



智能网联汽车 自动驾驶 系统测试场景数据采集和 分析标准需求研究报告

全国汽车标准化技术委员会
智能网联汽车分技术委员会
自动驾驶标准工作组

2023 年 11 月

目 录

前言..... I

第一章 自动驾驶场景库建设的必要性分析..... 1

1.1 自动驾驶发展现状..... 1

1.1.1 自动驾驶政策法规..... 1

1.1.2 自动驾驶关键技术..... 3

1.1.3 产品发展现状..... 5

1.1.4 现状总结..... 6

1.2 驾驶场景库建设的重要意义..... 7

1.2.1 对自动驾驶技术及产品开发的意义..... 7

1.2.2 对自动驾驶系统测试的重要意义..... 8

第二章 自动驾驶测试场景数据采集和分析技术发展现状..... 9

2.1 测试场景库发展现状..... 9

2.1.1 国外发展现状..... 9

2.1.2 国内发展现状..... 16

2.1.3 目前存在问题..... 22

2.2 测试场景数据采集技术发展现状..... 23

2.2.1 国外发展现状..... 23

2.2.2 国内发展现状..... 26

2.2.3 目前存在问题..... 28

2.3	测试场景数据分析技术发展现状.....	29
2.3.1	国外发展现状.....	29
2.3.2	国内发展现状.....	31
2.3.3	目前存在问题.....	33
第三章	自动驾驶系统测试场景政策及标准现状.....	34
3.1	自动驾驶系统测试政策及标准现状.....	34
3.1.1	我国自动驾驶系统测试地方政策.....	34
3.1.2	国内外自动驾驶系统测试标准.....	36
3.1.3	我国道路交通及基础设施标准.....	43
3.1.4	政策和标准总结.....	46
3.2	测试场景标准发展现状.....	46
3.2.1	国际自动驾驶系统测试场景标准发展现状.....	46
3.2.2	国内自动驾驶系统测试场景标准发展现状.....	48
3.2.3	采集及分析标准总结.....	53
3.2.4	术语定义.....	53
3.3	测试场景采集及分析标准化的意义.....	54
第四章	自动驾驶数据采集技术.....	56
4.1	概述.....	56
4.2	采集流程.....	57
4.3	采集需求研究.....	58
4.4	数据质量要求.....	59

4.4.1	空间配准质量要求.....	59
4.4.2	数据采集质量要求.....	59
4.5	采集区域选择.....	63
4.6	采集方式选取.....	65
4.6.1	实车采集方式.....	65
4.6.2	路侧设施采集方式.....	68
4.6.3	无人飞行器采集方式.....	70
4.6.4	采集方式差异化分析.....	71
4.7	采集设备部署及标定.....	72
4.7.1	采集设备部署.....	72
4.7.2	时间同步.....	78
4.7.3	空间同步.....	79
4.8	数据检查要求.....	80
4.8.1	通用要求.....	80
4.8.2	常规数据检查.....	81
4.8.3	数据同步性检查.....	83
4.9	数据上传要求.....	84
4.9.1	车端传输要求.....	84
4.9.2	路侧设备传输要求.....	88
4.10	数据存储要求.....	89
4.10.1	通用要求.....	89

4.10.2	采集端存储要求.....	90
4.10.3	云端存储要求.....	91
4.11	总结.....	94
第五章	自动驾驶系统测试场景数据分析技术标准化研究.....	95
5.1	场景数据分析流程.....	95
5.2	场景数据清洗技术.....	95
5.2.1	场景数据清洗流程.....	95
5.2.2	场景数据清洗方法.....	96
5.3	场景标注技术.....	98
5.3.1	场景标注流程.....	98
5.3.2	标注标签定义.....	99
5.3.3	自动标注方法.....	101
5.3.4	人工校验及补充标注要求.....	105
5.4	场景统计技术.....	106
5.4.1	场景统计流程.....	106
5.4.2	确定统计内容.....	107
5.4.3	场景数据提取.....	107
5.4.4	场景指标计算.....	109
5.4.5	统计结果描述.....	109
5.5	测试用例提取技术.....	110
5.5.1	测试用例提取要求.....	110

5.5.2	测试用例提取流程.....	112
第六章	自动驾驶系统测试场景数据采集和分析标准化建议.....	114
6.1	自动驾驶系统测试场景数据采集和分析标准化框架建议.....	114
6.2	自动驾驶系统测试场景数据采集和分析标准化需求研究总结.....	115

汽标委智能网联汽车分委会

前言

智能网联是汽车改革和竞争的“下半场”，智能网联汽车已经成为全球汽车发展的重要方向，乘用车领域正在经历从高级辅助驾驶到高级自动驾驶的阶段。今年我国四部门联合发布了《关于开展智能网联汽车准入和上路通行试点工作的通知》，其中通过场景的测试验证也是自动驾驶汽车国内外法规审批的重要安全要求，与自动驾驶相关的测试场景和安全验证等标准的研制需求迫切。

自动驾驶是一个数据密集型的领域，基于场景数据驱动的开发测试成为行业的广泛共识。自然驾驶场景数据是重要的数据来源，通过大量的数据采集、处理和分析，让自动驾驶系统能够自主学习和适应各种复杂的驾驶场景，从而实现数据驱动的自动驾驶技术成熟应用。因此，标准化的场景数据采集和分析，对于自动驾驶技术和产品的开发，以及自动驾驶系统测试具有重要的支撑作用。

国内外场景采集和分析领域的技术标准和规范尚未建立，各机构的场景数据内容、格式、技术要求、质量要求、采集流程和处理方法不统一。因此，集合行业主流技术服务商、车企和高校科研机构，通过对智能网联汽车自动驾驶系统测试场景数据采集和分析标准需求研究，梳理本领域的标准框架和思路，明确了标准化的重点方向及相关内容，可为相关标准制定的顶层设计提供参考，有助于统一关于测试场景数据采集和分析的意见，联合行业共同力量，推动相关标准工作的开展，建立汽车行业测试场景标准化、规范化、共享化的良好生态。

衷心感谢参与研究报告编写的各个单位和组织：中国第一汽车集团有限公司、北汽（北京）智能网联汽车研究院有限公司、中国汽车研究中心有限公司、宁波吉利汽车研究开发有限公司、上海临港绝影智能科技有限公司、中国汽车工程研究院股份有限公司、招商局检测车辆技术研究院有限公司、吉林大学、北京赛目科技有限公司、启明信息技术股份有限公司、北京百度智行科技有限公司、北京四维图新科技股份有限公司、深圳元戎启行科技有限公司、重庆长安汽车股份有限公司、极

蔚来汽车（宁波杭州湾新区）有限公司、信通院车联网创新中心（成都）有限公司、一汽解放、北京汽车研究总院有限公司、北汽福田汽车股份有限公司、高通无线通信技术（中国）技术有限公司、清华大学苏州汽车研究院、蔚来汽车科技（安徽）有限公司、东软睿驰汽车技术（沈阳）有限公司、上汽集团创新研究开发总院。

主要编写人：郑建明、王瑶、华一丁、张建军、张晋崇、孙良、刘瑞强、宋方方、房科、张庆余、丁雪聪、王子煜、张奥博、徐月云、覃斌、窦红玉、蒋慧、郑文博、纪东方、唐宇、曾杰、张玉新、白智敏、王树达、孙磊、程周、唐烈萍、李娜、涂文天、刘鑫、陈星筑、王猛、卢帅、李玉彬、付瀚龙、唐诚成、白宁、于腾飞、田俊涛、洪波盛、程汉、贾元辉、魏岩、徐优志、吴羽熙、贾浩、查仁方。

