Operations Research and Fuzziology 运筹与模糊学, 2025, 15(5), 13-25
Published Online October 2025 in Hans. https://www.hanspub.org/journal/orf
https://doi.org/10.12677/orf.2025.155227

Hans汉斯

基于风险评估的自动驾驶变道决策强化学习模型

李　琦[1]，周鲁露[2]

[1]上海电科智能系统股份有限公司，上海
[2]上海市公安局交通管理总队，上海

收稿日期：2025年7月21日；录用日期：2025年8月5日；发布日期：2025年9月29日

摘　要

在自动驾驶技术快速发展的背景下，智能体的安全、高效变道决策成为提升驾驶安全性与通行效率的核心挑战。现有基于深度强化学习的变道决策方法往往忽略了周围车辆驾驶风格等动态微观信息，且奖励函数设计单一，导致决策鲁棒性不足、训练过程不稳定。为解决上述问题，本文以带经验回放的深度Q网络(DQN)算法为基础架构，提出融合驾驶风格感知的自动驾驶变道决策优化方法。该方法的核心创新体现在两方面：一是突破传统DQN仅依赖宏观运动学信息的局限，通过量化邻车激进型、保守型等驾驶风格并构建风险系数，形成融合微观驾驶风格的风险状态(Risk)表示，提升智能体对动态环境风险的感知精准度；二是针对单一奖励目标导致的决策偏差问题，设计融合安全性、效率与规则遵守性的多目标奖励函数，通过权重调整引导智能体学习均衡驾驶策略。同时，借助经验回放机制保障训练过程的稳定性。为验证算法性能，本文在SUMO仿真平台中，将所提算法与传统DQN及Double DQN算法展开对比实验。结果表明，本文提出的算法在变道成功率、碰撞率及平均通行效率等关键指标上均展现出显著优势，为自动驾驶场景下的智能体决策提供了更安全、高效的解决方案。

关键词

自动驾驶，变道决策，经验回放，深度强化学习

A Reinforcement Learning Model for Autonomous Lane Change Decision Based on Risk Assessment

Qi Li[1], Lulu Zhou[2]

[1]Shanghai Searl Intelligent System Co., Ltd., Shanghai
[2]Traffic Police Corps of Shanghai Municipal Public Security Bureau, Shanghai

Received: July 21, 2025; accepted: August 5, 2025; published: September 29, 2025

文章引用: 李琦, 周鲁露. 基于风险评估的自动驾驶变道决策强化学习模型[J]. 运筹与模糊学, 2025, 15(5): 13-25.
DOI: 10.12677/orf.2025.155227

Abstract

Against the backdrop of rapid development of autonomous driving technology, the safe and efficient lane-changing decision-making of intelligent agents has become a core challenge to improve driving safety and traffic efficiency. The existing lane-changing decision methods based on deep reinforcement learning mostly ignore the issues of environmental dynamics and sample correlation, resulting in insufficient decision robustness unstable training process. To address the above issues, this paper proposes an optimization method for autonomous driving lane-changing decision making with driver style perception based on the deep Q-Network (DQN) algorithm with experience replay. The core innovations of this method are reflected in two aspects: one is to break through the limitations of traditional DQN, which relies on macro kinematic information, and to form a risk state representation by quantifying the driving styles such as aggressive and conservative neighboring vehicles and constructing a risk coefficient (risk), so as to improve the accuracy of the agent's perception of dynamic environmental risks; the other is to design a multi-objective reward function integrating safety, efficiency and with rules, so as to guide the agent to learn a balanced driving strategy by adjusting the weights, and at the same time, to ensure the stability of the training process by the replay mechanism. In order to verify the performance of the algorithm, a comparative experiment was conducted with the traditional DQN and Double DQN algorithms in the SUMO simulation platform. The experimental results show that the algorithm has significant improvements in lane changing success rate, collision rate, and average traffic efficiency compared to traditional DQN and rule-based methods, providing an effective solution for intelligent agent decision-making in autonomous driving scenarios.

Keywords

Autonomous Driving, Lane-Change Decision, Experience Replay, Deep Reinforcement Learning

1. 引言

在现代交通体系中，汽车保有量的急剧攀升引发了一系列严峻问题。交通拥堵状况日益加剧，交通事故发生率也随之显著提高。据相关研究显示，约 10%的高速公路撞车事故由车辆换道行为所引发，而在这些换道相关事故中，高达 75%是由于驾驶员对换道条件的判断失误造成的[1]。车辆变道作为一种基本且常见的驾驶行为，其过程较为复杂。当驾驶员产生变道意图后，需要仔细观察周边车辆的实时状态，精准选择安全的目标车道，同时合理调整自身车辆的驾驶状态，以维持与周边车辆的安全距离。只有当目标车道出现满足安全间隙条件的空隙时，驾驶员才会实施变道操作。

相较于相对简单的跟驰行为，变道操作所蕴含的安全风险更为复杂。这一过程需要满足诸多约束条件，例如目标车道前后安全间距的动态阈值需精准把控，车辆与周边车辆的轨迹安全性要充分保障，相邻车辆行驶状态的稳定性也不容忽视。变道行为中包含的纵向速度协调与横向位移控制存在多维运动耦合机制，这与道路交通安全状况紧密相关。而且，变道行为通过改变车辆间的微观交互模式，会引发交通流参数的非线性波动，进而对路网宏观通行效率产生影响。不当的变道行为可能导致道路通行能力下

降，对其他车辆的行驶安全构成威胁。在匝道区域，频繁的变道行为甚至可能形成高速交通瓶颈，严重时引发交通崩溃，极大地降低道路安全性。

为有效解决车辆换道决策问题，研究人员提出了多种方法，目前主要可分为四类：基于传统规则的方法、基于博弈论的方法、基于传统机器学习的方法以及基于强化学习的方法。然而，现有的这些换道决策方法均存在一定的局限性。传统规则方法对环境变化的适应能力较差，人工设计的规则难以应对复杂的车辆交互场景，且在扩展任务方面表现不佳，缺乏自主学习能力，难以根据实际情况灵活调整决策策略。博弈论方法在实际应用场景中，由于模型往往庞大复杂，分析难度较大，求解最优解或纳什均衡需要消耗大量的计算资源。并且，在动态变化的交通环境中，该方法难以有效处理策略随时间的演化问题。传统机器学习方法高度依赖大量的标注数据，其性能受数据质量与多样性的制约明显，而且该方法的可解释性较差，不利于深入理解决策过程。强化学习中的 Q-learning 方法仅适用于较为简单的驾驶环境以及低状态量的场景，在复杂场景下表现欠佳。深度强化学习(DRL)虽然具备自主学习策略的能力，但在高维状态空间中面临训练不稳定、样本相关性过高等问题，这些问题限制了其在实际交通场景中的应用效果。

