文章编号:1005-0523(2026)01-0082-11

基于深度强化学习的智能车辆风险评估决策模型

范泽敏,吴翊恺,王晨菡

(南京理工大学自动化学院,江苏 南京 210094)

摘要:为解决高速公路环境下车辆的安全驾驶决策问题,提出了一种基于深度强化学习与风险评估的智能车辆决策模型。首先,提出一种基于贝叶斯理论的位置不确定性量化方法,用于驾驶风险的建模与量化;然后,在决策模型中引入自注意力机制,帮助车辆感知复杂场景下的潜在危险,避免执行危险决策;最后,在Highway-env仿真平台构建仿真环境,通过仿真实验对模型进行训练和测试,并设计多种实验对比。结果表明,提出的RA-PPO-Mul模型实现了98%的无碰撞安全率和更高的行车效率,优于传统强化学习模型和仅引入单一模块的模型。

关键词:自动驾驶;深度强化学习;决策模型;风险评估;注意力机制

中图分类号:U461.91 **文献标志码:**A

本文引用格式:范泽敏,吴翊恺,王晨菡. 基于深度强化学习的智能车辆风险评估决策模型[J]. 华东交通大学学报,2026,43(1):82-92.

Intelligent Vehicle Risk Assessment Decision Model Based on Deep Reinforcement Learning

Fan Zemin, Wu Yikai, Wang Chenhan

(School of Automation, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract: A smart vehicle decision-making model based on deep reinforcement learning and risk assessment is proposed to solve the problem of safe driving decisions for vehicles in highway environments. Firstly, a Bayesian based position uncertainty quantification method is proposed for modeling and quantifying driving risks; Then, a self attention mechanism is introduced into the decision model to help vehicles perceive potential dangers in complex scenes and avoid dangerous decision execution; Finally, a simulation environment was constructed on the Highway env simulation platform, and the model was trained and tested using simulation experiments. Multiple experimental comparisons were also designed. The results show that the proposed RA-PPO-Mul model achieves 98% safety rate and higher driving efficiency, which is superior to the traditional reinforcement learning model and the model that only introduces a single module.

Key words: autonomous driving; deep reinforcement learning; decision model; risk assessment; attention mechanism

Citation format: FAN Z M, WU Y K, WANG C H. Intelligent vehicle risk assessment decision model based on deep reinforcement learning[J]. Journal of East China Jiaotong University, 2026, 43(1): 82-92.

收稿日期:2025-03-03

基金项目:河南省科技攻关项目(182102310004)

研究表明，全球约90%以上的交通事故由人为错误引发，换道操作是其中的高危行为之一。随着车辆智能化程度的提升，实现智能车辆在复杂交通环境下的安全高效换道，已成为当前自动驾驶领域的研究重点。换道作为一种典型的驾驶行为，其决策与操作的准确性直接关系到交通安全与效率，因此研究如何实现智能车辆的安全换道决策具有重要的现实意义。

目前研究主要集中在基于规则、基于博弈论和基于学习等决策方法[1-4]。传统的基于规则的决策模型包括Gipps模型和元胞自动机模型[5-6]等，主要通过预先设定的逻辑规则处理换道问题，当面对复杂的动态交通环境时，其适应性存在局限[7]。基于博弈论的方法将车辆之间的换道行为建模为一种博弈过程，利用纳什均衡解决换道冲突[8]。这类方法在极端或突发情况下，不能保证模型的鲁棒性。基于学习的决策方法，特别是强化学习（reinforcement learning，RL）和深度强化学习（deep reinforcement learning，DRL），成为优化自动驾驶车辆行为的主要手段[9]。相比传统基于规则或博弈论的决策模型，强化学习和深度强化学习通过与环境的交互，自主学习最优策略，在应对复杂、多变的交通场景方面展现了巨大潜力。众多学者已将该项技术应用在无人驾驶领域，Cheng等[10]通过标记的MSA（motion-sensitive area）数据集学习驾驶策略，验证了强化学习在优化换道决策中的优越性。Wang等[11]基于深度Q网络（deep Q-network，DQN）方法进行自动驾驶车道变换决策任务研究。MO等[12]利用双深度Q网络（double deep Q-network，DDQN）对智能车辆的纵向速度和换道决策进行学习训练。然而，现有方法通常缺乏对驾驶风险的量化建模，可能引发高风险驾驶行为。

为实现多车道场景下的安全驾驶，本文提出一种结合深度强化学习与风险评估的智能车辆决策模型。首先，提出一种基于贝叶斯理论的位置不确定性量化方法用于建模和量化驾驶风险；其次，在决策模型中引入自注意力机制，帮助车辆感知复杂场景下的潜在危险；最后，使用Highway-env[13]仿真平台搭建三车道高速环境，采用近端策略优化（proximal policy optimization，PPO）算法实现风险最小化策略的学习，并在仿真中验证了方法和模型的有效性。

1　理论概述和模型构建

基于深度强化学习和驾驶风险评估，构建一种面向多车道换道场景的智能车辆决策模型，以获得具有风险意识的驾驶决策策略。

1.1　问题描述

在DRL中，智能体在随机环境中按时间步骤中执行动作并根据反馈（即奖励）进行学习，以最大化累积奖励，该过程通常被建模为马尔可夫决策过程（Markov decision processes，MDP），如式（1）所示

$$M=\langle S,\ A,\ P,\ R,\ \gamma\rangle \tag{1}$$

式中：S为有限状态集；A为有限动作集；P为状态转移概率；R为奖励空间；γ为折扣因子，用来衡量未来奖励在当前累计奖励中的权重。DRL的目标是通过学习找到最优策略，使期望累计奖励最大化，如式（2）所示

$$\pi^{*}(S)=\arg\max_{\pi}\mathbb{E}_{\pi}\left[\sum_{i=0}^{+\infty}\gamma^{i}r_{t+i}\middle|S_{t}=S\right] \tag{2}$$

式中：$\pi^{*}(S)$为最佳策略；r_{t+i}为时间$t+i$的奖励。在传统的强化学习中，可引入Q值函数来指导策略改进，函数通常只考虑奖励的累积

$$Q_{\pi}(S,a)=\mathbb{E}_{\pi}\left[\sum_{i=0}^{+\infty}\gamma^{i}r_{t+i}\middle|S_{t}=S,a_{t}=a\right] \tag{3}$$

