DOI:10.3969/j.issn.1003-5060.2023.07.001

基于深度强化学习的轨迹跟踪横向控制研究

张炳力, 佘亚飞
(合肥工业大学 汽车与交通工程学院,安徽 合肥 230009)

摘 要:针对自动驾驶的轨迹跟踪问题,为实现性能优异且具有实际应用价值的控制器,文章将双延迟深度确定性策略梯度(twin delayed deep deterministic policy gradient,TD3)的深度强化学习算法应用于轨迹跟踪的横向控制。对车道线保持的应用场景进行控制器设计,首先基于TD3算法对神经网络结构及其参数进行设计,并依据人类驾驶员的行为方式定义状态空间和动作输出,使其具有较快的训练速度以及较好的控制执行效果;然后设计一种奖励函数,将跟踪精度和舒适度同时作为控制器性能的优化方向;最后,根据ISO 11270:2014(E)标准在Prescan中搭建多种使用场景进行仿真实验,验证所设计的控制器性能。通过与当前主流轨迹跟踪解决方案实验结果的对比,分别从跟踪精度和舒适度两方面证明了该控制器可以满足使用要求并且控制性能更加优异,具有的较高应用价值。

关键词:自动驾驶;轨迹跟踪;深度强化学习;双延迟深度确定性策略梯度(TD3)算法;奖励函数
中图分类号:TP242.6;U461 **文献标志码**:A **文章编号**:1003-5060(2023)07-0865-08

Research on lateral control of trajectory tracking based on deep reinforcement learning

ZHANG Bingli, SHE Yafei
(School of Automobile and Traffic Engineering, Hefei University of Technology, Hefei 230009, China)

Abstract: In order to explore a controller with better performance and practical application value for the trajectory tracking of autonomous driving, this paper applies the deep reinforcement learning algorithm of twin delayed deep deterministic policy gradient (TD3) to the lateral control of trajectory tracking. The controller design is based on the application scenario of lane line keeping. Firstly, the neural network structure and its parameters are designed based on the TD3 algorithm, and the state space and action output are defined according to the behavior of the human driver, so that it has higher training speed and better control effect. Then, a reward function is designed, which takes tracking accuracy and comfort as the optimization direction of controller performance at the same time. Finally, in order to verify the performance of the designed controller, a variety of simulation experiment scenarios were set up in Prescan to conduct simulation experiments according to the ISO 11270:2014(E) standard. In addition, the comparison with the experimental results of the current main trajectory tracking solutions proves that the controller can meet the application requirements and has better control performance in terms of tracking accuracy and comfort, and has high application value.

Key words: autonomous driving; trajectory tracking; deep reinforcement learning; twin delayed deep deterministic policy gradient (TD3) algorithm; reward function

收稿日期:2022-05-27;**修回日期**:2022-07-18
基金项目:安徽省科技重大专项计划资助项目(JZ2022AKKZ0111)
作者简介:张炳力(1968—),男,安徽合肥人,博士,合肥工业大学教授,博士生导师.

针对自动驾驶的横向控制,当前的主流方法大致分为基于运动学的控制算法和基于动力学的控制算法两类[1]。基于运动学的方法是利用目标路径与车辆运动过程中的几何关系,其中使用较为广泛的算法有 Pure Pursuit、Stanley 等。Pure Pursuit 算法最早由文献[2]提出,具有易于实现、计算成本低并且控制效果较为稳定等特点,在自动驾驶相关研究中被广泛应用[3-4];Stanley 算法由文献[5]提出,部署该控制算法的无人车取得了当年 DARPA 挑战赛冠军,Stanley 算法在大多数驾驶环境下尤其是弯道行驶时具有更好的跟踪性能,因此被广泛应用于自动驾驶技术的研究[6]。

相较于基于运动学的方法,基于动力学的方法在横向控制性能上表现更加优秀,其充分考虑了车辆的动力学特性,被越来越多地应用于轨迹跟踪问题,目前主流的控制算法为模型预测控制(model predictive control,MPC)。文献[7]设计了基于 MPC 的车道保持系统转向控制策略,验证其在轨迹跟踪的横向控制上具有较好的适应性和鲁棒性,引入的车辆动力学模型可以通过模型的等效约束转化减少规划与控制的计算量,提高系统的实时性;文献[8]考虑轮胎的非线性特性,对状态矩阵和控制矩阵进行了修正,调整了代价函数的权重,并通过实验验证了设计的控制器在真实环境中控制精度更高并且鲁棒性强;文献[9]设计了一种自主切换控制模型的 MPC 控制器,在稳态工况下以速度航向偏差作为跟踪误差,而在瞬态工况下以车辆横向偏差作为跟踪误差,从而大幅提高了跟踪精度。

