文章编号:1673-0291(2026)03-0152-12 DOI:10.11860/j.issn.1673-0291.20250168

网联环境下基于PPO的车队轨迹与信号配时协同优化方法

邢凯然
(北京交通大学 交通运输学院,北京 100044)

摘 要:既有基于深度强化学习的车辆轨迹优化方案在处理网联混行交通时,难以实现车队的动态合并与拆分,而传统的信号配时优化以周期或相位为单位,存在响应滞后问题,二者间的割裂控制限制了交叉口通行效率的进一步提升.针对车辆轨迹灵活控制与信号配时实时响应的协同优化需求,提出一种基于近端策略优化(Proximal Policy Optimization, PPO)的车队轨迹与信号配时协同优化方法(Fleet Trajectory and Signal timing based on Waiting Offsetting, FT-SIWO).该方法首先将车辆轨迹优化问题构建为一个马尔可夫决策过程,设计了一个融合局部车辆动态与全局交通状态变量的状态空间、一个包含跟车与多种加速度模式的动作空间及一个综合了驾驶舒适度、安全性等多维度的奖励函数,使智能体能够全面感知环境并学习高效策略.然后,采用PPO-clip算法进行策略更新,以提升训练过程的稳定性与样本效率.最后,引入一种基于候时对冲(Waiting Offsetting, WO)的信号配时实时优化机制,通过动态计算并平衡绿灯相位车辆的"时间亏损"与红灯相位车辆的"时间盈余",实现以秒为粒度的绿灯时长动态调整,从而与车辆轨迹优化形成深度协同.研究结果表明:在SUMO与Matlab联合仿真平台上,相较于无优化的基准方案,FT-SIWO使交叉口车辆平均延误减少28%~32%,平均等待时间降低30%~35%,同时多种污染物(CO_2, CO, HC, NO_x, PM_x)排放总量降低34%~38%,燃油消耗减少19%;与国内外同类先进的分布式控制或车路协同方案相比,FT-SIWO在控制成效上进一步提升了6%~10%.该方法通过车队智能编组与信号配时的秒级联动,有效提升了绿灯时间利用率和交通流整体运行效率,为智慧城市交通管理系统应对混合交通流、缓解拥堵与降低排放提供了可行的协同优化解决方案.

关键词:智能交通;协同优化;车队轨迹;信号配时;近端策略优化
中图分类号:U491 **文献标志码**:A

Collaborative optimization method for platoon trajectory and signal timing based on proximal policy optimization in connected vehicle environments

XING Kairan
(School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China)

Abstract: Existing vehicle trajectory optimization methods based on deep reinforcement learning

收稿日期:2025-10-16;**修回日期**:2025-12-03
基金项目:国家重点研发计划(2022YFB4301305)
Foundation item: National Key R&D Plan (2022YFB4301305)
第一作者:邢凯然(2003—),男,河北唐山人,硕士生.研究方向为交通运输规划与管理.**email**:15932259865@163.com.

引用格式:邢凯然.网联环境下基于PPO的车队轨迹与信号配时协同优化方法[J].北京交通大学学报,2026,50(3):152-163.
XING Kairan. Collaborative optimization method for platoon trajectory and signal timing based on proximal policy optimization in connected vehicle environments[J]. Journal of Beijing Jiaotong University, 2026, 50(3): 152-163. (in Chinese)

struggle to achieve the dynamic merging and splitting of platoons in connected, mixed-traffic environments. Simultaneously, conventional signal timing optimization, which operates at the cycle or phase level, is hindered by response lag. The decoupled control between these two elements restricts further improvements in intersection efficiency. To address the necessity for the collaborative optimization of flexible vehicle trajectory control and real-time signal timing, this paper proposes a joint optimization method for platoon trajectories and signal timing based on Proximal Policy Optimization (PPO), termed FT-SIWO (Fleet Trajectory and Signal timing based on Waiting Offsetting). First, the vehicle trajectory optimization problem is formulated as a markov decision process. A state space integrating local vehicle dynamics and global traffic state variables, an action space encompassing car-following and multiple acceleration modes, and a multi-dimensional reward function considering factors such as driving comfort and safety are designed. This formulation enables the agent to comprehensively perceive the environment and learn efficient policies. Subsequently, the PPO-clip algorithm is employed for policy updates to enhance both training stability and sample efficiency. Finally, a real-time signal timing optimization mechanism based on Waiting Offsetting (WO) is introduced. By dynamically calculating and balancing the "time loss" of vehicles in the green phase and the "time surplus" of vehicles in the red phase, this mechanism enables the dynamic adjustment of the green light duration at a second-level granularity, thereby achieving deep coordination with the vehicle trajectory optimization. Simulation results on a joint SUMO and Matlab platform demonstrate that, compared to a non-optimized baseline scheme, FT-SIWO reduces the average vehicle delay at intersections by 28%-32% and the average waiting time by 30%-35%. Additionally, it decreases the total emissions of various pollutants (CO_2, CO, HC, NO_x, and PM_x) by 34%-38% and reduces fuel consumption by 19%. Furthermore, when compared to state-of-the-art distributed control or vehicle-infrastructure cooperation schemes, FT-SIWO achieves an additional 6%-10% improvement in control effectiveness. By enabling intelligent platoon formation and a second-level linkage with signal timing, this method effectively enhances green time utilization and the overall operational efficiency of traffic flow. It provides a viable collaborative optimization solution for smart urban traffic management systems to handle mixed traffic flows, alleviate congestion, and reduce emissions.

Keywords: intelligent transportation; collaborative optimization; platoon trajectory; signal timing; proximal policy optimization

在现代交通系统中，交通拥堵、能源消耗和污染排放是亟待解决的三大难题.随着人工智能和通信技术的快速发展，智能网联交通系统（Intelligent Connected Transportation Systems, ICTS）应运而生.该系统通过车间通信（Vehicle-to-Vehicle, V2V）和车辆与基础设施间通信（Vehicle-to-Infrastructure, V2I）技术，为上述问题提供了创新的解决方案[1].然而，当前ICTS面临网联人工驾驶车辆（Connected Human-driven Vehicles, CHV）和网联自动驾驶汽车（Connected Autonomous Vehicles, CAV）混行的复杂交通模式.在此模式下，CHV由人驾驶，而CAV则能够完全受系统控制，精准调整位置、速度和加速度等参数.如何有效优化CAV与CHV的轨迹，以提升交通流运行的效率和安全，已成为ICTS领域的一大挑战.

在此背景下，车辆轨迹与交通信号的协同优化，被认为是提升交通效率和安全性的重要手段.车辆跟驰模型作为优化的基础，通过智能化调控车辆间距与行驶速度，能够有效提高乘客舒适度、缓解交通拥堵并减少环境污染.如文献[2]提出了适用于CAV、自动驾驶车辆和常规人工驾驶车辆混行的交通流跟驰模型；文献[3]则探讨了混合交通环境下如何通过CAV跟驰模型和优化策略来提高行车效率和交通流稳定性的问题；文献[4]提出了基于加速度的连接巡航控制方法，旨在通过V2V提升道路交通的流动性.

现有模型由于算法相对固定，对混合交通流中的突发情况响应能力较差，且传统跟驰模型无法满足实时、动态的车辆控制需求.深度强化学习（Deep Reinforcement Learning, DRL）的引入为解决这一

问题提供了新的思路.DRL通过与环境的交互学习最优策略,并结合深度学习的强大特征提取能力,能够在高维环境状态中寻求最优驾驶策略.如文献[5]通过多智能体强化学习算法,协同优化了无信号交叉口的混合交通流;文献[6]利用DRL改善了混合交通环境中的行车稳定性和能源效率;文献[7]探讨了基于DRL的自动驾驶环保策略,而文献[8]提出了CAV在孤立信号交叉口的分散式轨迹规划方法.