式中：$Q_{\pi}(S,a)$为从状态S开始遵循策略π并采取行动a的预期累计奖励。

1.2　模型框架

复杂的高速公路多车道驾驶场景下，智能车辆决策模型需要具备高效感知环境、准确评估风险和实时优化策略的能力。基于此，本文提出的模型框架如图1所示，主要包括仿真环境模块、风险评估模块、驾驶策略模块以及动作空间设计。仿真环境模块负责模拟高速公路环境，包括主车及周围车辆的动态信息（位置、速度、车距等），实时更新环境状态，为模型提供高动态输入和反馈。风险评估模块基于安全性指标和贝叶斯推断量化驾驶风险，输出风险概率分布，为策略优化提供依据，降低潜在安全隐患。自注意力机制在此模块中，将多源异构的车辆状态数据编码为特征向量序列，通过计算不同车辆间的相关性权重，强化高风险车辆的影响。状态空间由描述车辆运动学信息及相对位置关系的5个特征构成，用于描述主车与环境的交互。驾驶策略模块通过自注意力机制实现策略的双重优化；目标策

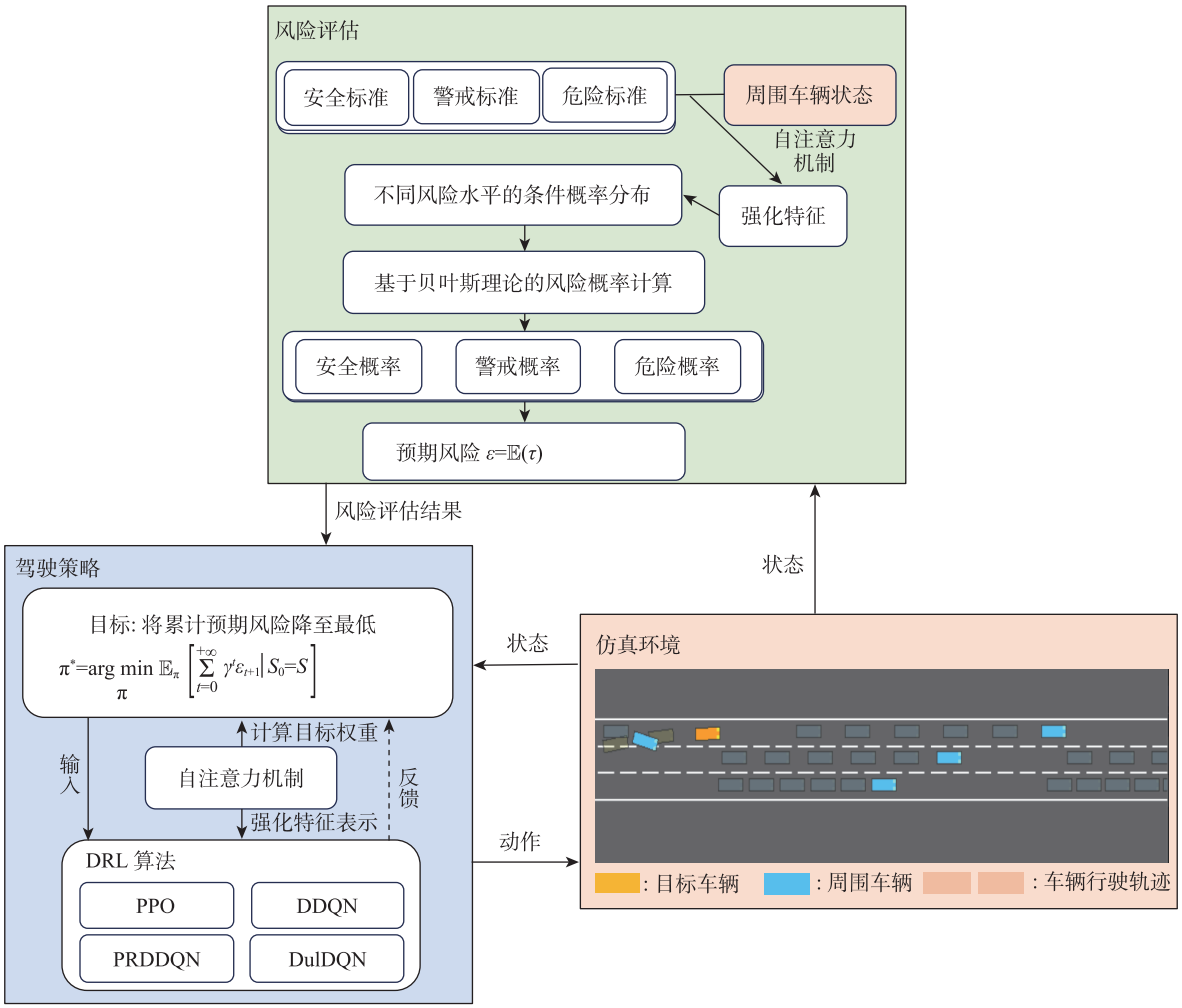

图1 驾驶策略整体框架

Fig. 1 Overall framework of driving strategy

略部分处理安全、效率等驾驶目标数据，动态调整目标权重（如高风险场景强化安全优先级）；DRL算法部分解析风险评估后的环境特征，捕捉多维度特征依赖关系（如车距、速度、车道位置），优化驾驶策略，实现安全性与效率的平衡。动作空间由保持车道、换道、加速和减速4种指令组成，覆盖典型驾驶行为，指导车辆在复杂场景中安全高效行驶。

1.3 风险评估方法

不同于传统的二元潜在风险预测方法，本文的风险评估方法通过估算不同风险水平下的概率分布来进行建模。具体的风险等级定义如下

$$\Omega=\{\text{危险，警戒，安全}\}=\{dangerous,attentive,safe\}=\{D,A,S\} \tag{4}$$

式中：Ω 为风险等级；将 $\{2,1,0\}$ 的分数对应分配到 $\{D,A,S\}$。风险水平定义为

$$\tau\in\Omega=\{2,1,0\} \tag{5}$$

基于不确定理论进行风险建模时，考虑了车辆间的相对位置和不确定性因素，并采用基于分布的安全度量来计算不同风险水平下的条件概率。然后，使用贝叶斯推断评估每一状态下的风险水平。安全度量计算如下

$$P(d|\tau=D)=\begin{cases}1, & d<d_{\mathrm{D}}\\ \mathrm{e}^{-\frac{\Delta d_{\mathrm{D}}^2}{2\sigma^2}}, & d\geqslant d_{\mathrm{D}}\end{cases} \tag{6}$$

$$P(d|\tau=A)=\mathrm{e}^{-\frac{\Delta d_{\mathrm{A}}^2}{2\sigma^2}} \tag{7}$$

$$P(d|\tau=S)=\begin{cases}\mathrm{e}^{-\frac{\Delta d_{\mathrm{S}}^2}{2\sigma^2}}, & d\leqslant d_{\mathrm{S}}\\ 1, & d>d_{\mathrm{S}}\end{cases} \tag{8}$$

$$\Delta d_i = \left| d - d_i \right|,\ \ i \in \Omega \tag{9}$$

式中：d 为主车(HV)和其他车辆(OVs)之间的相对距离；$d_{\mathrm{D}}, d_{\mathrm{A}}, d_{\mathrm{S}}$ 为预先定义的驾驶风险评估的阈值；σ 为标准差，控制概率分布的衰减速度。图2给出了方程的定性表示，展示了相对距离与潜在风险的关系，其中较短的距离意味着较高的风险，反之亦然。风险模型中的主要超参数 $(d_{\mathrm{D}}, d_{\mathrm{A}}, d_{\mathrm{S}}, \sigma)$ 用于控制不同风险曲线的光滑性，这些超参数的值是通过调整自身来确定的，以获得具有合理分布的平滑风险曲线[14-17]。