深度 Q 网络(DQN)作为 Q-learning 的重要拓展，近年来在交通领域得到了广泛应用。DQN 作为 Q-learning 的一个重要延伸，近些年被广泛应用，多数实验证明[2]，在面对动态且不确定的交通环境时，不仅可以实现安全、高效的换道，还能够自适应地处理各种复杂交通场景，例如多车道、高密度车流等。大量实验研究表明，在面对动态且充满不确定性的交通环境时，DQN 不仅能够实现安全、高效的换道决策，还能够自适应地处理诸如多车道、高密度车流等各种复杂交通场景。在此基础上，研究人员进一步提出了多种改进算法，如优先经验回放算法(Prioritized Experience Replay) [3]、深度确定性策略梯度(DDPG)算法[4]、双延迟深度确定性策略梯度(Twin Delayed Deep Deterministic Policy Gradient Algorithm, TD3)算法[5]等强化学习算法，这些算法在换道决策任务中均展现出了良好的性能。其中，DQN 算法具有独特优势，它无需预设复杂规则，能够通过与环境的大量交互数据不断优化自身策略，在灵活性与鲁棒性方面明显优于传统的基于规则的方法。

然而，尽管标准的 DQN 算法在很多场景下表现出色，但直接应用于复杂的变道决策时仍面临挑战。一方面，传统的 DQN 模型通常采用宏观的物理信息，距离、速度等作为状态输入，缺乏对周围车辆微观驾驶行为的感知能力，驾驶员的激进或保守程度等，这在人机混驾环境中可能导致对风险的误判。另一方面，其奖励函数的设计往往侧重于单一目标，安全目标即避免碰撞，容易使智能体学到过于保守或鲁莽的片面策略，难以在安全和效率之间达到理想的平衡。

为解决上述问题，本文以带经验回放的 DQN 算法为基础架构，着重从以下两个方面进行创新性改进：

1. 提出一种融合驾驶风格感知的状态表示方法：区别于传统方法仅使用宏观运动学信息，本研究通过量化邻车的驾驶风格，激进型、保守型，并构建风险系数(Risk)，使智能体能够预测性地评估变道风险，提升了对复杂人机混驾环境的感知深度。

2. 设计了面向变道场景的多目标奖励函数：为克服单一奖励目标导致的决策偏差，过度保守或鲁莽，本文精心设计了一个融合安全性、效率和规则遵守性的多目标奖励函数，并通过权重调整，引导智能体学会在不同要素间做出权衡，实现更均衡、类人的决策。

本文通过仿真实验验证了算法的有效性，旨在为自动驾驶智能体的动态决策提供理论与技术支撑，实现更均衡、类人的决策。本文的主要贡献在于，首次将驾驶员风格的量化风险引入到 DQN 的状态表示中，并通过实验证明了这种微观层面的信息感知，能够系统性地提升宏观决策的安全性与效率，为解决复杂人机混驾环境下的自动驾驶决策难题提供了新的思路与有效方案。

2. 相关研究背景

2.1. 深度强化学习在变道决策中的应用

深度强化学习(DRL)通过将深度神经网络强大的环境状态表征能力与强化学习的决策能力相结合，已广泛应用于自动驾驶决策任务[6]。深度 Q 网络(DQN)作为 DRL 的经典算法，其核心思想是利用深度神经网络来估计动作价值函数(Q 值)，在 Atari 游戏等领域取得了突破性的成功[7]。

这些进展推动了 DQN 等算法在自动驾驶交通场景中的应用[8]。有研究者利用 DQN 处理车辆换道等战术决策，通过直接输入摄像头图像或激光雷达数据来实现端到端的控制[9]。在训练这类模型时，如果直接按顺序使用连续采集的数据样本，其时序上的高度相关性会违反神经网络训练时样本独立同分布的假设，进而对训练过程的稳定性造成负面影响[10]。

后续改进算法如 Double DQN、Dueling DQN 通过修正 Q 值高估或分离价值函数结构提升性能，但在变道场景中，样本分布的动态性(如邻车行为突变)仍可能导致训练不稳定。

2.2. 重放记忆机制的研究进展

经验回放(Experience Replay)是解决 DQN 样本相关性问题的关键技术，其核心思想是将智能体与环境交互的经验存储于回放池，训练时随机采样以打破时序关联。Mnih 等[9]在原始 DQN 中引入该机制，显著提升了 Atari 游戏的学习稳定性。在交通领域，经验回放被用于优化信号控制与路径规划，但在变道决策中，如何设计回放池的样本筛选策略，优先保留关键变道经验等仍需深入研究。

2.3. 变道决策中的状态感知与奖励设计

现有变道决策研究多采用车间距、相对速度等宏观因素状态作为输入，但忽略了微观驾驶行为差异，比如激进型驾驶员的加减速特性就会显著影响变道时机选择。此外，奖励函数设计常侧重单一目标，如不碰撞目标，导致智能体过度保守，降低通行效率。因此，需构建融合微观行为特征的状态表示与多目标奖励机制。

3. 模型设计

本章聚焦于智能车辆换道决策模型的构建，旨在通过合理设计状态空间、动作空间、奖励函数及算法框架，实现智能体在复杂交通环境下的安全、高效换道决策。具体设计如下。

3.1. 状态空间定义

为使智能体全面感知变道环境，状态空间包含三部分信息：

自身车辆状态：当前速度 v、加速度 a、车道位置 pos 及目标车道距离 d_{target}，反映智能体的运动状态与变道意图；

周围车辆状态：前后 50 米范围内邻车的相对距离 d_i、相对速度 v_i 及车道位置，共包含左前、左后、右前、右后 4 个关键方位的车辆信息；

驾驶风格感知的风险系数：本文的核心创新之一。我们通过邻车的历史速度标准差与加速度峰值等特征，将其在线划分为三种风格：将其划分为激进型(权重 $w_{style}=1.21$)、普通型($w_{style}=1.0$)、保守型($w_{style}=0.8$)。一个综合的周围环境风险系数 R_{risk}，定义为所有邻车风险的加权和，公式如下：

$$R_{risk}=\sum_{i=1}^{n}\frac{w_{style}(i)}{d_i} \tag{1}$$

其中 n 为周围车辆数量，d_i 为与第 i 辆车的距离，$w_{style}(i)$ 是根据第 i 辆车被识别出的驾驶风格所对应的风险权重。

该系数的核心思想是，对于激进型车辆，即使距离稍远，其潜在风险也应被放大，从而使智能体能够做出更具前瞻性的安全决策；而保守型车辆则相对安全，其风险权重较低。在我们的仿真环境中，社会车辆的驾驶风格是根据预设的分布随机生成的。该系数量化邻车行为对变道安全性的影响。仿真环境中社会车辆的具体风格分布，详见 4.1 节实验设置。

