动态场景下自动驾驶运行时安全保障的自适应方法

徐丙凤[1]，陈嘉玲[1]，杨帅领[1]，何高峰[2]
（1. 南京林业大学信息科学技术学院、人工智能学院，江苏 南京 210037；2. 南京邮电大学物联网学院，江苏 南京 210003）

摘　要：针对已有自动驾驶运行时安全控制方法难以根据车辆实际运行环境进行动态调整导致车辆通行效率降低的问题，提出了一种动态场景下自动驾驶运行时安全保障的自适应方法，给出了运行时安全自动控制模型（RTA-AutoSafe）。在该模型中，针对性能控制器提出了一种基于自适应双缓冲优先经验重放机制的深度Q网络算法，通过增强动态交通环境下决策策略的适应性以优化通行效率；针对实际驾驶动态特征设计了自适应责任敏感安全（ARSS）模型，以增强安全判定的动态适应性；同时基于ARSS模型构建了一种结合车辆实时反馈和交通自适应的动态双向切换逻辑和安全控制器，用以实现车辆的实时安全保障和双控制器间的动态调控。仿真实验结果表明，与其他安全控制方法相比，所提方法在动态交通环境中降低了安全冗余控制对通行效率的限制，实现了安全实时响应与高效运行策略的动态兼容。

关键词：自动驾驶；深度强化学习；运行时保证；责任敏感安全模型；优先经验重放
中图分类号：TP391
文献标志码：A
DOI：10.11959/j.issn.1000–436x.2025120

Adaptive method runtime safety assurance for autonomous driving in dynamic scenarios

XU Bingfeng[1], CHEN Jialing[1], YANG Shuailing[1], HE Gaofeng[2]
1. College of Information Science and Technology & College of Artificial Intelligence, Nanjing Forestry University, Nanjing 210037, China
2. School of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

Abstract: An adaptive method runtime safety assurance for autonomous driving in dynamic scenarios was proposed to address the issue that existing safety control approaches lack the ability to adjust dynamically to actual driving environments, resulting in reduced traffic efficiency. A runtime safety assurance autonomous control model (RTA-AutoSafe) was presented for autonomous driving. In this model, a deep Q-network algorithm based on an adaptive dual-buffer prioritized experience replay mechanism was designed for the performance controller to enhance the adaptability of decision-making strategies in dynamic traffic environments and to optimize traffic efficiency. An adaptive responsibility-sensitive safety (ARSS) model was designed based on vehicle dynamics to improve the adaptability of safety assessments. Based on this ARSS model, a dynamic bidirectional switching logic and safety controller integrating real-time vehicle feedback and traffic adaptation was constructed to achieve real-time safety assurance and dynamic coordination between dual controllers. Simulation results show that, compared with other safety control methods, the proposed method reduces the limitation of safety redundancy control on traffic efficiency in dynamic environments and achieves dynamic compatibility between real-time safety response and efficient operational strategies.

Keywords: autonomous driving, deep reinforcement learning, runtime assurance, responsibility sensitive safety model, prioritized experience replay

收稿日期：2025–04–23；**修回日期：**2025–06–23
通信作者：何高峰，hegaofeng@njupt.edu.cn
基金项目：国家自然科学基金资助项目（No.62372240, No.62372239）；江苏省网络与信息安全重点实验室基金资助项目（No.BM2003201）
Foundation Items：The National Natural Science Foundation of China (No.62372240, No.62372239), Jiangsu Provincial Key Laboratory of Network and Information Security (No.BM2003201)

0　引言

自动驾驶是汽车产业与人工智能、物联网、高性能计算等新一代信息技术深度融合的产物[1]。依据美国汽车工程协会标准，自动驾驶技术可划分为L0~L5共6个级别[2]。当前各大汽车厂商主要采用L2级别的高级驾驶辅助系统（ADAS, advanced driver assistance system）。在L2级别自驾中，车辆主要提供诸如闸机自动通行、停车场寻位泊车等辅助驾驶功能，复杂路况以及紧急情况处理仍需要由驾驶员进行决策和操控。未来在L3以及更高级别的自动驾驶中，车辆需要在复杂的道路环境中完全依靠自身的决策系统进行驾驶而无须驾驶员干预，这种高度自动化的驾驶模式，对自动驾驶系统的安全性与可靠性有着更高的要求[3]。因此，需要设计有效的自动驾驶系统安全保障方法，确保车辆在复杂路况下能够自主识别并解决行驶安全威胁。

为了保障自动驾驶控制决策的安全性，研究人员尝试在自动驾驶的决策训练过程中引入安全约束目标[4-5]，即确保驾驶安全所必须满足的规则和限制。例如，文献[6]提出一种并行约束策略优化算法，将强化学习问题转化为约束优化问题，引入预期风险函数防止策略更新超出可接受风险水平，确保自动驾驶系统在学习和执行过程中不超出安全限制。文献[7]提出了一种基于先验知识的风险项并将其纳入价值函数，通过修改奖励函数惩罚不安全行为，并限制策略在风险区域的探索，以提高策略的安全性。这些方法通过优化奖励结构或策略更新过程来管理风险，能够降低安全风险发生的概率，属于软安全约束的范畴[8]。而对于自动驾驶这类存在高风险和关键任务的领域，需要严格的硬安全约束和确定性的安全保障。为此，文献[9-10]引入了控制理论中的控制障碍函数（CBF, control barrier function），并将其作为安全约束集成到强化学习的训练过程中以改变策略更新方向，通过限制训练过程中智能体的动作探索或状态演变，防止系统进入不安全状态。但这类方法都是通过在基于学习的控制算法的基础上增加安全约束来保证系统的安全性，在实际应用中通常仅适用于单一控制任务和场景[11]。随着驾驶场景时空复杂性的提升，安全约束的设计需覆盖更广的潜在场景与状况，导致约束的数量和形式显著增加。这不仅增大了学习过程中的计算复杂性，还可能在策略优化中引发任务冲突或不稳定现象，难以在多任务、多场景的动态环境中自适应。因此，基于安全约束的强化学习方法在应对高度复杂且多变的驾驶任务时仍面临挑战。

为此，目前已有一部分研究工作将基于运行时保证（RTA, runtime assurance）思想[12-13]的Simplex架构[14]应用到自动驾驶领域，通过在智能化系统的运行过程中提供额外的监控和备用功能，修正其可能带来的潜在风险确保系统的行为始终处于安全范围内。例如，文献[15]针对自动驾驶汽车安全速度调节问题，整合牵引力控制系统和防抱死制动系统提出了S$\mathscr{L}$1-Simplex框架，进行纵向车辆避障控制。文献[16]提出的基于RTA的高性能与安全保障的简单系统驱动框架（Simplex-Drive），采用最优互避碰撞法作为备用控制器，基于速度大小和方向实现了分布式底层避障。这些方法虽然增强了系统的安全性，但是在引入基于Simplex架构的RTA技术后，自动驾驶车辆的通行效率受到影响。为缓解车辆通行效率降低的问题，已有研究通过基于预设参数的安全模型来设计性能与安全控制器间的双向切换逻辑，以平衡车辆效率和安全。例如，文献[17]使用行车风险场模型设计切换逻辑，通过预设的安全场强阈值和车辆之间的实时场强值判断车辆是否需要切换控制器。文献[18]采用明确动态车辆间安全和危险界限的责任敏感安全（RSS, responsibility sensitive safety）模型来设计性能和安全控制器之间的双向切换逻辑。然而，这类固定安全参数设定下的切换逻辑对于实时变化的动态场景的响应能力有限，使切换逻辑难以根据实时驾驶场景自适应，会在规避预期风险时过度约束车辆性能，从而制约行驶效率。

因此，为使运行时安全保障方法能够依据车辆实际运行环境进行动态自适应调整以提升车辆的通行效率和安全，本文提出了一种动态场景下自动驾驶运行时安全保障的自适应方法。以运行时协同控制为核心思路，构建了包括性能控制器、安全控制器以及监控与决策模块的运行时安全自动控制模型（RTA-Autosafe, runtime safety assurance autonomous control model），融合了专家行为引导的策略学习机制、自适应安全边界计算与动态控制权切换机制，实现性能与安全的平衡。本

文的贡献如下。

1) 设计了一种自动驾驶运行时安全自动控制模型，并对该模型进行了形式化定义。该模型通过性能控制模块和安全控制模块的实时协同，并结合基于环境感知的动态双向切换逻辑，能够在动态的交通环境中自适应调整安全约束边界与控制策略，实现自动驾驶运行时安全与效率的双重保障。

2) 在RTA-AutoSafe的性能控制器设计中，提出一种基于自适应双缓冲优先经验重放机制的深度Q网络算法（E-DQN），通过双缓冲经验重放机制的设计提升策略学习的效率与稳定性，以获得对动态环境具有更强适应性的控制策略。