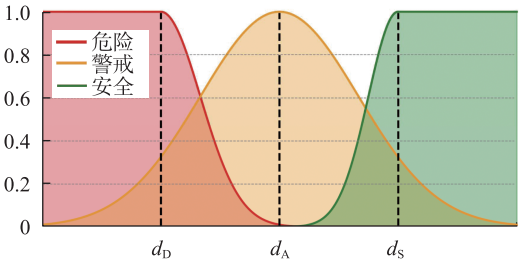

图2　不同相对距离的方程先验分布

Fig. 2　Prior distribution of equations with different relative distances

采用贝叶斯推断计算特定风险水平 τ 的后验概率，公式如下

$$P(\tau|d) = \frac{P(d|\tau)\cdot P(\tau)}{\sum_{\tau\in\Omega} P(\tau)\cdot P(d|\tau)} \tag{10}$$

式中：$P(\tau|d)$ 为给定状态 d 下特定风险水平的概率；$P(d|\tau)$ 为条件概率；$P(\tau)$ 为每个风险等级的先验概率。在本研究中，假设各风险水平具有相同的先验概率，且约束条件 $\sum_{\tau\in\Omega} P(\tau) = 1$。

为实现最小化安全驾驶风险的策略，需要将风险评估结果引入DRL的方法中。然而，离散风险评估结果不能直接应用于DRL。因此，本文定义一个连续的风险系数 ε 来量化风险水平 τ，如式(11)所示

$$\varepsilon = \mathbb{E}(\tau) = \sum_{\tau\in\Omega} \tau\cdot p(\tau|d) = \sum_{\tau\in(2,1)} \tau\cdot p(\tau|d) \tag{11}$$

式中：τ 为式(5)中描述的离散风险水平；ε 为评估风险的期望。通过该连续风险系数，可以表示具有最小预期风险的策略，如式(12)所示

$$\pi^*(S) = \arg\max_{\pi} \mathbb{E}_{\pi}\left[\sum_{i=0}^{+\infty} \gamma^i (\max\varepsilon - \varepsilon_{t+i}) \middle| S_t = S\right] \tag{12}$$

式中：$\max\varepsilon$ 为定义的最大风险值。经比较，式(12)与式(2)格式相同，代表将最小风险决策与传统的DRL策略优化相结合能找到期望风险最小最佳策略。对应Q值函数为

$$Q_{\pi}(S,a) = \mathbb{E}_{\pi}\left[\sum_{i=0}^{+\infty} \gamma^i (\max\varepsilon - \varepsilon_{t+i}) \middle| S_t = S, a_t = a\right] \tag{13}$$

1.4　自注意力机制

自注意力机制是一种通过将输入序列内部不同元素之间的关系进行对比和加权，从而增强局部特征表达准确性的注意力方法[18]。该方法无需人工标注，能够自动学习数据间的依赖关系，从而降低对外部信息的依赖。在动态驾驶场景中，自注意力机制能够帮助模型过滤无关信息，聚焦潜在危险目标，特别是在复杂交通环境中，提高模型安全感知能力并降低碰撞风险。该机制基于transformer框架[19]，采用多头注意力结构，如图3所示。目标车辆及其周围车辆的状态变量 $S_i \in R^{1\times 5}$ 经过共享的编码器(encoder)，通过多层感知机(MLP)和层归一化，映射为高维嵌入表示 $e_i \in R^{1\times d_x}$。这些嵌入特征用于描述车辆间的动态关系，并进一步输入到多头注意力层。其中目标车辆的嵌入表示作为查询(query)，周围车辆的嵌入表示作为键(key)和值(value)。

每个注意力头的注意力向量通过softmax函数计算，进一步结合残差网络生成强化的目标车辆动态特征表示，最终作为输入嵌入到决策模型中指导驾驶决策。

1.5　深度强化学习

PPO算法是一种基于演员-评论家(actor-critic，AC)框架的策略梯度算法，其通过多次小批量更新优化策略，相较于前身信任区域策略优化(trust region policy optimization，TRPO)算法更易于实现。因此，选择PPO算法作为智能车辆决策模型的核心算法，并结合风险最小化策略对其进行改进。

通过在传统优势函数 $A(S,a)$ 基础上引入风险调整项，将风险最小化融入PPO策略的优化过程。传统的优势函数通常由状态-动作值函数 $Q(S,a)$ 与状态值函数 $V(S)$ 的差值构成