但是基于模型预测的方法需要已知车辆的动力学参数,若车辆的动力学参数未知,则难以进行控制器的设计;除此之外,对于不同车辆的适用性不强。当所设计的控制器部署到其他车辆上时,由于不同车辆动力学参数不同,控制器的控制性能会受到很大影响。为了解决在不同车辆上的适用性问题并且探索更为有效的控制方案,一些学者提出了基于深度强化学习的方案[10-13]。文献[14]分别使用深度 Q 网络(deep Q-network,DQN)和深度确定性策略梯度(deep deterministic policy gradient,DDPG)算法设计轨迹跟踪横向控制器,通过对比实验结果证明了 DDPG 算法在车辆横向控制上的表现更优秀;文献[15]也验证了自动驾驶是一个连续控制问题,不能简单地将连续问题分解成离散问题。DDPG 算法因为已经被证实适用于连续控制场景,所以逐渐成为基于强化学习的主流控制方法[16]。然而,DDPG 算法本身在某些情况下可能会出现值函数过估计、训练时间较长并且容易陷入局部最优的问题。针对这些问题,文献[17]改进 DDPG 算法,并提出了双延迟深度确定性策略梯度(twin delayed deep deterministic policy gradient,TD3)算法,该算法相较于 DDPG 算法极大地提高了训练速度和训练效果,在决策和控制领域具有很大的应用前景[18-19]。

目前,针对轨迹跟踪的研究仍存在一些问题,无论是基于几何跟踪或者基于模型预测等传统控制方案,还是基于强化学习的方案,横向控制的性能优化大多聚焦于跟踪精度,而较少考虑乘客的舒适度。

针对以上问题,本文所做的研究如下:

1) 根据车道线保持的使用场景提出一种基于深度强化学习的横向控制方案,重新设计其状态空间以及动作空间。对输入和输出进行归一化处理,从而克服在不同应用情况的场景下输入和输出取值范围不同的问题。

2) 采用基于 TD3 的方案优化神经网络结构,设计 Critic 和 Actor 网络结构以及参数,使其具有较快的训练速度以及稳定的训练过程。

3) 设计奖励函数,将跟踪精度和乘客舒适度作为性能指标平衡考虑。

4) 选择仿真平台,并根据 ISO 11270:2014(E)[20]标准搭建仿真环境进行实验验证。

1 状态与动作定义

1.1 状态空间定义

当被控车辆即将进入曲率变化的区域时,如果不能得到足够多前方目标轨迹的信息,那么车辆可能会出现横向控制的输出值更新不及时的情况,此时控制输出往往会出现控制结果超调的现象,并且伴随着控制输出不稳定。

为了使车辆在前方目标轨迹的曲率发生变化时可以提前得到前方目标轨迹的信息,从而较快做出控制输出的更新,而不是在要进入曲率变化的区域时再改变横向控制的输出值,结合下文所述的奖励函数设计,本文选取状态空间 $l_1 \sim l_{12}$,具体如图 1 所示。

图 1 中:l_1、l_2 表示在车辆前轴处汽车中心线与车道线左边缘以及车道线右边缘的距离;l_3、l_4 表示在车辆前轴前方 2 m 处汽车中心线与车道线左边缘以及车道线右边缘的距离;l_5、l_6 表示在

车辆前轴前方 6 m 处汽车中心线与车道线左边缘以及车道线右边缘的距离；l_7、l_8 表示在车辆前轴前方 10 m 处汽车中心线与车道线左边缘以及车道线右边缘的距离；l_9、l_{10} 表示在车辆前轴前方 14 m 处汽车中心线与车道线左边缘以及车道线右边缘的距离；l_{11}、l_{12} 表示在车辆前轴前方 18 m 处汽车中心线与车道线左边缘以及车道线右边缘的距离。

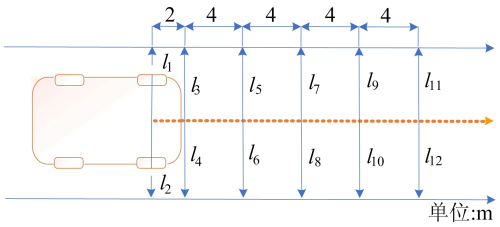


图 1　状态空间定义

不同应用情况的场景下，定义状态空间 l_1～l_{12} 的取值范围乃至数量级不同，会导致训练的智能体在某些应用场景下失效。为确保所设计的控制器具有较强的场景泛化能力，将定义状态的各距离减去最小值，之后再除以最大值与最小值之差得到 l_1～l_{12}，从而将定义的状态归一化至以下区间，即 $l_i \in [0,1]$。