最终状态向量表示为：

$$S=\left[v,a,pos,d_{target},\{d_i,v_i\},R_{risk}\right] \tag{2}$$

3.2. 动作空间定义

智能体的变道动作分为三类：保持当前车道(a_0)、向左变道(a_1)、向右变道(a_2)。动作执行时需满足交通规则(如禁止连续变道、实线禁止变道)，并通过平滑控制(如梯度调整转向角度)确保驾驶舒适性。

3.3. 奖励函数设计带重放记忆的 DQN 算法框架

奖励函数需平衡安全性、效率与合规性，定义为：

$$R=\alpha R_{safety}+\beta R_{efficiency}+\gamma R_{rule} \tag{3}$$

R_{safety}：安全性奖励，与邻车最小距离的正值，距离越近，奖励越低；碰撞时给予−100 的惩罚。

$R_{efficiency}$：效率奖励，与智能体当前速度正相关，速度越接近期望速度，奖励越高。

R_{rule}：合规性奖励，遵守交通规则的奖励，实线变道时给予−50 惩罚，成功变道至目标车道给予+30 奖励。

奖励函数的权重设置为 $\alpha=0.5,\beta=0.3,\gamma=0.2$。这些值是通过初步的网格搜索和实验调试确定的。

这些值的设定遵循了自动驾驶领域的“安全优先”的设计哲学。具体而言，我们赋予安全性 R_{safety} 最高的权重 $\alpha=0.5$，以确保智能体的所有决策都将规避碰撞作为首要任务。在保证安全的基础上，我们认为通行效率 $R_{efficiency}$ 是第二重要的目标，因为它直接关系到交通流的宏观表现，因此赋予其 $\beta=0.3$ 的权重。最后，规则遵守 R_{rule} 作为基本约束，其权重 $\gamma=0.2$ 相对最低，因为在大多数情况下，智能体遵守规则是默认行为，只有在违规时才需要通过惩罚进行修正。为进一步验证该参数组合的合理性与鲁棒性，我们在第 4.4 节中进行了详细的参数敏感性分析。

3.4. 带重放记忆的 DQN 算法框架

DQN (Deep Q-Network)算法作为基于值函数的强化学习算法的代表，是一种较好的解决自动驾驶汽车换道决策问题的算法，使用经验回放(Experience Replay)和目标网络(Target Network)来提高学习的效率和稳定性，原理图见图 1 所示。经验回放用于存储和重复利用之前的经验样本，从而减少样本间的相关性。目标网络用于稳定训练过程，通过固定一段时间更新目标网络的参数，以减少目标值的波动性。

算法流程基于经典 DQN 改进，核心步骤如下：

经验存储：智能体每步交互生成经验元组(S_t,a_t,R_t,S_{t+1})，存入容量为 $N=10^5$ 的经验回放池，优先保留高奖励样本，例如成功规避碰撞的变道经验。

采样与训练：每次训练从回放池随机抽 $B=3$ 个样本，通过目标网络计算目标 Q 值：

$$y_t=R_t+\gamma\max_{a'}Q\left(S_{t+1},a';\theta^-\right) \tag{4}$$

其中 θ^- 为目标网络参数，每 $C=100$ 步从主网络复制更新，确保训练稳定性。

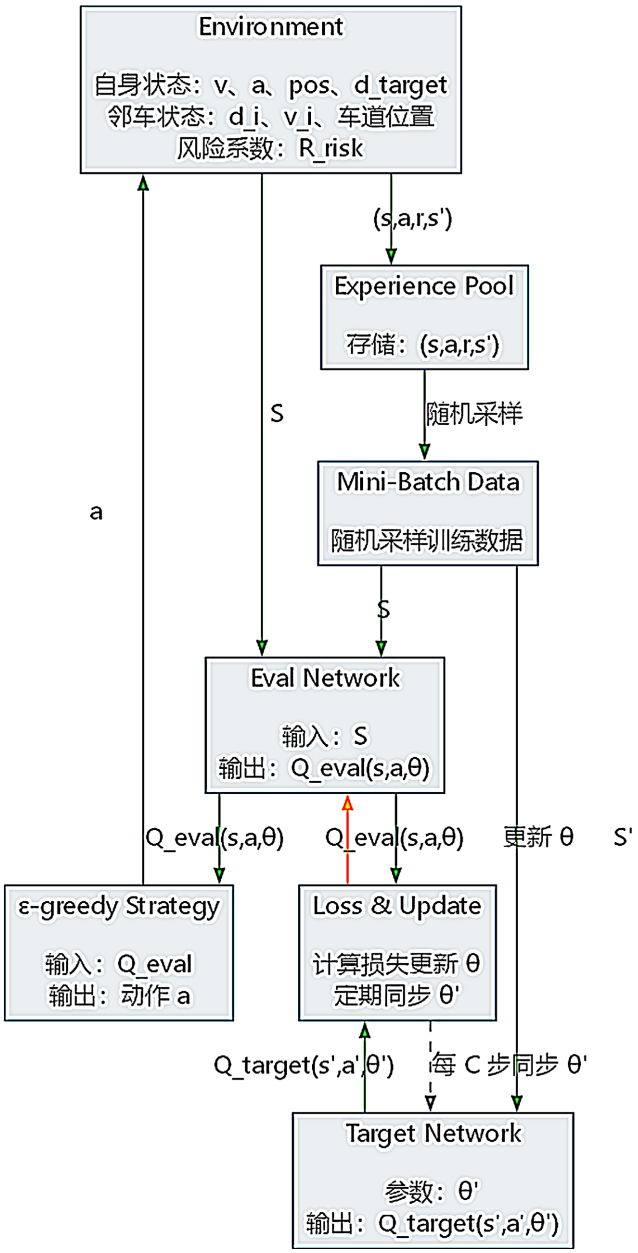

Figure 1. DQN algorithm diagram
图 1. DQN 算法原理图

参数优化：主网络通过最小化损失函数更新参数 θ：

$$L(\theta)=\mathbb{E}\left[\left(y_t-Q(S_t,a_t;\theta)\right)^2\right] \tag{5}$$

