文章编号：2095-2163(2024)01-0180-06 中图分类号：U491.2 文献标志码：A

融入驾驶员感知模型的智能车辆左转决策研究

邱思远，王孝兰，张伟伟
（上海工程技术大学 机械与汽车工程学院，上海 201620）

摘 要：智能车辆左转是最为复杂的驾驶任务之一，现有的基于强化学习的智能车辆左转决策模型，将上流感知模块的输出直接作为左转决策算法的状态空间，与人类驾驶员相比缺乏对感知信息的筛选和推理，使得决策算法很难收敛到一个较优的决策策略。本文通过对驾驶员感兴趣区域建模，将其融入基于近端策略优化（PPO）强化学习的决策算法中，过滤掉冗余的感知信息，使智能车辆关注对决策有用的信息区域；为了充分考虑周边车辆的交互关系，将注意力机制加入基于 PPO 的强化学习决策算法中；从智能车辆的安全、高效和舒适的 3 个方面设计奖励函数，以引导智能车辆的学习方向。实验结果表明，融入驾驶员感兴趣区域模型和注意力机制的决策框架优于其他的基线框架，收敛速度更快，提高了智能车辆决策的安全性和通行效率。

关键词：智能车辆；强化学习；PPO；决策模型

Research on left-turn decision of intelligent vehicles integrated into driver perception model

QIU Siyuan, WANG Xiaolan, ZHANG Weiwei
(School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

Abstract: Intelligent vehicle left turn is one of the most complex driving tasks, and the existing intelligent vehicle left turn decision-making model based on reinforcement learning takes the output of the upstream perception module as the state space of the left turn decision-making algorithm, and lacks the screening and reasoning of the perception information compared with the human driver, which makes it difficult for the decision-making algorithm to converge to a better decision-making strategy. In this paper, by modeling the driver's area of interest, this paper integrates it into the decision-making algorithm based on proximal policy optimization (PPO) reinforcement learning, filters out redundant perception information, and allows intelligent vehicles to pay attention to the information area useful for decision-making. In order to fully consider the interaction relationship of surrounding vehicles, the attention mechanism is added to the PPO-based reinforcement learning decision-making algorithm. And design the reward function from the three aspects of safety, efficiency and comfort of intelligent vehicles to guide the learning direction of intelligent vehicles. Experimental results show that the decision-making framework integrated into the driver's area of interest model and attention mechanism is better than other baseline frameworks, and the convergence speed is faster, which improves the safety and traffic efficiency of intelligent vehicle decision-making.

Key words: intelligent vehicles; reinforcement learning; proximal police optimization; decision model

0 引 言

在城市交叉路段，无保护左转的智能车辆与周边其他交通参与者存在不可避免的冲突，既要考虑安全通行，又要保证通行效率。智能车辆无保护安全左转是实现无人驾驶的关键之一，无人驾驶行为决策方法目前主要分为基于规则的决策方法与基于学习的决策方法。

基于规则的决策方法是根据行驶场景、交通法规和驾驶经验，人工建立驾驶行为规则库，依据驾驶场景确定智能车辆的驾驶行为。Montemerlo 等[1]采用有限状态机（FSM）作为 Junior 智能车辆的行为决策模型，根据驾驶场景人工建模驾驶行为规则，用于指导不同驾驶状态之间切换；Furda 等[2]将 FSM 与多标准的驾驶决策相结合，从一组可行的驾驶策略中选择最优的驾驶策略。中国科学技术大学杜明博等[3]研发的“智能先锋Ⅱ”，使用有限状态机切换不同的驾驶场景，再使用 ID3 （Iterative Dichotomiser

作者简介：邱思远（1997-），男，硕士研究生，主要研究方向：计算机视觉与交通安全。
通讯作者：王孝兰（1985-），女，博士，讲师，主要研究方向：智能车辆导航控制。Email：xiaolanwang@sues.edu.cn
收稿日期：2022-11-25