第一章 自动驾驶场景库建设的必要性分析

1.1 自动驾驶发展现状

1.1.1 自动驾驶政策法规

近年来，自动驾驶技术成为了全球范围内热议的话题之一。众所周知，自动驾驶技术的普及将为人们带来更加便捷和安全的出行方式，不仅能够减少交通事故的发生率，降低人为驾驶带来的交通堵塞，还可提高交通效率及减少能耗排放。自动驾驶已成为全球汽车产业发展的重要方向，当前世界各主要国家均对自动驾驶发展给出了明确的时间及技术等级规划，如图 1.1 所示。旨在推动自动驾驶汽车安全可靠地量产落地，各国政策均强烈关注基于场景库及数据驱动的安全验证及检测认证等关键环节。



图 1.1 全球主要国家自动驾驶技术路线规划

联合国 2019 年 6 月通过《自动驾驶汽车框架文件》，旨在确立 L3 及更高级别

自动驾驶的安全性和安全防护的关键原则，文件中对“系统安全”、“失效保护响应”、“人机交互界面/操作者信息”、“目标事件探测与响应”及“设计运行域”等多项内容进行了定义，确定了未来自动驾驶发展的基础^[1]。2021 年发布了首个 L3 级自动驾驶上路法规 R157，即《自动车道保持系统 ALKS》^[2]，R157 中的最小测试集涵盖了车道保持、拥堵跟车避障、他车切入、进入或退出高速、高速 Y 形道路及隧道中自车被部分堵塞等主要用户使用的关键场景。

欧盟 2019 年发布了《欧盟自动驾驶车辆许可豁免流程指南》，重点为 L3、L4 级自动驾驶汽车的量产准入进行了规定^[3]，豁免指南中明确规定整车企业应向认证机构申明自动驾驶汽车运行域，即车辆适宜运行的时间和空间条件。设计运行域应至少包含道路条件、地理区域、环境条件、通行速度范围，以及其他与自动驾驶运行安全密切相关的条件。

德国 2021 年修订的《自动驾驶法》允许 L4 级自动驾驶汽车在德国公共道路指定区域常态化运营，并对其技术要求、准入条件、数据处理规则等作了规定^[4]。即具有自动驾驶功能的机动车应在指定的运行区域独立执行驾驶任务，且“无需驾驶员驾驶车辆”。

英国 2022 年发布的《互联和自动出行 2025：在英国释放自动驾驶汽车的效益》从自动驾驶汽车安全、产业和市场发展、社会福祉三个方面提出了详细的计划^[5]。基于自动驾驶汽车比肩“胜任且谨慎的人类司机”的安全愿景，提出了包含五个主要维度的自动驾驶监管框架，包括审批、授权、操作者许可、使用监管和事故调查，构建了自动驾驶汽车的监管闭环。授权框架中明确指出车辆必须能识别自动驾驶功能的运行是否超出其运行设计域。

美国 2020 年发布的《自动驾驶汽车总体规划》提出了“安全第一”的自动驾驶发展愿景^[6]，美国交通运输部重点在法规修订、消除有关汽车功能、创新设计和运营模式、使用工具和框架开发及自动驾驶技术安全性能评估等方面贯彻安全第一的理念。其中，在有条件自动驾驶中，该法规要求自动驾驶辅助系统应在一组特定场景下执行完整的动态驾驶任务，且驾驶员可应系统要求随时接管。

日本自 2020 年《道路交通法》修订版给出了 L3 级自动驾驶车辆的定义和准入依据后，在 2022 年提交的最新法案又加入了特定条件下 L4 级自动驾驶车辆和自动行驶机器人的上路许可^[7]。2022 年 7 月发布了《特定自动运行许可制度》，正式开启了 L4 级机器人（包含 L4 级自动驾驶汽车、L4 级可自主行驶的机器人等）的商用服务。审批内容主要包括机器人是否符合特定自动运行标准、特定自动运行路线是否满足行驶环境条件、机器人是否会明显影响交通状况等。

韩国 2020 年发布的《自动驾驶汽车安全标准》^[8]。主要针对自动泊车、变换车道等自动驾驶技术提出安全标准，并逐步推出先导性的安全规范及标准。标准中应至少涵盖自动驾驶汽车面对高速公路出口、前方道路施工等应急情况下的主要测试场景及人类接管情况。

2020 年，我国发改委联合科技部、工信部等 11 部委联合发布《智能汽车创新发展战略》，指出急需建立健全智能汽车测试评价体系及测试基础数据库，以满足我国复杂道路环境和驾驶行为的测试需要^[9]。2022 年工信部发布了《关于开展智能网联汽车准和上路通行试点工作的通知(征求意见稿)》，拟遴选符合条件的道路机动车辆生产企业和具备量产条件的搭载自动驾驶功能的智能网联汽车产品，开展准入试点，目的为在保障安全的前提下，促进智能网联汽车产品的功能、性能提升和产业生态的迭代优化^[10]。其中针对汽车产品的测试要求中明确提出，为有效地验证自动驾驶系统安全，仿真测试场景至少包括充分、合理的标称场景、危险场景和边缘场景，车辆应能通过封闭场地典型场景及实际道路连续场景测试的安全性验证。

综上所述，基于场景的自动驾驶汽车安全验证已成为各国政策法规的关键组成部分，然而高复杂、高危险的测试场景缺失已成为自动驾驶市场准入及示范推广的重大阻碍，目前急需一套完善的自动驾驶场景库，以此高速推动自动驾驶汽车落地及商业化使用。

1.1.2 自动驾驶关键技术

自动驾驶汽车研发正从传统模式向数字化方向快速演进，自动驾驶汽车是一个

数据密集型的领域，数字化研发是支撑自动驾驶汽车不断进化的核心，随着自动驾驶汽车的快速发展和大规模落地需求，基于场景数据驱动的开发测试成为行业的广泛共识，如图 1.2 所示。场景数据驱动主要是通过大量的数据采集、处理和分析，让自动驾驶系统能够自主学习和适应各种复杂的驾驶场景，从而实现数据驱动的自动驾驶技术成熟应用。