上述研究集中于单一车辆的轨迹优化,若能以车队为单位进行控制,将显著提升交通效率.故此,文献[9]建立了部分检测条件下信号交叉口车辆编队的DRL模型;文献[10]研究了基于DRL的车辆编队控制和交通信号同步,以优化延迟和燃油消耗.同时当前信号配时优化多在周期或相位层面进行,显然有一定的局限性.如文献[11]提出的CycLight框架,虽采用周期级决策结构,但在其实验分析中明确指出:“周期级动作更新在高波动场景下会产生显著延迟,平均响应滞后达15~20s,限制了控制性能的进一步提升”,这说明即使先进的DRL模型,在周期约束下仍难以实现即时响应.类似地,文献[12]在分析集中式评论家与分布式执行器架构时发现,当干预频率从“每周期一次”提升至“每相位阶段一次”时,交叉口平均延误可进一步降低12.7%,证明更细粒度的动作执行频率对控制效果具有实质性影响.

然而,现有研究中的车辆编队机制仍不够灵活,无法实现智能化的车队合并与拆分.此外,信号配时优化以周期或相位为单位进行调整,存在一定的滞后性,若以秒级进行优化则能使优化效果更进一步.因此,本文采用DRL控制的灵活车辆编队轨迹与信号控制协同优化的方式,构建包含11个变量的状态空间,为智能体提供丰富环境信息;动作空间划分为跟车和头车两类,减少变道策略降低维度,提升学习效率;综合舒适度、驾驶安全等5个维度设计奖励函数,引导智能体学习最优策略.同时采用PPO-clip算法优化策略,引入基于候时对冲的信号配时优化,实现了更高效的信号配时决策.

1 问题描述

在混合交通环境下,CAV与CHV的共存对交通控制提出了双重挑战.首先,现有DRL在车辆轨迹优化中难以实现车队的动态合并与拆分,导致编队控制的灵活性不足.既有车辆编队模型多采用固定编队策略,缺乏对交通流动态变化的适应性[13].在信号交叉口场景中,当车队长度与信号相位时间不匹配时,无法智能拆分车队以实现“脉冲式”通行,导致绿灯时间利用率显著下降.其次,传统信号配时优化以固定周期为调整单元,存在显著的响应滞后性[14].依赖历史数据或固定周期调整的配时方案,难以适应实时变化的交通需求[15].此外,现有研究多将轨迹优化与信号控制割裂处理,缺乏双向反馈机制.车辆轨迹优化未考虑信号相位剩余时间对编队策略的影响,而信号配时调整亦未融合车队动态特征,这种单向控制模式导致协同优化域受限[16].

聚焦于网联混行交通中车队轨迹动态控制与信号配时实时优化的协同机制,提出一种车队轨迹与信号配时的协同优化方法(Fleet Trajectory and Signal timing based on Waiting Offsetting, FT-SIWO),实现车辆编队轨迹优化并制定实时的信号配时优化策略,旨在通过协同优化车队轨迹与信号配时,提高交叉口通行效率,总体优化思路如图1所示.

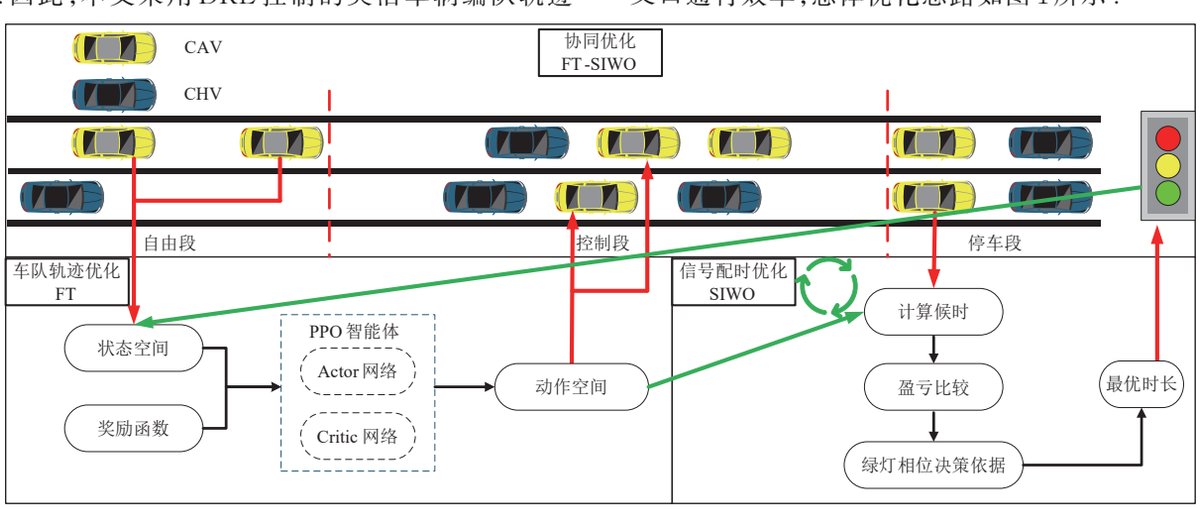


图1 总体优化流程

Fig.1 Overall optimization flowchart

2　优化方法

2.1　马尔可夫过程

网联交通环境下，车辆轨迹优化问题可以转化为马尔可夫决策过程（Markov Decision Process，MDP）[17]. MDP是一种描述具有时间序列的离散随机过程的数学模型. 为应用MDP，设置状态空间、动作空间和奖励函数.

2.1.1　状态空间

在ICTS中，车辆轨迹优化问题中的状态空间是影响智能体决策的关键环境变量集合. CAV能够实时获取周围车辆的状态及交通信号信息，根据这些信息作出驾驶决策. 因此，状态空间由车辆的局部变量和全局变量组成，即

$$U=\left\{U_{\text{local}},U_{\text{global}}\right\} \tag{1}$$

式中：U_{local}为局部特征变量的集合；U_{global}为全局特征变量的集合.

CAV的局部变量集合U_{local}包括自车和邻近车辆的动态数据，具体为

$$U_{\text{local}}=\left\{l,v,s,s_{\text{f}},d_{\text{f}},c_{\text{f}},\Delta v_{\text{f}}\right\} \tag{2}$$

式中：l表示车辆当前所在车道；v表示瞬时速度；s表示车辆的位置；s_{f}表示前车的位置；d_{f}和Δv_{f}分别表示CAV与同车道前车的间距和速度差；c_{f}为前车的类别，CHV或CAV.

全局特征变量集合U_{global}表达式为

$$U_{\text{global}}=\left\{d_{\text{q}},t_{\text{q}},t_{\text{r}},\phi\right\} \tag{3}$$

式中：d_{q}表示车辆当前时刻距离下游停车段的距离；t_{q}表示CAV以当前速度、强加速度a_2（2 m/s²）行驶下，到达停车段所需的预期时间；t_{r}表示当前信号相位的剩余时间；ϕ表示当前时刻的交通信号相位. 可得d_{q}的计算式为

$$d_{\text{q}}=s_{\text{q}}-s \tag{4}$$

式中：s_{q}为停车段末位车辆的位置，$s_{\text{q}}=s_{\text{inter}}$，即$s_{\text{q}}$与停止线位置$s_{\text{inter}}$相同. 通过速度学公式反解可得$t_{\text{q}}$计算式为

$$t_{\text{q}}=\frac{-v+\sqrt{v^2-2a_2d_{\text{q}}}}{a_2} \tag{5}$$

状态空间由7个局部变量和4个全局变量表示，共11个特征变量，并随时间不断更新. 如图2所示，这些特征变量为智能体提供了丰富的环境信息，使其能够在复杂的交通环境中作出最优的驾驶决策.

