# 强化学习在无人车领域的应用

1953729 吴浩泽

## 摘要

无人车(UGV)可替代人类自主地执行民用和军事任务，对未来智能交通及陆军装备发展有重要战略意义。随着人工智能技术的日益成熟， 采用强化学习技术成为了无人车智能决策领域最受关注的发展趋势之一。本文简要分析总结了强化学习在无人车智能决策中的研究进展，包括障碍物规避、变道与超车、车道保持和道路交叉口通行四种典型场景。

## 在避障问题中的应用

在避障问题中无人车根据自车和障碍物的位置和状态信息，在满足乘坐舒适性和行驶安全性的条件下，输出转向、制动和油门指令控制车辆规避障碍物。

Arvind 提出基于 MLP-SAＲSA 和基于 MLP-Q 学习的避障算法。设计了以车载的7 个超声波雷达的感知数据为输入量，输出离散的制动、转向和加速动作的端对端决策模型，将多层感知机( multi-layer perceptron，MLP) 引入到对 Q 函数的预测中，以提高避障策略的收敛速度。车辆在包含多 个动态障碍物的仿真环境下实现自主避障，且无碰撞通行的成功率达96% 。

Chae 提出复杂城市场景下基于DQN 的主动制动算法，如图 1 所示。使用 6 层的深度神经网络架构，采用障碍物相对于主车的横向和纵向的位置和速度作为 DQN 网络输入，输出无制动、弱制动、中制动和强制动四个不同强度等级的制动动作。在奖励函数的设计中，考虑车辆的乘坐舒适性和安全性，对过早的制动行为和与障碍物发生碰撞进行惩罚。经过 2000 次的迭代训练，无人车能有效地处理行人横穿马路等随机突发事件，但面对碰撞时间( time to collision，TTC) 等于 1．4 s 的紧急工况仅有 74% 的避障成功率。