1.2　动作空间定义

动作空间是所设计的轨迹跟踪控制器的输出值，其选取仿照人类驾驶员，定义动作空间为前轮转角 φ_{steer}。动作输出值的取值范围较大会导致 Actor 网络学习效果不佳，因此将动作输出的上下限限制为±1，将动作输出的前轮转角 φ_{steer} 限制在区间[−1,1]。为了让前轮转角适用于真实车辆的控制场景，在动作输出后设计增益，乘以 180°/π，即将前轮转角 φ_{steer} 放大至区间 [−57.3°, 57.3°]。

2　轨迹跟踪控制器的神经网络设计

2.1　Actor 网络

本文所设计的 Actor 网络如图 2 所示，学习率设置为 0.000 1。

从图 2 可以看出，Actor 网络由 1 个输入层（State）、2 个隐藏层（Actor FC1、Actor FC2）和 1 个输出层（Actor Output）组成。Actor 网络将状态向量作为输入，2 个隐藏层使用 relu 激活函数，输出层采用 tanh 激活函数输出前轮转角，使预期累积长期回报最大化，从而实现确定性策略。

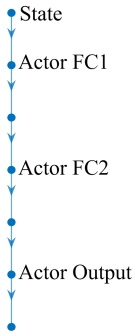


图 2　Actor 网络

2.2　Critic 网络

本文所设计的 Critic 网络如图 3 所示。

从图 3 可以看出，Critic 网络由 2 个输入层（State、Action）、4 个隐藏层以及 1 个输出层（Critic Output）组成。Critic 网络实现对 Q 值的近似，2 个 Critic 网络将状态向量和动作分别输入，并选择两者中较小的 Q 值输出。考虑到学习率低会使训练需要很长时间，而学习率高则可能会达到局部最优结果或者发散，因此学习率根据经验设置为 0.000 1。

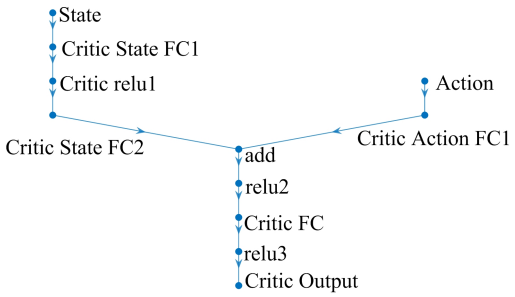


图 3　Critic 网络

Actor 网络与 Critic 网络互相作用，首先环境会给智能体一个状态，智能体将此状态输入到 Actor 网络并输出相应的动作。将此时刻下状态和动作输入 Critic 网络得到相应的 Q 值，并利用较小的 Q 值更新网络。

3　奖励函数设计

在转弯行驶工况下，转向过快可能会造成车辆在一定程度上丧失稳定性，从而使横向加速度变化剧烈导致舒适度较低。

为了使智能体尽可能地避免陷入局部最优，在接近最优情况的过程中奖励值增加得应该更快，奖励函数采用负指数形式。因此，结合选取的状态空间并且综合考虑跟踪精度与舒适性，本文设计的奖励函数如下：

$$r_t(s_t,a_t)=\begin{cases}e^{-\boldsymbol{\eta}\cdot\boldsymbol{\xi}}+e^{a_{\text{lat}}},\text{正常行驶};\\ -C,\qquad\text{车辆超出道路边界}\end{cases}\tag{1}$$

其中:a_{lat}为车辆横向加速度;C 为碰撞惩罚系数,这里取 10;$\boldsymbol{\eta}=[\alpha_1\quad\alpha_2\quad\alpha_3\quad\alpha_4\quad\alpha_5\quad\alpha_6]$为系数向量;$\boldsymbol{\xi}=[\Delta l_1\quad\Delta l_2\quad\Delta l_3\quad\Delta l_4\quad\Delta l_5\quad\Delta l_6]^{\mathrm{T}}$ 为距离向量。$\boldsymbol{\eta}$、$\boldsymbol{\xi}$ 的元素分别定义如下:

$$\alpha_i=(6-i)^2\Big/\sum_{i=1}^{6}i^2,\quad i=1,2,3,4,5,6\tag{2}$$

$$\Delta l_i=|\,l_{2i-1}-l_{2i}\,|,\quad i=1,2,3,4,5,6\tag{3}$$

$l_1\sim l_{12}$为状态空间中定义的观测量,当车辆前方横向距离在车道线中心线时,$\Delta l_i=0$;而车辆前方横向距离偏离车道线中心线越多,Δl_i 会随之增大,直到超过车道线边缘线,此时达到最大值 $\Delta l_i=1$。