3) 在RTA-AutoSafe的安全控制器的设计中，融合车辆响应特性和车辆密度提出了一种自适应责任敏感安全（ARSS）模型以动态计算安全边界，并在此基础上设计动态双向切换逻辑和安全控制器指令生成算法，在保证车辆安全的情况下提升通行效率，减少引入外部安全机制后对驾驶性能带来的影响。

4) 仿真实验结果表明，在不同车辆密度下，RTA-AutoSafe的碰撞率显著低于其他对比方法，特别是在高车辆密度环境下，能有效保障车辆安全；性能控制器能够在保持较低碰撞率的同时维持较长的回合时间和较高的平均速度，通行效率得到保障；在不同车辆密度下动态双向切换逻辑在不影响车辆安全的同时均减少了安全控制器的介入时间，提升了车辆的通行效率。

1　RTA-AutoSafe

本节首先提出了一个自动驾驶运行时安全自动控制模型，然后详细介绍了该模型的3个关键模块：性能控制器、安全控制器以及监控与决策模块，并对各模块中的核心设计进行了详细说明。

1.1　RTA-AutoSafe架构

RTA-AutoSafe主要由性能控制器、安全控制器、监控与决策模块和被控对象组成，如图1所示。该模型旨在通过性能控制器优化通行效率，并通过安全控制器在关键时刻确保车辆遵守安全约束，从而实现被控对象性能与安全的动态平衡。为了更清晰地对模型构件及其关系进行描述，下面对该模型进行形式化定义。

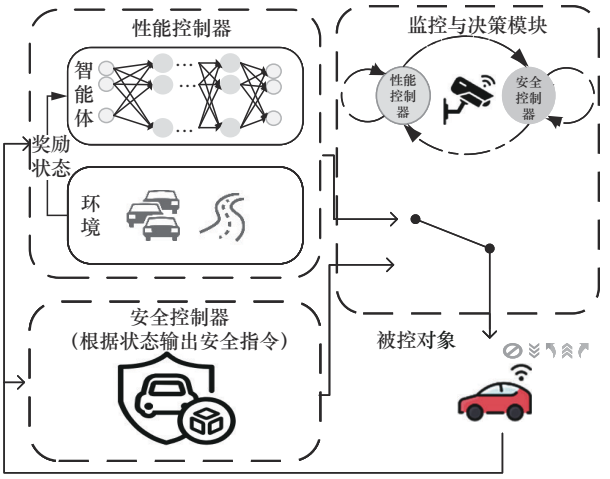

图1　RTA-AutoSafe架构

定义1　RTA-AutoSafe是一个四元组$M=\langle \mathrm{PC},\mathrm{SC},V,\mathrm{MD}\rangle$，各变量含义如下。

1) PC表示性能控制器，其目标是生成优化的控制信号a^{PC}，以提高车辆的通行效率。

2) SC表示安全控制器，其目标是输出控制信号a^{SC}，以确保车辆始终遵守安全约束，避免交通事故发生。

3) V表示被控对象（自动驾驶车辆）。车辆的状态表示为$\boldsymbol{S}=\left[x,y,v^x,v^y\right]^{\mathrm{T}}$，$x$和$y$分别表示车辆在纵向方向和横向方向上的位置，$v^x$和$v^y$分别表示车辆在纵向方向和横向方向上的速度。车辆的动作表示为A，车辆动态遵循运动学模型[19]，将车辆的状态变化表示为$\boldsymbol{S}_{t+\Delta t}=f\left(\boldsymbol{S}_t,A_t\right)$，其中$\Delta t$表示一个控制周期，$\boldsymbol{S}_t$和$\boldsymbol{S}_{t+\Delta t}$分别表示当前车辆状态和经过当前动作指令控制$A_t\in\left\{a_t^{\mathrm{PC}},a_t^{\mathrm{SC}}\right\}$后的车辆状态。

4) MD表示监控与决策模块，由监控器和决策模块组成。监控器监控车辆状态和环境信息，实时评估驾驶安全；决策模块基于监控评估结果和切换逻辑，在必要时进行PC和SC之间的切换，实现性能与安全的动态平衡。

RTA-AutoSafe通过MD实时评估车辆状态和环境信息，判断在当前状态下使用PC输出的控制信号是否能保持车辆安全。如果安全，则使用PC输出的控制信号进行车辆驾驶；如果无法保证安全，则切换至SC。通过这种机制，RTA-AutoSafe对性能与安全进行动态权衡，确保自动驾驶车辆能在始终满足安全约束的同时提高通行效率。

1.2　性能控制器设计

性能控制器的设计通常基于强化学习框架，智能体将在探索过程中产生的状态、动作、奖励等信息存储于经验池中，通过随机采样的方式训练模型，以手动方式设计奖励函数，主要针对固定环境或场景定义明确的优化目标[17,20]。这类设计方法难以有效应对动态驾驶环境，因此，本节设计了一种新的性能控制器，具体的设计框架如图2所示。

该性能控制器的主要设计思想是通过自适应双缓冲优先经验重放机制，增强自动驾驶系统在动态交通环境中的适应性。具体而言，性能控制器引入双重经验缓冲区动态融合专家经验和实时探索经验数据，利用优先经验重放（PER, prioritized experience replay）机制提升关键经验的利用率，并结合碰撞经验优先增益和实时经验优先增益提高策略适应性。此外，通过逆软Q学习（IQ-Learn, inverse soft-Q learning）算法[21]从专家数据中学习软Q函数和策略网络，并通过软贝尔曼方程隐式推导奖励函数$r(s,a)$。

$$r(s,a)=Q_{\text{soft}}(s,a)-\gamma \mathrm{E}_{s'}\left[\ln\sum_{a'}\exp\left(Q_{\text{soft}}(s',a')\right)\right] \tag{1}$$

其中，$Q_{\text{soft}}(s,a)$为软Q函数，γ是折扣因子，s和a分别表示当前状态和动作，s'和a'分别表示下一状态和动作。

1.2.1　自适应双缓冲优先经验重放机制

受软Q模仿学习（SQIL, soft Q imitation learning）[22]中混合专家演示和智能体自身探索经验相结合的学习理念的启发，本节设计了如图3所示的双重经验缓冲结构，该结构包括一个实时经验池B_{online}和一个动态更新的专家经验池B_{expert}，分别用于存储智能体自主探索过程中生成的经验数据与从专家控制器中获取的次优演示数据。为实现专家知识与自适应策略优化的高效融合，在专家经验池中引入了自适应更新机制：专家经验池除了存储初始专家经验外，在智能体探索过程中，当实时经验池中产生的某条经验的奖励值高于当前专家经验池的平均水平时，该经验将作为实时专家经验被写入专家经验池，并替换其中奖励最低的经验样本。通过这种奖励驱动的经验更新策略，在保留专家经验对策略学习引导的同时，引入了智能体自主探索的高价值动态补充，有助于提高控制策略对复杂环境变化的适应能力和整体决策效率。在双重经验缓冲区中，一条经验数据通常代表一次状态转移过程，本文将经验数据e_t形式化表示为

$$e_t=(s_t,a_t,r_t,s_{t+\Delta t},\text{done}_t,\text{id}) \tag{2}$$

其中，id为该经验产生时的回合数，$\text{done}_t\in\{\text{FALSE},\text{TRUE},\text{crashed}\}$为回合结束状态，默认为FALSE；如果在学习过程中某回合以碰撞结束，则