2.1.2　动作空间

在行驶过程中，CAV会通过调整加速度来控制自身的速度、位置等参数以实现跟车、最短时间通过交叉口等行为[18]. 为使用离散的深度确定策略梯度算法，将动作空间设计为两个主要类别：跟车和头车. 车辆通过变速与前车保持饱和车头时距的距离后，可以进入跟车模式，通过网络环境传递前车的速度及加速度从而实现精准复制前车状态来跟随前车；若为头车，选择匀速（0 m/s²）、弱加速（1 m/s²）、强加速（2 m/s²）、弱减速（−1 m/s²）、强减速（−2 m/s²）之一，主要考虑到乘客的舒适加速度均值为1.23 m/s²[19]. 默认车辆在自由段已完成变道，如此将降低动作空间的维度，提高学习效率. 因此，算法的动作空间为

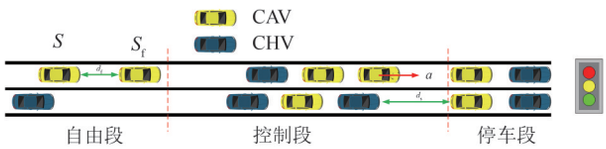


图 2　状态空间参数

Fig.2　State space parameters

$$A=\left\{\text{cf},0,1,2,-1,-2\right\} \tag{6}$$

式中：cf为进入跟车模式.

2.1.3　奖励函数

在DRL中，奖励函数是智能体学习行为准则的关键因素. 一旦执行动作，智能体根据环境反馈接收奖励（或惩罚）. DRL的目标是通过最大化未来折扣奖励的期望来学习最优策略[20]. 因此，不同的奖励函数会影响智能体的学习性能. 考虑到CAV与交通环境之间的复杂互动，考虑5个因素设计奖励函数.

因素1：舒适度. 过于频繁的加速度变化会降低乘客的舒适度，故应尽量减少加速度变化[21]. 其奖励函数为

$$r_{\text{comfort}}=\begin{cases}0 & A'=A_0\\ -1 & A'\neq A_0\end{cases} \tag{7}$$

式中：A_0为当前采取的策略；A'为接下来采取的策略.

因素2：驾驶安全. 驾驶过程中的跟车行为需要确保车辆的安全，即CAV不与其他环境车辆相撞. 其奖励函数为

$$r_{\text{safe}}=\begin{cases}0 & d_{\text{f}}>1,\ c_{\text{f}}=\text{CAV}\\ 0 & d_{\text{f}}>1,\ c_{\text{f}}=\text{CHV}\\ -10 & d_{\text{f}}\leqslant 1,\ c_{\text{f}}=\text{CAV}\\ -10 & d_{\text{f}}\leqslant 1,\ c_{\text{f}}=\text{CHV}\end{cases} \tag{8}$$

因素3：交通规则. CAV不得在红灯时驶入交叉口，而是停车等待，否则是闯红灯行为. 其奖励函数为

$$r_{\text{rules}}=\begin{cases}0 & t_{\text{leave}}\in\cos\left[\pi t_{\text{l}}/(T_{\text{r}}+T_{\text{y}})\right]\\ -10 & t_{\text{leave}}\notin\cos\left[\pi t_{\text{l}}/(T_{\text{r}}+T_{\text{y}})\right]\end{cases} \tag{9}$$

式中：t_{leave}为CAV离开交叉口时刻的时间；t_{l}表示与当前时间对应的信号周期中的时间点；T_{r}、T_{y}分别为红灯和黄灯的持续时间.

因素4：最短时间.相位周期内车队最大数量通过交叉口问题可以简化为头车最短时间内通过交叉口的问题[22].其奖励函数为

$$r_{\text{time}}=\begin{cases}-0.1t_{\text{leave}} & s=s'_{\text{inter}}\\ 0 & \text{其他}\end{cases} \tag{10}$$

式中：s'_{inter}表示车辆驶离交叉口时的位置.

因素5：跟车策略车辆进入控制段后，若以当前速度、强加速度状态通过停止线时，绿灯相位还剩余3 s以上，鼓励车辆加速与前车连成车队.其奖励函数为

$$r_{\text{follow}}=\begin{cases}1 & \cos\left[\pi\left(t+t_{\text{q}}\right)/\left(T_{\text{r}}+T_{\text{y}}\right)\right]>0,\\ & t_{\text{r}}>3,\ s_{\text{f}}-s\in[5,10]\\ 0 & \text{其他}\end{cases} \tag{11}$$

综合考虑5个因素，最终的奖励函数为

$$r=r_{\text{comfort}}+r_{\text{safe}}+r_{\text{rules}}+r_{\text{time}}+r_{\text{follow}} \tag{12}$$

2.2 PPO算法结构

近端策略优化（Proximal Policy Optimization，PPO）算法是一种强化学习算法，主要用于策略优化问题.它旨在改进策略梯度方法，提高训练的稳定性和样本效率，解决策略梯度方法更新不稳定的问题.PPO算法有两个主要版本：PPO-penalty和PPO-clip.此处采用效率更高的PPO-clip算法.

对于PPO-clip，目标函数$L^{\text{CLIP}}(\theta)$的形式为

$$L^{\text{CLIP}}(\theta)=E_t\left[\min\left(r_t(\theta)\hat{A}_t,\text{clip}\left(r_t(\theta),\varepsilon\right)\hat{A}_t\right)\right] \tag{13}$$

$$\text{clip}\left(r_t(\theta),\varepsilon\right)=\begin{cases}1+\varepsilon & r_t(\theta)>1+\varepsilon\\ r_t(\theta) & 1-\varepsilon\leqslant r_t(\theta)\leqslant 1+\varepsilon\\ 1-\varepsilon & r_t(\theta)<1-\varepsilon\end{cases} \tag{14}$$

式中：E_t表示对时间步t的期望；$r_t(\theta)$是当前策略π_θ和旧策略$\pi_{\theta\text{old}}$在状态s_t下采取动作a_t的概率比；$\hat{A}_t$是优势函数的估计值，用于衡量在状态s_t下采取动作a_t相对于平均表现的好坏程度；ε是一个小的超参数，用于控制策略更新的范围，通常取值在0.1～0.3.

在目标函数中，$r_t(\theta)$表示新旧策略的概率比.如果不加以限制，当新策略与旧策略差异过大时，可能会导致训练不稳定.通过clip函数将$r_t(\theta)$限制在$[1-\varepsilon,1+\varepsilon]$范围内，使得策略更新的幅度不会过大，从而提高训练的稳定性.