$$A(S,a) = Q(S,a) - V(S) \tag{14}$$

引入风险调整，重新定义 $Q(S,a)$，风险调整项的具体计算方法是结合贝叶斯推断获得的风险分布来量化当前驾驶场景的风险水平。表示为

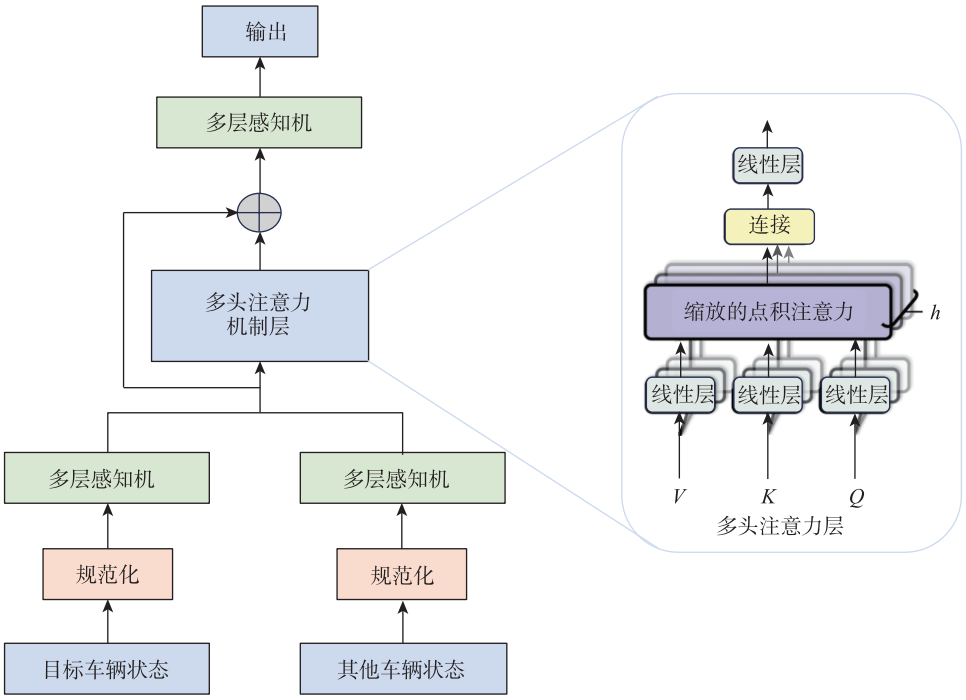

图3 自注意力安全机制网络结构

Fig. 3 Network structure of self attention security mechanism

$$Q_\pi(S,a)=\mathbb{E}_\pi\left[\sum_{i=0}^{+\infty}\gamma^i(r_{t+i}+(\max\varepsilon-\varepsilon_{t+i}))\middle|S_t=S,\ a_t=a\right] \tag{15}$$

为提高算法训练效率，PPO使用截断比率 $\rho_t(\theta)$ 来限制策略更新幅度。表达式为

$$\rho_t(\theta)=\frac{\pi_\theta(a_t|S_t)}{\pi_{\theta_{\text{old}}}(a_t|S_t)} \tag{16}$$

在此基础上的优化目标为

$$L_t^{\text{Clip}}(\theta)=\mathbb{E}_t\left[\min\begin{pmatrix}\rho_t(\theta)A_t(S,a),Clip(\rho_t(\theta),\\1-\varepsilon,1+\varepsilon)A_t(S,a)\end{pmatrix}\right] \tag{17}$$

式中：$\pi(a|S)$ 为策略在状态 S 下选择动作 a 的概率；θ 为策略参数；ε 为超参数，用于限制比率 ρ 的变化；$A_t(S,a)$ 为t时刻的优势函数；$Clip$ 函数的引入限制了策略的更新幅度。在计算优势函数时，直接使用可能引入较大方差，从而影响训练效果。因此，引入广义优势估计（generalized advantage estimation，GAE）以在偏差和方差之间实现权衡。GAE的表达式如下

$$A_t^{GAE(\gamma,\lambda)}=\sum_{l=1}^{\infty}(\gamma,\lambda)^l\left(r_t+\gamma V(S_{t+l+1})-V(S_{t+l})\right) \tag{18}$$

式中：$V(S)$ 为值函数，由网络Critic部分输出；γ 为折现因子，通常设为0.99；λ 为GAE中的权衡方差与偏差的参数，通常设为0.95。为了避免策略在训练过程中陷入局部最优，引入熵正则化（entropy regularization，ER），ER通过增加策略输出分布的熵来鼓励模型探索，防止策略过早收敛于局部最优解。$S(\pi_\theta)$ 是策略 π 的熵。c_1 和 c_2 是调节相应损失部分重要性的系数。熵计算如下

$$S(\pi_\theta)=-\sum_a\pi_\theta(a|S)\log\pi_\theta(a|S) \tag{19}$$

目标优化函数整体形式如下

$$L_t^{\text{Clip+VF+S}}(\theta)=\mathbb{E}_t\left[L_t^{\text{Clip}}(\theta)-c_1L_t^{\text{VF}}(\theta)+c_2S(\pi_\theta)\right] \tag{20}$$

式中：$L_t^{\text{Clip}}(\theta)$ 为通过剪切技术计算的目标函数部分，用于限制策略更新的大小；$L_t^{\text{VF}}(\theta)$ 为值函数损失，一般使用均方误差（MSE）。通过这一目标函数，PPO能够在动态交通环境中学习安全、高效的驾驶策略，特别是在引入风险调整的优势函数后，策略优化不仅关注行车效率，还显著加强了对安全性的约束。

1.6 状态空间与空间动作

选择基于数值信息的状态空间表示高速公路环境状态。该状态空间由目标车辆的绝对运动信息（位置、速度和航向角）及其与周边车辆的相对运动信息（相对距离、速度差和航向角）共同构成[20-21]。主车（HV）和其他车辆之间关系的状态表示如下

$$\boldsymbol{VAO}_0=\left[x_0,y_0,v_0^x,v_0^y,\varphi\right] \tag{21}$$

$$\boldsymbol{VAO}_i=\left[\Delta x_i,\Delta y_i,\Delta v_i^x,\Delta v_i^y,\varphi_i\right],i=1,2,\cdots,n \tag{22}$$

$$\boldsymbol{S}=\left[VAO_0,VAO_1,VAO_2,\cdots VAO_n\right],i=0,1,2,\cdots,n \tag{23}$$

如图4所示，同向3车道高速公路环境的状态可用车辆状态向量 $\boldsymbol{S}$ 表示。使用状态量 (x,y,v,φ) 来描述车辆的当前状态。$\boldsymbol{VAO}_0$ 由目标车辆的绝对运动信息构成；x_0,y_0 为目标车辆的横纵坐标；v_0^x,v_0^y 为横向速度和纵向速度；φ_0 为车身航向角；$\Delta x_i,\Delta y_i$ 为目标车辆与其他车辆之间的相对横纵距离；$\Delta v_i^x,\Delta v_i^y$ 为横纵速度差；φ_i 为车身航向角。

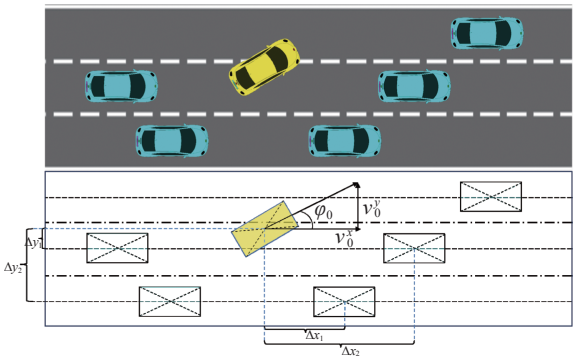

图4　车辆状态空间示意图

Fig. 4　Schematic diagram of vehicle state space

模型考虑用于横向控制的转向动作和用于纵向控制的油门动作，将动作空间 A 定义为5种典型驾驶行为

$$A=\left\{a_0,a_1,a_2,a_3,a_4\right\} \tag{24}$$

式中：$a_0=LT$ 表示车辆向左侧换道；$a_1=RT$ 表示车辆向右侧换道；$a_2=IL$ 表示车辆保持当前直行驾驶状态；$a_3=ILF$ 表示直行加速；$a_4=ILS$ 表示直行减速。