采用 ϵ-greedy 策略 ϵ 从 0.9 线性衰减至 0.1，平衡探索与利用。

4. 仿真实验

4.1. 实验设置

为在复杂且动态的交通环境中充分验证本文所提算法的有效性，我们在 SUMO 仿真平台上构建了一

个高挑战性的双向四车道高速公路场景。该场景的核心设置如下：高密度交通流，场景内共包含 100 辆随机行驶的社会车辆，以模拟繁忙时段的交通状况，增加决策的复杂性。

为模拟真实道路中驾驶员行为的多样性和不确定性，社会车辆被预设为三种驾驶风格：激进型(50%)、普通型(30%)和保守型(20%)。其中，高比例的激进型驾驶员会产生更多突发的、危险的驾驶行为，对智能体的风险感知和应对能力构成严峻考验。

任务目标：智能体需在 5 公里的路程内，安全、高效地完成 3 次指定的车道变换。

评价指标：平均奖励(Average Reward)，综合反映智能体策略的长期回报和总体性能；碰撞率(Collision Rate)，每公里发生的碰撞次数，是衡量安全性的核心负面指标；变道成功率(Success Rate)：成功完成目标变道的次数占比，衡量任务执行效率。

实验的科学性：为确保结论的统计显著性和可复现性，所有对比算法均在此环境下独立重复运行 5 次，最终结果以平滑处理后的均值曲线和标准差范围(阴影区域)的形式呈现。

4.2. 验结果与分析

为全面评估本文算法的性能，我们选取了三种主流的深度强化学习方法作为对比基线：DQN (w/Replay)、Double DQN (w/Replay)和 Dueling DQN (w/Replay)。所有算法均在同一高挑战性环境下进行训练和评估。

4.2.1. 合性能对比分析

四种算法的核心性能指标对比，包括平均奖励、碰撞率、换道成功率 3 个综合指标(5 次运行均值，平滑处理)。

从平均奖励可以看出，本文提出的融合风险感知的算法(My Algorithm (Proposed))展现出显著的优越性。本次实验的核心目标在于验证算法在复杂交通环境下的学习稳定性与策略鲁棒性，这对于安全攸关的自动驾驶系统至关重要。见图 2 所示的平均奖励曲线，平均奖励的绝对值上，在 350 轮之前本文算法优于其他三种算法，350 轮后总体差异并不明显，而在不同算法学习过程的方差中(由阴影区域表示)从图中可以看出，500 轮之后 Double DQN 及 Dueling DQN 这两种基线算法虽然取得了可观的平均奖励，但其学习过程表现出极高的不稳定性。其宽阔的方差区间表明，这些算法的最终性能高度依赖于随机的初始条件和动态的交通环境。通过碰撞率与变道成功率的仿真结果分析，这证明了通过感知和理解驾驶风格风险，智能体能够学习到一种获得更高长期回报的卓越策略。

碰撞率是评估安全性的核心指标。实验结果证明了本文算法在安全性上的巨大优势。见图 3 所示，本文算法的碰撞率在整个训练过程中均显著低于所有基线算法，并在训练后期基本收敛于 0.1 的水平。这强有力地表明，我们提出的风险感知模块 R_{risk} 能够让智能体有效识别并规避由激进驾驶员带来的潜在碰撞风险，从而实现了更高级别的主动安全。

变道成功率则反映了算法在保证安全前提下的决策效率。见图 4 所示，明显能观察到，在训练初期，所有算法的成功率都因环境的挑战性而有所下降。然而，在约 200 回合后，所有基线算法都陷入了难以提升的平台期，而本文算法的成功率开始逆势上扬，最终达到了近 90%的最高水平。这证明了本文算法并非通过保守，即放弃变道来换取安全，而是在对环境风险精准判断的基础上，更自信、更果断地抓住安全的变道时机，成功实现了安全性与效率的双重优化。

4.2.2. 奖励组件深入分析

为了深入探究本文算法取得优势的内在原因，我们对其学习过程中的奖励分量安全奖励、效率奖励、

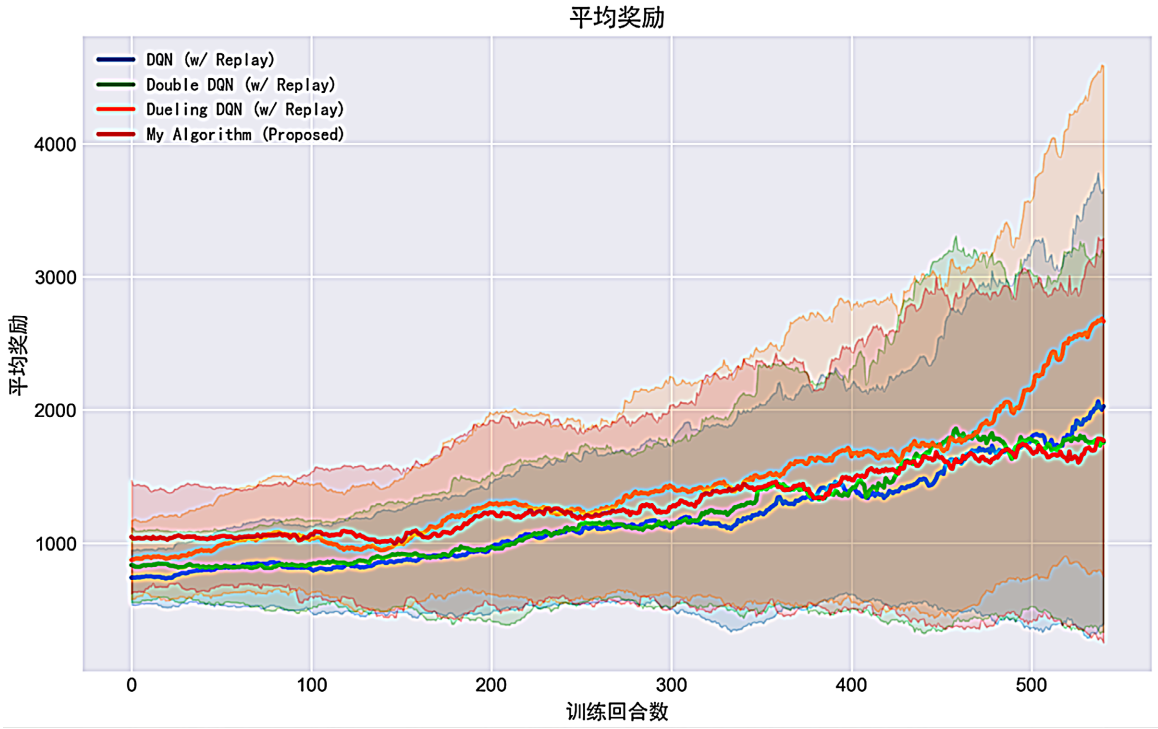

Figure 2. Comparison of average reward values of four algorithms
图 2. 四种算法的平均奖励值对比