2.3 基于WO的信号配时实时优化

针对信号配时实时优化问题，采取候时对冲（Waiting Offsetting，WO）的方法优化信号配时方案.WO方法主要聚焦于绿灯相位转换的关键节点.对于CHV，采用IDM（Intelligent Driver Model）模型模拟其行驶状态；对于CAV，则采用PATH实验室开发的CACC（Cooperative Adapted Cruise Control）模型[23]，以此构建贴合实际的车辆运行微观模拟环境，CACC模型的表达式为

$$\begin{cases}e=x_{i-1}-x_i-Tv_i\\ v_i=v_{i\text{prev}}+k_1e+k_2\dot{e}\end{cases} \tag{15}$$

式中：e为实际车间距离与期望车间距离的误差；x_i为跟驰车辆i的位移；T为安全车头时距，取0.6 s；$v_{i\,\text{prev}}$为跟驰车辆前一时刻的速度；$\dot{e}$为e的微分项；k_1为车间距误差权重，取0.45；k_2为车间距误差微分权重，取0.25.

当车队因红灯断开时，后方车队将以CAV为头车，限制后方车辆一同减速缓行，若根据等待时间计算出断开处为CHV，则CHV将会自行跟随前车行驶，并不会阻碍后方CAV打头的缓行车队前进.

绿灯期间，车辆以初速度匀速行驶通过交叉口，被视为基准情形.然而，车辆可能以强加速度行驶直至通过交叉口.这两种行驶方案每秒产生的位移存在差异，通过计算强加速度行驶与初速度匀速行驶的每秒距离差，并将其与初速度相除，即可得出每辆车在每个时间步长内节省下来的等待时间，将该部分等待时间视为“亏损”.相反，红灯期间，车辆会减速排队，等待时间增长，将增长的等待时间视为“盈余”.通过循环计算和动态累计，分别得到当前绿灯相位车辆等待时间的总亏损和下一相位车辆等待时间的总盈余，二者之和作为核心评估指标.计算过程共4步.

步骤1：计算车辆通过交叉口时间.强加速度下行驶通过交叉口的时间t_a为

$$t_a=\frac{-v\pm\sqrt{v^2-2a_2d_{\text{end}}}}{a_2} \tag{16}$$

$$t_{\text{f}}=\frac{d_{\text{end}}}{v} \tag{17}$$

式中：d_{end}是车辆到交叉口的距离；t_{f}是匀速行驶通过交叉口的时间.

步骤2：计算绿灯相位车辆等待时间“亏损”.设当前绿灯相位有n辆车，第i辆车到交叉口的距离为$d_{\text{end}i}$，初速度为v_i.第i辆车以强加速度a_2和初速度v_i行驶的每秒距离差为$\Delta s_i(t)$.

第i辆车在时刻t以强加速度a_2行驶的位移$s_{ai}(t)$、以初速度v_i匀速行驶的位移$s_{\text{f}i}(t)$分别为

$$s_{ai}(t)=v_it+\frac{1}{2}a_2t^2\quad t\leqslant t_a \tag{18}$$

$$s_{\text{f}i}(t)=v_it \tag{19}$$

则每秒距离差$\Delta s_i(t)$为

$$\Delta s_i(t)=s_{ai}(t)-s_{fi}(t)\quad t\leqslant t_{ai} \tag{20}$$

每秒距离差对应的时间节省$\Delta t_i(t)$为

$$\Delta t_i(t)=\frac{\Delta s_i(t)}{v_i}\quad t\leqslant t_{ai} \tag{21}$$

在车辆通过交叉口的时间t_{ai}内求和，得到第i辆车的等待时间亏损w_{loss_i}为

$$w_{\text{loss}_i}=\sum_{t=1}^{t_{ai}}\Delta t_i(t) \tag{22}$$

绿灯相位车辆等待时间总亏损W_{loss}为

$$W_{\text{loss}}=\sum_{i=1}^{n}w_{\text{loss}_i} \tag{23}$$

步骤3：计算红灯相位车辆等待时间“盈余”.设红灯相位有m辆车，第i辆车开始排队等待的时间为t_i，当前时间为t.第i辆车的等待时间盈余w_i为

$$w_i=\begin{cases}0 & t_i>t\\ t-t_i & t_i<t\end{cases} \tag{24}$$

红灯相位车辆等待时间总盈余W_{surplus}为

$$W_{\text{surplus}}=\sum_{i=1}^{m}w_i \tag{25}$$

步骤4：计算绿灯相位决策依据.车辆等待时间总和$W(t)$为

$$W(t)=W_{\text{loss}}-W_{\text{surplus}} \tag{26}$$

鉴于交通流的复杂性和不确定性，在实际运行过程中可能出现多个零点的情况.以负值峰值出现后的第一个零点作为最终绿灯持续时间的决策依据.同时，为保持交通信号的稳定性，设定每个绿灯相位的持续时间下限为5 s.最终绿灯时长t_{optimal}计算式为

$$W\left(t_{\text{optimal}}\right)=0\quad t_{\text{optimal}}>5 \tag{27}$$

3　仿真实验

3.1　实验设计

采用Matlab与SUMO软件构建联合仿真平台，对智能网联交通场景下的车辆行驶行为及信号控制策略进行模拟分析.仿真场景设定为两条双向四车道道路相交的十字路口.每进口方向两条车道，分别为直左、直右车道.每条车道的速度上限均为$v_{\max}=16$ m/s.

以交叉口为中心划分控制段、停车段与自由段3个区段.

区段1：控制段.覆盖交叉口停止线向各进口道延伸300 m的区域，该范围能够涵盖85%以上车辆的加减速行为[24].在此区域内，CAV执行编队重组与速度协同优化策略.

区段2：停车段.位于控制段内侧，因红灯而在停止线处开始排队的车辆长度，由停止线至末位停止车辆尾部.

区段3：自由段.延伸于控制段外侧300 m以外区域，车辆在此以进入路段的初速度行驶，不受协同控制策略干预，需要根据驶离方向自行完成变道.

交叉口采用两相位控制方案，采用经典的Webster优化周期公式确定初始信号周期，并结合相位绿信比分配原则设置各相位初始绿灯时长，初始信号参数为37 s、3 s、37 s、3 s，南北绿灯、南北黄灯、东西绿灯、东西黄灯.对于绿初损失时间问题，采用工程实践中广泛接受的处理方式——将其视为固定值纳入相位间隔设计.设定周期结构时已将绿初损失包含在相位切换间隔内，不参与动态优化，以避免引入驾驶员行为不确定性带来的模型振荡.

为确保智能体能够在多样化的交通场景中全面学习，实验共进行1 000次训练，每次训练包含1 000个时间步，且每个时间步长为0.1 s.在每次训练启动时，运用蒙特卡罗模拟方法对车辆的多项初始参数进行随机化处理[25].初始流量设定基于饱和度指标进行控制，假定各进口道流量服从泊松分布，平均到达率设为0.75倍饱和流量，对应轻度至中度交通负荷水平，故而车辆的生成位置统一设定在每条车道距离交叉口300 m处，每个方向生成的车辆数量介于15～20辆，初始速度取值范围为{5 m/s，6 m/s，7 m/s，8 m/s}.CAV的学习目标是从初始车道出发，顺利通过交叉口驶入目标车道，在此过程中不断优化自身的行驶策略.该设定参考NGSIM等公开数据集的典型运行状态，既能体现一定的排队现象以检验优化效果，又避免因极端拥堵导致系统不可控，从而保证训练过程的稳定性与收敛性.

针对未涵盖中度至重度拥堵场景的问题，需特别说明：本文虽以轻中度拥堵为主，但通过引入动态等待时间作为奖励函数的关键项，已使模型具备应对突发高负荷的能力.事实上，在部分仿真运行过程中，由于随机到达波动，局部时段实际饱和度短暂超过0.9，形成临时性严重排队.此时，WO机制能迅速识别累计等待时间增长趋势，主动延长绿灯以消解积压车辆，有效防止排队溢出.