1.7　奖励函数

深度强化学习的策略优化依赖于奖励函数的引导，因此，设计合理的奖励函数十分重要。传统DRL目标是找到一种最大化实现奖励函数的策略。然而，本文希望找到一种可以最小化预期驾驶风险的策略（见式(10)）。因此对评估的驾驶风险引入负号处理，使目标转变为最大化问题。为鼓励智能体学习稳健策略，奖励通常应是一个数值为正且具有明确物理含义的值[22]。因此在评估驾驶风险时，引入最大风险值 $\max\varepsilon$ 来描述冗余风险空间，以反映相应策略的安全等级。为找到风险最小策略，着重考虑风险评估结果 ε（见式(11)）。风险回报函数定义如下

$$R_{\text{risk}}=\max\varepsilon-\varepsilon_t \tag{25}$$

式中：R_{risk} 为风险评估的回报；ε_t 为瞬时时刻 t 的风险评估结果。在确保安全性的同时，奖励函数还考虑了驾驶效率和车道规范性，具体设计如下：

1）速度奖励函数：鼓励车辆在安全范围内接近理想速度。函数定义为

$$R_{\text{speed}}=\frac{v_{\max}-v_{\text{hv}}}{v_{\max}-v_{\min}} \tag{26}$$

式中：v_{hv} 为当前主车速度；$v_{\min}$ 和 $v_{\max}$ 分别为高速环境允许通行的最低和最高速度阈值。本文的速度范围设置为 [18,30] m/s。

2）最优目标车道奖励函数：引导车辆尽量保持在中间车道或超车道，避免占用最右侧车道。在环境中，车道编号自右向左依次为1、2、3车道，其中 *lane_id* 为车道编号。函数定义为

$$R_{\text{lane}}=\begin{cases}-1, & lane\_id=1\\ 1, & lane\_id=2,3\end{cases} \tag{27}$$

3）生存奖励函数：避免主车（HV）碰撞和越界。函数定义为

$$R_{\text{exist}}=\begin{cases}0.1, & \text{exist}\\ -1, & \text{otherwise}\end{cases} \tag{28}$$

式中：*exist* 表示不发生碰撞和冲出边界；函数奖励值0.1和−1是以往研究中常用的[23-24]。如果车辆发生碰撞或越界，立即给予大幅惩罚，促使主车在训练过程中学习并掌握规避碰撞及越界的策略。

4）换道惩罚函数：抑制频繁换道行为。函数定义为

$$R_{\text{LC}}=-0.1\cdot change\_penalty \tag{29}$$

式中：*change_penalty* 是一个二元值，表示当前是否发生换道操作。如果换道，值为1；反之为0。

5）车道中心奖励函数：鼓励车辆在车道中心行驶，提高稳定性。通常情况下，人类驾驶员为了驾驶安全，会将汽车行驶在车道中央。函数定义为

$$R_{\text{center}}=\mathrm{e}^{-\frac{\left(la_{\text{center}}-la_{\text{hv}}\right)^2}{2\sigma^2}} \tag{30}$$

式中：la_{center} 为当前车道中心位置，la_{hv} 为主车的横向位置。

综上，考虑到在自动驾驶中降低风险比其他奖励回报要更重要，且为避免显著取值范围差异问题，子奖励函数设计了合理的取值范围，按照强化

学习中常见的简化方法[25-26]，将各个子函数的权重设置为1。可得到完整的奖励函数为

$$R=R_{\text{risk}}+R_{\text{speed}}+R_{\text{lane}}+R_{\text{exist}}+R_{\text{LC}}+R_{\text{center}} \quad (31)$$

通过式(31)中的奖励设置，依据式(10)、式(11)、式(12)，可利用DRL方法找到考虑驾驶风险评估的驾驶策略。本文用于训练原始DRL方法(DDQN，Dueling DQN，PPO，PRDDQN，PPO-Mul)的奖励函数定义如下

$$R_{\text{l}}=R_{\text{speed}}+R_{\text{lane}}+R_{\text{exist}}+R_{\text{LC}}+R_{\text{center}} \quad (32)$$

奖励函数仅在训练阶段用于引导智能体学习。为公平比较，使用相 同超参数的风险模型和奖励函数来检验不同方法在换道场景下的有效性。

2 试验与分析

2.1 仿真环境及试验设计

大多数基于DRL的驾驶决策方法都是在仿真模型上进行训练和测试[27-28]，以避免真实场景中的高昂试错成本[29]。为验证所提模型的表现，设计了一系列试验，评估模型在复杂动态交通场景中的安全性与效率。

试验基于轻量化仿真平台Highway-env，构建了3车道高速公路驾驶场景。仿真环境包括自主决策的主车(HV)及周围的其他行驶车辆(OV)。主车在多样化的交通情境中执行换道、超车和保持车道等任务，以验证模型的性能。为模拟更真实复杂的动态交通环境，其他车辆(OV)的初始位置在不发生碰撞的前提下随机生成，采用智能行驶模型[30](intelligent driver model，IDM)及在此基础上扩展的激进驾驶模型(aggressive driving model，ADM)和保守驾驶模型(conservative driving model，CDM)对周围其他车辆进行行为决策和控制，模型参数均采用环境中的默认值。为模拟真实混合交通流，基于Ubiquitous Traffic Eye开放数据集的实际观测数据，激进驾驶行为、正常驾驶行为和保守驾驶行为的比例分别为33%、22%、45%。试验参考该实际比例设置分配情况，设定激进驾驶模型车辆占比30%，正常驾驶模型(采用IDM作为正常驾驶模型)车辆占比30%，保守驾驶模型车辆占比40%。

主车配备虚拟激光雷达传感器，感知范围为150 m，包括左右前方、正前、正后和左右后方6个方向。表1列出了仿真环境及车辆的关键参数。

在风险评估方法中，基于分布的安全度量计算涉及的多个超参数会影响模型性能与风险评估准确性。距离阈值 $d_{\text{D}},d_{\text{A}},d_{\text{S}}$ 依据中国交通法规设定，标准差 σ 遵循统计学原理控制不同风险区域过渡的平滑度，危险区域到警戒区域过渡的标准差 σ_{D} 和警戒区域到安全区域过渡的标准差 σ_{S} 计算公式如式(33)、式(34)

$$\sigma_{\text{D}}=(d_{\text{A}}-d_{\text{D}})/3 \quad (33)$$

$$\sigma_{\text{S}}=(d_{\text{S}}-d_{\text{A}})/3 \quad (34)$$

为验证风险评估模块和多头注意力机制的有效性，选择多个经典算法和改进模型作为基线，并将风险评估模块和注意力机制分别嵌入基线模型进行验证。具体测试模型如表2所示。

表1 仿真环境及车辆主要参数说明

Tab.1 Description of simulation environment and main vehicle parameters

编号	参数	值
1	车道宽度	4.0 m
2	车辆长度	5.0 m
3	车辆宽度	2.0 m
4	最大行驶速度	30.0 m/s
5	最小行驶速度	16.7 m/s
6	采样时间	1.0 s
7	仿真频率	15.0 Hz
8	相对危险距离阈值	50 m
9	相对警戒距离阈值	80 m
10	相对安全距离阈值	100 m