3)决策树进行行为决策。基于规则的智能车辆行为规划是确定的,但智能车辆无保护左转遇到的交通情况是复杂多变的,很难建立一个能够包含所有状况的规则库[4]。

为了克服基于规则的决策模型的限制,研究人员开始使用基于学习的方法进行自动驾驶决策,主要分为模仿学习(IL)和强化学习(RL)两大类。IL 通过监督学习的方式从专家演示中学习驾驶策略,将传感器观测值映射到控制命令[5];Yang 等[6]提出了一个多任务学习框架,以端到端的方式同时对转向角和速度进行控制。IL 以端到端的方式学习专家数据,已取得初步成功,但其性能受限于训练数据。相比之下,RL 通过与环境交互获取奖励来更新策略,不需要人工标注的数据集。Kendall 等[7]通过深度强化学习实现自动驾驶,减少对传统规则和地图的依赖;Hoel 等[8]提出了一种基于强化学习的决策模型,用于高速公路车辆的变道和速度规划;Isele 等[9]使用基于深度强化学习的驾驶策略应对无信号灯的交叉路口;Tram 等[10]通过将强化学习和模型预测控制 MPC 相结合来学习如何在十字路口进行决策。但现有基于强化学习的智能车辆决策模型,往往是直接将上流感知模块的输出作为强化学习的状态空间,缺乏人类驾驶员主观的观察和推理,使得状态空间中包括了许多冗余的信息,训练时很难快速收敛到一个最优的行为决策模型。

本文对驾驶员感兴趣区域建模,并将其融入基于强化学习的行为决策模型中,来对感知模块输入的信息进行筛选,降低强化学习环境的状态空间维度。为了充分考虑周边车辆的交互关系,将注意力机制加入强化学习的特征提取网络中,让智能车辆对决策有利的信息给予更高关注度。

1　驾驶员感兴趣区域建模

本文研究十字路口智能车辆的左转驾驶任务,在考虑智能车辆与周围交通参与者之间潜在碰撞风险的基础上,对驾驶员感兴趣区域进行建模。因为行人的运动空间相对较小,所以重点考虑了自车与周围车辆之间的碰撞风险,而未涵盖行人等其他行为参与者。同时为了方便驾驶员感兴趣区域建模,假设所有车辆都是沿固定轨迹行驶的,周边车辆都是以 9 m/s 的速度匀速行驶,受 Kolekar[11]中风险域的启发,在风险域的基础上对驾驶员感兴趣区域建模。首先对沿固定跟踪轨迹的车辆风险域进行建模,进而对车辆碰撞风险建模;计算自车与周边车辆的潜在冲突点处的风险场强度,通过设定车辆碰撞概率阈值;计算可能与自车发生碰撞的车辆边界位置,把车辆碰撞边界作为自车的感兴趣区域边界。

1.1　车辆风险域建模

本文研究的驾驶场景为无交通信号的十字路口,研究车辆在该场景下的左转驾驶任务,以车辆质心为原点,沿跟踪轨迹建立 Frenet 坐标系,考虑车辆速度和加速度对车辆风险强度影响,建立沿跟踪轨迹变化的车辆风险域函数,可表示为式(1):

$$r(s_\tau, d_\eta) = r_c \cdot r_l,\ s_\tau \in (0, k_1 v) \tag{1}$$

其中,r 为任意位置风险域强度;s_τ 为纵向坐标点;d_η 为横向坐标点;a 为车辆加速度;v 为车辆速度。

r_c 为纵向风险域强度,r_l 为横向风险域强度,可表示为式(2)和式(3):

$$r_c = (\alpha + 1)(s_\tau - k_1 v)^2 \tag{2}$$

$$r_l = e^{-k_2 d_\eta^2} \tag{3}$$

其中,k_1,k_2 为参数。

建立的车辆沿跟踪轨迹的风险域如图 1 所示。

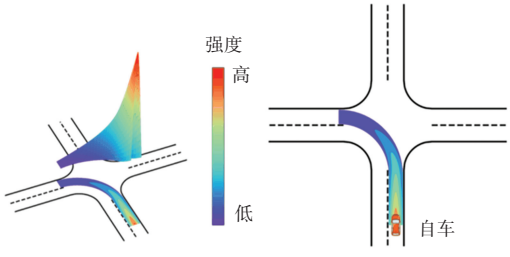

图 1　车辆沿跟踪轨迹的风险域

Fig. 1　Risk domain of the vehicle along the tracking trajectory

1.2　车辆碰撞风险建模

本文在风险域的基础上根据两辆车辆分别在冲突点处产生的风险强度的关系,来评估两辆车辆碰撞的风险,即车辆之间的风险强度越近,发生碰撞的概率就越大。如图 2 所示的自车与车辆 C 的风险强度更近,与车辆 A、B 相差较大,所以自车与 C 发生碰撞的可能性更大。