4 探索策略设计

对于连续动作信号来说,设定噪声以鼓励探索是十分重要的。自相关的奥恩斯坦-乌伦贝克(Ornstein-Uhlenbeck,OU)噪声与高斯噪声等独立的噪声相比,前者可以使控制的信号较为连续,后者会使前后两步相差较大[21],因此对于惯性系统来说 OU 噪声更为合适。而本文系统动作空间为前轮转角 φ_{steer},需要在原始动作上增加噪声。为了产生在原始动作附近邻域内的实际执行动作,增加的噪声均值应为 0。

综上可知,本文通过添加均值为 0 的 OU 噪声模型来增加智能体探索,其 OU 噪声离散形式的微分方程如下:

$$x(t+1)=x(t)-\theta(x(t)-\mu)T_{\text{s}}\tag{4}$$

其中:T_{s} 为每一步的时间;μ 为噪声均值;θ 决定了接近均值的速度。式(4)中各项参数均为常数,具体设计如下:$\mu=0$;$\theta=0.15$;$T_{\text{s}}=0.05$。

5 仿真实验奖励函数设计

5.1 仿真平台的选择

本文选择使用 Prescan 和 MATLAB 联合仿真的方法对所设计的控制器进行仿真验证,其仿真平台如图 4 所示。在 Prescan 中进行仿真环境的搭建,对车辆、道路以及传感器等进行创建,并将这些数据导入 MATLAB。在 MATLAB 中完成控制算法的实现,并把控制结果的车辆数据导入 Prescan 中进行更新,整个算法更新时间同上文所述的探索策略更新时间一致。此外,使用 Prescan 内部集成的 3D 可视化查看工具,以便于对运行结果进行直接观测。

图 4 Prescan 联合 MATLAB 仿真平台

5.2 仿真场景搭建

仿真场景依据 ISO 11270:2014(E)搭建,该国际标准适用于乘用车、商用车和公共汽车的车道线保持的性能测试。根据 ISO 11270:2014(E)车道线保持的性能评价试验分直道和弯道,并且要求整个测试过程中横向加速度不超过 3 m/s^2,因此分别搭建直道、左转弯道、右转弯道 3 个仿真场景。

5.2.1 使用车辆

使用型号为 AudiA8 的车辆进行仿真,该车辆的基本参数见表 1 所列。

表 1 车辆基本参数

参数	数值
长度/m	5.21
宽度/m	2.04
轴距/m	2.94
转向系统角传动比	20

5.2.2 直道

根据测试标准,测试车辆以 20~22 m/s 的速度沿直线道路直线行驶,并且允许车辆轮胎外边缘超过车道线边界的最大值为 0.4 m。直道仿真场景道路长度为 300 m,车辆行驶速度为 20 m/s。

5.2.3 弯道

根据测试标准,整个测试过程中车速应处在 20~22 m/s 之间,弯道中的行驶时间大于 5 s。设定:车辆行驶速度 $v=20$ m/s;弯道行驶时间 $t=10$ s;弯道中车辆横向加速度 $a_y=1$ m/s^2。则道路几何参数如下:

$$S=tv=200\text{ m}\tag{5}$$

$$R=\frac{v^2}{a_y}\tag{6}$$

仿真弯道设计如图 5 所示。弯道的测试过程由 2 个单独的测试组成,一次进入左曲线,一次进入右曲线。允许车辆轮胎外边缘超过车道边界的最大值为 0.4 m。

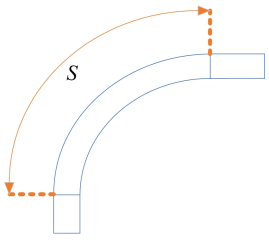


图 5 仿真弯道设计网络

5.3　训练智能体

5.3.1　单次运行终止条件

设定单次运行停止并对场景进行初始化开始下一次运行的条件如下：

1）运行过程中目标车辆的车轮外侧超出车道线边缘。

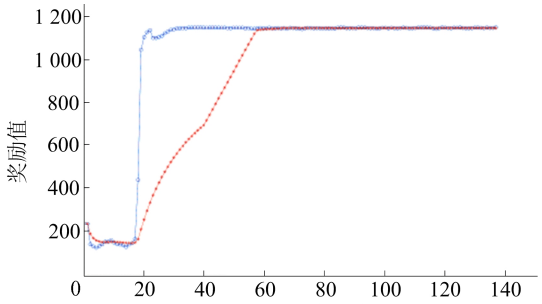
2）目标车辆到达目标地点。

5.3.2　训练终止条件

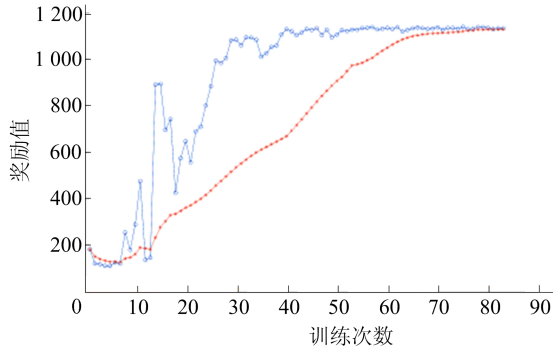
为了加强训练后智能体的稳定性，定义平均奖励，并取平均的窗口长度为 50 次运行结果。为了避免训练结果陷入局部最优，从而发生长时间训练但未能达到设置的平均奖励目标值，设置最大运行次数为 2 000。