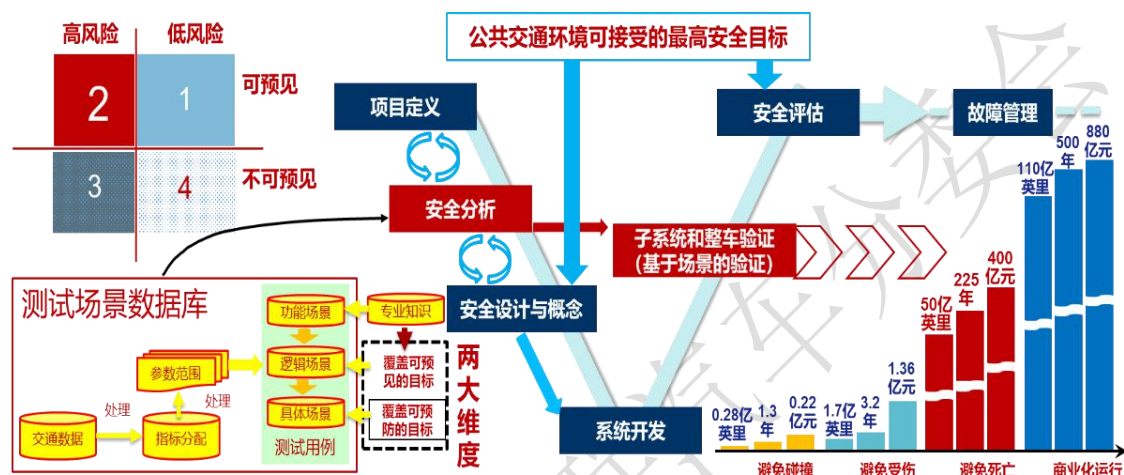


图 1.2 自动驾驶产品开发测试流程

在高等级自动驾驶技术中高精地图和融合感知是感知阶段重要的技术手段，高精地图具备较高的定位精度，但因其投入成本高、更新周期长、“鲜度”低，已成为现阶段量产应用的短板。相应地，激光雷达等传感器的搭载量在新发布的车型中占比逐渐增高，如埃安、领克、哪吒及“蔚小理”等。基于场景采集的大数据集及其数据结果可以高效训练车端融合感知及决策规划系统，进一步提升自动驾驶安全性，因此场景大数据采集种类对自动驾驶产品能力提升极其重要。交通车、行人、车道线及信号灯等的感知识别、基于场景大数据的控制策略优化及自动驾驶系统的接管模型等开发和测试均离不开场景大数据的支撑。

综上所述，数字化时代的自动驾驶技术开发强烈依赖场景数据驱动，不论是感知系统的算法训练，还是面对复杂交通环境时的决策和规划，均需要海量的场景数据“投喂”自动驾驶模型。尤其是目标物感知精度和识别准度，需要车端、路端、云端的数据采集、融合及训练，因此基于场景的开发测试已成为自动驾驶研发的关键，《2023 年中国汽车技术趋势报告》将自动驾驶汽车场景数据库评选为“未来

3-5 年中国汽车十大前沿技术”。

1.1.3 产品发展现状

自动驾驶技术当前正处于高速发展阶段，L3 级自动驾驶产品逐渐进入量产，L4 级开发及验证如火如荼。2022 年我国 L2 级智能网联乘用车渗透率已达 34.9%。预计到 2025 年我国 L2 级乘用车渗透率有望达到 50%，如图 1.3 所示。预计到 2040 年，自动驾驶车辆预计注册的总数量欧洲将达 550 万辆，美国/北美自由贸易区将达 740 万辆，中国在政府政策、法规及标准的强力推动下将达 1450 万辆。



图 1.3 中国智能汽车产品渗透率

国外有代表性的智能网联汽车企业主要有特斯拉、本田及奔驰等。特斯拉通过采用图像数据进行机器学习，实现了自动辅助导航驾驶、自动辅助变道、自动泊车、智能召唤、自动识别信号灯及标识，同时还具有城市街道辅助驾驶功能，但目前该功能仅在美国部分地区搭载。同期本田汽车获得了日本 L3 自动驾驶认证，德国联邦汽车运输管理局根据 UN-R157 法规批准了奔驰的 L3 级自动驾驶汽车，并于 2023 年 1 月并获得了美国监管机构的批准，可在内华达州运营 L3 级自动驾驶^[1]。

当前我国量产的智能网联汽车均以高等级辅助驾驶功能为主，如红旗、东风及长安等。小鹏、极狐等 NOA 功能实现城市场景覆盖；广汽计划 2024 年推出 L4 级

战略车型：“智能吉利 2025 战略”中提出 2025 年实现 L4 级商业化，多数整车企业未来几年产品均定位为有条件自动驾驶系统，一般是首先实现高速公路场景领航驾驶，逐步实现城市道路场景领航驾驶，也有部分企业直接以高度自动驾驶和完全自动驾驶为目标对产品进行规划设计，如图 1.4 所示。

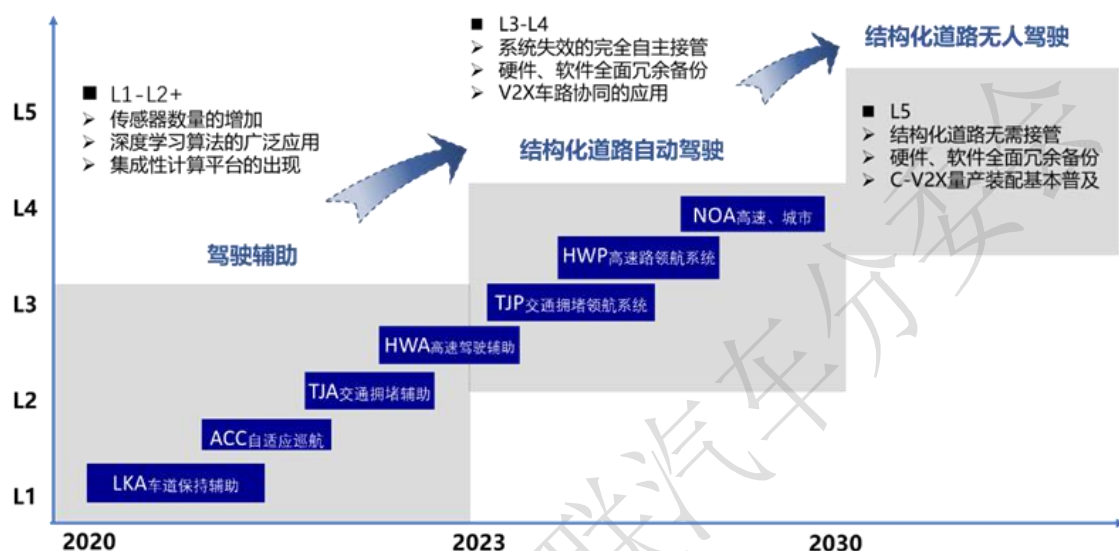


图 1.4 自动驾驶汽车产品现状及规划

1.1.4 现状总结

政策法规重点关注自动驾驶可能会面临的安全问题，尤其是基于测试场的设计运行域识别能力、动态驾驶任务和驾驶员接管；自动驾驶汽车正在经历从高级驾驶辅助向高等级自动驾驶的跨越，自动驾驶技术研究以及产品开发测试，均对场景数据提出了更高的要求；如图 1.5 所示，场景库是自动驾驶开发测试的基础数据资源，场景库关键技术（包含测试场景数据采集和分析技术）是自动驾驶“开发-测试”闭环迭代的重要支撑。