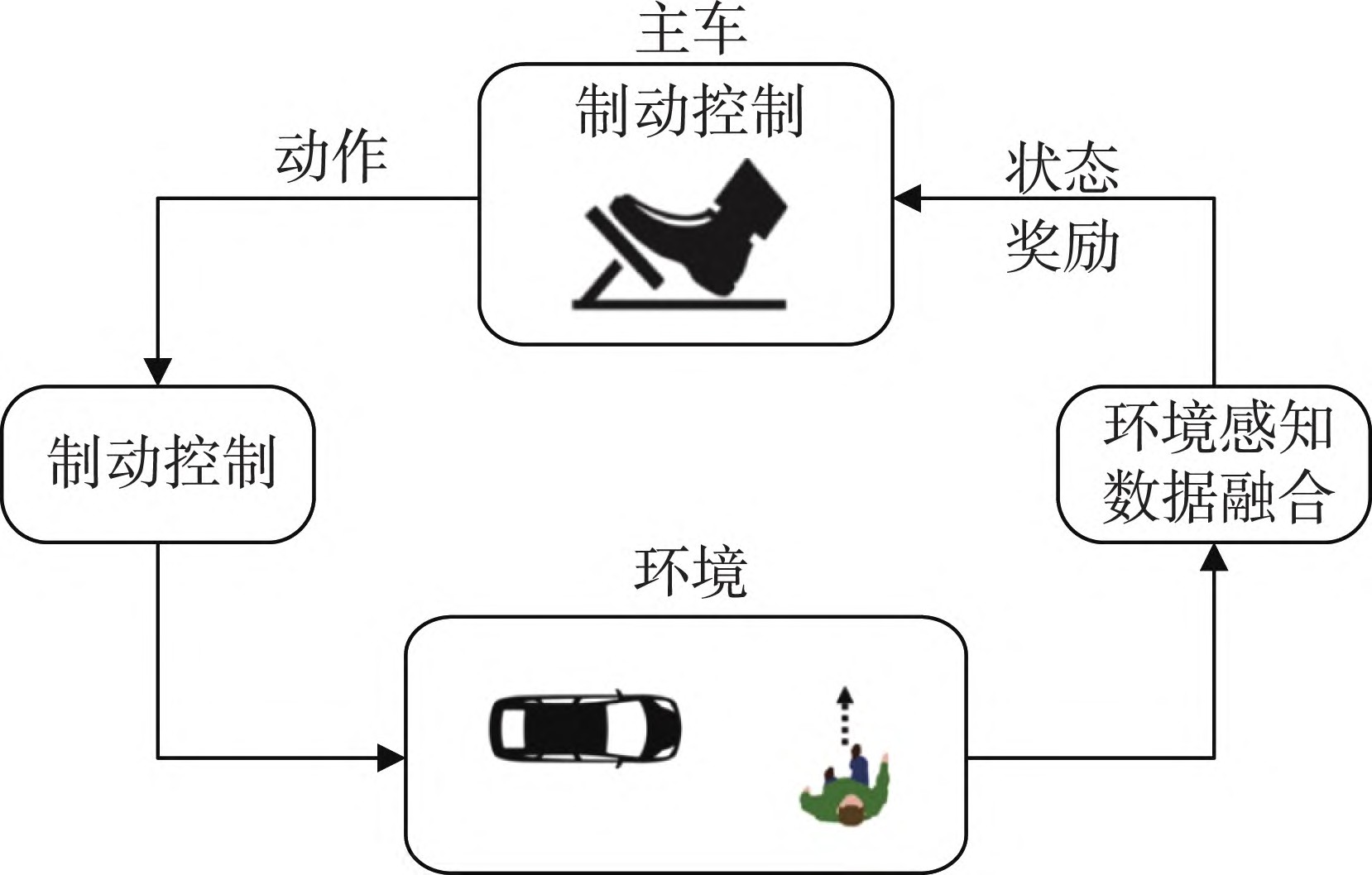


图 1 基于 DQN 的主动制动算法

## 在变道与超车问题中的应用

在变道与超车问题中，无人车根据自车和周围车辆状态、自车的期望速度和交通规则约束等，做出变道及超车决策，指导车辆超越前方低速车辆，以尽快地通过特定的交通流。

Loiacono 提出基于 Q 学习的超车决策算法，建立了包含主车和前方车辆相对距离、相对速度，主车和车道边缘横向距离等在内的离散状态，并以有限的离散动作驱动车辆完成超车。在TOＲCS 赛车模拟器中验证了该算法在直线赛道和弯道上的超车效果，在超车持续时间、超车时最高车速和超车成功率等指标上明显优于人类驾驶员。 针对求解连续空间下超车决策问题时 Q 学习存在的计算效率低的问题。Liu 提出基于 线性函数逼近强化学习的变道决策算法。作者将变道场景建立为状态、动作空间连续的 MDP 模型，将基于多核的最小二乘策略迭代法( multi-kernel LSPI， MKLSPI) 引入对 Q 函数的拟合中，并基于国防科技大学研制的红旗 HQ3 无人车采集的实车感知数据， 对决策算法开展离线测试工作，论证了算法的有效性和泛化能力。Min 利用非线性值函数逼近 的方法，提出基于 Dueling DQN 的超车决策算法，构建以卷积神经网络( convolutional neural networks， CNN) 和长短期记忆网络( Long short-term memory， LSTM) 提取的视觉图像和雷达点云的特征作为状态输入，输出横向的变道操作及纵向的车速变化的决 策模型。该算法改进 DQN 网络结构，利用 DNN 输出的状态值函数和动作优势函数近似拟合 Q 函数， 提高了策略学习的收敛速度。

An 提出车联网环境下基于 DDPG 的变道决策算法，网络结构如图 2 所示。该算法策略网络输入包含两部分，分别为由车载传感器获得的主车 状态信息和由 V2X 通信获得的前方车辆状态信息， 并通过 2 个全连接的隐藏层输出对主车油门和方向盘的控制。在 Airsim 软件中的仿真实验验证该算法的有效性，但由于输入层网络结构固定，其仅能处 理 2 个车辆交互这种简单场景，缺少对更为复杂交通场景的适应性。