3.2　训练过程

为实现对复杂交通场景的有效模拟与优化，需针对模型的DRL网络及其相关参数进行调优.

PPO模型中，在构建主网络架构时，初始设定包含一个输入层、一个输出层和两个隐藏层.具体而言，AC网络均采用全连接方式设计输入层，依据状态空间中所定义的状态向量，配置11个神经元；

隐藏层均设置为128个神经元;输出层则设置为6个神经元,对应智能体可执行的动作数量,以此实现对智能体行为的有效控制.

在激活函数的选择上,采用ReLU(Rectified Linear Unit)函数.ReLU计算效率高,仅需简单比较与取值操作,相比Sigmoid、Tanh能显著减少训练时间.

目标网络在架构设计上与主网络保持一致,旨在为训练过程提供稳定的目标参考.经过训练调试,模型中其他参数取值如表1所示.

表1 PPO训练参数

Tab.1 PPO training parameters

参数	描述	取值
M	经验回放大小	20 000
batch size	训练样本	128
ε	贪婪策略初始值	1
γ	折扣因子	0.97
α	学习率	0.000 5

3.3 训练结果

在PPO算法中,通过最大化累积奖励来探寻最优策略,其学习成效可直观体现为最大平均奖励值.因而,可借助对比各训练周期中智能体的平均奖励,对模型的学习效果予以评估.训练进程中的平均奖励曲线如图3所示.

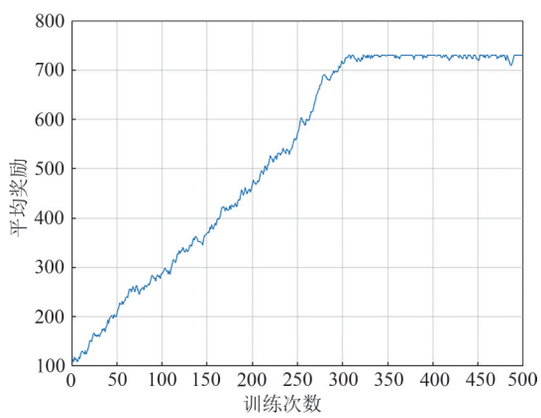


图3 平均奖励曲线

Fig.3 Average reward curve

由图3可知:在初始训练阶段,每个周期的平均奖励数值较低,因为此阶段智能体正处于环境探索时期,可利用的有效经验较为匮乏,极易受到奖励函数设定的惩罚因素影响;随着训练周期数目的逐步增多及ε值的不断减小,智能体能够获取并运用大量有效经验,平均奖励呈现出逐渐上升的趋势,并在历经300个周期后达到稳定状态,最终奖励值收敛至730.这一结果充分表明,智能体在探索与利用之间实现了良好的平衡,成功学习到了车队驾驶策略.

3.4 优化结果

3.4.1 车辆轨迹

将经单独训练完备的CAV智能体,引入CAV渗透率为60%的混合车流场景中.在单方向驶来20辆车且其中包含12辆CAV的情况下,通过对车流通行时遭遇红灯与未遭遇红灯这两种典型情形进行观测,发现CAV轨迹在车辆自由编队与断开方面展现出显著的优化效果.CAV的轨迹均能清晰呈现出车队的形成与拆分过程,基于PPO的轨迹优化与基准方案的车辆轨迹如图4所示.

由图4可知:①当车流通行过程中未遭遇红灯时,所有车辆均能在绿灯时段顺利驶过交叉口;从控制段开始,车辆启动编队行为,CAV轨迹的绿线在行驶进程中会适时加速并与前车构建车队,这种行为有效降低了道路空间的浪费.②当车流通行遭遇红灯时,车流首先需完成车队的组建;而后依据信号灯的可通过时间对车队进行拆分,由于红灯的限制,断开后的后方车队为有效减少等待时间,会开始减速缓行;在此过程中,借助DRL,车队能够寻找到距离交叉口适宜的位置,并从该位置加速行驶.

由于每个CAV都基于实时感知环境作出最优动作选择,因此当外部条件变化时(例如前方红灯即将结束、侧向车队接近同相位入口),智能体可通过自适应速度调节实现自然的车队分离或聚合,无需预设编队规则或中央控制器干预.更重要的是,WO通过延长高优先级相位的绿灯时间,间接鼓励符合条件的车辆集群“合并”成队列通过;反之,在低优先级相位,则促使车辆主动“拆分”以减少无效等待.

3.4.2 车辆速度

CAV的速度变化在遭遇红灯和未遭遇红灯两种场景下,均展现出明显的优化效果,如图5所示.

由图5可知,CAV的速度变化可归纳为3个阶段:①车辆加速追赶前车,直至抵达变速点.在此阶段,车辆通过合理加速,缩短与前车的间距,为后续的跟驰行为做准备.②车辆抵达变速点后,开始逐渐减速,直至车速高于前车车速1~2 m/s.此后,车辆持续贴近前车,直至与前车保持在车队跟驰范围外3 m处.这一阶段的速度调整,旨在确保车辆在安全的前提下,尽可能地与前车形成紧密的跟驰状态.③车辆开始震荡变速,在进入车队跟驰范围的同时,实现与前车速度的一致性.