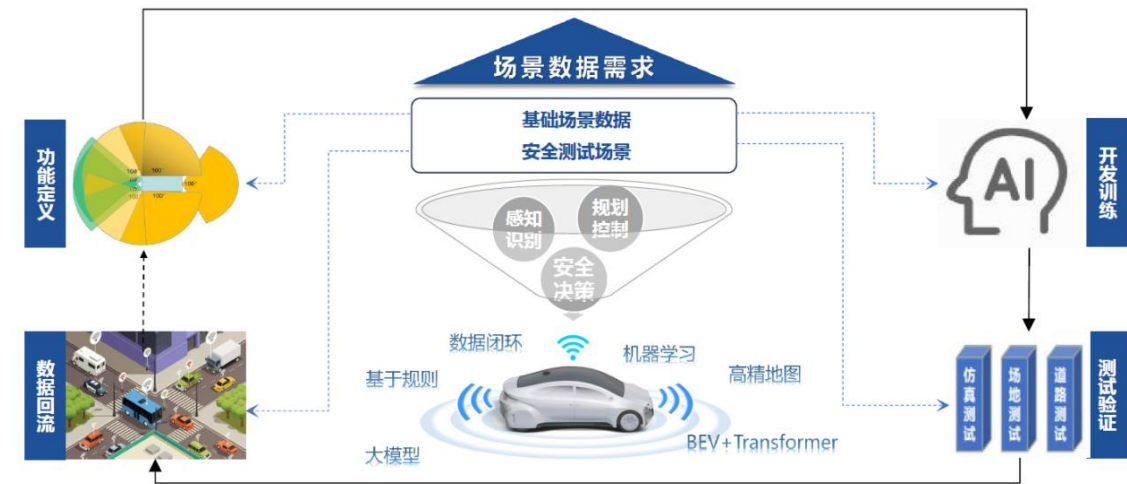


图 1.5 自动驾驶“开发-测试”闭环迭代过程

1.2 驾驶场景库建设的重要意义

1.2.1 对自动驾驶技术及产品开发的意义

如图 1.6 所示，场景数据贯穿自动驾驶产品开发全流程，对产品安全性及适用性提升具有重要意义：

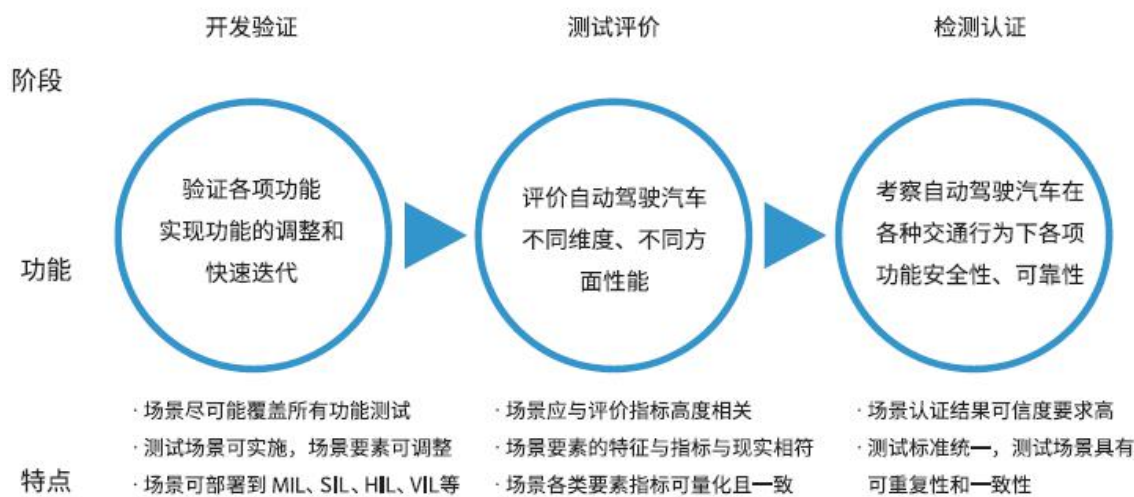


图 1.6 自动驾驶汽车不同阶段对场景的要求

- a) 有利于在开发早期阶段打通开发侧和用户侧的壁垒，将用户体验全程纳入开发过程，设定基于数据的工程目标，打造极致体验的高魅力产品；

- b) 有利于开发早期阶段**发现产品功能以及性能上的问题**，尤其是感知系统性能提升更为依赖场景数据的训练；
- c) 有利于为深度学习的感知、强化学习的规划算法奠定坚实的训练集基础，从而促进对算法的开发和迭代，**缩短自动驾驶产品开发周期**；
- d) 通过不断采集自然驾驶场景数据，可以持续提高自动驾驶系统的**安全性与适用性**，使其能够在不同的道路、环境和天气条件下**安全可靠**地运行。

1.2.2 对自动驾驶系统测试的重要意义

自动驾驶系统需经过的大量的场景测试，无论前期系统研发还是后期产品检测认证，均对场景数据有极大的需求。有针对性地生成兼容性高、通用性强、覆盖度广的场景数据，在自动驾驶产品的测试流程规范化以及安全性、稳定性、适用性测试验证具有重要意义。

- a) 规范自动驾驶系统**测试流程**，通过场景数据采集和分析可以生成多维度的测试场景数据，仿真、场地、道路测试法可以相互协同参考，确保从算法开发到系统集成，各阶段的自动驾驶系统功能和安全性满足预期设计要求。
- b) 可高效测试自动驾驶系统在**多类型、长里程测试场景下的稳定性和适用性**。通过不同类型的场景泛化和随机组合测试，提升自动驾驶系统的稳定性和适用性。
- c) 自动驾驶系统**安全验证**。通过场景采集并生成高危险度、复杂度的测试场景，验证自动驾驶汽车的安全性和可靠性，证明其具备在真实交通环境下应对复杂多变场景、规避风险的能力。

第二章 自动驾驶测试场景数据采集和分析技术发展现状

2.1 测试场景库发展现状

2.1.1 国外发展现状

国外场景库建设主要依托国家机构、科研机构及高校等开展，研究时间较早，如德国的 PEGASUS 项目、美国 Waymo Open Dataset 等，研究一般基于实车采集逐步开发仿真测试所需的场景数据集，有的甚至仅有图片集^[12]，国外场景库主要针对架构模型、数据集属性及场景要素等进行介绍，国外主要测试场景数据集相关项目的发展历程见图 2.1。

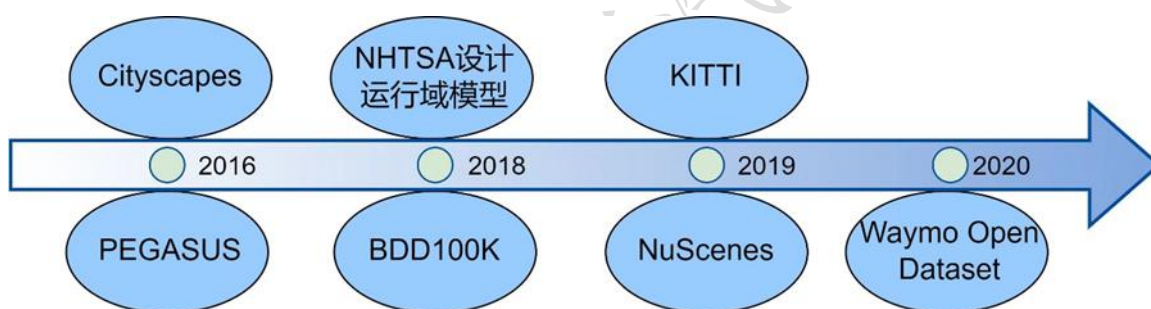


图 2.1 国外主要场景库及数据集发展历程

a) 测试场景架构模型

国外场景库项目对测试场景架构均有一定的研究，但研究及应用比较有代表性的是德国的 PEGASUS 项目及美国的 NHTSA “自动驾驶系统测试用例和场景框架”研究项目，具体测试场景架构建设情况如下：

1) 德国 PEGASUS 项目测试场景架构模型简介

2016 年，由德国联邦经济事务和能源部发起，以制定自动驾驶车辆系列测试标准为目的的 PEGASUS 项目正式立项，历时 42 个月完成。项目主要聚焦于 SAE L3 级别的高速公路场景的研究和分析，并将场景按照抽象级别划分为三个层次，即功能场景、逻辑场景和具体场景。项目从测试场景解构与重构的角度首次提出了 6 层

