

一、选题背景及意义

1.1 背景

1.1.1 无人驾驶介绍（定义概念、大概涉及到的技术）

1.1.2 深度学习介绍（）

1.2 意义

为什么要研究，自动驾驶的优点，基于深度学习的自动驾驶的优点：安全意义，便捷，

二、国内外研究现状

找的国内外公司研究的资料，谁在做，做了什么，效果

三、课题内容及研究

1.1 怎么做（综述）

1.2 关键技术（分点 1.2.3.）

1.3 技术算法运用、结合

四、仿真平台的介绍（测试）

选一个仿真平台，大概内容，搭建。

五、课题总结及改进

还可以改进什么什么

自动驾驶分级

SAE 将自动驾驶分为六个等级，0 级到 5 级。**L0** 所指的是预警和少量保护辅助功能，碰撞告警属于第 0 级的技术；**L1** 阶段车辆开始介入制动与转向其中一项控制，分担驾驶员的工作，自动防碰撞、定速巡航属于第 1 级的辅助驾驶；**L2** 阶段车辆开始接管纵向与横向的多个控制，但驾驶员需要监控驾驶环境也就是说司机注意力仍然要保持驾车状态，出现问题时以便随时接管车辆，驾驶员在车内的自动泊车功能属于第 2 级；**L3** 阶段自动系统既能完成某些驾驶任务，也能在某些情况下监控驾驶环境，但驾驶员必须准备好重新取得驾驶控制权，无法进行睡觉或者深度的休息，技术人员也通常将 **L2 和 L3 之间的分界线**，视作“辅助驾驶”和“自动驾驶”的区别所在；**L4** 阶段自动系统在某些环境和特定条件下，能够完成驾驶任务并监控驾驶环境，L4 的部署目前来看多数是基于城市的使用，可以是全自动的代客泊车，也可以是直接结合打车服务来做；**L5** 阶段自动系统在所有条件下都能完成的所有驾驶任务。

背景

传统的自动驾驶基本选择保守的“**专家策略**”。所谓自动驾驶的专家策略，就是根据专家已知的场景，采用 **if-then 的形式设计行驶规则**，如跟车，停止，变道等。但是这样的方式在高速的简单场景可以勉强应付，但面对普通公路上，双向车线，行人，信号灯，十字路口等**复杂场景**，显然十分吃力。

近几年，**强化学习**逐渐在各大领域活跃。它**适合解决决策问题**，即自主进行决策，并且**可以做连续决策**。非常合适解决自动驾驶的决策问题，所以决策（路径规划）和强化学习（尤其是深度强化学习 DRL）一拍即合，近年越来越受到研究者的关注与青睐，**现阶段实际应用的更多还是高级辅助驾驶（ADAS）技术**。但是，**具体怎样应用强化学习到决策中，应用到决策模块的哪个位置？**例如是应用到类似于机器人的局部路径规划，还是应用于上层驾驶决策，还是 end-to-end 等等。还有一些人提出用**分层强化学习和逆向强化学习（模仿学习）**来做自动驾驶，研究者们各持己见，探索前行。

关于自动驾驶

实现全自动驾驶都需要哪些功能和技术呢？**高精度识别技术**(结合传感器：摄像头、超声波传感器、毫米波雷达 LiDAR（激光雷达）、高清动态地图)、**最佳决策和路径规划（AI 大脑）**、**通信互联技术**（将车辆的行驶状况统一收集到 Cloud 上，通过“车-车”间的通信、“路-车”间的通信，在传感器产生识别死角时也能确保安全）。

自动驾驶的三个模块：①**感知模块**，包括摄像头、传感器，即硬件方面，采集到的图像信息、视频信息或者传感器的数据反馈到了决策模块，也叫黑匣子，是无人驾驶技术当中具有决定性的模块，主要包括 Planning 和之后的预测。进行**信息的预处理**，主要包括对图像或者视频信息的分割、检测或者识别，如果识别的准确率更高可能会对之后的决策有比较好的优势。②**决策模块**，主要包括 GPU、CPU 等计算单元。经过很多训练，利用强化学习来做自动驾驶即像人考驾照的过程，学习怎样开车，最后达到上路的水平。③**控制模块**，主要是对自动驾驶的控制，比如制动和减速。预警系统，如果是突发情况，决策模块反应不过来就会直接给到预警系统，采取制动或者减速。（通过方向盘来对车辆进行横向运动操作、通过加速和减速来控制车辆的纵向运动。）