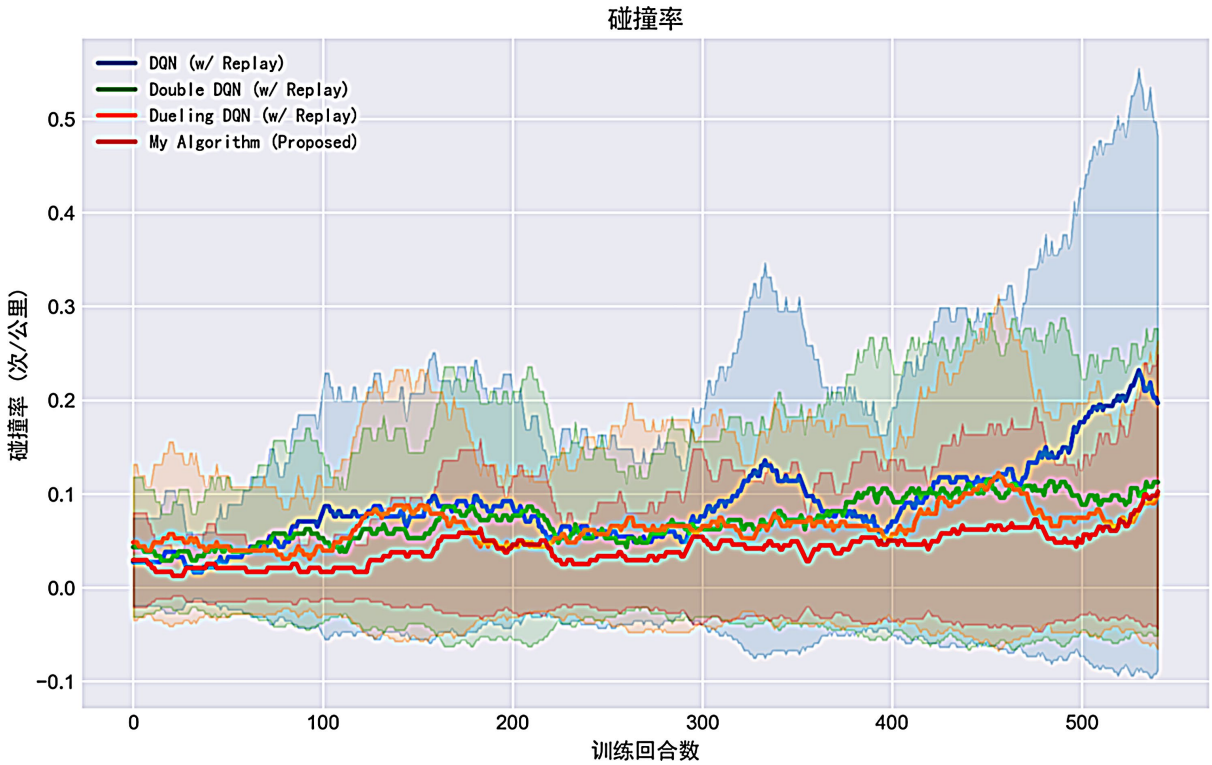

Figure 3. Comparison of collision rates of four algorithms
图 3. 四种算法的碰撞率对比

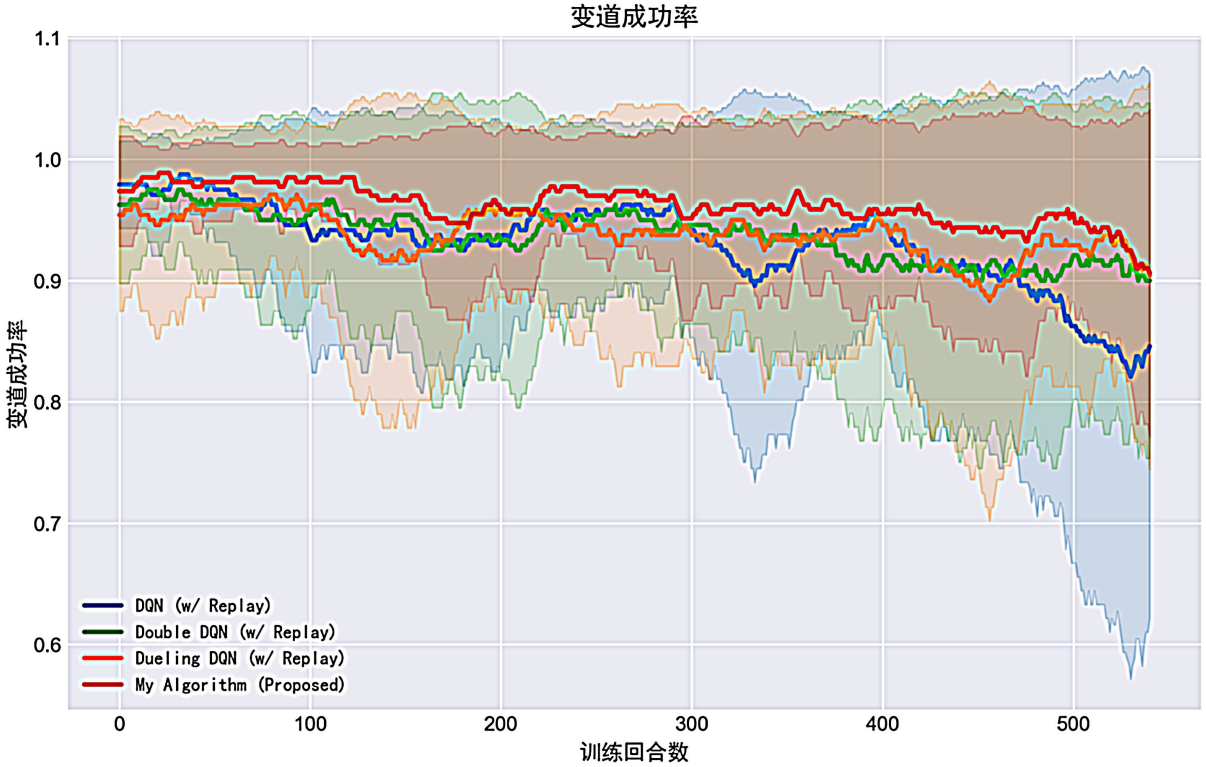

Figure 4. Comparison of the lane-changing success rates of four algorithms
图 4. 四种算法的变道成功率对比

