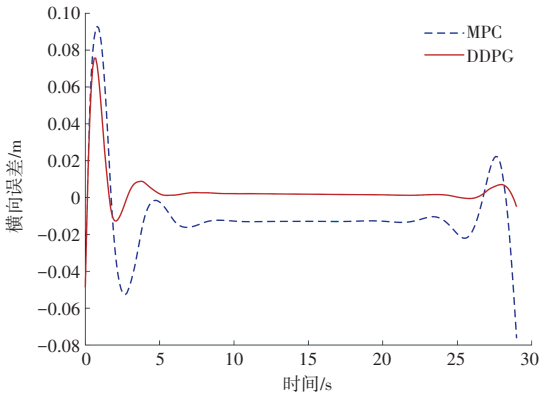


(a)左向弯道场景

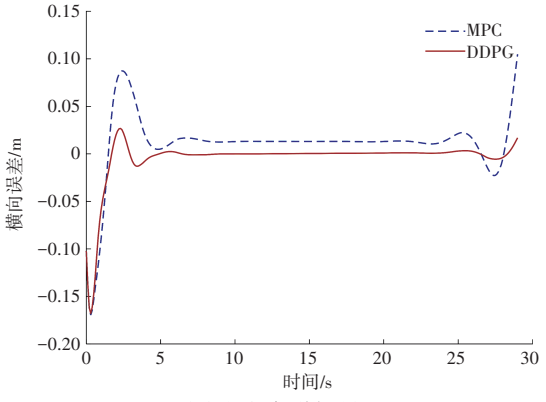


(b)右向弯道场景

图9　30 km/h车速条件下非固定曲率测试场景横向误差

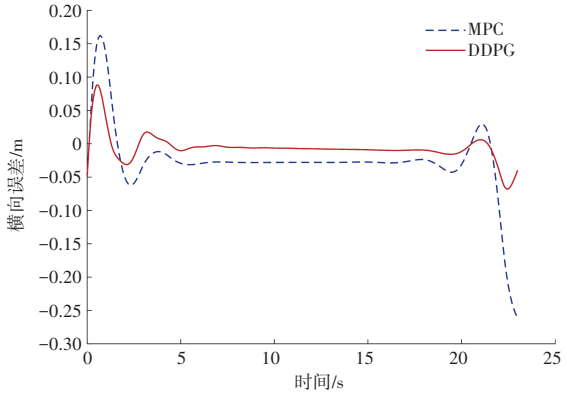


(a)左向弯道场景

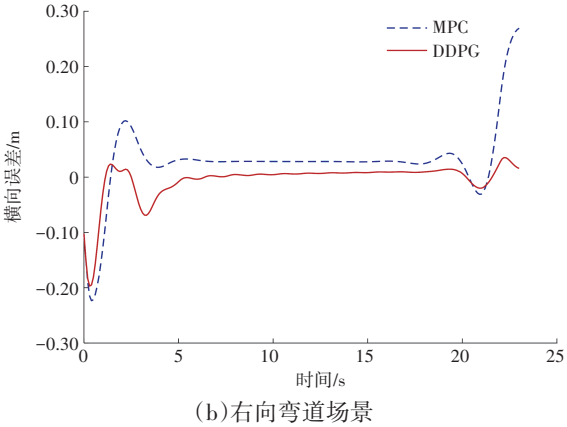


(b)右向弯道场景

图10　60 km/h车速条件下非固定曲率测试场景横向误差



(a)左向弯道场景



(b)右向弯道场景

图11　90 km/h车速条件下非固定曲率测试场景横向误差

表3　横向误差绝对值均值　　m

弯道场景	算法	横向误差绝对值均值		
		车速30 km/h	车速60 km/h	车速90 km/h
左向弯道	MPC	0.004 5	0.018 0	0.041 0
	DDPG	0.001 7	0.004 9	0.014 0
右向弯道	MPC	0.006 4	0.023 4	0.049 0
	DDPG	0.003 7	0.007 4	0.018 0

总体而言，尽管DDPG算法的横向误差随车速增加而有所增大，但其误差波动幅度始终保持在较低水平，且能够更快地将误差控制在较小范围内。这一特性体现了DDPG算法在不同工况下的显著优势，尤其是在高速、大曲率弯道等复杂场景中，其通过深度强化学习策略实现的非线性控制能力显著优于传统MPC方法。