关于强化学习

强化学习的核心思想就是，通过某种手段影响被试验者的行为。为了实现这个目标，实验者需要构建一个完整的实验环境，通过给予被试验者一定的**观测和回报**，让其产生实验者想要的结果。强化学习的**目标是最大化**所获得的回报。它有两个最大的**特点**，就是不断试错和看重长期回报。为了让更多人快速地应用强化学习，业界开放了一批可以公开访问搭建的**实验环境**，比如 Arcade Learning Environment、Box2D、MuJoCo、Gym、Baselines（关于 Gym：https://mp.weixin.qq.com/s/7Qj7RVG1-j7kQvKS7h_BeA）等等。

一些比较热门的强化学习算法：**DQN**（Deep Q-Learning Networks）以及他的变体 **Double DQN** 在离散型动作中应用效果比较好，但连续性动作当中表现效果并不好；**A3C 算法**（asynchronous advantage actor-critic）是对 DQN 的一种改进，采用 actor-critic 的一种模式；**DDPG**（Deep Deterministic Policy Gradient）也是对 DQN 的改进，采用 actor-critic 机制和 reply buffer（经验缓冲区）的形式，算法在处理连续动作时表现就要好的多；**TRPO**（trust region policy optimization）算法（信任域的策略优化）对前面的算法做了一些改进，主要包括对状态分布进行处理，利用重要性采样对动作分布进行的处理及在约束条件当中，用平均 KL 散度代替最大 KL 散度；**PPO**（Proximal Policy Optimization）也是最近比较热门的一种深度强化学习算法，分为 N 个 Actor，同时进行一些工作，这样平均分配给很多个 actor，合作来做的话效率会更高，而且会节省更多的时间。他在样本复杂性和易于调优之间取得平衡，试图在每一步最小化成本函数来计算更新，确保与先前策略偏差较小；**HER**（Hindsight Experience Replay）算法做第二次实验或者工作的时候吸取了前面的经验，然后再进行下面的训练或者工作的话，就会避免一些错误，如无人驾驶撞车了，上次为什么撞车了呢？第二次需要避免这个错误，即不让它撞车。Mobileye 的 **RSS 模型**的做法是不主动、不拒绝、不负责，不去主动撞你，但你要是撞我的话我没有任何办法。

数据集

Waymo 于 2019 年 8 月 21 日宣布，在公共道路上获得的自动驾驶汽车数据集将公开，免费供研究人员使用，数据集可以访问 <https://waymo.com/open>。包括 Aptiv、Uber 和 Cruise 等均在近期对外公开了自动驾驶相关数据。

目前的一些研究状况

现在解决自动驾驶技术问题有两种方法：一种是低精度定位+低精度地图+高准确识别率，另一种是高精度定位+高精度地图+更准确的识别率。一种是依赖于高精度的地图做控制策略，百度和 DeepMap 这家公司做的就是这方面的内容，需要采集所有全景的图像来做上传，最后再和百度地图结合，用激光点云制定比较好的高精度地图，之后就可以基于这个地图知道所处位置以及周边情况，进而做行程规划。这种方式成本会比较高，而且非常依赖于地图的精确性。还有一种就是采用深度强化学习算法根据高识别率的信息做决策，不太依赖于高精度地图。

利用深度强化学习做自动驾驶的几个团队有：WAYVE 团队（英国剑桥大学两位机器学习博士创立的英国自动驾驶公司）、本田研究院团队（本田研究院、宾夕法尼亚大学、乔治亚理工学院合作）、堪萨斯州立大学团队（KSU）、韩国汉阳大学团队（机器监测和控制实验室博士生）。