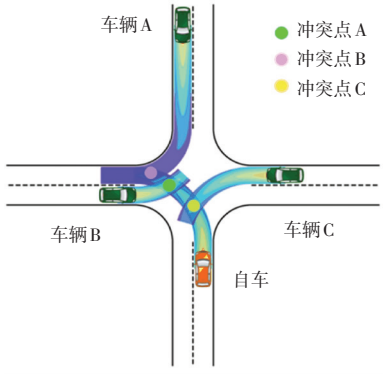

图 2　车辆碰撞风险图

Fig. 2　Vehicle collision risk diagram

本文通过两个车辆在冲突点处产生的风险强度之差 Δr，对两辆车辆的碰撞风险 p_c 进行建模，如式(4)所示：

$$p_c = \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{(\Delta r)^2}{2}\right)} \tag{4}$$

1.3 驾驶员感兴趣区域边界

自车左转时与左边路口、右边路口和对象路口来车存在潜在的冲突点，冲突点的个数与自车在路口的位置相关。自车左转时与左边路口来车的潜在冲突点有两处，一处是左边路口车辆直行与自车产生的碰撞冲突；另一处是左边路口车辆左转与自车产生的碰撞冲突。同理可知其他路口来车与自车的冲突点，各车道车辆与自车产生的冲突点如图 3 所示。

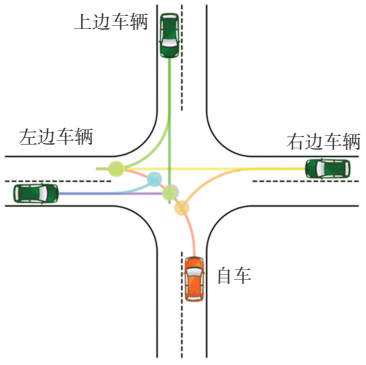

图 3 自车与周边车辆的冲突点

Fig. 3 The point of conflict between ego and the surrounding vehicles

本文假设周边车辆都是沿固定轨迹做匀速行驶，并根据交通法规假设行驶速度为 9 m/s，选取相对较大的速度，以确保驾驶员感兴趣区域边界有一定的感知冗余空间，同时假设周边车辆和自车的跟踪轨迹方程由直线和圆弧组成分段函数 $tr(x)$，这些假设使得周边车辆产生沿着车辆跟踪轨迹的固定风险域，相反自车的风险域是随自身运动状态动态变化的。

在已知自车位置、运动状态、跟踪轨迹时，就可得到当前时刻自车的风险域。获取冲突点位置 $C_i(x,y)$，i 为冲突点的数目，根据式(5)将冲突点坐标转换为自车 Frenet 坐标系的坐标点 $\bar{C}_i(s,0)$，将 $\bar{C}_i$ 代入式(1)可得冲突点在自车的风险域中的风险强度 $r_i(s,0)$。根据设定的碰撞阈值 p_c 和式(4)、式(6)，计算不同路口沿不同跟踪轨迹的周边车辆与自车冲突的边界风险强度 r_i^b；根据周边车辆的风险域和轨迹方程，计算不同路口周边车辆到冲突点处的边界轨迹长度，由式(5)和轨迹函数 $tr(x)$ 将其转换为路口中心坐标系下的边界点 $b_i(x,y)$。

$$s = \int_{x_v}^{x_c} \sqrt{1 + tr'(x)}\, dx \tag{5}$$

$$r_i^b = r_i + \Delta r \tag{6}$$

从路口各个边界点中选取驾驶员感兴趣区域边界点，通过建立自车的车身坐标系，取边界点到 y 轴最远的点为左右边界点，并与默认值 6 相比，取最大值；同理取边界点中到 x 轴最远的点为上边界点，与默认值 20 相比，取最大值；最终确定感兴趣区域 3 个方向的边界点 b_r, b_l, b_{up}，如图 4 所示。