3.5 成效分析

3.5.1 基准方案对比

将使用PPO算法训练出的车辆轨迹优化模型与FT-SIWO的优化效果进行对比.在两相位交叉

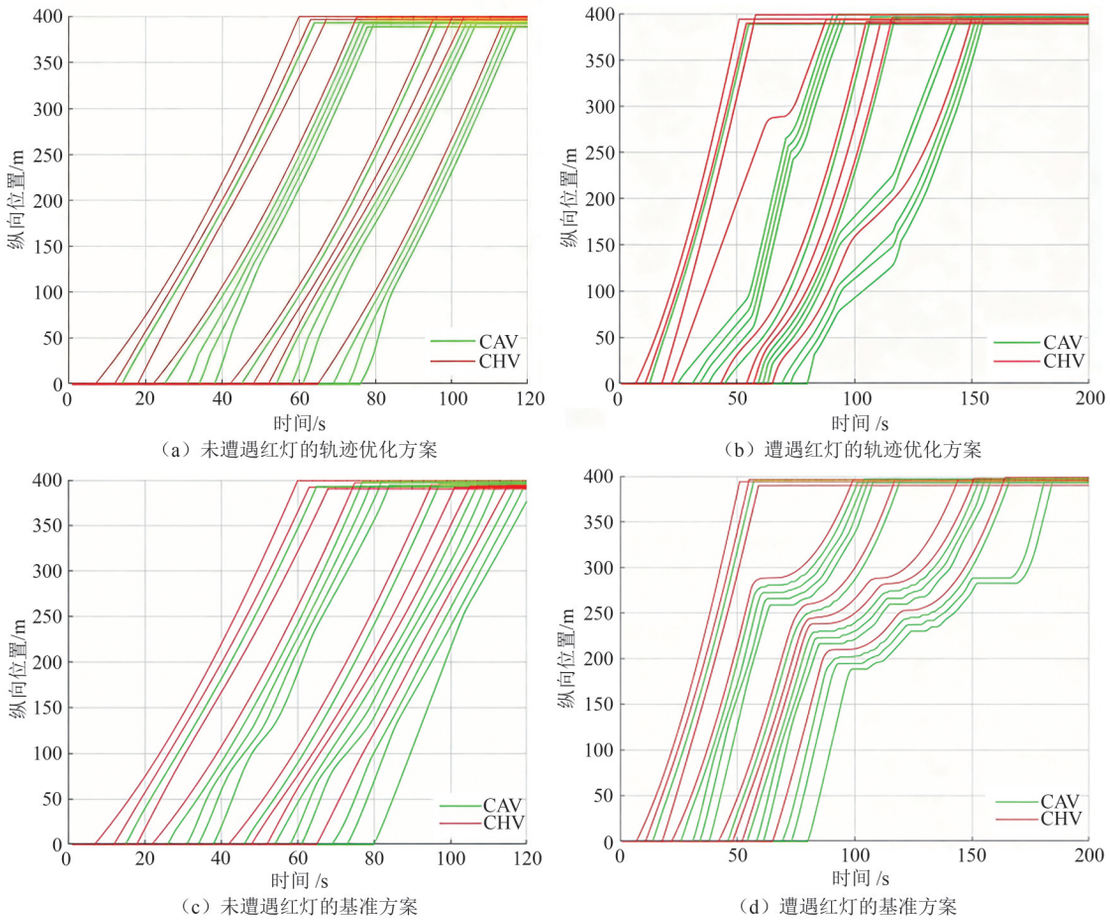


图 4　CAV 车辆轨迹变化

Fig.4　CAVs trajectory changes

口中，60% 的 CAV 渗透率、每个方向来车 40 辆的条件下运行 10 次，可以得到基准方案、轨迹优化方案、协同优化方案的运行结果．结果分为通行效率与生态环境两方面，分别通过 SUMO 平台自带的 duration-log.statistics 模块、emissions-output 模块获取原始数据求平均值得出．

通行效率选取平均延误、平均等待时间，结果如图 6 所示．

由图 6 可知：对比基准方案，轨迹优化方案在平均延误、平均等待时间上分别降低了 18%～23%、22%～26%；协同优化方案则分别减少了 28%～32%、30%～35%．

生态环境选取 CO_2、CO、HC、NO_x、PM_x、燃油消耗量，结果如表 2 所示．

由表 2 可知：对比基准方案，轨迹优化方案在 CO_2、CO、HC、NO_x、PM_x 等污染物排放上减少了 25%～29%，燃油消耗量降低了 15%；协同优化方案则分别降低了 34%～38% 和 19%；考虑到 CAV 渗透率的提升有助于车队控制优化，进行 CAV 渗透率的敏感性分析，仿真了 CAV 渗透率 60%、80%、100% 生态环境结果的变化，得出优化效果分别提升约 4% 和 5%．

以燃油消耗量与 CO_2 排放量为例，探究 CAV 渗透率对基准方案、轨迹优化方案及协同优化方案的影响．燃油消耗量 P_{fuel} 与 CO_2 排放量 P_{CO_2} 的计算式为

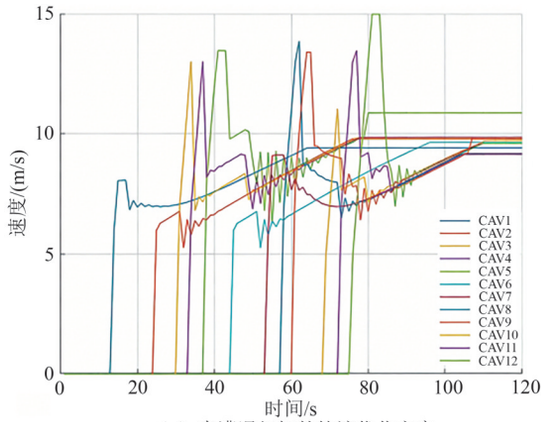
$$P_{\text{fuel}} = \max\left(P_0 + \alpha v + \beta v^2 + \gamma v^3 + \kappa a v, 0\right) \tag{28}$$

$$P_{CO_2} = P_{\text{fuel}} \cdot c_{CO_2} \tag{29}$$

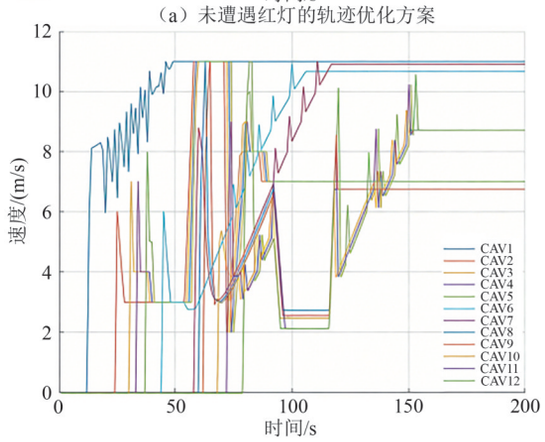
式中：P_0 为怠速燃油消耗率，取 0.3 mL/s；α、β、γ 为速度相关系数，分别取 0.02、0.000 3、0.000 004；κ 为加速度修正系数，取 0.001；c_{CO_2} 为燃油-CO_2 转换因子，汽油车取 2.32 g/mL．

CAV 渗透率对 CO_2 排放量及燃油消耗量的影响如图 7 所示．

由图 7 可知：①CO_2 排放量随渗透率提升呈非线性下降趋势，在 10%～30% 与 70%～100% 渗透率区间内降幅趋缓，而在 40%～60% 区间内降速显著提升，表明该阶段协同优化效应最为显著．10%～30% 渗透率时，CAV 数量较少，仅能局部优化单个车队的加速行为，无法形成大规模协同编队，减排效果主要源于 CAV 自身的节能驾驶，因此降速平缓；



（a）未遭遇红灯的轨迹优化方案

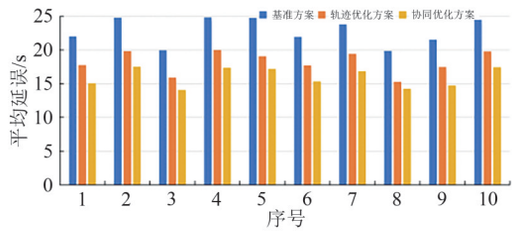


（b）遭遇红灯的轨迹优化方案

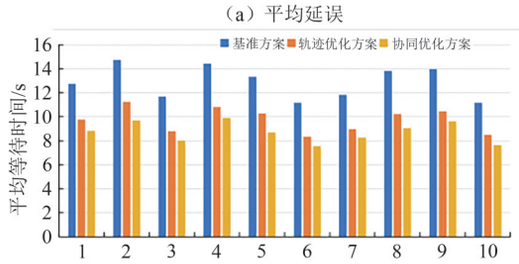
图 5 CAV 车辆速度变化

Fig.5 CAVs speed variations

40%～60% 渗透率时，CAV 数量达到临界规模（渗透率≥40%），可构建连续车队并实现动态编队拆分，此时，优化的协同效应显著增强；70%～100% 渗透率时，CAV 主导交通流，编队控制趋于饱和，优化潜力边际递减.②燃油消耗量随渗透率增加呈近似线性递减趋势，但在 50%～60% 渗透率区间内因车队编组效率跃升，降幅较其他区间增加 12%.燃油消耗的近似线性趋势反映 CAV 渗透率对能耗的全局优化作用.



（a）平均延误



（b）平均等待时间

图 6 通行效率

Fig.6 Traffic efficiency

3.5.2 同类方法对比

为验证 FT-SIWO 的可用性，选取文献[26]提出的基于深度强化学习的混合交通编队中网联自动驾驶汽车分布式控制策略（方案 A）和文献[27]建立的网联环境下信号交叉口车速控制策略及优化方法（方案 B）进行对比分析.