表2 测试模型及说明

Tab.2 Test Model and Explanation

编号	模型	说明
1	DDQN	改进自DQN，引入两个网络
2	PRDDQN	结合优先经验回放的改进型DDQN
3	Dueling DQN	将Q值函数分解为状态值和优势函数的DQN
4	PPO	近端策略优化算法
5	RA-DDQN	加入风险评估模块的DDQN基线模型
6	RA-PRDDQN	加入风险评估模块的PRDDQN基线模型
7	RA-Dueling DQN	加入风险评估模块的Dueling DQN基线模型
8	RA-PPO	加入风险评估模块的PPO基线模型
9	PPO-Mul	加入多头注意力机制的PPO基线模型
10	RA-PPO-Mul	结合风险评估模块与多头注意力机制的综合模型

为全面评估模型性能，定义以下关键指标。这些指标从安全性、效率和合理性3个方面，量化了模型的性能表现。具体指标定义如表3所示。

表3　评价指标及说明

Tab. 3　Evaluation indicators and explanations

编号	指标名称	说明
1	时间安全率	无碰撞时间占比
2	无碰撞安全率	无碰撞次数占比
3	平均速度	完成任务的平均行驶速度
4	换道次数	每回合的平均换道次数

2.2　试验验证及试验分析

训练阶段试验结束后，对模型进行测试与分析，评估其在复杂动态交通场景中的实际表现。根据试验结果，从安全性、效率和决策合理性3个方面分析模型表现。

训练阶段，仿真环境设置最大仿真时长为40 s，训练步数为30万次，以确保模型充分学习动态场景中的驾驶策略。图5展示了不同模型在训练阶段的时间安全率变化曲线，反映了风险评估模块（RA）和多头注意力机制的作用。

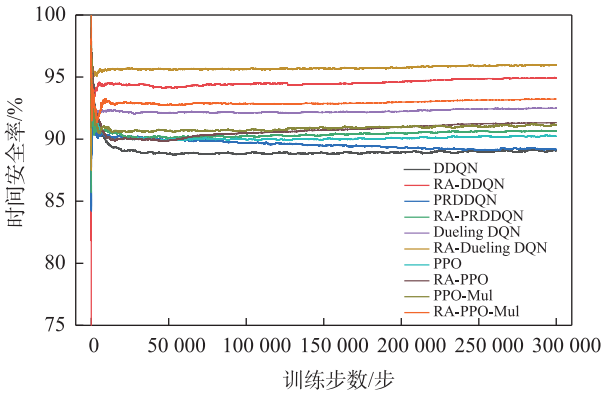

图5　训练阶段的时间安全率

Fig. 5　Time safety rate during the training phase

1）风险评估模块：从图中看到，所有基线模型加入风险评估模块后，时间安全率都有显著提升，收敛速度也更快。例如，RA-DDQN、RA-Dueling DQN和RA-PPO-Mul的时间安全率最终稳定在95%~97%。

2）多头注意力机制：PPO-Mul和RA-PPO-Mul模型的时间安全率和收敛速度均优于未引入自注意力机制的模型前身PPO和RA-PPO。表明多头注意力机制通过增强环境感知能力，帮助模型在训练阶段更有效地识别潜在风险，提高了学习效率。

在测试阶段，对训练好的模型进行验证。主车（HV）需在2 km的固定行驶距离内完成驾驶任务。为确保结果的稳定性与可靠性，测试重复多次并取平均值。测试过程中，随机初始化周围车辆（OVs）的行为和位置，以模拟复杂的动态交通场景。主要评价指标包括无碰撞安全率、平均速度和换道次数。图6展示了不同模型在测试阶段的无碰撞安全率。引入RA模块的模型（如RA-DDQN、RA-Dueling DQN和RA-PPO-Mul）在所有对比中均表现出色，其安全率显著高于未引入RA模块的原始模型。RA-DDQN的无碰撞安全率提升至88%，相较于DDQN（70%），提升了18个百分点，证明风险评估模块有效降低了模型在测试场景中的碰撞风险。同时RA-PPO-Mul模型的无碰撞安全率达98%，高于PPO（84%）和PPO-Mul（91%），验证了多头注意力机制进一步提升了模型对复杂场景的感知能力。

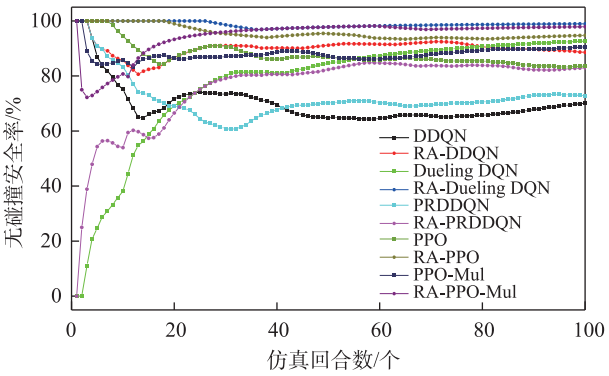

图6　测试阶段的无碰撞安全率

Fig. 6　Collision free safety rate during the testing phase

图7和图8分别展示了不同模型在测试阶段的平均速度和换道次数。结合图6可知，PRDDQN和DDQN尽管速度较高（分别为28.29 m/s和27.58 m/s），但换道次数也最高（9.72次和8.35次），且无碰撞安全率低于75%，表明高频换道和过高速度显著增加了碰撞风险，导致整体安全性较低。Dueling DQN和RA-Dueling DQN无碰撞安全率很高，后者接近99%，但平均速度较低（约19 m/s），其极低的换道次数表明模型采取了极端保守的策略，不符合高效驾驶需求。RA-PPO在测试阶段以24.58 m/s的平均速度和5.30次的换道次数实现了95%的无碰撞安全率，展现了效率与安全的良好平衡。而RA-