规则奖励实验仿真结果进行详细的分析进行了分析。

安全奖励的分析揭示了一种高级的策略演进。见图 5 所示，本文算法在训练后期，其策略目标从最大化单一安全指标，转向了优先保障全局任务的完成。它学会了为实现更高的任务成功率而接受“可计算的风险”，这是过度保守的 Dueling DQN 未能实现的有效权衡。此外，算法后期扩大的方差反映了其更高的策略复杂度，而非不稳定性。其核心优势在于维持了更高的“安全下限”，即最差情况下的表现依然安全；相比之下，Dueling DQN 的策略则表现出偶发灾难性失效的脆弱性。本质上，我们的算法展现了卓越的多目标优化能力，学会了主动地管理风险，而不仅仅是规避风险。

效率奖励的对比结果尤为引人注目。见图 6 所示，本文算法的效率奖励曲线在短暂的探索后，形成了一条完美的、持续上扬的收敛曲线，最终远远甩开了所有基线算法。基线算法的效率则在波动后趋于一个较低的平庸水平。这一结果的内在逻辑是：基线算法因无法准确判断风险，常常被迫在不合适的时机减速或长时间跟驰，导致通行效率低下；而本文算法则能凭借其风险性预判，看清路况并选择进入风险更低、通行更顺畅的车道，从而能够以更高的平均速度行驶，获得了最高的效率奖励。

规则奖励综合反映了任务完成情况和行为的合规性。见图 7 所示，本文算法的规则奖励同样是最高的，这说明它能更频繁、更顺利地完成任务目标，获得任务成功的正奖励，同时因其更安全的驾驶行为而受到更少的负面惩罚。

综上所述，本文算法的优越性是系统性的。其核心的风险感知能力首先保证了安全性的提升；而卓越的安全性又赋予了智能体在高密度车流中进行高效决策的“信心”和“能力”，从而带来了效率和任务成功率的提升；最终，安全与效率的协同优化，使其获得了远超所有基线算法的综合平均奖励。