5.3.3　训练结果

基于 2 种算法设计的智能体训练结果如图 6 所示。



(a) 本文设计的智能体



(b) 基于DDPG算法设计的智能体

图 6　基于 2 种算法设计的智能体训练结果

从图 6a 可以看出：蓝色线表示每次运行结果的累计奖励值，该值随着不断训练整体上呈上升趋势；红色线表示平均窗口内累计奖励值的平均值，该值训练过程中呈上升趋势并逐渐收敛于最优情况的累计奖励值；单次运行的累计奖励值在运行 20 次时第 1 次达到最优值附近；平均奖励在运行 40 次后逐渐收敛，并且在 60 次后达到稳定。由图 6 可知，相较于图 6b 所示的 DDPG 算法的训练过程，本文算法训练过程可以更快达到收敛，并且训练过程较为稳定，训练过程中奖励值不会出现剧烈波动。

训练后运行的每一步奖励值如图 7 所示，其平均值为 1.98。而设计的奖励函数最大值为 2，即在每一步都采取最优行动时奖励值为 2。训练后的奖励平均值达到了最大奖励函数的 99%，可以认为此时已经基本达到最优情况。

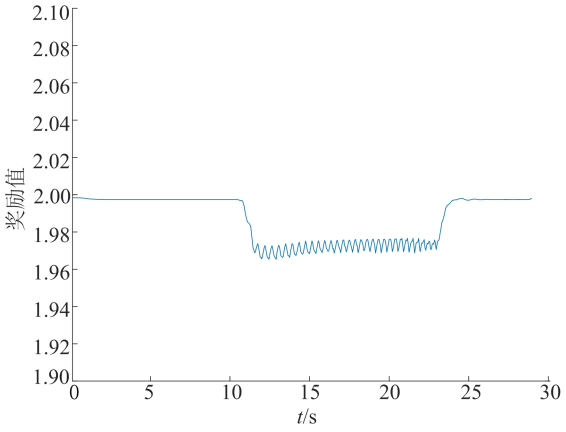


图 7　训练后单步奖励值

5.4　定义评价指标

本文采用横向误差 e 作为衡量跟踪精度的量化指标。横向误差定义如图 8 所示，定义为车辆质心到车道线中心线的距离。此外，使用车辆横向加速度来评价乘客舒适度的性能[22]。

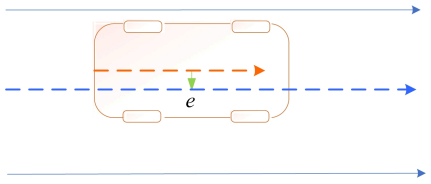


图 8　横向误差定义

6　仿真实验结果对比

以跟踪精度和乘客舒适度作为控制器的性能指标，将本文所设计的控制器与文献[3]设计的 Pure Pursuit 控制器、文献[6]设计的 Stanley 方法控制器、文献[8]设计的 MPC 控制器以及文献[18]基于 DDPG 算法设计的控制器的仿真结果进行对比。除本文算法外，其他算法的参数与其对应文献中设置的参数一致，包括本文算法在内的所有算法仿真环境设置完全相同。