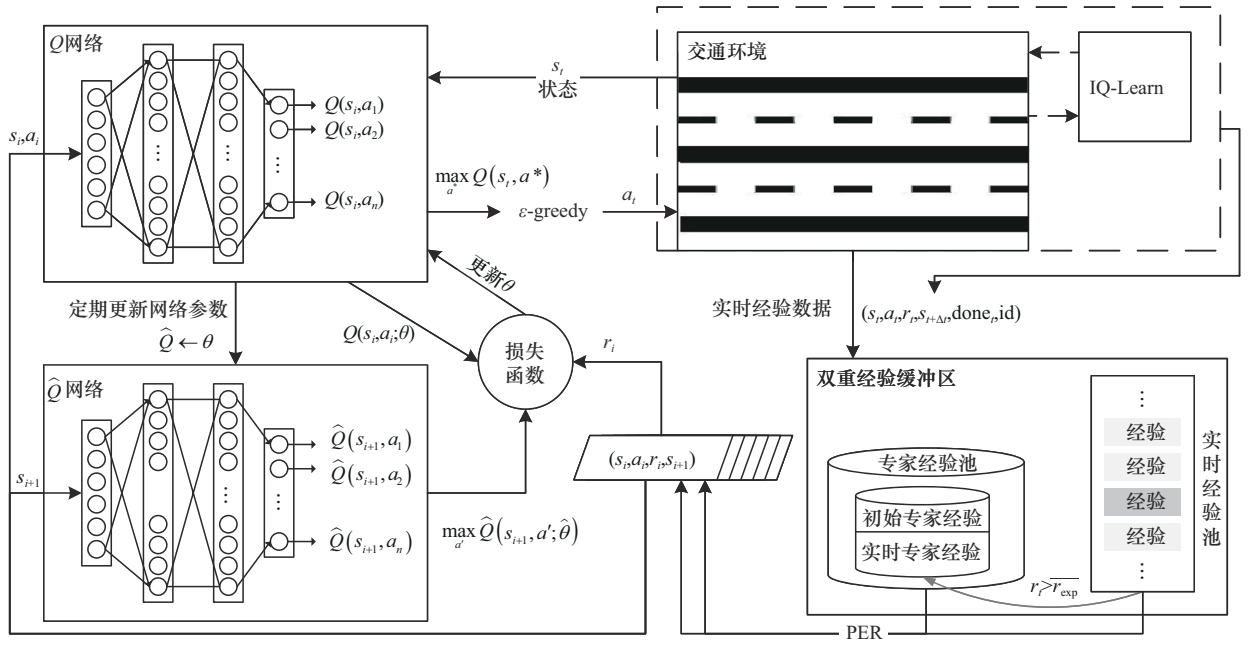

图2　性能控制器设计框架

设置 done_t = crashed；如果回合正常结束，则设置 done_t = TRUE。

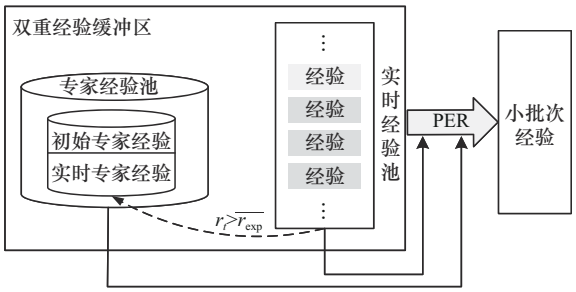

图3　双重经验缓冲结构

尽管双重经验缓冲区的动态更新机制能够筛选高质量实时经验，但在训练过程中不同经验对策略优化的价值存在显著差异。尤其当环境中出现碰撞事件或智能体获得异常高额奖励时，有关经验往往蕴含着关键的环境动态信息或策略优化方向，而传统的均匀采样方式难以充分发挥此类经验的学习价值。为此，本文引入了优先级经验重放机制[23]，通过增加关键经验的抽样概率，提升学习效率与模型适应性。在PER中，按照优先级对所有经验进行采样，这些优先级根据时序差分误差（TD Error, temporal difference error）确定。TD Error表示当前 Q 值与目标 Q 值之间的差值，其计算式为

$$\text{TD Error} = r + \gamma \max_{a'} \hat{Q}(s',a') - Q(s,a) \tag{3}$$

在自动驾驶控制任务中，存在一些TD Error不显著，但对安全性和策略优化至关重要的关键经验。为此，在PER的基础上，本节设计了2种新的经验优先增益，包括碰撞经验优先增益 η 和实时经验优先增益 ζ，以进一步优化经验的采样策略，突出关键经验的作用。碰撞经验优先增益 η 用于强调出现碰撞时有关经验的重要性，当经验缓冲区中存储进 done_t = crashed 的经验数据时，为对应回合的全部经验添加碰撞经验优先增益 η。

$$\eta_j = \frac{j}{n}, j = 1,2,\cdots,n \tag{4}$$

其中，j 表示出现碰撞回合的第 j 条经验。由式(3)和式(4)可得，实时经验池中第 i 条经验的优先级 p_i 计算式为

$$p_i = \left|\text{TD Error}_i\right| + \eta_i \tag{5}$$

为了进一步对专家经验池中经验数据的优先级进行计算，设计了实时经验优先增益 ζ，强调智能体自身探索的实时优秀高奖励经验的重要性，ζ 的计算式为

$$\zeta = \frac{1}{1 + \mathrm{e}^{-\lambda \cdot \text{id}}} \tag{6}$$

其中，λ 是时间衰减速率。由式(3)和式(6)可得，专家经验池中第 i 条经验的优先级 p_i^e 的计算式为

$$p_i^e = \left|\text{TD Error}\right| + \zeta_i \tag{7}$$

基于上述2种优先级计算方式，经验数据的采样概率 $P(i)$ 为

$$P(i) = \frac{p_i^\alpha}{\sum\limits_{m \in M} p_m^\alpha} \tag{8}$$

其中，M 表示当前用于采样的经验池集合（可为实时经验池或专家经验池），α 代表优先经验重放的使用程度。

1.2.2　E-DQN

针对深度Q网络（DQN, deep Q-network）算法中经验重放机制对关键经验利用率不足且策略适应性有限导致其在动态交通环境中适应性不足的问题，本节基于1.2.1节提出的双重经验动态更新、碰撞优先增益与实时优先增益，提出E-DQN，具体描述如算法1所示。

算法1　E-DQN

初始化双重经验缓冲区 $D = B_{\text{expert}} \cup B_{\text{online}}$，初始化主要 Q 网络的参数 θ，初始化目标 $\hat{Q}$ 网络的参数 $\hat{\theta} = \theta$，设置训练回合数 E 和单次训练回合最大步长 T

1)for 回合 episode = 1,2,⋯,E do

2)　迭代次数更新 episode = episode + 1

3)　更新当前回合数 id = episode

4)　重置环境获取初始状态 s_1

5)　for 步数 t = 1,2,⋯,T do

6)　给定状态 s_t，基于 ε-greedy 选择并执行动作 a_t

$$a_t = \begin{cases} \text{随机动作}, & \text{以概率}\varepsilon \\ \arg\max\limits_a Q(s_t,a;\theta), & \text{以概率}1-\varepsilon \end{cases}$$

7)　获得反馈 r_t、下一状态 s_{t+1} 和是否触发终止回合条件 done_t

8)　将实时经验数据 $(s_t,a_t,r_t,s_{t+1},\text{done}_t,\text{id})$ 存储到 B_{online} 中

9)　if $\text{done}_t \neq$ TRUE then $t = T$

10) end if

11) 计算B_{expert}中所有经验的平均奖励$\overline{r_{\text{exp}}}$

12) if $r_t > \overline{r_{\text{exp}}}$ then 将实时经验从B_{online}中转存到B_{expert}

13) end if

14) 根据式(5)和式(8)计算B_{online}中经验数据的优先级和采样概率，根据式(7)和式(8)计算B_{expert}中经验数据的优先级和采样概率

15) if NUM(D) > Batchsize then

16) 根据优先级从B_{online}采样N_1条经验数据，从B_{expert}采样N_2条经验数据

17) 对样本数据计算目标值

$$y_i = \begin{cases} r_i, t+1\text{时刻回合结束} \\ r_i + \gamma \max\limits_{a'} \hat{Q}\left(s_{i+1}, a'; \hat{\theta}\right), \text{其他} \end{cases}$$

18) 使用梯度下降优化包含权重ω的加权损失函数

$$L(\theta) = \frac{1}{m}\sum_{i=0}^{m} \omega_i \left(y_i - Q\left(s_i, a_i; \theta\right)\right)^2$$

19) end if

20) 每c个时间步更新目标$\hat{Q}$网络的参数: $\hat{\theta} \leftarrow \tau\theta + (1-\tau)\hat{\theta}$

21) end for

22)end for

算法1通过动态融合专家经验与实时探索经验数据，基于经验优先增益策略优化经验采样概率，增强了DQN在动态环境中的学习效率和模型适应性，从而更有效地生成优化的控制策略，提高自动驾驶车辆的驾驶效率。该算法主要包括初始化、环境交互、优先经验重放、模型更新4个步骤。

①初始化：设定主要Q网络和目标$\hat{Q}$网络的初始参数，初始化专家经验数据，使用$r(s,a)$重新计算并覆盖专家数据中每个状态−动作对的相应奖励，初始化双重经验缓冲区并预存专家经验；②环境交互：在当前状态下，智能体基于ε-greedy策略执行动作并收集交互数据，动态更新双重经验缓冲区，对应算法1中的步骤6)~步骤13)；③优先经验重放：结合碰撞增益与实时增益计算经验优先级，按权重采样训练数据，对应算法1中的步骤14)~步骤16)；④模型更新：基于步骤③的经验样本，计算TD Error优化网络参数，目标网络参数定期软更新保持稳定，对应算法1中的步骤17)~步骤20)。