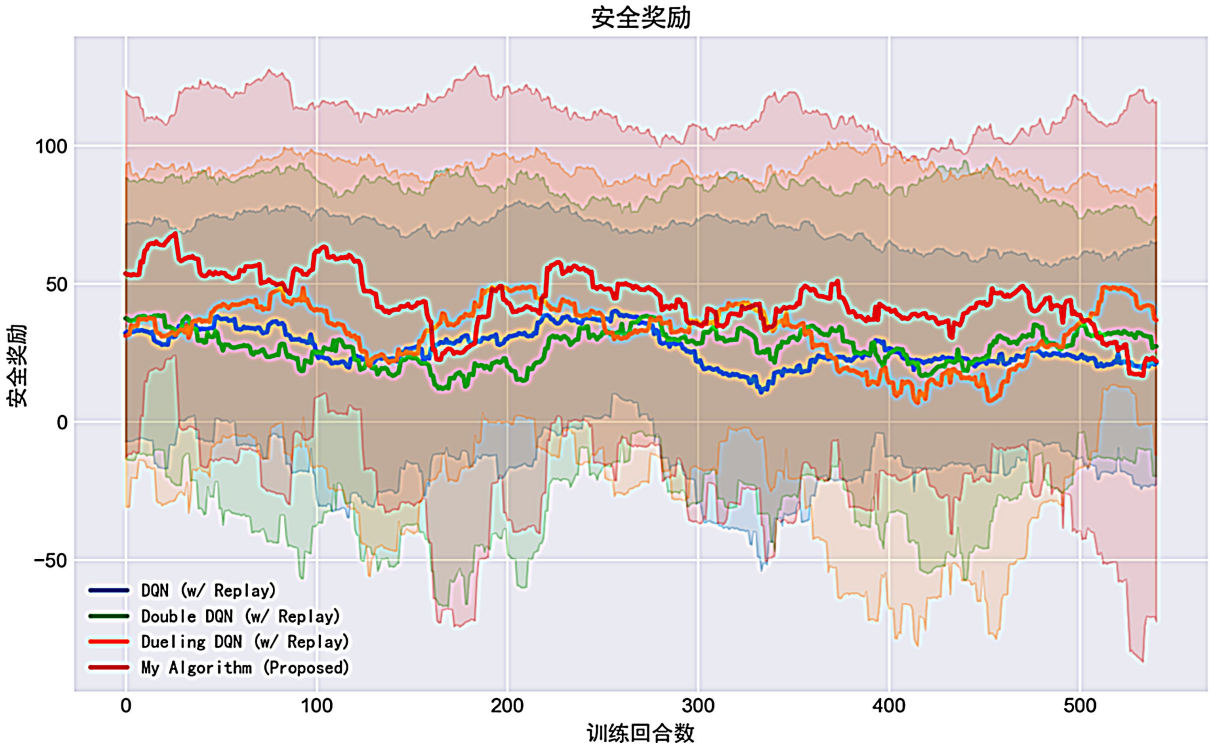

Figure 5. Comparison of safe rewards for four algorithms
图 5. 四种算法的安全奖励对比

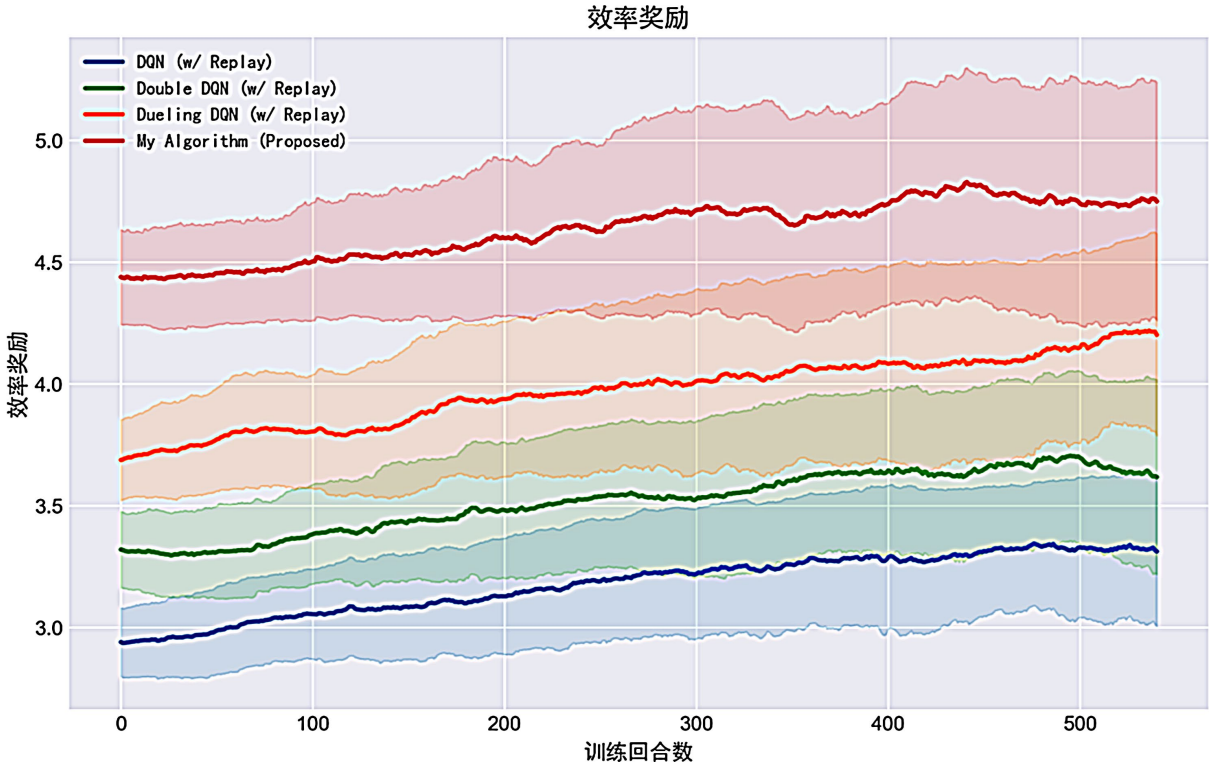

Figure 6. Comparison of efficiency rewards of four algorithms
图 6. 四种算法的效率奖励对比

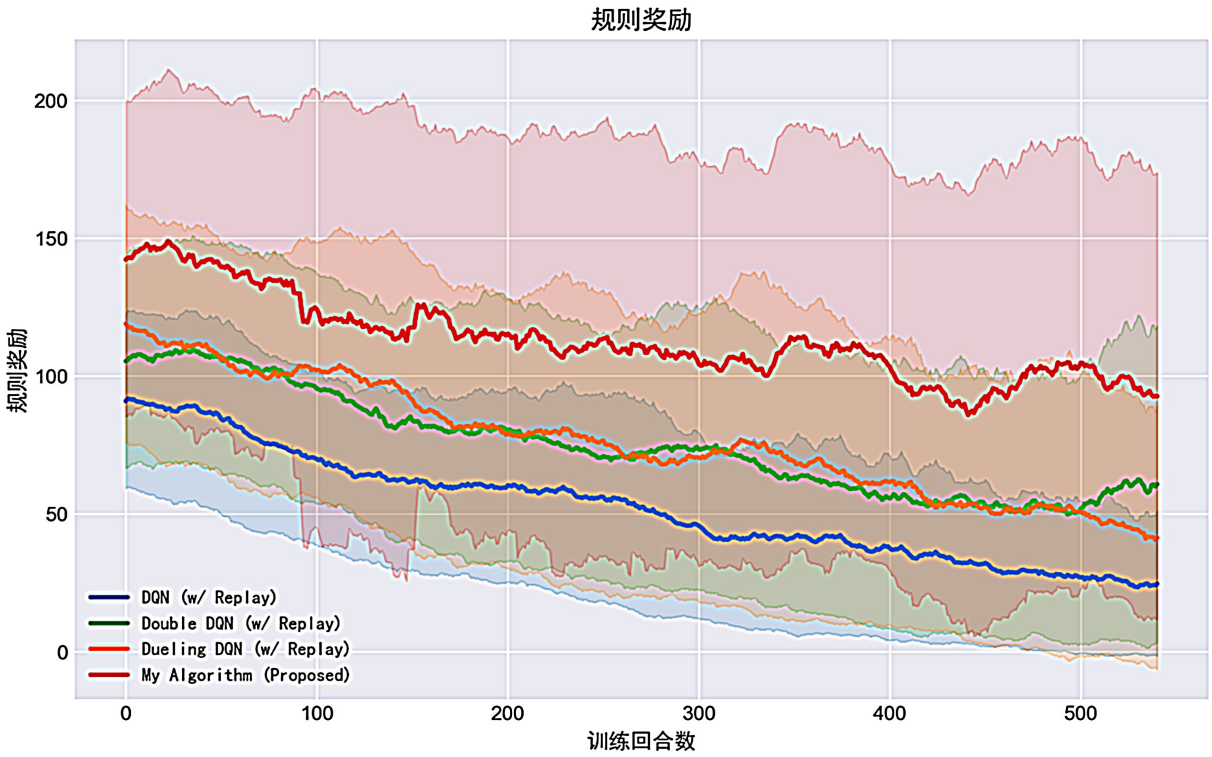

Figure 7. Comparison of rule rewards of four algorithms
图 7. 四种算法的规则奖励对比

4.3. 性能指标对比

为更直观地量化各算法的最终性能，我们将训练最后 100 个回合的关键指标进行了平均，结果如表 1 所示。

Table 1. Comparison table of the final performance indicators of four algorithms
表 1. 四种算法最终性能指标对比表

算法	变道成功率	碰撞率(次/公里)	平均奖励
DQN (w/Replay)	85.75 ± 34.96	18.18 ± 44.77	1895.96 ± 2281.81
Double DQN (w/Replay)	91.00 ± 28.62	10.08 ± 32.20	1773.46 ± 1982.29
Dueling DQN (w/Replay)	92.00 ± 27.13	8.28 ± 28.18	2387.14 ± 2286.67
本文算法(Proposed)	92.80 ± 24.42	7.76 ± 26.45	1732.61 ± 2140.01