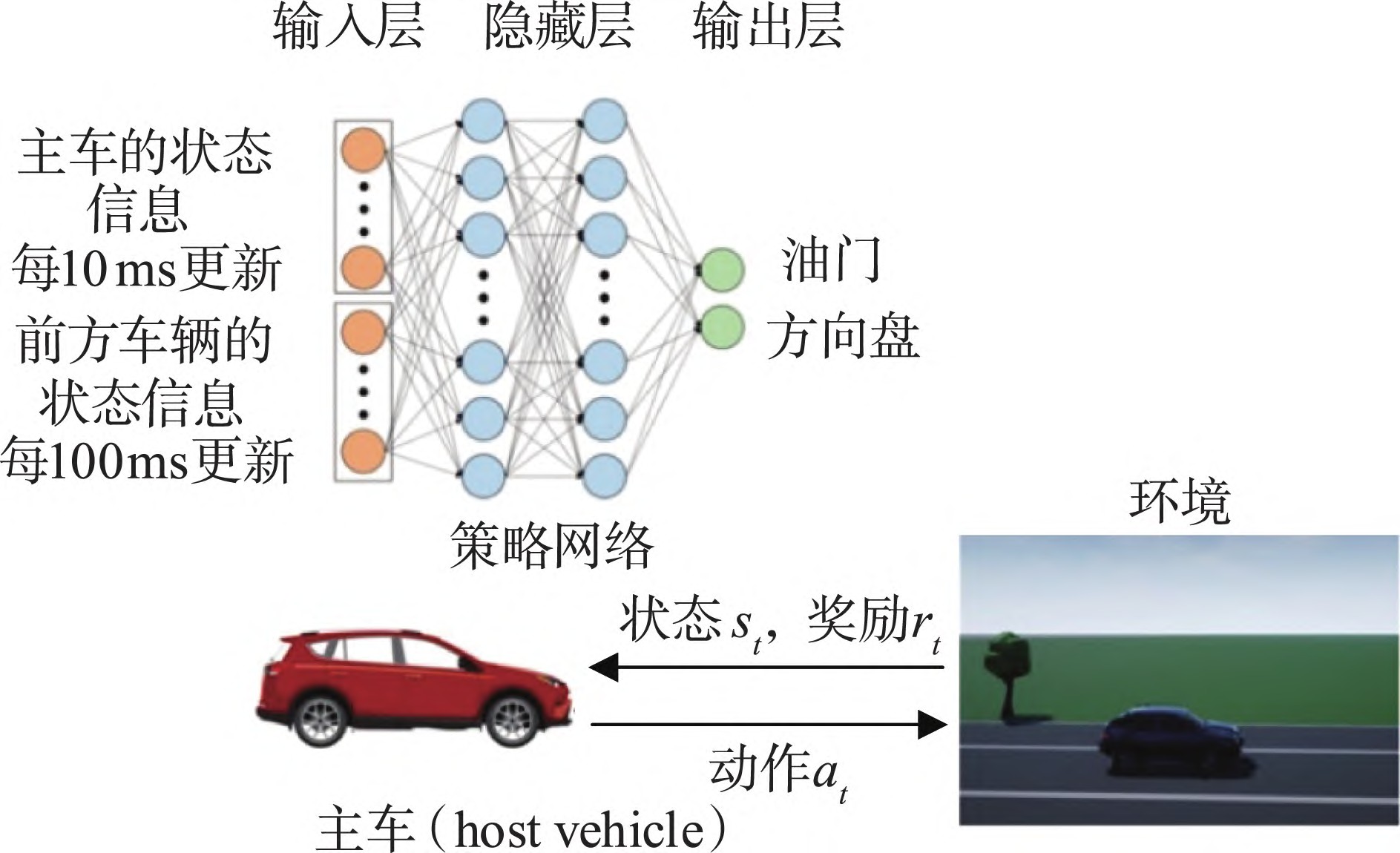


图 2 车联网环境下变道决策算法框架

## 在车道保持问题中的应用

在车道保持问题中，无人车根据车载传感器获 得的车道线信息，输出车辆方向盘转角控制指令， 以使车辆在车道中心线附近行驶。

视觉感知是检测车道线的最有效手段。方川 提出基于 Double DQN 的车道保持算法，以原始的 ＲGB 图像作为网络输入，分别利用当前 Q 网络和目标Q 网络处理方向盘控制动作选择和目标Q 函数预测。在仿真试验中，车辆在直线车道及大曲 率弯道的车道保持任务中均表现出良好的性能。Kendall 等［48］提出视觉感知数据输入下基于 DDPG 的车道保持算法( 如 图 3) ，并将在虚拟环境中训练好的算法网络结构和参数迁移到实车上，车辆仅依 靠单目相机的 ＲGB 图像完成了 250 m 的车道保持路测。然而该方法忽略视觉传感器抗干扰能力差、易受光照影响等缺点，且决策模型场景遍历的深度 不足，难以完成特殊天气条件下的车道保持任务。