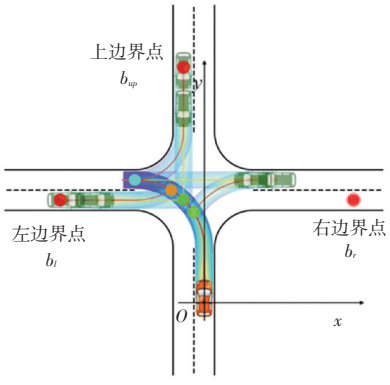

图 4 感知边界点

Fig. 4 Sense boundary points

1.4 驾驶员感兴趣区域

本文在自车的车身坐标系下对驾驶员感兴趣区域(Driver Sense Interest Field -DSIF)进行建模，选取二次多项式作为驾驶员感兴趣区域的上边界、左边界和右边界函数，如式(7)所示，其中左右边界函数关于 y 轴对称，a_1, a_2, b, c_1, c_2 为参数，y_i 为上边界函数与左右边界函数交点的 y 值，x_l^b, x_r^b 为左右边界点的 x 值。本文根据风险域函数、车辆尺寸结构和驾驶场景的特点对参数进行了简化，b 设为 6，c_2 设为 3。通过左右感知边界点可以得到参数 a_2，再根据上边界点和左右边界函数可以求解 a_1 和 c_1。这些参数是由自车的位置、速度和加速度决定的，驾驶员感兴趣区域边界与自车的运动状态相关，如图 5 所示。

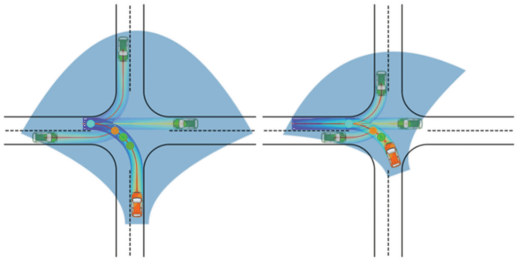

图 5 驾驶员感兴趣区域

Fig. 5 The driver's area of interest

$$\begin{cases} y_u = a_1 x^2 + c_1, & x \in (x_l^b, x_r^b) \\ x_r = a_2 (y-b)2 + c_2, & y \in (-5, y_i) \\ x_l = -a_2 (y-b)2 - c_2, & y \in (-5, y_i) \end{cases} \quad (7)$$

2　基于 PPO 强化学习的智能车辆左转决策框架

本文使用 PPO 强化学习算法构建智能车辆的左转决策框架，如图 6 所示。将驾驶员感兴趣区域模型融入基于 PPO 强化学习算法的智能车辆左转模型中，减少无关信息对车辆决策的影响。同时，为了充分考虑周边车辆的交互关系，本文用基于注意力机制的策略网络和价值网络对周边车辆的信息进行提取。

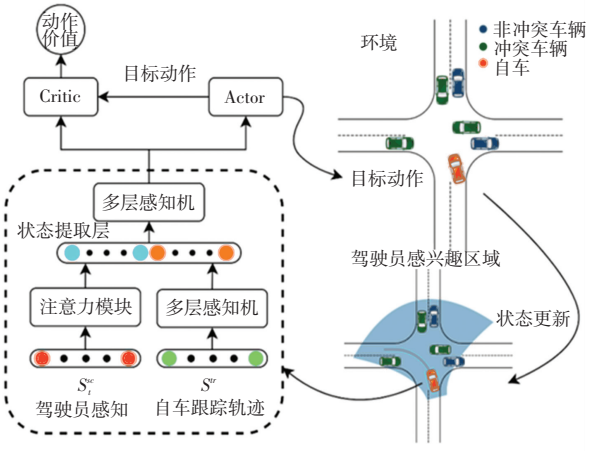

图 6　智能车辆的左转决策框架

Fig. 6　Left-turn decision framework for smart vehicles

2.1　PPO 算法介绍

近端策略优化算法 PPO（Proximal Policy Optimization）是基于 Actor-Critic 框架的强化学习算法，相比信赖域策略优化算法（TRPO）更加易于实现，对超参数不敏感，对连续和离散控制问题都有很好的性能表现。PPO 算法有 PPO-Penalty 和 PPO-Clip 两种实现方式，PPO-Penalty 算法通过引入新旧两个策略分布的 Kullback-Leibler（KL）散度作为目标函数的惩罚项，在训练时自动调整惩罚系数，可以解决 TRPO 算法的硬约束问题。PPO-Clip 算法通过对目标函数进行裁剪，从而得到更保守的更新。本文使用 PPO-Clip 算法作为决策模型的基础算法，目标函数如式（8）所示。

$$L^{CLIP}(\theta) = \hat{E}_t[\min(r_t(\theta)\hat{A}_t, clip(r_t(\theta), 1-\varepsilon, 1+\varepsilon)\hat{A}_t] \quad (8)$$