在综合平均奖励方面，Dueling DQN 凭借其高效的价值函数结构，获得了最高的数值。然而，值得注意的是，其高奖励伴随着比本文算法更高的碰撞风险。相比之下，本文算法虽然在平均奖励上不占优势，但它实现了一种在自动驾驶领域中更为理想的性能平衡。它通过其独特的风险感知能力，主动规避了那些虽然可能带来高回报、但同样伴随着高碰撞风险的激进决策。因此，本文算法学习到的是一种将安全置于首位的、更具鲁棒性的驾驶策略，这从其全场最低的碰撞率指标中得到了充分的证明。这种以牺牲部分极限奖励为代价换取决定性安全提升的策略，更符合真实世界自动驾驶的应用要求。

4.4. 参数敏感性分析

为验证本文所选奖励函数超参数($\alpha = 0.5$，$\beta = 0.3$，$\gamma = 0.2$)的合理性，我们对其中最为关键的安全

权重α和效率权重β进行了参数敏感性分析。我们采用“独立因子观点”的控制变量法，即在分析一个权重时，保持其他权重不变。实验在一个简化的环境中进行，独立重复运行 3 次，以最终的平均奖励和碰撞率作为核心评价指标。实验结果总结如表 2 所示。

从表 2 对安全权重α的分析中可以看出，α的取值对智能体的安全性能有决定性影响。当α从 0.3 提升至 0.5 时，碰撞率大幅下降了约 31.7%，证明了强化安全信号的有效性。然而，当α被过度放大至 0.7 时，智能体表现出过度保守的行为倾向，虽然其平均奖励有所提高，但其策略的鲁棒性下降(标准差增大)，且碰撞率不降反升。因此，$\alpha = 0.5$被确认为能够在最大化安全(实现最低碰撞率)与保证任务完成效率之间取得最佳平衡的优选值。

类似地，对效率权重β的分析揭示了效率与安全之间的经典权衡关系。随着β值的升高，智能体的驾驶风格变得更为激进，导致碰撞率呈现单调上升的趋势。有趣的是，当β从 0.3 增加到 0.5 时，平均奖励并未继续增长，反而有所下降。这表明，过高的效率激励会诱使智能体采取高风险策略，从而因更频繁的碰撞惩罚而损害其长期总体收益。因此，$\beta = 0.3$被证明是在激励高效驾驶与控制安全风险之间的最佳平衡点。

Table 2. Sensitivity analysis results of reward weights α and β
表 2. 奖励权重α和β的敏感性分析结果

变化参数	参数组合	平均奖励	奖励标准差	碰撞率	碰撞率标准差
	$\alpha=0.3,\beta=0.3,\gamma=0.2$	281.076	9.37642	0.0409357	0.00472586
α	$\alpha=0.5,\beta=0.3,\gamma=0.2$ (本文)	375.383	30.311	0.0284043	0.00312586
	$\alpha=0.7,\beta=0.3,\gamma=0.2$	456.928	39.8233	0.0434419	0.00472586
	$\alpha=0.5,\beta=0.1,\gamma=0.2$	345.604	8.3405	0.0426065	0.00409271
β	$\alpha=0.5,\beta=0.3,\gamma=0.2$ (本文)	375.383	30.311	0.0284043	0.00312586
	$\alpha=0.5,\beta=0.5,\gamma=0.2$	366.829	36.0179	0.047619	0.0141776

综上所述，本文所选的超参数组合($\alpha = 0.5$，$\beta = 0.3$，$\gamma = 0.2$)并非随意设定，而是通过参数敏感性分析验证的、能够使模型综合性能达到最优的合理选择，从而增强了本文所提模型设计的严谨性。

5. 结论

本文提出的带重放记忆的 DQN 算法通过融合驾驶风格感知、设计多目标奖励函数并结合经验回放机制，有效提升了智能体变道决策的安全性与效率。通过详尽的参数敏感性分析，我们验证了模型超参数选择的合理性，进一步增强了设计的严谨性。实验表明，该算法能适应不同驾驶风格的交通环境。实验表明，该算法能适应不同驾驶风格的交通环境，减少碰撞风险并提高通行速度，为自动驾驶动态决策提供了可行方案。未来的研究可从以下方面展开：1) 将当前的单智能体决策框架扩展为多智能体协作模型，研究在匝道汇入、无信号交叉口等典型冲突场景下的博弈与协调策略；2) 探索基于逆强化学习(Inverse Reinforcement Learning)的方法，从人类驾驶数据中自动学习奖励函数的权重，以替代当前的手动调参，从而实现更高级别的自适应能力。

参考文献

[1] 蒲龙忠. 驾驶员驾驶车辆变道行为原因综述[J]. 交通科技与管理, 2021(17): 215-215+94.

[2] Yu, S.J., Ma, C. and Chen, J.Z. (2023) Research Progress of Automatic Driving Lane Change Decision Algorithms Based on Learning. *Automobile Applied Technology*, **48**, 189-194.

[3] 程诺. 基于改进优先经验回放的 DDPG 路径规划算法研究[D]: [硕士学位论文]. 济南: 山东交通学院, 2024.

[4] 张斌, 何明, 陈希亮, 等. 改进 DDPG 算法在自动驾驶中的应用[J]. 计算机工程与应用, 2019, 55(10): 264-270.

[5] 裴晓飞, 莫烁杰, 陈祯福, 等. 基于 TD3 算法的人机混驾交通环境自动驾驶汽车换道研究[J]. 中国公路学报, 2021, 34(11): 246-254.

[6] Grigorescu, S., Trasnea, B., Cocias, T. and Macesanu, G. (2020) A Survey of Deep Learning Techniques for Autonomous Driving. *Journal of Field Robotics*, **37**, 362-386. https://doi.org/10.1002/rob.21918

[7] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., *et al.* (2015) Human-Level Control through Deep Reinforcement Learning. *Nature*, **518**, 529-533. https://doi.org/10.1038/nature14236

[8] Kiran, B.R., Sobh, I., Talpaert, V., Mannion, P., Sallab, A.A.A., Yogamani, S., *et al.* (2021) Deep Reinforcement Learning for Autonomous Driving: A Survey. *IEEE Transactions on Intelligent Transportation Systems*, **23**, 4909-4926. https://doi.org/10.1109/tits.2021.3054625

[9] Fosgerau, M., Melo, E., de Palma, A. and Shum, M. (2020) Discrete Choice and Rational Inattention: A General Equivalence Result. *International Economic Review*, **61**, 1569-1589. https://doi.org/10.1111/iere.12469

[10] Mnih, V., Kavukcuoglu, K., Silver, D., *et al.* (2013) Playing Atari with Deep Reinforcement Learning. arXiv:1312.5602.