场景架构模型，如图 2.2 所示，即第 1 层（L1）-道路层、第 2 层（L2）-交通基础设施层、第 3 层（L3）-L1 与 L2 的临时操作层、第 4（L4）-对象层、第 5 层（L5）-环境层、第 6 层（L6）-数据信息层。6 层场景架构模型实现了实车运行场景要素的清晰分层，初步提出了三种测试场景格式标准，即 OpenDRIVE、OpenSCENARIO 及 OpenCRG。为场景库相关技术及标准建设奠定了基础，在国内外场景库建设中得到了重大推广使用^[13]。



图 2.2 PEGASUS 6 层模型场景

2) NHTSA “自动驾驶系统测试用例和场景框架”研究项目

2018 年，美国高速公路交通安全管理局（National Highway Traffic Safety Administration, NHTSA）在 PEGASUS 6 层架构的基础上首次提出了与本车相关的场景要素特征，主要包括车辆运动行为、设计运行域、事件感知与决策行为、失效模式行为 4 部分，同时将车辆动态运行任务重新划分为 6 层，即物理设施、操作限制、目标对象、连接性、环境条件和区域，如图 2.3 所示。旨在开发一套标准化的测试用例和场景框架，以评估自动驾驶系统的安全性和性能^[13]。

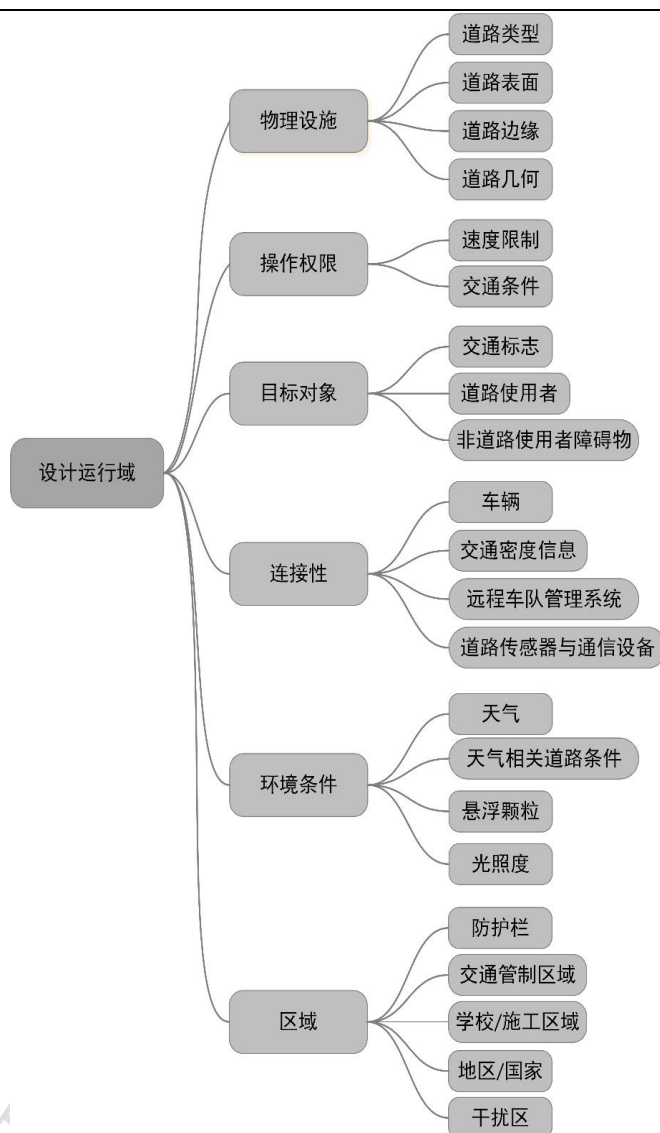


图 2.3 NHTSA 6 层设计运行域模型

PEGASUS 项目基于高速公路采集初步建立了三种测试场景格式、提出了 6 层场景架构模型；NHTSA 研究项目在 PEGASUS 6 层场景架构的基础上，重点完善了与本车动态驾驶任务相关的重点场景要素，明确了测试场景中交通参与者的行为，为测试场景构建中动态交通参与者行为设计提供了依据，但二者关键数据处理分析方法及最终测试场景等均没有公开。

b) 测试场景库及数据集

1) KITTI 场景数据集

2019 年，德国卡尔斯鲁厄理工学院和丰田美国技术研究院共同发布了 KITTI

数据集，该数据集用于训练和测试面向自动驾驶场景下的计算机视觉算法，主要包括评价目标检测、目标跟踪及路面分割等计算机视觉技术在车载环境下的性能^[14]。KITTI 包含市区、乡村和高速公路等场景采集的真实图像数据，每张图像中最多达 15 辆车和 30 个行人，还有各种程度的遮挡与截断。



图 2.4 KITTI 数据集

2) Waymo Open Dataset 场景数据集

2019 年，Waymo 发布了 Waymo Open Dataset 开放数据集，旨在促进自动驾驶技术的研究和发展。该数据集包含了大量的传感器数据，如激光雷达、摄像头、雷达高精度地图及车辆行驶轨迹等信息。数据集主要用于训练和验证面向自动驾驶场景的计算机视觉算法、决策规划算法^[15]。目前，Waymo 公开数据集已成为全球最大、最多样化的自动驾驶数据集之一。

3) nuScenes 场景数据集

2019 年，Motional 团队发布了包含 1000 多个场景的完整 nuScenes 数据集，旨在支持计算机视觉和自动驾驶的公共研究。nuScenes 是第一个提供自动驾驶汽车整个传感器套件（6 个摄像头、1 个激光雷达、5 个雷达、GPS、IMU）数据的大型数据集。与 KITTI 相比，nuScenes 数据集包含了 7 倍多的对象注释^[16]。

4) Cityscapes 场景数据集

2016 年,梅赛德斯-奔驰集团发布了 Cityscapes 数据集,如图 2.5 所示,即城市景观数据集,旨在提供高质量的城市街景图像和注释数据,提供无人驾驶环境下的图像分割数据集,用于评估视觉算法在城区场景语义理解方面的性能。该数据集是一个大规模数据集,包含来自 50 个不同城市的街道场景中记录的多种立体视频序列。Cityscapes 数据集共有 fine 和 coarse 两套评测标准,前者提供 5000 张精细标注的图像,后者提供 5000 张精细标注外加 20000 张粗糙标注的图像^[17]。

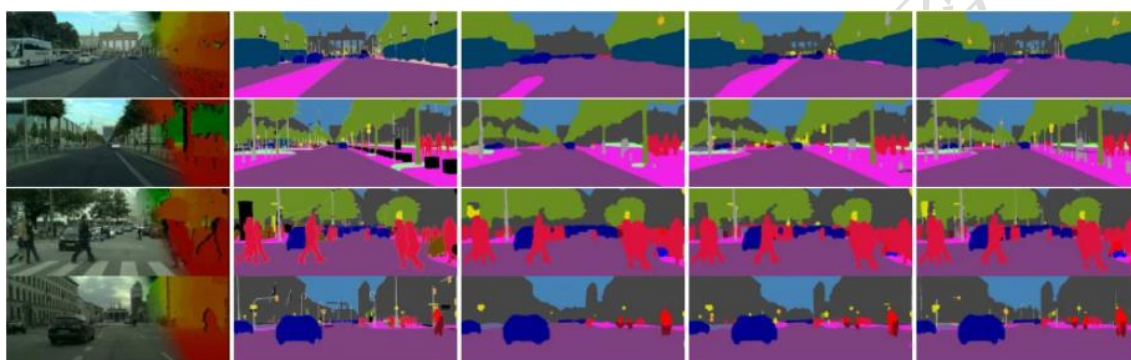


图 2.5 Cityscapes 数据集

5) high D 场景数据集

high D 数据集是使用无人机采集自德国高速公路上的自然车辆轨迹,如图 2.6 所示,空中视角克服了已建立的交通数据收集方法的典型局限性,如遮挡。数据集涵盖德国科隆附近的六个不同地点,包括超过 110500 辆汽车,每个车辆的轨迹,包括车辆类型、尺寸和操纵均可自动提取。该数据集使用最先进的计算机视觉算法,定位误差小于 10cm。尽管该数据集是为高度自动化车辆的安全验证而创建的,但它也适用于许多其他任务,如交通模式分析、驾驶员模型参数化分析等^[18]。