Wayve 采用从感知到控制的 end-to-end 深度强化学习算法来解决车道跟踪问题（深度确定性策略梯度：deep deterministic policy gradients, **DDPG**）（论文地址：<https://arxiv.org/abs/1807.00412>）（论文注解：<https://mp.weixin.qq.com/s/w3SsadgKaL8-tIzYLvMm-A>）模型输入是单目摄像头图像，系统经过了探索、优化和评估三个过程的迭代，输出是车辆的转角和速度命令。只用了不到 20 分钟，就能够教会一辆汽车从零开始沿着一条车道行驶。缺点是没有感知的预处理过程和控制单元的预警与防碰撞系统。

本田研究院的团队采用 TTC（time to collision 碰撞时间）模式（TTC 一般都是 2.7s），使车辆在远处能够提前知道岔路口的状态，预测到达这个岔路口的时间是不是有危险，有没有足够的时间进行制动，从而计划如何通过岔路口。团队存在的缺陷就是因为就是 DQN 存在的问题，即在离散动作当中表现优异，在连续性动作中表现不好，对于在高速行驶的情况下如何应用这个问题，解决得并不是太好。

如何在对抗性极强的情况下，对碰撞避免机制的行为进行训练，使系统进入不安全预警状态，**堪萨斯州立大学团队**提出了一种基于深度强化学习的新框架（融合 DDPG、TRPO、A3C），用于对自动驾驶汽车的碰撞避免机制的行为进行基准测试，但是有一个缺点：仅在控制单元做了一些研究，没有研究感知和决策单元，只能说单方面效果比较好。

韩国汉阳大学团队使用了传统的驾驶辅助系统和 DQN 结合，是在高速公路的模拟场景当中驾驶，而且采用的是两个 DQN 的模式作为输入，看一看是不是在车道变化情况下有一个很好的效果，超车的时候是不是也会有更好的效果，存在的问题其实也就是 DQN 的问题，离散性会更好一些，连续性并不是太好。（论文：Deep Q Learning Based High Level Driving Policy Determination）

Mobileye 的处理方式是前部感知模块可以进行独立处理，而后续的局部路径规划模块根据模型输出的决策结果，阻断不确定因素，进行有针对性的路径选择。模型的输入包括：车道线的位置，无障碍区域，本车 100 米范围内所有车的位置，速度，方向。（论文地址：<https://arxiv.org/abs/1610.03295> 好像比较老了）（**Mobileye** 新动作：<http://baijiahao.baidu.com/s?id=1605311713413449399&wfr=spider&for=pc>）

图森：货运汽车、卡车的自动驾驶

一个示例（礼让自动驾驶、大连交通大学）

感知模块

感知部分主要包括检测、识别和图像分割等方面。**检测**当中用的比较多的有 YOLO 算法，大的速度还是比较快的，如果车速特别快的话就需要快速的检测，然后再去做一些决策

方面的任务。**识别**方面使用 VCG 模型，模型结构简单而且，识别效果也是比较不错的。**分割**当中有局部分割、语义分割和全景分割，现在应用最多的是语义分割和全景分割。

感知模块借鉴 Andreas Geiger 的思想，将地图、三维传感器、二维传感器中的信息给到“**世界模型**” (world model)，把感知部分所有信息汇总到一个地图当中，做成一个 Map，相当于解除了我们对于高精度地图的高度依赖感，同时可以理解每个时刻的不同物体，相对于地面和道路这些位置，并且可以做之后的预测，相当于之后的路径规划问题。

决策模块

决策部分主要**借鉴 DDPG** 的思想， Actor-Critic 机制选择的一些 policy 通常都是最优的， reply buffer 能够存储之前驾驶的经验 and 其所得到的 Policy，之后的驾驶任务当中遇到类似的问题直接可以采用这种 policy，不需要再做其它的训练和学习。

此外还加入了一些**礼让的驾驶概念** (“安全行车、礼让三先”:先让，先慢，先停，我不去撞别人，别人撞我的时候要先避让一下，避免发生撞击的情况。)，就是我们在遇到问题的时候要首先想到先做避让，也就是**主动避让**的情况，连续动态的情况下可以让自动驾驶汽车避免发生碰撞。“礼让”这一词最早起源于机器人，但机器人的速度会比较慢，如果转移到车辆方面其实还是有些难度的，而且高速当中的礼让应该还是比较困难的问题，所以这也是日后工作的难点。