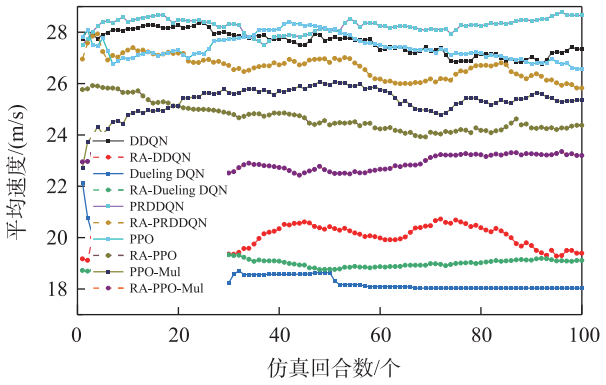

图7 测试阶段的平均速度

Fig. 7 Average speed during the testing phase

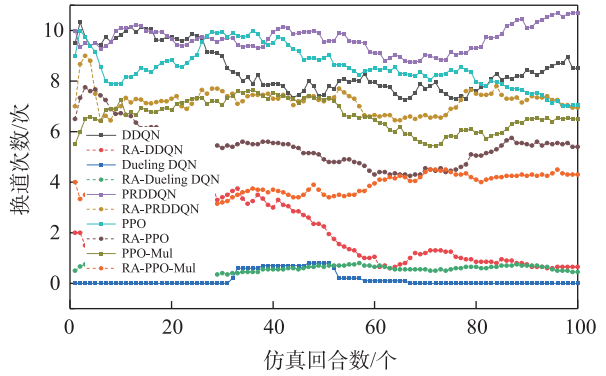

图8 测试阶段的换道次数

Fig. 8 Number of lane changes during the testing phase

PPO-Mul模型在此基础上进一步优化，以稍低的速度和更低的换道频率实现了98%的无碰撞安全率。这表明多头注意力机制使模型能够更加全面地评估周围环境，在复杂场景中更好地感知风险并采取更谨慎的决策。

表4总结了各模型测试阶段的无碰撞安全率、平均速度和平均换道次数。风险评估模块的引入显著提升了各模型的无碰撞安全率，减少了换道次数，并适度降低了驾驶速度，使得决策更为谨慎合理。特别是RA-PPO-Mul模型，在安全性方面表现最佳，达到98%的高安全率，显示出风险评估模块是确保自动驾驶任务中决策安全性和稳定性的关键机制。同时，RA-PPO-Mul通过降低速度和减少换道次数，在效率与安全之间实现了最佳平衡，表明多头注意力机制使得模型能够动态感知周围车辆行为的变化，更准确地评估潜在风险，从而减少不必要的换道行为并优化行驶策略。

表4 各模型测试结果

Tab.4 Test results of each model

模型	无碰撞安全率/%	平均速度/（m/s）	平均换道次数/次
DDQN	70.00	27.58	8.35
RA-DDQN	88.00	24.61	4.96
Dueling DQN	93.00	19.56	2.18
RA-Dueling DQN	99.00	19.03	0.57
PRDDQN	73.00	28.29	9.72
RA-PRDDQN	83.00	26.54	7.18
PPO	84.00	27.34	8.42
RA-PPO	95.00	24.58	5.30
PPO-Mul	91.00	25.48	6.69
RA-PPO-Mul	98.00	22.86	3.77

此外，为探究超参数取值对模型决策安全性的影响，设计了参数敏感性试验。保持其他条件不变，仅调整 d_D,d_A,d_S 的取值，计算对应标准差 σ_D,σ_S，分析其对RA-PPO-Mul性能的影响。试验设置如表5所示。

表5 超参数试验取值

Tab.5 Value of over parameter test

编号	相对危险距离阈值/m	相对警戒距离阈值/m	相对安全距离阈值/m	标准差 σ_D	标准差 σ_S
1	50	80	100	10	6.67
2	30	60	80	10	6.67
3	70	100	120	10	6.67

其中，编号1为基准组，编号2为实验组A，编号3为实验组B。表6总结了RA-PPO-Mul在不同超参数下的测试结果。实验组A缩小危险区域之后，无碰撞安全率下降5%，速度提升10.8%，换道增加45%，表明 d_D 缩小后，模型对中距离风险的容忍度提高，决策更激进，效率提升但安全性下降。实验组B在扩大危险区域之后，无碰撞安全率上升1%，速度下降9.6%，换道减少68%，表明 d_D 扩大后，模型对中距离风险过度规避，决策过度保守，效率降低。

由表6可知，超参数取值对模型安全性和效率

表6 改变超参数的模型测试结果

Tab.6 Model test results for changing super parameters

组合	无碰撞安全率/%	平均速度/（m/s）	平均换道次数/次
基准组	98.00	22.86	3.77
实验组A	93.00	25.32	5.48
实验组B	99.00	20.65	1.19

影响显著。d_D 与安全率正相关,与速度和换道次数负相关。实际应用需依场景合理调整超参数。

最后,不同驾驶策略场景也会影响模型决策。改变分配情况,设定原场景为场景1;场景2为激进驾驶占比50%,正常驾驶占比30%,保守驾驶占比20%;场景3为激进驾驶占比10%,正常驾驶占比40%,保守驾驶占比50%。由表7可知,面对不同驾驶策略场景,模型表现出良好的适应性。

表7　不同驾驶场景的模型测试结果

Tab.7　Model test results of different driving scenarios

场景	无碰撞安全率/%	平均速度/(m/s)	平均换道次数/次
场景1	98.00	22.86	3.77
场景2	96.00	21.57	5.83
场景3	99.00	24.81	2.30

上述试验表明,相比于传统模型和仅引入单一模块的模型,RA-PPO-Mul 模型优势显著。该模型表现出良好的场景适应性,借助风险评估和多头注意力机制,能在复杂动态场景中精准感知周围车辆潜在风险,做出安全高效决策。

3　结论

1）基于位置不确定性和相对距离的风险评估方法,降低了模型的高风险操作概率,试验结果验证了其在提升决策安全性方面的有效性。

2）自注意力机制通过捕捉关键环境特征,提升了模型对复杂场景的动态感知和适应能力。

3）RA-PPO-Mul 模型在复杂场景下的安全性和稳定性均优于对比基准模型,验证了将注意力机制与风险评估模块结合的有效性,表明其在动态驾驶环境中具备安全高效的决策能力和较大的实际应用潜力。

参考文献：

[1] LI G F, LIAO Y, GUO Q Q, et al. Traffic crash characteristics in Shenzhen, China from 2014 to 2016[J]. International Journal of Environmental Research and Public Health, 2021, 18(3): 1176.