6.1　直线行驶

直道仿真实验结果如图 9 所示。

从图 9 可以看出,在直线行驶工况下,各方案的控制器控制性能基本相同,横向加速度接近于 0,横向误差也都在 0.01 m 以内,跟踪精度与舒适性均比较高。

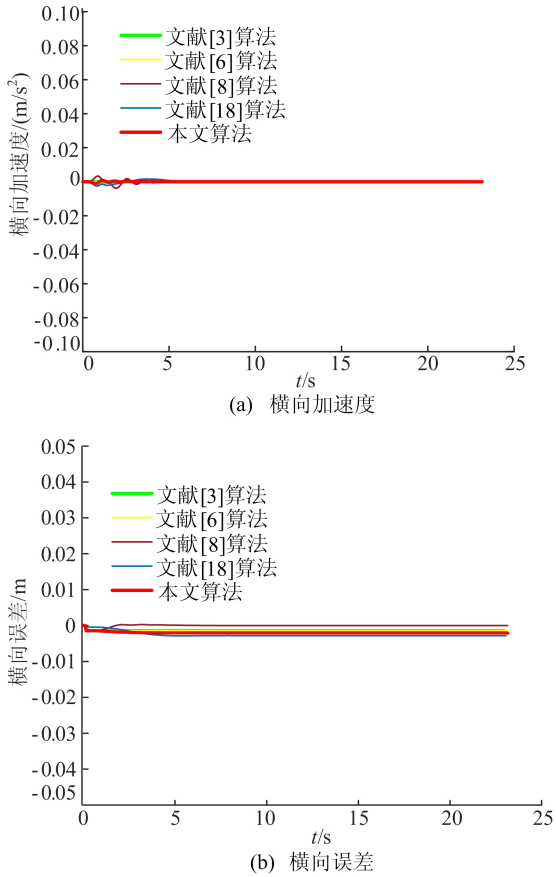


图 9 直道仿真实验结果

6.2 右转弯行驶

右转弯道仿真实验结果如图 10 所示。

从图 10 右转弯道的仿真实验结果可以看出,对于横向加速度,本文控制器下横向加速度波动最小,稳定性最高;其次是使用文献[18]中 DDPG 算法的控制器。这 2 种基于强化学习方案的横向加速度稳定性要明显优于轨迹跟踪的其他 3 种传统控制方案。

右转弯道不同算法的横向误差均方根值见表 2 所列。

由表 2 可知,在右转弯道场景中,本文算法的横向误差均方根值比其他主流控制方案(文献[3]算法、文献[6]算法、文献[8]算法、文献[18]算法)分别小 74%、63%、36%、50%。

表 2 和图 10b 中的数据表明:对于横向误差,本文控制器的仿真结果最趋近于 0,表明其跟踪精度最高;控制结果较优的是文献[8]MPC 方法和文献[18]算法的控制器,在控制精度上要优于文献[3]和文献[6]基于运动学设计的控制器。