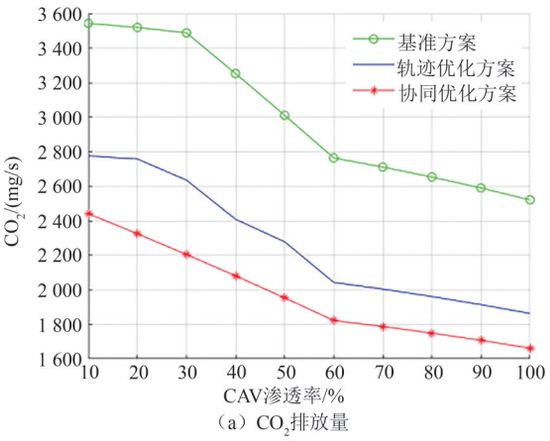
在实验设计的仿真场景中对 3 种方案分别进行仿真，仿真结果构建了 CO_2 排放量-车均延误气泡图（气泡大小表征 CAV 渗透率）见图 8.

由图 8 可知：①FT-SIWO 较方案 A 具有 6%～7% 的优势.方案 A 采用分布式 CAV 控制，虽然能够优化单车轨迹，但由于缺乏车队-信号协同机制，存在明显的局限性.一方面，固定周期信号配时无法响应车队动态，致使 23%～28% 的绿灯时间浪费（仿真中 A 方案绿灯利用率仅为 72%，而 FT-SIWO 方案达到 95%）；另一方面，未实现车队智能拆分，在渗透率 60% 时，因编队长度不匹配信号周期，额外产生 14% 的加速能耗.②相较于方案 B，FT-SIWO 具有 8%～

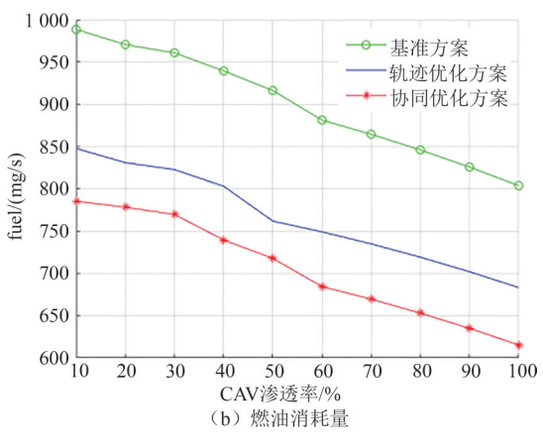
表 2 生态环境结果

Tab.2 Ecological environmental results mg/s

CAV 渗透率/%	方案	CO_2	CO	HC	NO_x	PM_x	fuel
60	基准方案	2763.23	84.79	0.45	1.11	0.05	881.37
	轨迹优化方案	2042.72	61.04	0.32	0.83	0.04	749.16
	协同优化方案	1822.08	54.38	0.29	0.71	0.03	684.42
80	基准方案	2652.70	81.40	0.43	1.07	0.05	846.11
	轨迹优化方案	1961.01	58.60	0.31	0.80	0.03	719.20
	协同优化方案	1749.19	52.21	0.28	0.68	0.03	653.04
100	基准方案	2520.06	77.33	0.41	1.02	0.04	803.81
	轨迹优化方案	1862.96	55.67	0.29	0.76	0.03	683.24
	协同优化方案	1661.74	49.60	0.27	0.65	0.03	615.39



（a）CO_2排放量



（b）燃油消耗量

图 7　CAV 渗透率对控制成效的影响

Fig.7　Impact of CAVs penetration rate on control effectiveness

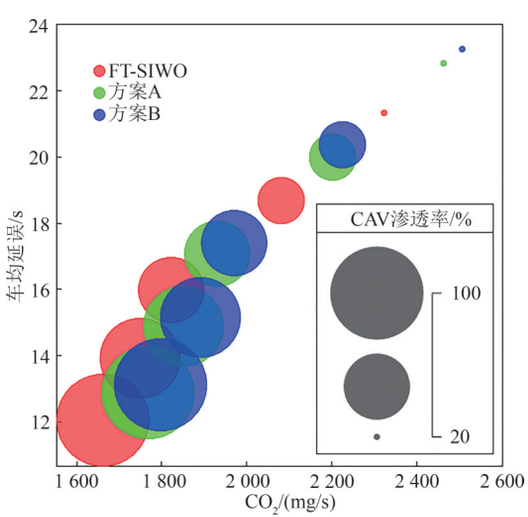


图 8　同类方法结果对比

Fig.8　Comparison of results with similar methods

10% 的优势．方案 B 侧重车路协同控制，然而其基于规则的拆分策略存在固有缺陷．其一，采用固定距离（如 200 m）触发车队拆分的静态拆分阈值，在动态交通流中会导致 19%～24% 的过早或过晚拆分；其二，仅以延误为优化目标的奖励函数单一性，忽略了加速度平滑度约束，致使 CAV 频繁调速，间接增加了 8%～11% 的燃油消耗．

FT-SIWO 通过车辆编队与信号配时的深度耦合机制，突破了传统方法在异质交通流协同优化中的局限性，为智能网联交通系统提供了更具通行效率的解决方案．

4　结论

1）在网联混合交通流环境下，所提出的 FT-SIWO 能够有效提升交叉口通行效率，仿真结果表明，与无优化基准方案相比，该方法可使车辆平均延误减少 28%～32%，平均等待时间降低 30%～35%．

2）该方法在提升通行效率的同时，显著降低了交通系统的环境影响，能够使多种污染物（CO_2，CO，HC，NO_x，PM_x）排放总量降低 34%～38%，并减少 19% 的燃油消耗．

3）与国内外同类先进的分布式控制或车路协同方案相比，FT-SIWO 通过车队智能编组与信号配时的秒级协同，在控制成效上实现了进一步 6%～10% 的提升，且随着 CAV 渗透率的增加，其优化效果更为明显．

未来研究可朝以下方向深入：一是推动算法在真实道路场景下的实车测试与验证，评估实际通信延迟、感知误差等因素对控制效果的影响；二是深化对更复杂异质交通流（如包含非网联车辆、弱势道路使用者等）的研究，以提升方法的泛化能力；三是将协同优化框架从单一交叉口扩展至区域路网，研究车队在连续交叉口及干线绿波协调下的全局最优通行策略．

参考文献（References）：

[1] 杨澜，赵祥模，吴国垣，等．智能网联汽车协同生态驾驶策略综述[J]．交通运输工程学报，2020，20(5)：58－72.
YANG Lan，ZHAO Xiangmo，WU Guoyuan，et al. Review on connected and automated vehicles based cooperative eco-driving strategies [J]. Journal of Traffic and Transportation Engineering，2020，20(5)：58－72.（in Chinese）

[2] 宗芳，石佩鑫，王猛，等．考虑前后多车的网联自动驾驶车辆混流跟驰模型[J]．中国公路学报，2021，34(7)：105－117.
ZONG Fang，SHI Peixin，WANG Meng，et al. Connected and automated vehicle mixed-traffic car-following model considering states of multiple front and rear vehicles [J]. China Journal of Highway and Transport，2021，34(7)：105－117.（in Chinese）

[3] 彭佳力，上官伟，柴琳果，等．混合交通环境下网联自

动驾驶车辆跟驰模型与优化策略[J]. 交通运输工程学报, 2023, 23(3): 232－247.
PENG Jiali, SHANGGUAN Wei, CHAI Linguo, et al. Car-following model and optimization strategy for connected and automated vehicles under mixed traffic environment[J]. Journal of Traffic and Transportation Engineering, 2023, 23(3): 232－247. (in Chinese)

[4] GE J I, OROSZ G. Dynamics of connected vehicle systems with delayed acceleration feedback[J]. Transportation Research Part C: Emerging Technologies, 2014, 46: 46－64.