[2] KUEFLER A, MORTON J, WHEELER T, et al. Imitating driver behavior with generative adversarial networks[C]// 2017 IEEE Intelligent Vehicles Symposium (IV), June 11-14, 2017, Los Angeles, CA, USA. New York: IEEE, 2017: 204-211.

[3] GONZÁLEZ D, PÉREZ J, MILANÉS V, et al. A review of motion planning techniques for automated vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(4): 1135-1145.

[4] 王靓喆. 基于博弈论的智能网联自动驾驶车辆换道行为研究[D]. 长春: 吉林大学, 2022.
WANG L Z. Research on lane-changing behavior of intelligent networked autonomous vehicles based on game theory [D]. Changchun: Jilin University, 2022.

[5] GIPPS P G. A model for the structure of lane-changing decisions[J]. Transportation Research Part B: Methodological, 1986, 20(5): 403-414.

[6] 邓建华, 冯焕焕. 基于换道决策机理的多车道元胞自动机模型[J]. 交通运输系统工程与信息, 2018, 18(3): 68-73.
DENG J H, FENG H H. Multilane cellular automaton model based on the lane-changing mechanism[J]. Journal of Transportation Systems Engineering and Information Technology, 2018, 18(3): 68-73.

[7] 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.
PEI X F, MO S J, CHEN Z F, et al. Lane changing of autonomous vehicle based on TD3 algorithm in human-machine hybrid driving environment[J]. China Journal of Highway and Transport, 2021, 34(11): 246-254.

[8] TALEBPOUR A, MAHMASSANI H S, HAMDAR S H. Modeling lane-changing behavior in a connected environment: a game theory approach[J]. Transportation Research Procedia, 2015, 7: 420-440.

[9] LOPEZ-MARTIN M, CARRO B, SANCHEZ-ESGUEVILLAS A. Application of deep reinforcement learning to intrusion detection for supervised problems[J]. Expert Systems with Applications, 2020, 141: 112963.

[10] CHENG S, WANG Z, YANG B, et al. Convolutional neural network-based lane-change strategy via motion image representation for automated and connected vehicles[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(9): 12953-12964.

[11] WANG P, CHAN C Y, DE LA FORTELLE A. A reinforcement learning based approach for automated lane change maneuvers[C]//2018 IEEE Intelligent Vehicles Symposium (IV), June 26-30, 2018, Changshu, China. New York: IEEE, 2018: 1379-1384.

[12] MO S J, PEI X F, CHEN Z F. Decision-making for oncoming traffic overtaking scenario using double DQN[C]//2019 3rd Conference on Vehicle Control and Intelligence (CVCI). September 21-22, 2019. Hefei, China. New York: IEEE, 2019: 1-4.

[13] EDOUARD L. An Environment for Autonomous Driving Decision-Making[Z].https://github.com/eleurent/highway-env, 2018.

[14] LI G F, CHEN Y Y, CAO D P, et al. Extraction of descriptive driving patterns from driving data using unsupervised algorithms[J]. Mechanical Systems and Signal Processing, 2021, 156: 107589.

[15] LI G F, LIAO Y, GUO Q Q, et al. Traffic crash characteristics in Shenzhen, China from 2014 to 2016[J]. International Journal of Environmental Research and Public Health, 2021, 18(3): 1176.

[16] LI G F, YANG Y F, ZHANG T R, et al. Risk assessment based collision avoidance decision-making for autonomous vehicles in multi-scenarios[J]. Transportation Research Part C: Emerging Technologies, 2021, 122: 102820.

[17] LI G F, LI S L, LI S, et al. Continuous decision-making for autonomous driving at intersections using deep deterministic policy gradient[J]. IET Intelligent Transport Systems, 2022, 16(12): 1669-1681.

[18] 李岚欣. 面向自然语言处理的注意力机制研究[D]. 北京: 北京邮电大学, 2019.
LI L X. Research on attention mechanism for natural language processing[D]. Beijing: Beijing University of Posts and Telecommunications, 2019.

[19] 周双喜, 杨丹, 潘远, 等. 基于注意力机制的YOLOv5路面裂缝检测与识别[J]. 华东交通大学学报, 2024, 41(2): 56-63.
ZHOU S X, YANG D, PAN Y, et al. Detection and recognition of YOLOv5 pavement cracks based on attention mechanism[J]. Journal of East China Jiaotong University, 2024, 41(2): 56-63.

[20] CHEN Y F, EVERETT M, LIU M, et al. Socially aware motion planning with deep reinforcement learning[C]//2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), September 24-28, 2017, Vancouver, BC, Canada. New York: IEEE, 2017: 1343-1350.

[21] GALCERAN E, CUNNINGHAM A G, EUSTICE R M, et al. Multipolicy decision-making for autonomous driving via changepoint-based behavior prediction: theory and experiment[J]. Autonomous Robots, 2017, 41(6): 1367-1382.

[22] DUAN J L, EBEN LI S, GUAN Y, et al. Hierarchical reinforcement learning for self-driving decision-making without reliance on labelled driving data[J]. IET Intelligent Transport Systems, 2020, 14(5): 297-305.

[23] LONG P X, FAN T X, LIAO X Y, et al. Towards optimally decentralized multi-robot collision avoidance via deep reinforcement learning[C]//2018 IEEE International Conference on Robotics and Automation (ICRA), May 21-25, 2018, Brisbane, QLD, Australia. New York: IEEE, 2018: 6252-6259.

[24] LI G F, YANG Y F, QU X D. Deep learning approaches on pedestrian detection in hazy weather[J]. IEEE Transactions on Industrial Electronics, 2020, 67(10): 8889-8899.

[25] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.

[26] QI X W, LUO Y D, WU G Y, et al. Deep reinforcement learning enabled self-learning control for energy efficient driving[J]. Transportation Research Part C: Emerging Technologies, 2019, 99: 67-81.

[27] YE Y J, ZHANG X H, SUN J. Automated vehicle's behavior decision making using deep reinforcement learning and high-fidelity simulation environment[J]. Transportation Research Part C: Emerging Technologies, 2019, 107: 155-170.

[28] KIRAN B R, SOBH I, TALPAERT V, et al. Deep reinforcement learning for autonomous driving: a survey[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(6): 4909-4926.

[29] GRIGORESCU S, TRASNEA B, COCIAS T, et al. A survey of deep learning techniques for autonomous driving[J]. Journal of Field Robotics, 2020, 37(3): 362-386.

[30] TREIBER M, HENNECKE A, HELBING D. Congested traffic states in empirical observations and microscopic simulations[J]. Physical Review E, 2000, 62(2): 1805-1824.

通信作者:范泽敏(1999—),男,研究方向为自动驾驶决策算法。E-mail:zmfan@njust.edu.cn。

(责任编辑:李根)