下面从空间复杂度和时间复杂度2个方面对算法1进行分析。空间复杂度主要由存储经验数据和网络参数决定。双重经验缓冲区的空间复杂度为$O\left(B_{\text{expert}} + B_{\text{online}}\right)$，网络参数的空间复杂度为$O(W)$，其中$W$为网络参数的数量，其他变量如当前状态、动作、奖励等所需空间为常数级别$O(1)$，总体空间复杂度为$O(B_{\text{expert}} + B_{\text{online}} + W)$。时间复杂度主要由关键步骤的迭代操作决定。环境交互阶段的单步操作（动作选择、存储经验以及专家池更新判断）复杂度为$O(1)$，整体线性依赖于单回合长度T。优先经验重放阶段单次采样复杂度为$O(\text{lb}\,B_{\text{expert}} + \text{lb}\,B_{\text{online}})$，总时间复杂度与回合训练数$E$呈正比，最终整体规模为$O\left(E \times T \times \left(\text{lb}\,B_{\text{expert}} + \text{lb}\,B_{\text{online}}\right)\right)$。

1.3 基于ARSS模型的安全控制机制设计

为了保障自动驾驶车辆在动态场景下的行驶安全，本节设计了一种基于ARSS模型的安全控制机制。该机制由监控与决策模块和安全控制器组成，通过实时监控车辆状态与环境信息，动态评估驾驶安全性，并在识别到潜在风险时通过动态切换逻辑切换控制权至安全控制器以确保自动驾驶车辆的安全。为提升传统RSS模型在实际交通场景中的适应能力，ARSS模型引入了车辆实时加速度与车辆密度等因素，对安全距离进行动态调整，在保障安全的同时提升模型的动态适应性。

1.3.1 ARSS模型

为防止自动驾驶车辆发生交通事故，摩安视智能科技（Mobileye）公司提出了RSS模型，该模型给出了自动驾驶车辆必须保持的纵向最小安全距离$d_{\text{long}}^{\text{safe}}$和横向最小安全距离$d_{\text{lat}}^{\text{safe}}$[24]。然而，RSS模型中最小安全距离的计算方式基于最坏情况下预设的安全参数，未充分考虑车辆动态响应能力和交通的实时特征，导致模型的动态适应能力较差。因此，本节提出一种融合车辆实时动态参数与交通动态特征的ARSS模型，针对RSS模型的动态适应性做出调整，构建了如下所示的动态纵向最小安全距离$d_{\text{adaptive}}^{\text{long}}$和横向最小安全距离$d_{\text{adaptive}}^{\text{lat}}$。

$$d_{\text{adaptive}}^{\text{long}}=(1+k\rho)\cdot\left[v_{\text{r}}t_{\text{rec}}+\frac{1}{2}\alpha_{\text{current}}^{\text{long}}t_{\text{rec}}{}^{2}+\frac{\left(v_{\text{r}}+t_{\text{rec}}\alpha_{\max}^{\text{long}}\right)^{2}}{2\beta_{\min}^{\text{long}}}-\frac{v_{\text{f}}{}^{2}}{2\beta_{\max}^{\text{long}}}\right]_{+}\tag{9}$$

其中，v_{f}和v_{r}分别表示前车纵向速度和后车纵向速度，$\alpha_{\max}^{\text{long}}$表示最大纵向加速度，$\beta_{\max}^{\text{long}}$和$\beta_{\min}^{\text{long}}$分别表示最大和最小纵向刹车加速度，$\alpha_{\text{current}}^{\text{long}}$表示车辆实时加速度，$t_{\text{rec}}$表示反应时间，$\rho$为车辆密度，$k$为密度影响系数。

$$d_{\text{adaptive}}^{\text{lat}}=(1+k\rho)\cdot\left[\frac{\left(v_1+v_{1,t_{\text{rec}}}\right)}{2}t_{\text{rec}}+\frac{v_{1,t_{\text{rec}}}{}^{2}}{2\beta_{\min}^{\text{lat}}}-\left(\frac{\left(v_2+v_{2,t_{\text{rec}}}\right)}{2}t_{\text{rec}}+\frac{v_{2,t_{\text{rec}}}{}^{2}}{2\beta_{\min}^{\text{lat}}}\right)\right]_{+}\tag{10}$$

其中，v_1表示左侧车辆的横向速度，v_2表示右侧车辆的横向速度，$\beta_{\min}^{\text{lat}}$表示最小横向刹车加速度，$t_{\text{rec}}$表示反应时间，$v_{1,t_{\text{rec}}}$和$v_{2,t_{\text{rec}}}$分别表示反应时间内的横向修正速度。

$$\begin{aligned}v_{1,t_{\text{rec}}}&=v_1+t_{\text{rec}}\cdot\alpha_{\max}^{\text{lat}}\\v_{2,t_{\text{rec}}}&=v_2-t_{\text{rec}}\cdot\alpha_{\max}^{\text{lat}}\end{aligned}\tag{11}$$

其中，$\alpha_{\max}^{\text{lat}}$表示最大横向加速度。

由于在实际驾驶场景中，车辆的加速度受工况限制无法持续保持理论最大值，因此与RSS中基于恒定最大加速度的静态预设不同的是，本文采用自车（即自动驾驶车辆）实时加速度$\alpha_{\text{current}}^{\text{long}}$值作为计算参数，使安全边界随车辆实时状态动态变化，让安全距离更贴合实际驾驶状况。

在驾驶环境中随着车辆密度的增加，车与车之间的平均距离缩短，此时任何一个车辆的紧急制动或变道动作都更容易导致连锁反应，需要更大的安全距离来降低碰撞风险。因此，在式(9)和式(10)中针对路段车辆密度ρ的动态变化，引入密度影响系数k对最小安全距离进行自适应缩放。这种交通密度系数补偿方式，使得在低密度场景中收缩安全边界以释放性能控制空间，而在高密度场景中扩展安全距离以增强容错能力，解决传统方法因静态参数无法响应环境波动导致的适应性差问题。

ARSS模型融合实时车辆加速度参数与环境车辆密度动态变化对安全距离的动态补偿机制，通过感知数据在线更新模型参数，建立了动态可调的安全边界约束。保留RSS模型在保障安全和碰撞责任界定优势的同时，降低了传统模型中静态预设与动态场景失配导致的适应性差的问题。

1.3.2 动态双向切换逻辑

监控与决策模块是自动驾驶运行时安全控制的关键组件，负责实时监测车辆状态与环境信息，并基于安全性指标进行双控制器之间的切换。监控与决策模块的核心机制是双向切换逻辑，然而切换逻辑中依赖预定义的切换标准，存在动态场景适配不足、过度限制控制器效能导致通行效率下降的问题。因此本节提出了动态双向切换逻辑，通过环境参数动态修正切换指标与轻量化决策结构，解决传统安全模型中参数固化与动态环境适配不足影响通行效率的问题。

本节引入1.3.1节ARSS模型中的动态纵向最小安全距离$d_{\text{adaptive}}^{\text{long}}$和横向最小安全距离$d_{\text{adaptive}}^{\text{lat}}$，以及RSS模型中纵向最小安全距离$d_{\text{long}}^{\text{safe}}$和横向最小安全距离$d_{\text{lat}}^{\text{safe}}$。根据这些最小安全距离，定义动态双向切换逻辑的切换参考依据。首先，定义自动驾驶车辆运行时的安全状态集合$\phi_{\text{safe}}\subseteq S$，表示满足安全要求$d_{\text{long}}\geqslant\min\{d_{\text{long}}^{\text{safe}},d_{\text{adaptive}}^{\text{long}}\}\wedge d_{\text{lat}}\geqslant\min\{d_{\text{lat}}^{\text{safe}},d_{\text{adaptive}}^{\text{lat}}\}$，即车辆在运行过程中与前车和邻车的距离均大于所定义的最小安全距离的车辆状态集合；其次，定义自动驾驶车辆运行时的安全预警状态集合$\phi_{\text{warning}}\subseteq\phi_{\text{safe}}$，表示在该车辆状态下使用性能控制器的控制指令后自动驾驶车辆将进入不安全状态，即$\boldsymbol{S}_{t+\Delta t}=f(\boldsymbol{S}_t\in\phi_{\text{warning}},a_t^{\text{PC}})\notin\phi_{\text{safe}}$。

为了使自动驾驶车辆在运行期间始终保持在ϕ_{safe}内，监控和决策模块中的监视器时刻监视自动驾驶车辆和环境的信息，一旦监测到车辆状态进入ϕ_{warning}就会激活决策模块，并按照动态双向切换逻辑在性能控制器和安全控制器之间进行切换。动态双向切换逻辑利用基于ARSS模型的安全状态集合ϕ_{safe}和安全预警状态集合ϕ_{warning}作为切换判断标准，具体切换逻辑为

$$A_t=\begin{cases}a_t^{\text{PC}},\ A_{t-\Delta t}=a_{t-\Delta t}^{\text{SC}}\wedge f(S_t,a_t^{\text{PC}})\in\phi_{\text{safe}}\\a_t^{\text{SC}},\ A_{t-\Delta t}=a_{t-\Delta t}^{\text{PC}}\wedge S_t\in\phi_{\text{warning}}\\A_{t-\Delta t},\ \text{其他}\end{cases}\tag{12}$$