其中 ε 为截断因子。

$r_t(\theta)$ 为新旧策略的比值，如式（9）所示：

$$r_t(\theta) = \frac{\pi_\theta(a_t \mid s_t)}{\pi_{\theta_{old}}(a_t \mid s_t)} \quad (9)$$

其中，π_θ 为随机策略。

$\hat{A}_t$ 为 t 时刻的优势函数，如式（10）所示：

$$\hat{A}_t = \sum_{i=0}^{m-1} \gamma^i r_{t+i} + \gamma^m V_\phi(s_{t+m}) - V_\phi(s_t) \quad (10)$$

其中，V_ϕ 为状态价值函数，γ 为奖励折扣因子。

2.2　状态表征

本文中智能车辆观测状态由两部分组成 $s_t = \{s_{hs}, s_{tr}\}$，s_{hs} 为驾驶员感兴趣区域下观测到的周边车辆和自车的运动状态，s_{tr} 为自车将要跟踪的轨迹。将智能车辆观测状态定义为 $s_{hs} = \{o_e, o_1, \cdots, o_{20}\} \in R^{21}$，其中 o_i 由周边车辆和自车的运动状态 $p = (x, y, \varphi, f)$ 转化而来，f 表示是否与自车有冲突关系。因为智能车辆决策受跟踪轨迹影响很大，所以本文将部分智能车辆的跟踪轨迹作为智能车辆的观测状态的一部分。智能车辆轨迹状态为 $s_{tr} = \{x_1, y_1, \varphi_1, \cdots, x_{20}, y_{20}, \varphi_{20}\} \in R^{60}$，由将要跟踪的 20 个轨迹点组成，采样间距为 0.5 m。

2.3　动作表征

由于智能车辆沿固定轨迹行驶，因此只需要对智能车辆的纵向速度进行规划。为了更好的考虑车辆的动力学和运动学特性，本文将智能车辆的加速度作为强化学习模型的动作，不考虑智能车辆加速和制动时的相应时间，用智能车辆的输出的加速度对车辆的运动状态进行更新。本文考虑到智能车辆的驾驶安全性、舒适性和驾驶场景，对智能车辆的加速度的范围进行限制，即 $a \in [-5, 2.5]$ m/s^2。

2.4　奖励函数

驾驶员通常希望安全、舒适、快速的通过十字路口。为了尽可能让智能车辆能够像人类驾驶员一样，智能车辆的奖励函数由安全奖励、舒适奖励和速度奖励组成，安全奖励是鼓励智能车辆与周边车辆保持一个相对安全的距离，防止发生碰撞；舒适奖励是希望自车输出的加速度和上一时刻相差不能太大，减少加速度，提高舒适度；速度奖励是为了鼓励智能车辆以[7-9 m/s]速度区间中的速度行驶，以尽快通过路口。奖励函数定义如式（11）~式（14）所示：

$$r_s = \begin{cases} -20(0.2 + \dfrac{v_e}{9}), & \text{if collision occurs} \\ -20e^{-t_c}, & \text{other} \end{cases} \quad (11)$$

$$r_v = \begin{cases} 0.4\dfrac{(v_e - v_{\min})}{(v_{\max} - v_{\min})}, & \text{if } v_e \leqslant v_{\max} \\ -0.2e^{(v_e - v_{\max})}, & \text{if } v_e \geqslant v_{\max} \end{cases} \tag{12}$$

$$r_c = \begin{cases} -|\ a_e^t - a_e^{t-1}\ |, & |\ a_e^t - a_e^{t-1}\ | > 0.5 \\ 0, & \text{other} \end{cases} \tag{13}$$

$$r = r_s + r_c + r_v \tag{14}$$

其中，r_s 为安全奖励；t_c 为自车与最近车辆碰撞时间；r_v 为速度奖励；r_c 为舒适奖励；v_e 和 a_e 为自车速度和加速度；$v_{\max}$ 和 $v_{\min}$ 为期望的最大速度和最小速度。