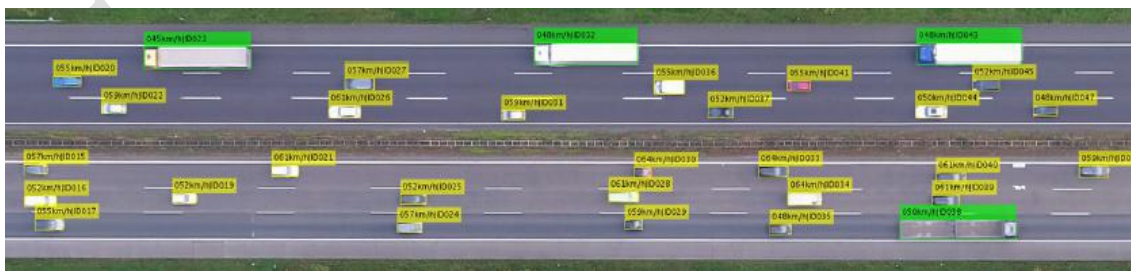


图 2.6 high D 数据集

为了更加清晰描述上述各大场景库及数据集的建设情况，本文从发布机构、采集方式、采集区域、道路类型、天气类型及采集时间等几个维度开展对比分析，具体见表 2.1。

表 2.1 各场景库及数据集数据分布对比

场景库及数据集	发布时间	发布机构	采集方式	采集区域	道路类型	天气类型	时间段
KITTI 数据集	2012	德国卡尔斯鲁厄技术研究院和丰田芝加哥技术研究院	车端采集	德国卡尔斯鲁厄市，以及周边城市和乡村地区	城市、高速、乡村、园区、住宅	—	白天
Cityscapes 数据集	2016	梅赛德斯奔驰集团股份公司	车端采集	德国和瑞士 50 个城市	城市街道、广场和建筑物等	良好、中等天气，不包含雨、雪等极端天气场景	白天
high D 数据集	2018	德国亚琛工业大学汽车工程研究所	无人机采集	德国科隆附近的六个不同地点	采集包含 2-3 条高速公路，视野中路段长为 400-420m	阴天、雨天	白天、夜晚
Waymo Open Dataset	2019	Waymo（谷歌旗下自动驾驶技术公司）	车端采集	亚利桑那州凤凰城、旧金山、西雅图、加州山景城、底特律和洛杉矶	城市街道、高速公路、乡村道路	晴天、雨天	白天、夜晚、黎明、黄昏
nuScenes 数据集	2019	Motional（原 nuTonomy）团队	车端采集	波士顿（海港和南波士顿）和新加坡（One North、荷兰村和皇后镇）	城市、高速、乡村	晴天、雨天	白天、夜晚、黄昏

表 2.2 为当前阶段各场景库及数据集的**数据规模对比**，重点介绍了各数据集的测试数据量，对自动驾驶算法测试具有一定的参考意义。

表 2.2 各场景库和数据集数据规模对比

场景库及数据集	数据规模
KITTI 数据集	a. 包含彩色图像数据（12GB）、点云数据（29GB）、相机矫正数据（16MB）和标签数据（5MB），均包含训练集和测试集两部分； b. 标签数据训练样本 7481 个，测试样本 7518 个；但由于测试样本不可见，一般将 7481 个训练样本划分为 3712 与 3769 分别作为训练集和测试集。
Cityscapes 数据集	a. 5000 张精细标注图像，其中 2975 张用于训练，500 张用于验证，1525 张用于测试；19 个分类类别； b. 20000 张非精细标注图像（标注位于 gtCoarse 文件夹），源图大小为 1024x2048 (hwxw)。
high D 数据集	a. 图像数据集：包含 90,000 张图像，分为训练集和测试集两部分，每张图像的分辨率为 1280x720； b. 视频数据集：包含了 90 个视频，每个视频的长度为 1 分钟，分辨率为 1280x720； c. 标注数据集：包含了路面、车道线、交通标志等信息的标注数据，可以用于自动驾驶无人机的目标检测和场景理解研究。
Waymo Open Dataset	a. Perception Dataset 有 2030 个场景，训练集解压缩后大小为 812.7GB，验证集为 204.9GB； b. Motion Dataset 有 103354 个场景，训练集占 70%，包括未来轨迹真值，验证集占 15%，测试集占 15%。
nuScenes 数据集	a. 包含 1000 个场景，每个场景 20s（850 个用于模型训练，150 个用于模型测试）； b. 大约 140 万张相机图像 c. 39 万个激光雷达扫描点云图像 d. 140 万个毫米波雷达扫描点云图像 e. 40 万个关键帧； f. 为 23 个对象类标注的 1400 万个 3D 标注框。

表 2.3 为各场景库/数据集在采集数据的**原始数据、标注信息**等方面的对比，通过对这些方面的比较分析，可以帮助选择适合特定需求的数据格式，并了解不同数据格式的优缺点。

表 2.3 各场景库和数据集数据格式对比

场景库及数据集	图像数据	激光雷达点云数据	毫米波雷达点云数据	GPS/IMU数据	轨迹信息	标注信息
KITTI 数据集	png	bin		txt	xml	txt
Cityscapes 数据集	png					json
high D 数据集					csv	csv
Waymo Open Dataset	TFRecord					
nuScenes 数据集	jpg	bin	pcd		json	json

综上所述，国外测试场景库/数据集的发展已经取得了显著的发展，为自动驾驶算法的训练和测试提供了重要的数据支撑，但中国的道路交通环境与国外存在较大差异，国外场景库/数据集不适用于中国自动驾驶场景的原因有以下几个方面：

- a) 交通标志、标线、交通信号灯等道路指示信息不同。中国的道路交通标识和标线与国外不同，例如中国的限速标志是白底黑图形、红圈，美国的限速标志是白底黑字、黑框。因此，使用国外数据集训练的自动驾驶系统可能会出现误判。
- b) 道路结构和地形不同：中国的道路结构和地形与国外也不同，例如中国的高速公路上通常会有较多的立交桥、隧道以及路堤，这些地形和结构直接影响自动驾驶系统的性能。
- c) 交通流量不同：中国的交通流量与国外也存在差异，例如中国的交通流量较大，道路上的车辆会更加拥挤，这也会影响自动驾驶系统的性能。
- d) 场景覆盖度不足：目前国外大多数场景库及数据集的采集数据仍主要来自于城市、高速场景以及良好的天气状况，对于一些乡间道路、停车场以及复杂天气状况的场景覆盖度严重不足。