若当前使用的控制动作指令由SC生成，且在下一控制周期内使用PC生成的控制指令不会将车辆带离ϕ_{safe}，则立即将车辆的控制权切换到PC；若当前使用的控制动作指令由PC生成，且下一控制周期内继续使用PC会将车辆带入ϕ_{warning}，则将车

辆的控制权切换到SC；除此之外，控制器间不进行切换，保持原有控制权。

动态双向切换逻辑通过融合ARSS模型中的动态安全边界调整机制，将切换规则与环境参数实时关联以适配不同驾驶场景，突破了以往切换判断标准对预设安全参数的依赖，有效避免了安全冗余监管对车辆通行效率的影响。

1.3.3　安全控制器指令生成算法

当动态双向切换逻辑将控制权移交给安全控制器后，安全控制器需要针对不同的安全预警情况做出具体的响应，以保证自动驾驶车辆在运行期间始终保持在ϕ_{safe}内。本节设计了一种基于ARSS模型的安全控制器指令生成算法，该控制方法分别针对纵向、横向预警以及其并发状况，实现了多维度安全评估与响应动作生成的融合，具体如算法2所示。

算法2　安全控制器指令生成算法

输入　t时刻自车与前车的距离d_{long}，自车与邻车的距离$d_{\text{lat}}^{\text{left}}$和$d_{\text{lat}}^{\text{right}}$

输出　安全控制指令a_t^{SC}

1)if $d_{\text{long}} < \min\{d_{\text{long}}^{\text{safe}}, d_{\text{adaptive}}^{\text{long}}\}$ then

2)　$a_t^{\text{SC}} = \text{Decelerate}$

3)end if

4)if $d_{\text{lat}}^{\text{left}} < \min\{d_{\text{lat}}^{\text{safe}}, d_{\text{adaptive}}^{\text{lat}}\}$ and $d_{\text{lat}}^{\text{right}} < \min\{d_{\text{lat}}^{\text{safe}}, d_{\text{adaptive}}^{\text{lat}}\}$ then

5)　$a_t^{\text{SC}} = \text{Decelerate}$

6)end if

7)if　$d_{\text{lat}}^{\text{left}} < \min\{d_{\text{lat}}^{\text{safe}}, d_{\text{adaptive}}^{\text{lat}}\}$　and　$(d_{\text{lat}}^{\text{right}} < \min\{d_{\text{lat}}^{\text{safe}}, d_{\text{adaptive}}^{\text{lat}}\}$ then

8)　if自车右（左）侧存在车道且右（左）侧车道上的其他车辆与自车的纵向距离都大于$\min\{d_{\text{long}}^{\text{safe}}, d_{\text{adaptive}}^{\text{long}}\}$ then

9)　$a_t^{\text{SC}} = \text{LaneRight}(\text{LaneLeft})$

10)　else

11)　$a_t^{\text{SC}} = \text{Decelerate}$

12)　end if

13)end if

算法2中采用了分层响应机制：当监控器发出纵向安全预警$d_{\text{long}} < \min\{d_{\text{long}}^{\text{safe}}, d_{\text{adaptive}}^{\text{long}}\}$，安全控制器通过连续调整制动压力使车辆进行减速；当监控器发出横向安全预警$d_{\text{lat}}^{\text{left}} < \min\{d_{\text{lat}}^{\text{safe}}, d_{\text{adaptive}}^{\text{lat}}\}$、$d_{\text{lat}}^{\text{right}} < \min\{d_{\text{lat}}^{\text{safe}}, d_{\text{adaptive}}^{\text{lat}}\}$时，安全控制器优先分析相邻车道可用空间、目标车道后方来车速度及距离等参数，结合ARSS安全约束条件动态选择最优策略，即执行变道规避或实施紧急减速。在每一控制周期内，安全控制器都会在监控器发出安全预警时对车辆进行控制，直到车辆完全停止或安全预警解除。需要强调的是，安全控制器的设计完全聚焦于安全保障维度，不考虑驾驶性能。通过上述基于ARSS模型的安全控制机制，能够保障自动驾驶车辆的安全性。

2　仿真实验

本节通过仿真实验全面验证了RTA-AutoSafe在保障驾驶安全的同时提升通行效率的能力。首先，通过设计安全评估实验，验证RTA-AutoSafe在保障自动驾驶车辆行驶实时安全方面的效果；其次，设计性能评估实验，验证RTA-AutoSafe通过性能控制器和动态双向切换逻辑多方面提升通行效率的能力。

2.1　实验设置

为了评估本文方法，使用了基于Gym的自动驾驶仿真环境（highway-env）[25]来模拟高速公路上自动驾驶的行驶过程和动作决策。highway-env作为底层模拟器，内置多种典型驾驶场景，能够直接模拟多车道高速行驶环境，包含车辆变道、超车、避障等核心交互行为，能有效验证自动驾驶策略在动态交通中的各种性能，是自动驾驶领域广泛应用的仿真平台。

本节在highway-env中构建了四车道同向高速公路仿真环境，所有车辆均从左向右行驶。自车的任务是尽可能快速地行驶，同时避免与其他交通参与者发生碰撞。图4展示了该环境中某一时刻自车与其他车辆的实时状态，当前状态下自车需通过向右方车道变道以避让周围车辆，避免潜在碰撞风险，从而保持行驶安全。在环境设置中，若自车发生碰撞或达到最大回合长度，当前回合终止，以两者中先发生者为准。环境中的总车辆数设置为50辆，包括1辆自车和49辆其他车辆，其他交通参与者的运动决策由模拟器默认给出，其他车辆初始位置和速度均随机，具体仿真环境参数如表1所示。

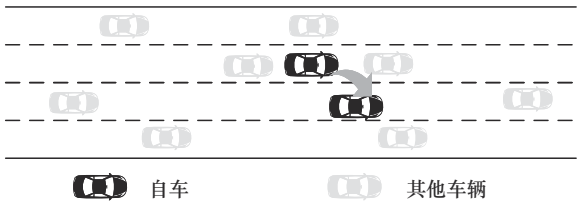

图4　自车在四车道高速公路上的运行示意

表1 仿真环境参数

环境配置参数	参数值
车辆速度范围/(m·s^{-1})	0~30
纵向最大加速度/(m·s^{-2})	5
纵向刹车加速度范围/(m·s^{-2})	−5~−3
横向最大加速度/(m·s^{-2})	2
横向最小刹车加速度/(m·s^{-2})	−0.2
目标车辆数/辆	10
反应时间t_{rec}/s	0.5
信息表示方式（absolute）	TRUE（全局坐标）

2.2 安全性评估

本节首先通过5次100回合的实验测试式(9)和式(10)中车辆密度影响系数k对安全和性能的影响，结果如表2所示。表2中，每组数据通过/分隔左右两侧数据，分别表示在车辆密度ρ = 1.5和ρ = 2.0下的实验结果。当k =0.45时，平均碰撞次数减少，驾驶安全性提升，同时平均速度最高，确保通行效率。此外，控制器之间切换次数最少，反映出系统的稳定性，而较低的SC平均控制时间步则意味着对性能影响较小。因此，k =0.45在安全性、效率和稳定性间取得了良好平衡，将其作为后续实验的参数值。

为评估RTA-AutoSafe在保障车辆实时安全方面的有效性，选择以下方法进行对比。

1) E-DQN：不考虑任何安全控制策略，仅使用性能控制器进行车辆控制，该方法代表强化学习在自动驾驶中的一个基础应用，提供了一个无安全控制的基准方法。

2) E-DQN+纵向安全：在性能控制器基础上，使用文献[18]中的RTA框架，结合了纵向安全控制机制，并通过RSS模型来确保车辆的纵向安全，提供了一个仅考虑纵向安全控制的基准方法。

3) E-DQN+纵向安全+横向安全1：在性能控制器基础上，采用文献[26]中基于RSS模型的RTA框架思想，引入RSS模型中的纵向和横向安全距离，评估RTA-AutoSafe相较于已有的基于RSS模型的安全控制框架对安全性的提升。

4) E-DQN+纵向安全+横向安全2：在性能控制器的基础上，使用文献[27]中基于RSS模型的安全掩码方法。在决策过程中加入了基于RSS模型的安全掩码机制，与同样在平衡自动驾驶性能和安全的方法相比，验证RTA-AutoSafe在保障安全性方面的优势。