5　结束语

本文基于DDPG算法搭建并改进了一种车道保持横向控制模型，针对不同工况下车道保持的精准性设计了综合奖励函数，并考虑任务完成度，进一步优化模型训练过程中的收敛速度，通过训练获得车道保持前轮转角的最优控制策略，最后针对学习获得的最优策略在不同弯道场景下进行验证。结果表明，DDPG算法在横向误差控制方面展现出显著优势：在30 km/h、60 km/h、90 km/h车速下，相比传统的模型预测控制方法，DDPG算法在不同弯道场景下的横向位置误差更小，其横向误差绝对值均值较MPC降低了42%~72%，特别是在90 km/h高速与大曲率弯道场景下，DDPG基于深度神经网络的非线性映射能力，有效抑制了横向误差的超调与稳态误差波动，体现了其对复杂驾驶场景的适应能力。

参　考　文　献

[1] WEI S J, PFEFFER P E, EDELMANN J. State of the Art: Ongoing Research in Assessment Methods for Lane Keeping Assistance Systems[J]. IEEE Transactions on Intelligent Vehicles, 2024, 9(9): 5853-5875.

[2] FANG Z W, WANG J X, WANG Z J, et al. Human-Machine Shared Control for Path Following Considering Driver Fatigue Characteristics[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(7): 7250-7264.

[3] 方川. 基于深度强化学习的无人驾驶车道保持决策的研究[D]. 南京: 南京大学, 2019.
FANG C. Research on Lane Keeping Decision for Autonomous Driving Based on Deep Reinforcement Learning[D]. Nanjing: Nanjing University, 2019.

[4] JAIN H, BABEL P. A Comprehensive Survey of PID and Pure Pursuit Control Algorithms for Autonomous Vehicle Navigation[EB/OL]. (2024-09-15) [2025-12-11]. https://arxiv.org/abs/2409.09848.

[5] NADIRI F, RAD A B. A Novel Lateral Control System for Autonomous Vehicles: A Look-Down Strategy[J]. Machines, 2025, 13(3).

[6] DIAB M K, AMMAR H H, SHALABY R E. Self-Driving Car Lane-Keeping Assist Using PID and Pure Pursuit Control[C]// 2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT). Sakheer, Bahrain: IEEE, 2020: 1-6.

[7] 王贵林. 基于模型驱动的车道保持辅助(LKA)系统算法研究与实现[D]. 十堰: 湖北汽车工业学院, 2023.
WANG G L. Research and Implementation of Algorithms for Model-Driven Lane Keeping Assist (LKA) System[D]. Shiyan: Hubei University of Automotive Technology, 2023.

[8] GUO L, WEI L R, GE P S, et al. Lane Keeping Controller Based on LQR Optimized by Genetic Algorithm[C]// 2022 6th CAA International Conference on Vehicular Control and Intelligence (CVCI). Nanjing, China: IEEE, 2022: 1-6.

[9] SAMUEL M, MOHAMAD M, HUSSEIN M, et al. Lane Keeping Maneuvers Using Proportional Integral Derivative (PID) and Model Predictive Control (MPC)[J]. Journal of Robotics and Control (JRC), 2021, 2(2): 78-82.

[10] MONTOYA J D G, TEIXEIRA E L S, MURILO A, et al. A Comprehensive Analysis of Model Predictive Control for Lane Keeping Assist System[J]. IEEE Access, 2023(11): 140216-140228.

[11] SALT DUCAJÚ J M, SALT LLOBREGAT J J, CUENCA Á, et al. Autonomous Ground Vehicle Lane-Keeping LPV Model-Based Control: Dual-Rate State Estimation and Comparison of Different Real-Time Control Strategies[J]. Sensors, 2021, 21(4).

[12] 张俊阳. 基于深度强化学习的端到端无人驾驶决策算法研究[D]. 成都: 电子科技大学, 2022.
ZHANG J Y. Research on End to End Autonomous Driving Decision Algorithm Based on Deep Reinforcement Learning [D]. Chengdu: University of Electronic Science and Technology, 2022.

[13] PÉREZ-GIL O, BAREA R, LÓPEZ-GUILLÉN E, et al. DQN-Based Deep Reinforcement Learning for Autonomous Driving[C]// Workshop of Physical Agents. Cham: Springer International Publishing, 2020: 60-76.

[14] RAN Q, LIANG C, LIU P. A Safe Lane-Changing Strategy for Autonomous Vehicles Based on Deep Q-Networks and

Prioritized Experience Replay[J]. Digital Transportation and Safety, 2025, 4(3): 170-174.