因此，为了能使自动驾驶系统更加适应中国的交通环境，获得更好的性能，需要构建中国本土化场景库，并形成统一的标准。

2.1.2 国内发展现状

目前国内已经有很多研究机构和企业正在进行仿真测试场景库的搭建工作，如中汽数据、中国汽研及百度 Apollo 等，为行业内仿真测试场景库研究提供一定的参

考。图 2.7 为国内主要测试场景及数据集相关项目的发展历程。

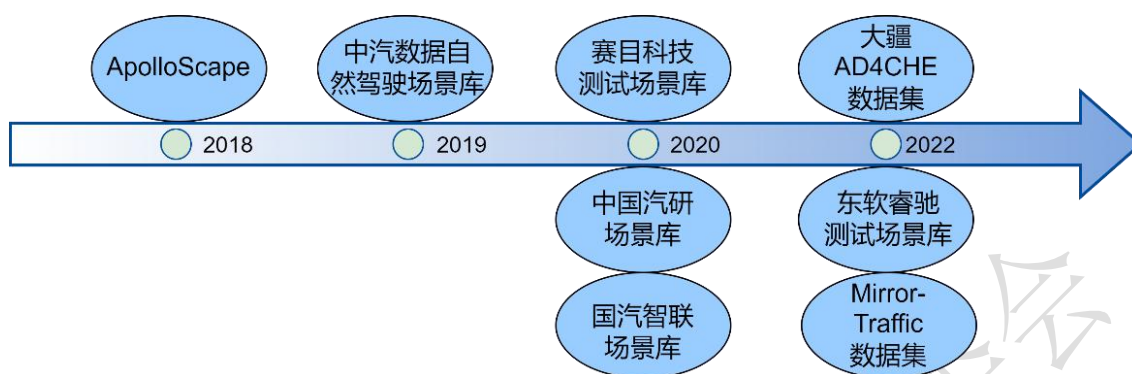


图 2.7 国内主要场景库发展历程

a) ApolloScape 场景数据集

2018 年，百度 Apollo 自动驾驶开放平台正式开放 ApolloScape 大规模自动驾驶数据集，如图 2.8 所示。ApolloScape 数据集旨在促进自动驾驶从感知、导航到控制等各个方面的创新。它可提供对语义注释（像素级）街景图像和支持用户定义策略的仿真工具的开放访问^[19]。

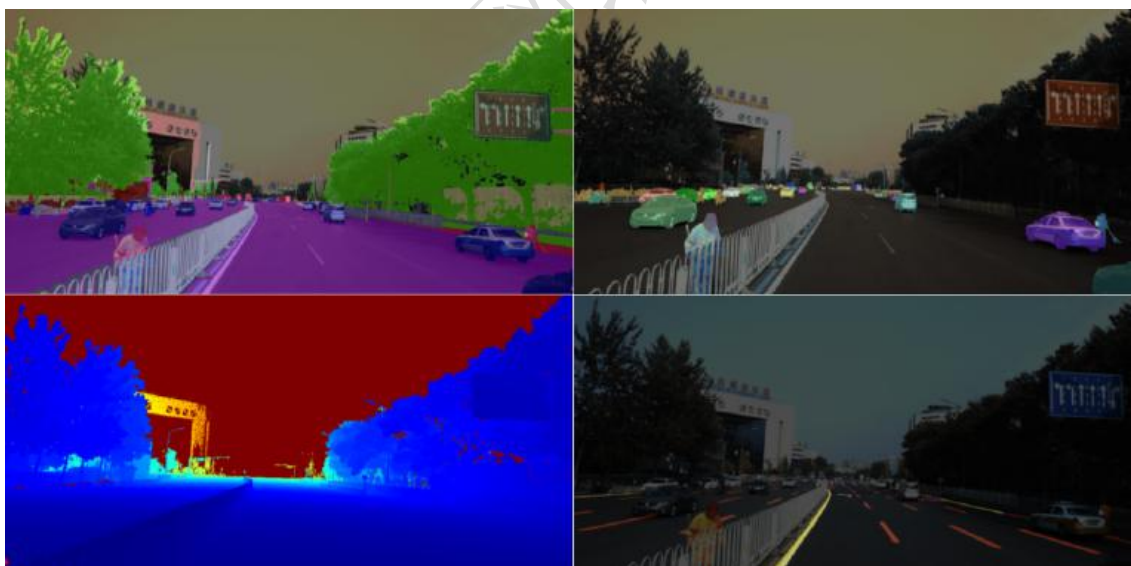


图 2.8 ApolloScape 数据集

b) 中汽数据场景库数据集

中汽数据有限公司（以下简称“中汽数据”）以汽车产业大数据为研究对象，深入开展驾驶场景大数据技术的研究与应用，中汽数据以积累的近 50 万公里高精

度真实世界数据为基础，从不同维度对数据库进行分类构建和更新迭代，地域覆盖北京、天津、上海等重点城市，构建形成自然驾驶场景库，进一步推动围绕驾驶场景各项业务的不断发展。

c) 中国汽研自然驾驶场景库

中国汽研已开展上百万公里自然驾驶数据采集，如图 2.9 所示，结合大数据分析技术，构建了覆盖中国高速公路、城市快速路、城市道路及市郊道路等典型道路类型以及晴、阴、雨、雾、雪等天气类型的自然驾驶场景库。



图 2.9 中国汽研自然驾驶场景库

d) 国汽智联自然驾驶场景库

国汽（北京）智能网联汽车研究院有限公司（以下简称“国汽智联”）依托国家重大研究项目，联合中国一汽、北汽研究院等 10 余家企业，在全国范围内共同采集了百万公里的自然驾驶数据，覆盖雾、雨、雪等多种天气与国内复杂交通状况，搭建了仿真场景模型，自主开发了场景数据管理平台，建立具备中国道路特色的 ICV 场景库。

e) 东软睿驰自然驾驶场景库

自然驾驶场景主要来源于车端采集，通过人工标记、截取等手段，获取复杂多

变的动态交通信息与静态环境信息，包括常规场景、危险场景及边缘场景，主要用于测试自动驾驶汽车在真实路况下的能力水平。

f) 赛目自然驾驶场景库

2020 年，北京赛目科技股份有限公司（简称“赛目科技”）自主研发适用于中国特色智能网联汽车测试、验证和认证的场景数据库。该场景库基于 OpenX 标准格式，满足多种工具链的接口需求。工况覆盖高速、城市、乡村、停车场等重点领域，环境覆盖晴天、雨天、雪天及雾霾等多种天气，具有无限性、扩展性、批量化、自动化等特点。

g) 中国拥堵场景数据集

2022 年吉林大学及大疆车载联合开发中国高速路、快速路拥堵场景数据集，采用无人机悬停航测，基于多种中国城市场景的部分高速路和快速路进行数据采集，如图 2.10 所示，主要用于自然交通参与者行为分析，为传感器、感知与决策算法提供真值，快速生成自动驾驶测试场景，衍生设计用于场地测试的目标障碍物等。



图 2.10 中国拥堵场景数据集

h) Mirror-Traffic 路侧场景库

清华大学苏州汽车研究院基于路侧视频以及无人机航测等方式获取的人类自然驾驶过程中车辆的行驶轨迹，发布了 Mirror-Traffic 场景库。该场景库不仅可以用于

于自动驾驶车辆测试过程中场景挖掘与研究，还可以应用到交通模型分析等方面。采集数据包含城市道路的交叉口、高速路汇入汇出匝道等多种道路类型，多种天气类型，总采集时长超过 2000 分钟，累计采集历程超过 10000 公里涉及多种交通参与物类型以及多种道路交通元素。