还有就是结合 PPO 与 HER 的思想，自动驾驶在高速运行的情况下也会需要一个快速决策的过程，所以选用 **PPO 算法的“分工合作”思想**使得速度能够提升。**HER 算法的“总结经验”思想**能够有效的避免重复犯一些错误，使得之后的驾驶过程当中少犯错误，尽量避免发生不必要的危险。

控制模块

控制方面主要还是 **RSS 模型上面做出一些改进**，因为不可能只是关注到前方的避让或者碰撞，**也要关注后方**，别人撞你的时候应该怎么办。再有就是采用的是双保险的机制，当决策失误或者突发情况发生，汽车自动制动。**激光雷达检测周围出现的车辆**，对车辆距离进行测量，当然如果传感器检测到有危险，，自动驾驶汽车会直接进入安全机制，或者是作出礼让的行为。

仿真平台

仿真平台 TORCS 属于 3D 赛车模拟游戏，是世界通用的赛车游戏，也是相对有说服力，效果会比较好一点，但是场景单一，不适合在复杂场景下做训练。框架的话是基于 TensorFlow 和 Keras。

困难

针对 end-to-end 的模型，它带来的**不确定性**是自动驾驶汽车无法预计的，更无法承担由不确定性导致事故带来的责任。

使用**模拟器训练和在现实场景训练是有一定差别的**。真实车辆在道路上的试错很难实现。而模拟器中的车辆动力学模型很难与真实车辆保持完全一致。

安全性问题。由于是黑箱模型，人类至今无法对它的输出做出有效的解释，在自动驾驶领域这套方法的使用受到了严格的政策限制。

自动驾驶技术框架

自动驾驶技术实现的基本原理是：**感知层**的各类硬件传感器捕捉车辆的位置信息以及外部环境（行人、车辆）信息。**决策层**的大脑（计算平台+算法）基于感知层输入的信息进行环境建模（预判行人、车辆的行为），形成对全局的理解并作出决策判断，发出车辆执行的信号指令（加速、超车、减速、刹车等）。最后**执行层**将决策层的信号转换为汽车的动作行

为（转向、刹车、加速）。

自动驾驶感知层传感器

自动驾驶汽车搭载的传感器类型一般分为两类——环境感知传感器和车辆运动传感器。

环境感知传感器的技术方案主要可以分为视觉主导和激光雷达主导。1) **视觉主导**的方案：摄像头（主导）+毫米波雷达+超声波雷达+低成本激光雷达，典型的车企是特斯拉。特斯拉最为激进，创始人马斯克坚持在其方案中不加入激光雷达；2) **激光雷达主导**的方案：低成本激光雷达（主导）+毫米波雷达+超声波传感器+摄像头，典型的代表是 Google Waymo。目前，谷歌 Waymo 自己组建团队研发激光雷达的硬件，把成本削减了 90% 以上，基本上是 7000 美金左右，同时也已经在美国凤凰城地区进行商业化的试运营。

高精度定位可以分为三类。第一类，基于信号的定位，代表就是 GNSS 定位，即全球导航卫星系统；第二类，航迹推算，依靠 IMU（惯性测量单元）等，根据上一时刻的位置和方位推断现在的位置和方位；第三类是环境特征匹配，基于激光雷达的定位，用观测到的特征和数据库中的特征和存储的特征进行匹配，得到现在车的位置和姿态。观察目前产业的主流方案，普遍采取融合的形式，大体上有：1) 基于 GPS 和惯性传感器的传感器融合；2) 基于激光雷达点云与高精地图的匹配；3) 基于计算机视觉技术的道路特征识别，GPS 卫星定位为辅助的形式。