3 实验设计和分析

3.1 实验环境

本文使用 highway-env 仿真环境来训练和评估提出的决策框架。本文将提出的驾驶员感兴趣区域模型并入 highway-env 中，在此基础上搭建一个无交通信号灯的双车道十字路口驾驶场景如图 7 中所示，其中绿色车辆为智能车辆，蓝色车辆为周边车辆，黄色区域为驾驶员感兴趣区域。为了增加场景多样性，通过调整周边车辆生成的数量来控制交通流，并随机设定周边车辆的相关属性，例如生成位置、行驶方向和速度等。在十字路口中呈现连续的周围交通环境，以模拟真实的交通情况，直接从仿真环境中获取周围车辆的位置信息作为智能车辆感知输出。仿真环境中每个回合的终止条件为智能车辆与其他车辆发生碰撞、智能车辆顺利通过路口和智能车辆通行时间超过 30 s。

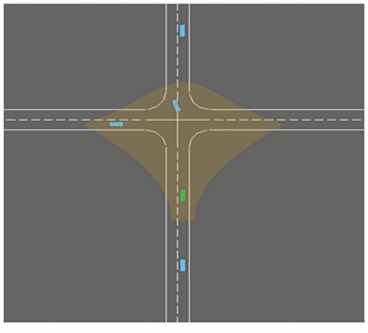

图 7 highway-env 仿真场景

Fig. 7 highway-env simulation scenario

3.2 结果分析

为了验证本文提出的驾驶员感兴趣区域模型和注意力机制对智能车辆决策的影响，将融入驾驶员感兴趣区域模型和注意力机制的 PPO 决策模型（DSIF-Attention-PPO）与其他两个基线策略进行对比试验。将不同模型都训练4 000回合，驾驶员感兴趣区域模型与基于注意力机制的策略网络或基于全连接的策略网络相互组合的 PPO 决策模型的平均累积奖励曲线如图 8 所示。

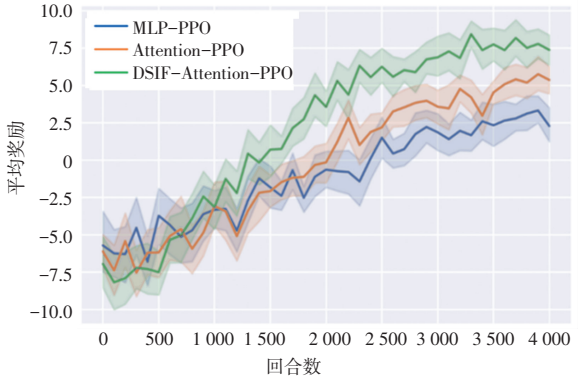

图 8 平均累计奖励曲线

Fig. 8 Average cumulative reward

由图 8 可知，基于注意力机制 PPO 决策模型的 Attention-PPO 和 DSIF-Attention-PPO 模型，在训练 2 500轮后奖励曲线都趋于收敛，而基于全连接层的 PPO 决策模型训练 3 400 轮后才开始缓慢收敛，这说明注意力机制的策略网络可以让智能车辆更关注有利于决策的信息，同时可以充分考虑周边车辆的交互关系，从而加快 PPO 决策模型的收敛，让智能车辆做出更优的决策。加入驾驶员感兴趣区域模型 DSIF-Attention-PPO 与不加相比，奖励曲线收敛更快，最终的收敛的奖励值也更高，说明驾驶员感兴趣区域模型可以很好去除感知模型输入的冗余信息，使智能车辆像驾驶员一样从物理空间上去关注对决策有利的信息，帮助智能车辆更好的做决策。

模型测试结果见表 1，可知本文提出的决策模型在成功率和平均速度都要优于其他的决策模型，融合驾驶员感兴趣区域模型的 PPO 后在成功率和平均速度上分别提高了 2.3%和 0.49 m/s，加入注意力机制后成功率和平均速度上分别提高了1.3%和 0.35 m/s。

表 1 模型测试结果

Table 1 Model test results

模型	成功率/%	平均速度/(m·s^{-1})
PPO + MPL	85.6	6.98
PPO + Attention	86.9	7.33
PPO + DSID + Attention	89.2	7.82

4 结束语

本文针对智能车辆在十字路口无保护左转驾驶

（下转第 190 页）