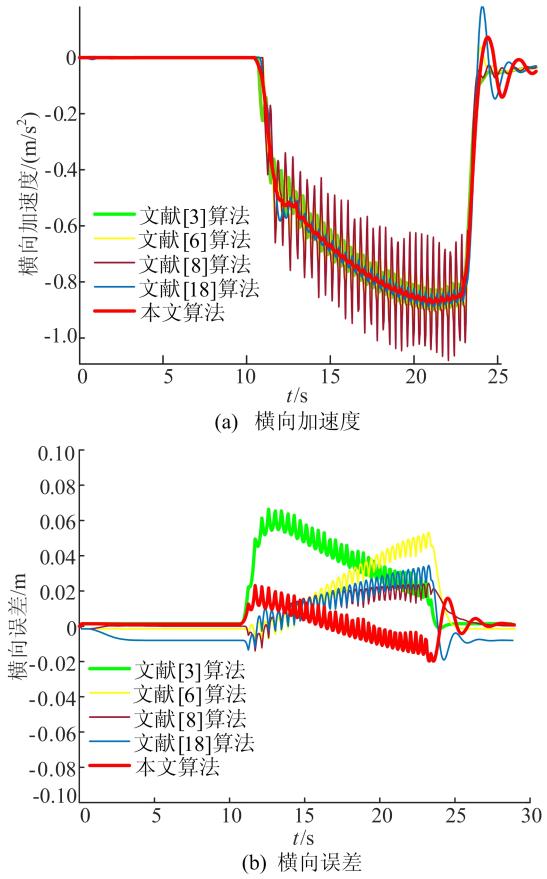


图 10 右转弯道仿真实验结果

表 2 不同算法的右转弯道横向误差均方根值

控制方案	均方根值/m
文献[3]算法	0.027
文献[6]算法	0.019
文献[8]算法	0.011
文献[18]算法	0.014
本文算法	0.007

6.3 左转弯行驶

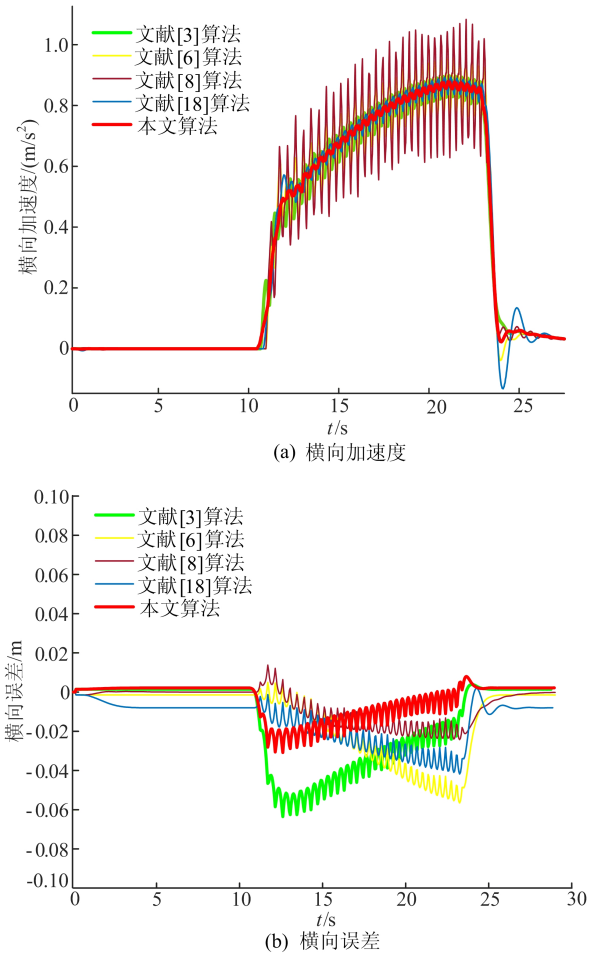
左转弯道仿真实验结果如图 11 所示。

从图 11 可以看出,左转弯道仿真实验结果与右转弯道仿真实验结果相似,验证了本文控制器具有场景泛化性。

左转弯道不同算法的横向误差均方根值见表 3 所列。

由表 3 可知,在左转弯道工况下,本文算法的横向误差均方根值比文献[3]算法、文献[6]算法、文献[8]算法、文献[18]算法的方案分别小 64%、59%、18%、43%。实验结果表明本文控制器比其他控制方案的横向加速度波动小,稳定性表现更

优异，同时控制器跟踪精度也最高。



(a) 横向加速度

(b) 横向误差

图 11　左转弯道仿真实验结果

表 3　左转弯道横向误差均方根值

控制方案	均方根值/m
文献[3]算法	0.025
文献[6]算法	0.022
文献[8]算法	0.011
文献[18]算法	0.016
本文算法	0.009

6.4　复杂工况

为了验证本文所设计控制器的泛化性，将其在复杂工况中进行仿真。

仿真测试场景如图 12 所示，由 4 段弯道以及若干段直道组成。

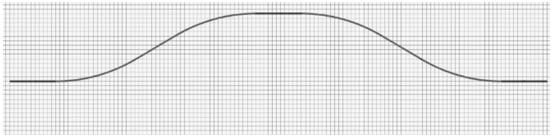


图 12　长距离连续转弯

仿真测试结果如图 13 所示，本文控制器可以将横向误差控制在 0.1 m 之内，在复杂工况使用场景下表现出良好的鲁棒性。

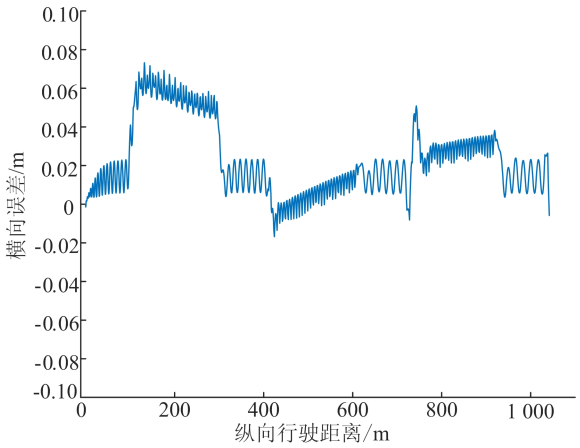


图 13　长距离连续转弯测试结果

7　结　　论

本文将 TD3 算法应用于轨迹跟踪的横向控制，设计出一种基于深度强化学习的控制器，旨在解决自动驾驶中的轨迹跟踪问题。通过仿真实验，验证了本文所设计的控制器具有以下特点：

1）在不同使用场景下均可以将横向误差控制在可接受范围之内，可以满足实际使用中的功能要求。相较于目前主流的控制器方案，尤其在复杂环境中的行驶工况（如大曲率的弯道行驶）下，控制精度更加优秀。

2）可以在大曲率弯道行驶中保持横向加速度的稳定，相较于其他主流方案，本文控制器横向加速度较小同时也较为稳定，最大程度上保证乘客的舒适度。

3）具有较强的场景泛化性，可以满足车辆在复杂工况下的使用要求，可以较大程度上保证车辆不会偏离车道中心线。

本文设计并利用仿真实验验证了深度强化学习在轨迹跟踪问题上的效果，探索了强化学习算法应用在自动驾驶上的实现方式。所提出的控制器方案相较于当前的主流解决方案有着更优异的性能，具有较高的实际应用价值。

［参　考　文　献］

［1］ PARK M W，LEE S W，HAN W Y. Development of lateral control system for autonomous vehicle based on adaptive pure pursuit algorithm［C］//2014 International Conference

on Control, Automation and Systems. [S. l.]: IEEE, 2014: 1443-1447.