计算平台（主控芯片）

高等级自动驾驶的本质是 AI 计算问题，车载计算平台的计算力需求至少在 20T 以上。从最终实现的功能来看，计算平台在自动驾驶中主要负责解决两个主要问题：1) 处理输入的信号（雷达、激光雷达、摄像头等）；2) 做出决策判断、给出控制信号：该加速还是刹车？该左转还是右转？

目前运用于**自动驾驶的芯片架构**主要有 4 种：CPU、GPU、FPGA（现场可编程门阵列）和 ASIC（专用集成电路），未来芯片有望迎来全新的设计模式——**应用场景决定算法，算法定义芯片**。如果说过去是算法根据芯片进行优化设计的时代（通用 CPU+算法），现在则是**算法和芯片协同设计的时代（专用芯片 ASIC+算法）**，这一定程度上称得上是「AI 时代的新摩尔定律」。具体而言，自动驾驶核心计算平台的研发路径将是根据应用场景需求，设计算法模型，在大数据情况下做充分验证，待模型成熟以后，再开发一个芯片架构去实现，该芯片并不是通用的处理器，而是针对应用场景，跟算法协同设计的人工智能算法芯片。

自动驾驶算法

1) **感知层算法核心任务**——是将传感器的输入数据最终转换成计算机能够理解的自动驾驶车辆所处场景的语义表达、物体的结构化表达，具体可以包括：物体检测、识别和跟踪、3D 环境建模、物体的运动估计；

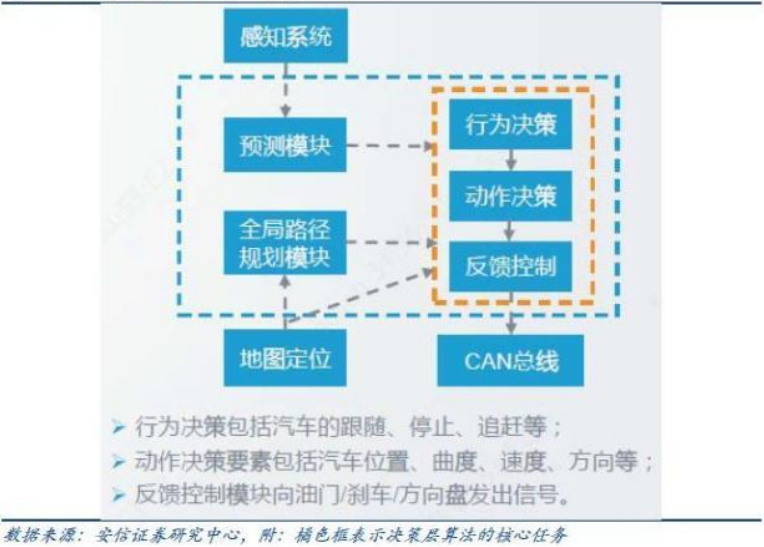
图 29：感知层算法的核心任务



数据来源：亿欧，安信证券研究中心

2) **决策层算法的核心任务**——是基于感知层算法的输出结果，给出最终的行为/动作指令，包括行为决策（汽车的跟随、停止和追赶）、动作决策（汽车的转向、速度等）、反馈控制（向油门、刹车等车辆核心控制部件发出指令）。

图 30：决策层算法的核心任务



整体来看，不同等级的自动驾驶算法的焦点不同。L3 级别的自动驾驶，侧重于替代人的环境感知能力，因此感知层算法将是核心。L4 级别的自动驾驶，除了环境感知能力之外，侧重点更在于复杂场景的决策算法的突破。

算法的验证及迭代需要**路测+仿真**。按照产业普遍观点，车企需要 100 亿英里的试驾数据来优化其自动驾驶系统，仿真测试成为大多数公司的共同选择。从产业来看，为了更高效的迭代和验证自动驾驶算法，仿真系统已经逐渐成为标配，Waymo、百度、腾讯将仿真系统研发作为头等大事；AutoX、Roadstar.ai、Pony.ai 等诸多自动驾驶初创公司也在自主研发仿真环境；业内开始出现 CARLA、AirSim 等开源式自动驾驶仿真平台。

高精度地图

自动驾驶 OS

HMI（人机交互）