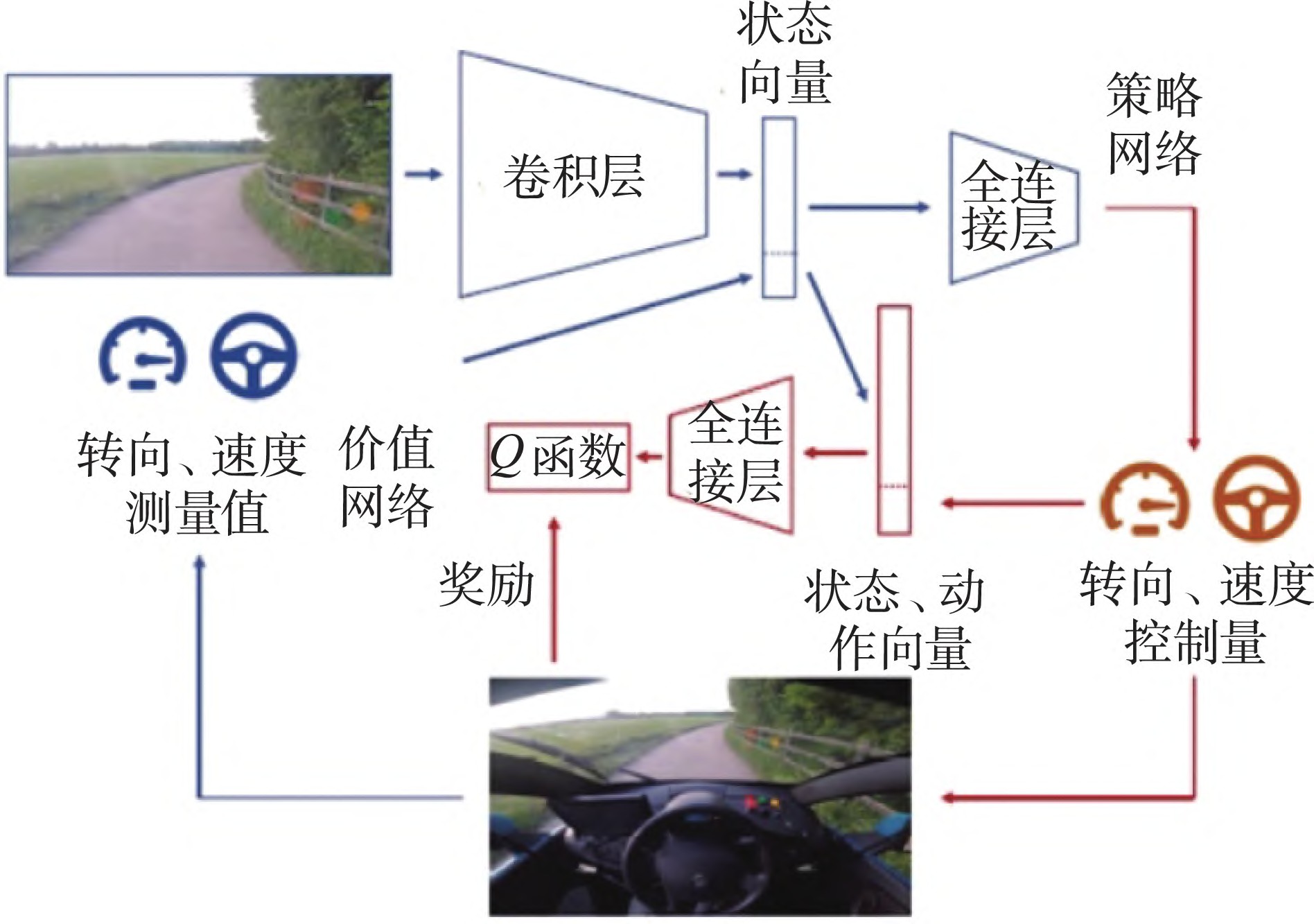


图 3 基于 DDPG 的车道保持算法框架

## 原始视觉图像包含大量与决策无关的环境细节，而细微的环境变化易导致决策模型错误，进而 引发车辆驶出车道等危险驾驶行为。针对此问题， Wolf［49］利用机器视觉剔除无关环境信息，提取车道 线的灰度化图像，构建由灰度化的视觉图像到车辆 方向盘的端对端决策，降低细微的环境亮度变化及 无关环境细节对决策模型的影响。并利用经验回放 机制降低训练样本的相关性，以减轻 DQN 算法处理高维图像数据时存在的不稳定性。

## 在道路交叉口通行问题中的应用

在道路交叉口通行问题中，无人车根据交叉口各车道上车辆位置、速度及交通规则等，输出执行机构控制指令，以控制车辆无碰撞地通过交叉口。

无交通信号灯的交叉口通行是最具挑战性的交通场景，学者们对基于强化学习的通行决策方法进行大量研究。Saxena 设计基于近端策略优化( proximal policy optimization，PPO) 的通行决策算法。作者利用由数据驱动的仿真训练建立交叉口中无人车周围车辆间交互的隐式模型，并通过设置车辆的加速度和转向角度阈值，减少不良的加速和转向动作，提高乘坐舒适性。Qiao 提出课程式学习( curriculum learning，CL) 和 DＲL 结合的交叉口决策算法。作者利用 CL 自动生成若干由简单到复杂的样本，引导 DＲL 学习驶入并通过城市交叉路口的策略，仿真实验中通过交叉口的成功率达98．7% 。