[2] WALLACE R S, STENTZ A, THORPE C E, et al. First results in robot road-following[C]//Proceedings of the 9th International Joint Conference on Artificial Intelligence. [S. l.]: IJCAI, 1985: 1089-1095.

[3] LAL D S, VIVEK A, SELVARAJ G. Lateral control of an autonomous vehicle based on Pure Pursuit algorithm[C]// 2017 International Conference on Technological Advancements in Power and Energy. [S. l.]: IEEE, 2017: 1-8.

[4] 王亮,陈齐平,罗玉峰,等. 基于"Pure Pursuit"自动驾驶汽车的路径跟踪控制[J]. 汽车零部件,2021(8):1-7.

[5] DOMINA A, TIHANYI V. Path following controller for autonomous vehicles[C]//2019 IEEE International Conference on Connected Vehicles and Expo. [S. l.]: IEEE, 2019: 1-5.

[6] HOFFMANN G M, TOMLIN C J, MONTEMERLO M, et al. Autonomous automobile trajectory tracking for off-road driving: controller design, experimental validation and racing [C]//2007 American Control Conference. [S. l.]: IEEE, 2007: 2296-2301.

[7] 罗莉华. 基于MPC的车道保持系统转向控制策略[J]. 上海交通大学学报,2014,48(7):1015-1020.

[8] CHENG S, LI L, CHEN X, et al. Model-predictivecontrol-based path tracking controller of autonomous vehicle considering parametric uncertainties and velocity-varying[J]. IEEE Transactions on Industrial Electronics, 2021, 68(9): 8698-8707.

[9] SUN C, ZHANG X, ZHOU Q, et al. A model predictive controller with switched tracking error for Autonomous vehicle path tracking [J]. IEEE Access, 2019 (7): 53103-53114.

[10] 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.

[11] SHAN Y, ZHENG B, CHEN L, et al. A reinforcement learningbased adaptive path tracking approach for autonomous driving[J]. IEEE Transactions on Vehicular Technology, 2020, 69(10): 10581-10595.

[12] JIANG L, Y WANG, WANG L, et al. Path tracking control based on deep reinforcement learning in autonomous driving[C]//2019 3rd Conference on Vehicle Control and Intelligence. [S. l.]: IEEE, 2019: 1-6.

[13] CHEN I M, CHAN C Y. Deep reinforcement learning based path tracking controller for autonomous vehicle[J]. Proceedings of the Institution of Mechanical Engineers, 2021, 235(2/3): 541-551.

[14] WANG Q, ZHUANG W, WANG L, et al. Lane keeping assist for an autonomous vehicle based on deep reinforcement learning[C]//WCX SAE World Congress Experience. [S. l.]: SAE International, 2020: 1-7.

[15] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning[DB/OL]. [2021-04-02]. https://arxiv.org/abs/1509.02971.

[16] LIU M, ZHAO F, NIU J, et al. Reinforcement driving: exploring trajectories and navigation for autonomous vehicles [J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(2): 808-820.

[17] FUJIMOTO S, HOOF H V, MEGER D. Addressing function approximation error in Actor-Critic methods[C]// Proceedings of the 35th International Conference on Machine Learning. [S. l.]: PMLR, 2018: 1587-1596.

[18] TIONG T, SAAD I, TEO K, et al. Deep reinforcement learning with robust deep deterministic policy gradient [C]//2020 2nd International Conference on Electrical, Control and Instrumentation Engineering. [S. l.]: IEEE, 2020: 1-5.

[19] OPALIC S M, GOODWIN M, LEI J, et al. A deep reinforcement learning scheme for battery energy management [C]//2020 5th International Conference on Smart and Sustainable Technologies. [S. l.]: IEEE, 2020: 1-6.

[20] International Organization for Standardization. Intelligent transport systems—Lane keeping assistance systems (LKAS) — Performance requirements and test procedures: ISO 11270: 2014(E)[S]. Switzerland: International Organization for Standardization, 2014: 3-11.

[21] WAWZYNSKI P. Control policy with autocorrelated noise in reinforcement learning for robotics[J]. International Journal of Machine Learning and Computing, 2015, 5(2): 91-95.

[22] LEE K, LI S E, KUM D. Synthesis of robust lane keeping systems: impact of controller and design parameters on system performance[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(8): 3129-3141.

(责任编辑 胡亚敏)