通过对比不同车辆密度下各方法的碰撞率来评估驾驶安全，实验结果如表3所示。

具体而言，对比表3中E-DQN与RTA-AutoSafe的实验数据，在加入安全控制模块的设计后，不同车辆密度下的碰撞率都有所降低，特别是在高密度车辆环境下，碰撞率分别从32.68%和47.88%下降到2.72%和3.88%。其次，当车辆密度较低时，本文方法、文献[18]中的RTA框架和文献[27]中的安全掩码方法均能实现零碰撞，但随着车辆密

表2 不同车辆密度影响系数对实验的影响

车辆密度影响系数k	车辆密度ρ	平均碰撞次数/(次·百回合$^{-1}$)	平均速度/(m·s^{-1})	控制器平均切换次数/(次·回合$^{-1}$)	SC平均控制时间步/(步·回合$^{-1}$)
0.40	1.5/2.0	14/25	26.15/25.23	8/16	16/26
0.45	1.5/2.0	8/22	26.63/25.48	4/10	15/22
0.50	1.5/2.0	9/24	26.37/24.80	6/10	14/26
0.55	1.5/2.0	8/23	26.21/25.17	5/9	16/23

表3 不同安全方法在不同车辆密度下的碰撞率

安全方法	车辆密度为1	车辆密度为1.25	车辆密度为1.5	车辆密度为1.75	车辆密度为2
E-DQN	1.8%	9.56%	22.36%	32.68%	47.88%
文献[18]	0	0	5.40%	11.84%	18.32%
文献[26]	0	0	1.20%	2.92%	4.08%
文献[27]	0	0	4.00%	9.76%	16.72%
RTA-AutoSafe	0	0	1.32%	2.72%	3.88%

度增加至1.5和2.0，文献[18]中RTA框架的碰撞率分别上升至5.40%和18.32%，文献[27]中安全掩码方法的碰撞率上升至4.00%和16.72%，本文方法则保持在1.32%和3.88%。最后，对比文献[26]基于RSS的RTA-AutoSafe和本文基于ARSS的RTA-AutoSafe的实验数据，基于ARSS的RTA-AutoSafe在不同车辆密度下的碰撞率与基于RSS的RTA-AutoSafe的碰撞率基本吻合，且在车辆密度上升时表现得更好。这表明ARSS模型在增加动态适应性的同时稳定维持了传统RSS模型的安全基线。

总体来看，RTA-AutoSafe在减少交通事故方面具有明显优势，特别是在交通密度较高的情况下，其安全优势更为突出，这表明该框架可能更适合应用于交通动态变化的驾驶场景。

2.3　性能评估

本节将从性能控制器和动态双向切换逻辑2个方面评估RTA-AutoSafe在动态交通环境下对通行效率的提升效果。

2.3.1　性能控制器效果评估

为验证E-DQN的有效性，本节从消融实验角度设计对比实验，选取3种典型深度强化学习算法作为基线：基础DQN[28]、双重深度Q网络（DDQN）[17]，以及融合专家经验的SQIL[22]。DQN和DDQN代表了当前离散动作空间下强化学习在自动驾驶决策问题中的主流方法，用于衡量本方法在基本策略学习能力方面的性能提升；SQIL则代表了融合静态专家演示数据的方法[22]，评估单纯引入专家经验对控制策略的提升作用。通过与这些方法对比，表明E-DQN的自适应双缓冲优先经验重放机制对整体性能的提升效果，从而验证本文方法在动态交通环境下的适应性和通行效率的提升效果。

为了将上述方法应用于自动驾驶高性能控制策略的学习中，并实现自动驾驶车辆在尽可能快速通行的同时避免与其他交通参与者发生碰撞的控制任务，使用马尔可夫决策过程（MDP, Markov decision process）[29]将高速公路环境下自动驾驶的控制指令生成过程建模为一个五元组MDP $=(S,A,p,R,\gamma)$，其中，S为状态空间，A为动作空间，p为状态转移概率，R为奖励函数，$\gamma\in(0,1)$为折扣因子。具体而言，将自动驾驶车辆在仿真环境中实时感知到的自身状态与周围交通参与者的状态信息映射为一组状态向量，用于构建MDP中的状态空间：$S=(\boldsymbol{s}_i)_{i\in[0,N-1]}$，$\boldsymbol{s}_i=[x_i,y_i,v_i^x,v_i^y]^{\mathrm{T}}$。其中，$\boldsymbol{s}_i$表示道路上自车能检测到的第$i$辆车的状态信息，共有$N$辆车，记$\boldsymbol{s}_0$为自车的状态信息。其次，为实现自动驾驶车辆在动态环境中进行速度控制和变道控制决策，实验采用高层、离散动作空间[30]：$A\in\{a^{\mathrm{PC}},a^{\mathrm{SC}}\}$，$a^{\mathrm{PC}},a^{\mathrm{SC}}\in\{\mathrm{LaneLeft},\mathrm{LaneRight},\mathrm{Accelerate},\mathrm{Decelerate},\mathrm{Cruise}\}$完成对自车的横向控制和纵向控制。其中，LaneLeft和LaneRight分别表示将车辆左移和右移至相邻车道，Accelerate和Decelerate分别表示提高和降低车辆的速度，Cruise表示匀速（即不改变车辆的速度和车道）。

E-DQN的网络结构使用多层感知机作为基础架构，由输入层、两层隐藏层和输出层组成。输入层维度包括智能体自身（自动驾驶车辆）的状态信息维度（为5）和其他智能体的相对状态信息维度（为$(N-1)\times5$）。两层隐藏层的神经元数量均为256，并使用ReLU作为激活函数。输出层的维度与动作空间一致（为5）。本文中的专家经验数据来自近端策略优化（PPO, proximal policy optimization）模型。该模型通过仿真器默认的奖励函数进行训练，生成的控制轨迹具有较高安全性和稳定性，作为后续策略训练的专家演示数据来源。专家轨迹被用于构建奖励函数以及作为专家经验数据参与双缓冲经验重放机制，以引导E-DQN的策略优化过程，其他具体的实验训练参数按照表4进行设置。

表4　　E-DQN训练参数

参数	值
训练回合数E/回合	3 500
单次回合最大步长T/(步·回合$^{-1}$)	40
学习率	0.000 5
初始探索率ε	0.99
时间衰减速率λ	0.01
折扣因子γ	0.9
缓冲区大小/条	20 000，其中B_{expert}=5 000，B_{online}=15 000
采样批次大小/条	64，其中N_1=48，N_2=16
目标网络更新步数c/步	50
软更新参数τ	0.01

在相同场景下，通过训练E-DQN与其他3种基准方法进行了性能对比分析。图5展示了不同方

法在高速公路驾驶环境中的训练曲线。通过图5中的内容可以分析出，E-DQN训练初期奖励值高于DQN和DDQN，说明双重经验缓冲区能够提升初期学习效果；此外，在相同的训练回合下，E-DQN能够达到较高的奖励值，尤其与SQIL相比，说明在训练过程中用实时高奖励经验更新专家经验的机制保证了学习过程中示范样本的时效性和高质量，能有效提升策略质量。

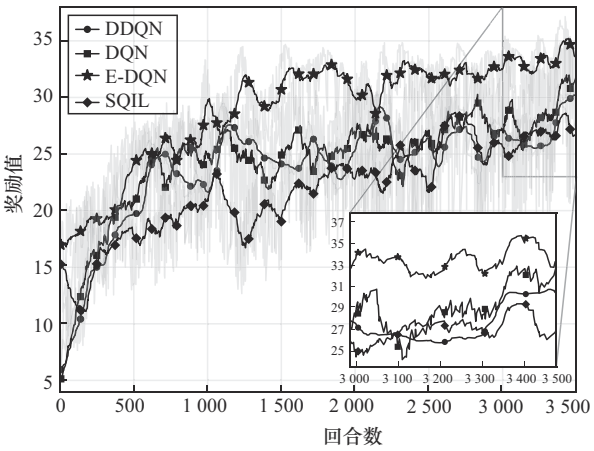

图5　高速公路环境下不同算法的训练曲线

在模型测试阶段引入如下4个指标：碰撞率、回合平均速度、回合长度和回合奖励。其中，碰撞率评估安全水平，回合平均速度评估通行效率水平，回合长度同时评估安全水平和通行效率，回合奖励对车辆的整体性能进行全面的评估，其计算方式参考文献[18]。本节将经过3 500回合训练的各方法模型在不同车辆密度下分别进行5次1 000回合的测试，各个模型平均性能对比如表5所示。