Müller 等［53］提出基于视觉场景理解的决策算法，引入编码器—解码器网络来提取 ＲGB 视觉图像更细化的语义特征，实现原始图像到多场景通用的语义分割图像的映射，将语义图像作为决策模型输入，输出车辆期望的轨迹。其后，作者将训练好的决策模型迁移至小型卡车上，车辆可在多个驾驶场景( 晴朗、阴天、雨雪) 自主地通过交叉路口。该方法通过模块化和抽象语义分割的方法降低真实场景传感器噪声等对决策的影响，提高决策算法的迁移能力。

无交通信号灯的交叉口中车辆缺少交通规则约束。无人车无法获悉其他车辆的驾驶意图，因而无法预判其行驶轨迹，且因车辆间的相互遮挡易造成无人车的感知盲区，给决策的安全性带来巨大隐患。Isele 等［54 － 55］ 利用卡尔曼滤波( Kalman filte- ring，KF) 预测可能与无人车发生碰撞车辆的行驶轨迹，并根据预测结果约束 DQN 决策算法的动作空间，提高车辆在交叉口通行的安全裕度。Gruber等［56］设计基于 ＲA 的在线安全验证方法，利用 ＲA 建立其他车辆未来时间在交叉口所有可达集，以验证决策的安全性。其后，Lauer 等［57］提出基于 ＲA和责任敏感安全模型( responsibility-sensitive safety， ＲSS) 的验证方法，解决了 ＲA 因考虑最危险情况下周围车辆的占用空间而导致的无人车在交叉口驾驶策略过度保守的问题。Stiller 提出一种风险认知 DQN 的交叉口决策算法，在奖励函数中引入风险项度量感知盲区内的车辆对决策安全性的程度，减少无人车采取冒进决策行为的概率。

无交通信号灯的交叉口的复杂程度高，且事故风险隐患多，给无人车决策的安全性带来巨大挑战。基于强化学习的决策模型无法有效预估事故风险，结合行驶轨迹预测、安全性验证等方法对提高决策安全性具有重要意义。

### 互评人：2053512 孙博闻 互评成绩：8分