[15] 张奔. 基于深度强化学习的车辆车道保持算法研究[D]. 南京: 南京理工大学, 2021.
ZHANG B. Research on Vehicle Lane Keeping Algorithm Based on Deep Reinforcement Learning[D]. Nanjing: Nanjing University of Science and Technology, 2021.

[16] TERVEN J. Deep Reinforcement Learning: A Chronological Overview and Methods[J]. AI, 2025, 6(3).

[17] IBRAHIM S, MOSTAFA M, JNADI A, et al. Comprehensive Overview of Reward Engineering and Shaping in Advancing Reinforcement Learning Applications[J]. IEEE Access, 2024, 12: 175473-175500.

[18] LU C B, SHI L X, CHEN Z W, et al. Overcoming the Curse of Dimensionality in Reinforcement Learning Through Approximate Factorization[EB/OL]. (2024-11-12) [2025-12-11]. https://arxiv.org/abs/2411.07591.

[19] 贺伊琳, 宋若旸, 马建. 基于强化学习DDPG的智能车辆轨迹跟踪控制[J]. 中国公路学报, 2021, 34(11): 335-348.
HE Y L, SONG R Y, MA J. Intelligent Vehicle Trajectory Tracking Control Based on Reinforcement Learning DDPG [J]. China Journal of Highway and Transport, 2021, 34(11): 335-348.

[20] 李文礼, 邱凡珂, 廖达明, 等. 基于深度强化学习的高速公路换道跟踪控制模型[J]. 汽车安全与节能学报, 2022, 13(4): 750-759.
LI W L, QIU F K, LIAO D M, et al. Highway Lane Change Tracking Control Model Based on Deep Reinforcement Learning[J]. Journal of Automotive Safety and Energy Efficiency, 2022, 13(4): 750-759.

[21] 孙腾超, 陈焕明. 基于深度强化学习的自主换道控制模型[J]. 农业装备与车辆工程, 2024, 62(4): 30-34.
SUN T C, CHEN H M. Autonomous Lane Changing Control Model Based on Deep Reinforcement Learning[J]. Agricultural Equipment & Vehicle Engineering, 2024, 62 (4): 30-34.

（责任编辑　斛　畔）

修改稿收到日期为2025年12月11日。

（上接第19页）

[89] TIAN Y H, QIN H, XIA F, et al. Syntax-Driven Approach for Semantic Role Labeling[C]// International Conference on Language Resources and Evaluation. Marseille, France: European Language Resources Association, 2022: 7129-7139.

[90] CHEN L, WU P H, CHITTA K, et al. End-to-End Autonomous Driving: Challenges and Frontiers[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(12): 10164-10183.

[91] 岳超, 钟佳儒, 宁启立, 等. 新技术形势下“车能路云”融合发展战略研究[J]. 中国工程科学, 2024, 26(1): 45-58.
YUE C, ZHONG J R, NING Q L, et al. Development Strategy of Vehicle-Energy-Road- Cloud Collaboration in the New Technological Situation[J]. Strategic Study of CAE, 2024, 26(1): 45-58.

[92] XU R S, XIANG H, XIA X, et al. OPV2V: An Open Benchmark Dataset and Fusion Pipeline for Perception with Vehicle-to-Vehicle Communication[C]// 2022 International Conference on Robotics and Automation (ICRA). Philadelphia, PA, USA: IEEE, 2021: 2583-2589.

[93] LI Y M, MA D K, AN Z Y, et al. V2X-Sim: Multi-Agent Collaborative Perception Dataset and Benchmark for Autonomous Driving[J]. IEEE Robotics and Automation Letters, 2022, 7(4): 10914-10921.

[94] XU R S, XIANG H, TU Z Z, et al. V2X-ViT: Vehicle-to-Everything Cooperative Perception with Vision Transformer [C]// 17th European Conference on Computer Vision (ECCV 2022). Tel Aviv, Israel: Springer, 2022: 107-124.

[95] YU H B, LUO Y Z, SHU M, et al. DAIR-V2X: A Large-Scale Dataset for Vehicle-Infrastructure Cooperative 3D Object Detection[C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 21329-21338.

[96] XIANG H, ZHENG Z L, XIA X, et al. V2X-Real: A Large-Scale Dataset for Vehicle-to-Everything Cooperative Perception[EB/OL]. (2024-12-16) [2025-10-30]. https://arxiv.org/abs/2403.16034v2.

[97] ZIMMER W, WARDANA G A, SRITHARAN S, et al. TUMTraf V2X Cooperative Perception Dataset[C]// 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2024: 22668-22677.

（责任编辑　斛　畔）

修改稿收到日期为2025年7月14日。