从表5中可以看出，整体上E-DQN在各个指标上的表现都比基线方法更好。E-DQN在低车辆密度下，碰撞率仅为1.81%，显著低于其他方法，这表明E-DQN在低密度场景下能够有效避免碰撞。当车辆密度增加到1.5时，E-DQN的碰撞率为22.35%，DQN的碰撞率高达58.36%，DDQN为67.33%，SQIL为56.33%。即使在高车辆密度下，E-DQN的碰撞率达到47.89%，仍然显著低于DQN的82.18%、DDQN的85.89%和SQIL的75.95%。结果表明，E-DQN在面对更高复杂性的交通环境时，不仅能够保持较低的安全风险，还展现出更好的稳定性和适应性，整体表现优于其他对比方法。这验证了碰撞优先增益与高奖励优先采样机制强化了对关键决策片段的关注，使智能体在面对风险或高效状态时能更快调整策略，从而提升整体决策质量和安全性。

在低车辆密度下，E-DQN的平均回合长度为39.295，明显高于其他方法，表明E-DQN能够在保持低碰撞率的同时，维持更长的回合时间，这意味着E-DQN策略更加稳健。在中等车辆密度下，E-DQN的回合长度为31.418，依旧优于其他方法，尽管碰撞率增加，但其控制策略能够保证车辆在较长时间内保持安全驾驶。在高车辆密度下，E-DQN的回合长度为21.640，比SQIL和DQN表现更优，这说明即使面临复杂的交通环境，E-DQN仍然能够保

表5　不同车辆密度下各个算法的性能对比

方法	车辆密度ρ	碰撞率	回合长度/步	回合奖励	回合平均速度/(m·s^{-1})
DQN	1.0	19.39%	32.570	29.351	28.10
	1.5	58.36%	17.240	14.135	26.00
	2.0	82.18%	8.730	6.460	24.47
DDQN	1.0	24.56%	30.420	27.81	28.52
	1.5	67.33%	13.740	11.160	26.40
	2.0	85.89%	7.644	5.660	24.90
SQIL	1.0	16.87%	33.810	30.798	27.58
	1.5	56.33%	19.330	15.618	24.88
	2.0	75.95%	10.657	7.790	22.94
E-DQN	1.0	1.81%	39.295	33.450	28.65
	1.5	22.35%	31.418	27.580	27.00
	2.0	47.89%	21.640	17.640	25.90

证较长的回合时间，展示了其良好的适应性。

为进一步评估本文奖励函数结构对性能控制器控制策略的影响，本节使用与训练专家数据相同的奖励函数对 E-DQN 进行训练，并将训练得到的策略（记为 E-DQN2）与使用从专家轨迹中提取的奖励结构训练的策略（记为 E-DQN1）进行比较，实验结果如图 6 所示。

由图 6 可以看出，E-DQN1 的碰撞率始终低于 E-DQN2，体现出以专家数据为基础构建奖励信号在安全性能优化方面的优势。在回合长度方面，E-DQN1 策略在低密度环境下表现更稳定，且在高密度下波动更小，展现出更好的稳定性。在回合奖励方面，E-DQN1 在中高密度条件下仍保持较高水平，说明其策略能更有效地避免低奖励事件。在平均速度上，E-DQN1 和 E-DQN2 在不同车辆密度下的表现整体较为接近，尽管 E-DQN2 在低密度场景下略占优势，但其安全性不足。相较之下，E-DQN1 策略以较高的平均速度达成更优的安全和效率平衡。由实验结果可以看出，本文提出的通过 IQ-Learn 从专家数据中获取的奖励函数在策略训练方面是有效的，训练得到的策略不仅能保持通行效率，还提升了策略的安全，实现了更优的性能和安全平衡。

2.3.2　动态双向切换逻辑效果评估

RTA-AutoSafe 在性能控制器外引入了冗余安全控制机制，这种外部安全控制的设计会影响驾驶效果，因此本文设计了基于 ARSS 的动态双向切换逻辑以调控双控制器之间的切换，从而降低安全控制器对通行效率的影响。为了进一步分析基于 ARSS 的动态双向切换逻辑在降低安全冗余干预方面的效果，设计了在不同切换逻辑下安全控制器介入时长对比实验。通过分别对比基于 RSS 的双向切换逻和本文中基于 ARSS 的动态双向切换逻辑，统计 2 种切换逻辑下每回合安全控制器的运行步长，量化评估切换逻辑对通行效率的影响，实验结果如图 7 所示。

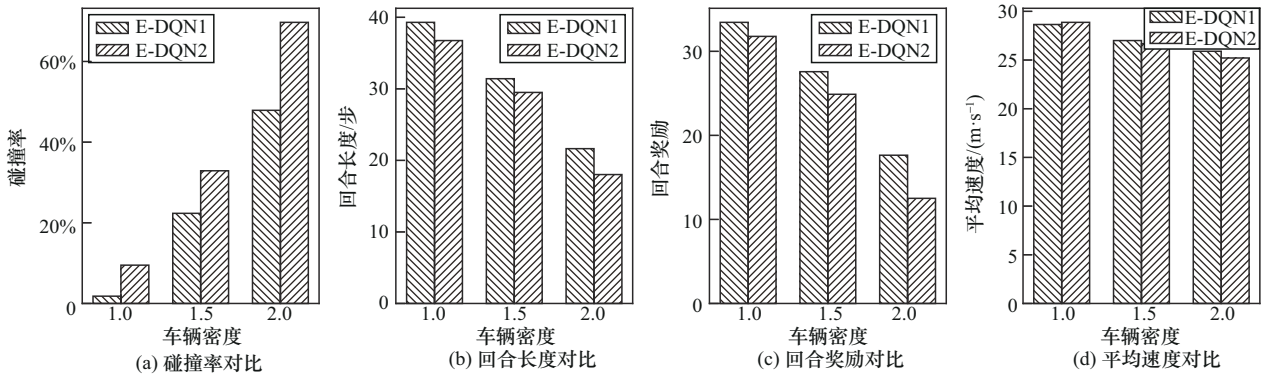

图 6　不同奖励下 E-DQN 的性能对比

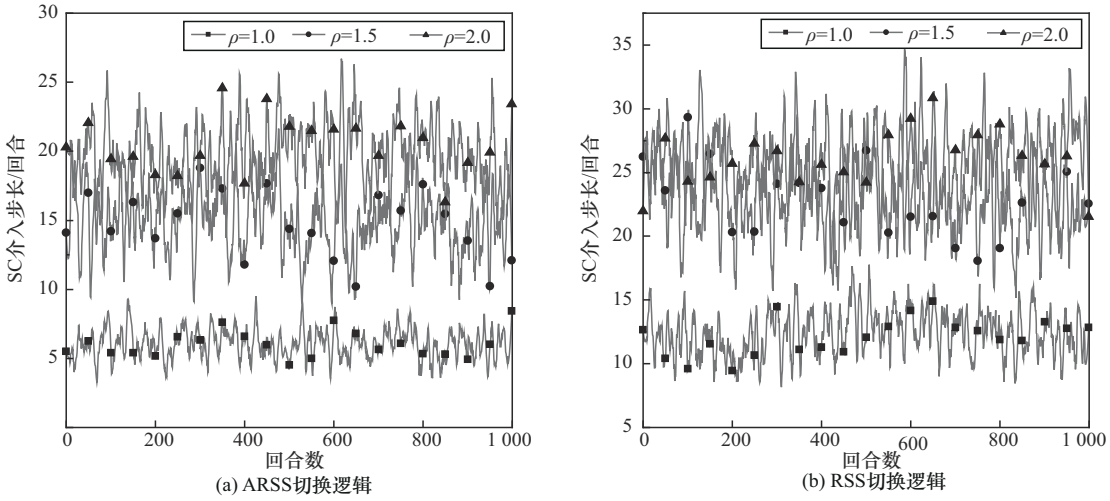

图 7　不同车辆密度下 2 种切换逻辑与安全控制器控制步长的关系

由图7可知，在基于RSS的双向切换逻辑的调控下，不同车辆密度下安全控制器的介入步长范围分别为9~18、15~30、18~35；本文在基于ARSS的动态双向切换逻辑的调控下，不同车辆密度下安全控制器的介入步长范围分别为4~9、10~20、14~24。总体而言，基于ARSS的动态双向切换逻辑在不同车辆密度下能够减少安全控制器的介入时间。这说明ARSS通过自适应安全边界缩放有效减少了因预设保守规则触发的非必要切换，释放了更多性能控制空间，提升了自动驾驶系统的整体运行效率。

综合结果表明，在基于ARSS的动态双向切换逻辑的动态调控下，RTA-AutoSafe能够在保持与保守安全方法同等安全水平的前提下，通过自适应机制减少保守安全控制介入，成功缓解传统运行时保证技术中安全与效率的对立矛盾，为复杂场景下的自动驾驶系统提供了更优的平衡方案。