[5] ZHUANG H B, LEI C F, CHEN Y H, et al. Cooperative decision-making for mixed traffic at an unsignalized intersection based on multi-agent reinforcement learning[J]. Applied Sciences, 2023, 13(8): 5018.

[6] SHI H T, ZHOU Y, WU K S, et al. Connected automated vehicle cooperative control with a deep reinforcement learning approach in a mixed traffic environment[J]. Transportation Research Part C: Emerging Technologies, 2021, 133: 103421.

[7] YANG Z W, ZHENG Z D, KIM J, et al. Eco-driving strategies using reinforcement learning for mixed traffic in the vicinity of signalized intersections[J]. Transportation Research Part C: Emerging Technologies, 2024, 165: 104683.

[8] MA C Y, YU C H, YANG X G. Trajectory planning for connected and automated vehicles at isolated signalized intersections under mixed traffic environment[J]. Transportation Research Part C: Emerging Technologies, 2021, 130: 103309.

[9] TRINH H T, BAE S H, TRAN D Q. Deep reinforcement learning for vehicle platooning at a signalized intersection in mixed traffic with partial detection[J]. Applied Sciences, 2022, 12(19): 10145.

[10] GAO H, YEN C C, ZHANG M. DRL based platooning control with traffic signal synchronization for delay and fuel optimization[J]. Transportation Research Part C: Emerging Technologies, 2024, 163: 104655.

[11] HAN G Y, LIU X H, HAN Y, et al. CycLight: Learning traffic signal cooperation with a cycle-level strategy[J]. Expert Systems with Applications, 2024, 255: 124543.

[12] WANG M N, CHEN Y R, KAN Y H, et al. Traffic signal cycle control with centralized critic and decentralized actors under varying intervention frequencies[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(12): 20085－20104.

[13] 李永福, 周发涛, 黄龙旺, 等. 基于深度强化学习的网联车辆队列纵向控制[J]. 控制与决策, 2024, 39(6): 1879－1887.
LI Yongfu, ZHOU Fatao, HUANG Longwang, et al. Longitudinal control of connected vehicle platoon based on deep reinforcement learning[J]. Control and Decision, 2024, 39(6): 1879－1887. (in Chinese)

[14] 王润民, 张心睿, 赵祥模, 等. 混行环境下网联信号交叉口车路协同控制方法[J]. 交通运输工程学报, 2022, 22(3): 139－151.
WANG Runmin, ZHANG Xinrui, ZHAO Xiangmo, et al. Vehicle-infrastructure cooperative control method of connected and signalized intersection in mixed traffic environment[J]. Journal of Traffic and Transportation Engineering, 2022, 22(3): 139－151. (in Chinese)

[15] 韩磊, 张轮, 郭为安. 混合交通流环境下基于改进强化学习的可变限速控制策略[J]. 交通运输系统工程与信息, 2023, 23(3): 110－122.
HAN Lei, ZHANG Lun, GUO Weian. Variable speed limit control based on improved dueling double deep Q network under mixed traffic environment[J]. Journal of Transportation Systems Engineering and Information Technology, 2023, 23(3): 110－122. (in Chinese)

[16] 黄浩, 胡智群, 王鲁晗, 等. 基于Sumtree DDPG的智能交通信号控制算法[J]. 北京邮电大学学报, 2021, 44(1): 97－103.
HUANG Hao, HU Zhiqun, WANG Luhan, et al. Intelligent traffic signal control algorithm based on sumtree DDPG[J]. Journal of Beijing University of Posts and Telecommunications, 2021, 44(1): 97－103. (in Chinese)

[17] 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 Yilin, SONG Ruoyang, MA Jian. Trajectory tracking control of intelligent vehicle based on DDPG method of reinforcement learning[J]. China Journal of Highway and Transport, 2021, 34(11): 335－348. (in Chinese)

[18] JIANG L M, XIE Y C, EVANS N G, et al. Reinforcement Learning based cooperative longitudinal control for reducing traffic oscillations and improving platoon stability [J]. Transportation Research Part C: Emerging Technologies, 2022, 141: 103744.

[19] DE WINKEL K N, IRMAK T, HAPPEE R, et al. Standards for passenger comfort in automated vehicles: Acceleration and jerk[J]. Applied Ergonomics, 2023, 106: 103881.

[20] 刘春禹, 刘永红, 罗霞, 等. 混合交通流环境下单交叉口自动驾驶车辆轨迹优化[J]. 交通运输系统工程与信息, 2022, 22(2): 154－162.
LIU Chunyu, LIU Yonghong, LUO Xia, et al. Trajectory optimization of connected vehicles at isolated intersection in mixed traffic environment [J]. Journal of

Transportation Systems Engineering and Information Technology, 2022, 22(2): 154－162. (in Chinese)

[21] ZHU M X, WANG Y H, PU Z Y, et al. Safe, efficient, and comfortable velocity control based on reinforcement learning for autonomous driving[J]. Transportation Research Part C: Emerging Technologies, 2020, 117: 102662.

[22] 李传耀，张帆，王涛，等. 基于深度强化学习的道路交叉口生态驾驶策略研究[J]. 交通运输系统工程与信息，2024，24(1)：81－92.
LI Chuanyao, ZHANG Fan, WANG Tao, et al. Signalized intersection eco-driving strategy based on deep reinforcement learning [J]. Journal of Transportation Systems Engineering and Information Technology, 2024, 24(1): 81－92. (in Chinese)

[23] 秦严严，王昊，王炜，等. 自适应巡航控制车辆跟驰模型综述[J]. 交通运输工程学报，2017，17(3)：121－130.
QIN Yanyan, WANG Hao, WANG Wei, et al. Review of car-following models of adaptive cruise control [J]. Journal of Traffic and Transportation Engineering, 2017, 17(3): 121－130. (in Chinese)

[24] 中华人民共和国住房和城乡建设部. 城市道路交叉口规划规范：GB 50647—2011[S]. 北京：中国计划出版社，2012.
Ministry of Housing and Urban-Rural Development of the People's Republic of China. Code for planning of intersections on urban roads: GB 50647—2011[S]. Beijing: China Planning Press, 2012. (in Chinese)

[25] GU X Y, WU S Y, ZHAO S L, et al. Generation of signalized intersection test scenarios based on traffic participant model [J]. Automotive Innovation, 2025, 8(4): 882－895.

[26] SHI H T, CHEN D J, ZHENG N, et al. A deep reinforcement learning based distributed control strategy for connected automated vehicles in mixed traffic platoon[J]. Transportation Research Part C: Emerging Technologies, 2023, 148: 104019.

[27] 刘显贵，王晖年，洪经纬，等. 网联环境下信号交叉口车速控制策略及优化[J]. 交通运输系统工程与信息，2021，21(2)：82－90.
LIU Xiangui, WANG Huinian, HONG Jingwei, et al. Speed control strategy and optimization of signalized intersection in network environment[J]. Journal of Transportation Systems Engineering and Information Technology, 2021, 21(2): 82－90. (in Chinese)