3 结束语

针对现有自动驾驶运行时安全控制方法无法根据车辆实际运行环境进行动态调整导致通行效率降低的问题，本文提出了一种动态场景下自动驾驶运行时安全自动控制模型，在Simplex架构的基础上引入自适应特性，通过动态调整性能控制器与安全控制器的协同工作机制，使模型能更好地在动态交通环境中同时满足高效驾驶与运行安全的需求。此外，设计了E-DQN，构建双重经验缓冲区融合专家经验和实时探索数据，同时利用优先经验重放机制优化关键经验采样概率，提升了车辆的适应能力和通行效率。考虑车辆实时加速度和车辆密度给出了一种自适应责任敏感安全模型，通过动态调整安全距离的计算提高了模型的适应性，在此基础上实现了安全控制器并提出了动态双向切换逻辑，有效降低了安全控制对效率的过度制约，实现了性能控制与安全保障的动态平衡。实验结果显示，RTA-AutoSafe在各项指标上均优于对比方法，能够有效提升运行效率与安全性，并具备动态平衡运行效率与安全性的能力。

在未来的研究工作中，还可以考虑对方法进行真实环境验证，将性能控制器与安全控制器嵌入真实自动驾驶车辆控制系统，并结合激光雷达、毫米波雷达、摄像头等车载传感器获取的实时信息，在开放道路或封闭测试场中进行实地验证，进一步验证模型的环境适应能力。

参考文献：

[1] KATIYAR D N, SHUKLA D A, CHAWLA D N. AI in autonomous vehicles: opportunities, challenges, and regulatory implications[J]. Educational Administration Theory and Practices, 2024,30(4): 2148-2403.

[2] LU X, TAN H Q, ZHANG H D, et al. Triboelectric sensor gloves for real-time behavior identification and takeover time adjustment in conditionally automated vehicles[J]. Nature Communications, 2025, 16: 1080.

[3] 陈亚男, 李昂, 吴丹. 基于六维语义空间的自动驾驶风险评估研究[J]. 通信学报, 2024, 45(1): 77-93.
CHEN Y N, LI A, WU D. Risk assessment of autonomous vehicle based on six-dimensional semantic space[J]. Journal on Communications, 2024, 45(1): 77-93.

[4] WU J D, HUANG C, HUANG H L, et al. Recent advances in reinforcement learning-based autonomous driving behavior planning: a survey[J]. Transportation Research Part C: Emerging Technologies, 2024, 164: 104654.

[5] NIU Y C, WANG Y J, XIAO M, et al. Reliable safety decision-making for autonomous vehicles: a safety assurance reinforcement learning[J]. Transportmetrica B: Transport Dynamics, 2025, 13(1): 2439997.

[6] WEN L, DUAN J L, LI S E, et al. Safe reinforcement learning for autonomous vehicles through parallel constrained policy optimization[C]// Proceedings of the 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC). Piscataway: IEEE Press, 2020: 1-7.

[7] LI G F, YANG Y F, LI S, et al. Decision making of autonomous vehicles in lane change scenarios: deep reinforcement learning approaches with risk awareness[J]. Transportation Research Part C: Emerging Technologies, 2022, 134: 103452.

[8] CEUSTERS G, PUTRATAMA M A, FRANKE R, et al. An adaptive safety layer with hard constraints for safe reinforcement learning in multi-energy management systems[J]. Sustainable Energy, Grids and Networks, 2023, 36: 101202.

[9] ZHANG Z L, HAN S Y, WANG J W, et al. Spatial-temporal-aware safe multi-agent reinforcement learning of connected autonomous vehicles in challenging scenarios[C]//Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2023: 5574-5580.

[10] WANG C Q, WANG Y. Safe autonomous driving with latent dynamics and state-wise constraints[J]. Sensors, 2024, 24(10): 3139.

[11] GU S D, YANG L, DU Y L, et al. A review of safe reinforcement learning: methods, theories, and applications[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(12): 11216-11235.

[12] HOBBS K L, MOTE M L, ABATE M C L, et al. Runtime assurance for safety-critical systems: an introduction to safety filtering approaches for complex control systems[J]. IEEE Control Systems Magazine, 2023, 43(2): 28-65.

[13] 董磊, 王琦, 陈曦, 等. 运行时保证技术的研究现状与发展综述[J]. 计算机应用, 2025, 45(3): 1003-1015.
DONG L, WANG Q, CHEN X, et al. Survey of research status and development of runtime assurance technology[J]. Journal of Computer Applications, 2025, 45(3): 1003-1015.

[14] SHA L. Using simplicity to control complexity[J]. IEEE Software, 2001, 18(4): 20-28.

[15] MAO Y B, GU Y L, HOVAKIMYAN N, et al. S$\mathscr{L}_1$-simplex: safe velocity regulation of self-driving vehicles in dynamic and unforeseen environments[J]. ACM Transactions on Cyber-Physical Systems, 2023, 7(1): 1-24.

[16] CHEN S D, SUN Y W, LI D C, et al. Runtime safety assurance for learning-enabled control of autonomous driving vehicles[C]//Proceedings of the 2022 International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2022: 8978-8984.

[17] PENG Y F, TAN G Z, SI H W, et al. DRL-GAT-SA: deep reinforcement learning for autonomous driving planning based on graph attention networks and simplex architecture[J]. Journal of Systems Architecture, 2022, 126: 102505.

[18] PENG Y F, TAN G Z, SI H W. RTA-IR: a runtime assurance framework for behavior planning based on imitation learning and responsibility-sensitive safety model[J]. Expert Systems with Applications, 2023, 232: 120824.

[19] POLACK P, ALTCHÉ F, D'ANDRÉA-NOVEL B, et al. The kinematic bicycle model: a consistent model for planning feasible trajectories for autonomous vehicles? [C]//Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV). Piscataway: IEEE Press, 2017: 812-818.

[20] LIN X Y, HUANG J W, ZHANG B, et al. A velocity adaptive steering control strategy of autonomous vehicle based on double deep Q-learning network with varied agents[J]. Engineering Applications of Artificial Intelligence, 2025, 139: 109655.

[21] GARG D, CHAKRABORTY S, CUNDY C, et al. IQ-Learn: Inverse soft-Q learning for imitation[J]. Advances in Neural Information Processing Systems, 2021, 34: 4028-4039.

[22] REDDY S, DRAGAN A D, LEVINE S. SQIL: imitation learning via reinforcement learning with sparse rewards[J]. arXiv Preprint, arXiv: 1905.11108, 2019.

[23] HASSANI H, NIKAN S, SHAMI A. Traffic navigation via reinforcement learning with episodic-guided prioritized experience replay[J]. Engineering Applications of Artificial Intelligence, 2024, 137: 109147.

[24] REIMANN J, MANSION N, HAYDON J, et al. Temporal logic formalisation of ISO 34502 critical scenarios: modular construction with the RSS safety distance[C]//Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing. New York: ACM Press, 2024: 186-195.

[25] LI Y Y, YUAN W, ZHANG S A, et al. Choose your simulator wisely: a review on open-source simulators for autonomous driving[J]. IEEE Transactions on Intelligent Vehicles, 2024, 9(5): 4861-4876.

[26] KOBAYASHI T, BONDU M, ISHIKAWA F. Formal Modelling of-safety architecture for responsibility-aware autonomous vehicle via Event-B refinement[C]//International Symposium on Formal Methods. Berlin: Springer, 2023: 533-549.

[27] HE X K, HUANG W H, LV C. Toward trustworthy decision-making for autonomous vehicles: a robust reinforcement learning approach with safety guarantees[J]. Engineering, 2024, 33: 77-89.

[28] WANG Z X, YAN H, WEI C S, et al. Research on autonomous driving decision-making strategies based deep reinforcement learning[C]//Proceedings of the 2024 4th International Conference on Internet of Things and Machine Learning. New York: ACM Press, 2024: 211-215.

[29] ELALLID B B, BENAMAR N, HAFID A, et al. A comprehensive survey on the application of deep and reinforcement learning approaches in autonomous driving[J]. Journal of King Saud University-Computer and Information Sciences, 2022, 34(9): 7366-7390.

[30] IRSHAYYID A, CHEN J, XIONG G J. A review on reinforcement learning-based highway autonomous vehicle control[J]. Green Energy and Intelligent Transportation, 2024, 3(4): 100156.

[作者简介]

徐丙凤（1986– ），女，安徽安庆人，博士，南京林业大学副教授，主要研究方向为车联网安全、系统安全风险建模与分析、软件工程。

陈嘉玲（2001– ），女，湖北荆州人，南京林业大学硕士生，主要研究方向为车联网安全、系统安全风险建模及分析、安全强化学习等。

杨帅领（2002– ），男，河南虞城人，南京林业大学硕士生，主要研究方向为车联网安全、系统安全风险建模及分析、安全强化学习等。

何高峰（1984– ），男，安徽安庆人，博士，南京邮电大学副教授，主要研究方向为网络安全。