表 2.4 为国内主要测试场景库及数据集采集数据的分布情况，分别从发布机构、采集方式、采集区域、道路类型、天气类型以及采集时间段几个维度进行对比分析。

表2.4 各测试场景及数据集数据分布对比

场景库及数据集	发布时间	发布机构	采集方式	采集区域	道路类型	天气类型	时间段
ApolloScape 数据集	2018	百度 Apollo	车端采集	北京、上海、深圳三个站点周围 1000km 的城市交通轨迹	高速、城市、停车场、园区、机场等	良好、极端天气	白天
中汽数据自然驾驶场景库	2019	中汽数据（天津）有限公司	车端采集 无人机采集	北京/天津/上海等 20 多个重点城市	城市、高速、快速路、停车场等重点道路区域	晴、雨、雪、雾、霾等多种天气	白天、傍晚、夜晚等
中国汽研自然驾驶场景库	2020	中国汽车工程研究院股份有限公司	车端采集	中国 29 个省市自治区（除港澳台新疆西藏）	高速公路、国道/省道、城际高速、城市道路、乡村等	晴、阴、雨、雪、雾等	白天、黑夜、黄昏
国汽智联自然驾驶场景库	2020	国汽（北京）智能网联汽车研究院有限公司	车端采集	北京、天津、河北等，覆盖全国 25 个省、市、自治区	城市道路、高速公路、封闭道路、封闭区域、站台、涵洞、高架桥、高铁站	晴、阴、雨、雾等各种天气	白天、夜晚

表2.4 各测试场景及数据集数据分布对比（续）

场景库及数据集	发布	发布	采集	采集	道路	天气	时间
---------	----	----	----	----	----	----	----

自动驾驶系统测试场景数据采集和分析标准需求研究项目组

	时间	机构	方式	区域	类型	类型	段
东软睿驰自然驾驶场景库	2020	东软睿驰汽车技术（上海）有限公司	车端采集	中国一线、新一线、二线和三线城市 50+	城市（含快速路）55%、高速 35%、城际 10%一线，其中新一线城市间走省道，二线，三线城市间高速	晴天 80%、阴天 /多云 10%、雨天/雾天/雪天 10%	白天 70%、夜晚 20%、白天-逆光 5%、黄昏 5%
赛目自然驾驶场景库	2020	北京赛目科技有限公司	车端采集	长沙、杭州	高速、城市、乡村、停车场等重点领域	晴天、雨天、雪天、雾霾等多种天气	早晚高峰、平峰；傍晚、夜晚
中国高速路快速路拥堵场景数据集（AD4CHE）	2022	吉林大学及大疆车载联合开发	无人机采集	中国五个城市多条高速公路、快速路	高速路、快速路、城市拥堵路段、弯道及车道汇入/汇出	/	/
Mirror-Traffic 路侧场景库	/	清华大学苏州研究院	路侧采集以及无人机采集	/	城市道路的交叉口、高速路汇入汇出匝道、停车场等多种道路类型	晴天、阴天、多云等	早晨、中午等

表 2.5 为当前阶段国内各场景库及数据集关于采集数据规模的对比，重点介绍了自然驾驶场景的数据量，对自动驾驶算法测试具有一定的参考意义。

表 2.5 国内主要测试场景及数据集数据规模对比

场景库及数据集	数据规模
ApolloScape 数据集	轨迹数据集，三维感知激光雷达目标检测和跟踪数据集，包括约 100K 图像帧，80k 激光雷达点云和 1000km 城市交通轨迹。
中汽数据自然驾驶场景库	工程车采集自然驾驶数据超过 100 万+km，其中高精度数据里程 30 万+km、高速公路 50 万+km、结构化场景 10 万+段、停车场 500+个、标注图片 50 万+张；无人机采集数据涵盖 17 个城市，包括视频片段 5 万+、场景片段 10 万+；联合建立场景 2500+，可生成测试用例上万个；
中国汽研自然驾驶场景库	基于百万公里自然驾驶数据，形成 5000 类/114000 例中国典型道路场景
国汽智联自然驾驶场景库	切分、提取、泛化后得到场景数量 10 万+

表 2.5 国内主要测试场景及数据集数据规模对比（续）

场景库及数据集	数据规模
东软睿驰自然驾驶场景库	泊车场景 200h，行车场景 800h
赛目自然驾驶场景库	500+个逻辑场景，共分为 63 类
AD4CHE 数据集	中国五个城市多条高速路、快速路提取的 68 段数据采集轨迹共计 53761 条，采集轨迹长度共计 6540.7 公里，图像数据帧数 5529670；包含 10000 张车道线标注及 50000 张车辆标注
Mirror-Traffic 路侧场景库	总采集时长超过 2000 分钟，累计采集历程超过 10000 公里，目前已筛选出超过 1500 个轨迹数据包，涉及多种交通参与物类型以及多种道路交通元素

表 2.6 为国内场景库及数据集在采集数据的原始数据、标注信息、存储格式等方面的对比。

表 2.6 各测试场景及数据集数据格式对比

场景库及数据集	图像、视频数据	激光雷达点云数据	毫米波雷达点云数据	车辆 CAN 数据	车辆、轨迹信息	场景存储数据	场景地图	场景文件
ApolloScape 数据集								
中国汽研自然驾驶场景库					csv、txt	xls	xodr	xosc
国汽智联自然驾驶场景库	jpg	pcd	pcd	csv	rosvbag			
东软睿驰自然驾驶场景库	avi（支持 YUV 格式）			CAN、CANFD			xodr	xosc
赛目自然驾驶场景库					rosvbag		xodr	xosc
AD4CHE 数据集	png				csv			
Mirror-Traffic 路侧场景库					csv			

综上所述，国内测试场景库建设如火如荼，很多企业均构建了场景库，但目前国内场景库场景覆盖度仍较低，数据格式差异较大。因此为构建中国特色的自动驾驶场景库及数据集，亟需进行数据采集与数据格式的标准化研究。

2.1.3 目前存在问题

目前国内外研究人员从测试场景架构、测试场景构建等方面开展了诸多研究。

然而，场景库建设中存在的以下问题阻碍其进一步发展：

- a) **场景覆盖度不足**。目前已建场景库主要集中在城市、高速驾驶场景，且以晴天为主，而其他场景量级较小，如乡村及各种复杂天气条件等。
- b) **数据格式不统一**。各企业或机构实现自动驾驶技术路线存在差异，传感器硬件不一致导致像素、点云密度均不一致。此外，数据存储的格式标准不统一，因此构建场景库后也无法共享使用。
- c) **场景库数据共用困难**。测试场景库多由各个企业或机构独立采集和构建，采集数据质量差异很大，且单独构建成本较高。因此，为建设行业共享可用的场景库，必须推进场景库及相关技术标准的统一。

2.2 测试场景数据采集技术发展现状

自然驾驶数据作为典型场景和边角场景来源的主要基础，代表了约 80% 最常见的道路交通状况。现阶段，场景数据采集依赖于先进的摄像机、雷达等高精度传感设备，进行真实行驶环境高置信度的数字化映射。目前常见采集方式有 3 种：高精度采集车采集、路侧采集、无人机采集^[20]。

2.2.1 国外发展现状

场景数据采集系统一般将激光雷达、毫米波雷达、双目摄像头、广角摄像头、超声波传感器和全球导航卫星系统（GNSS）等多种传感器组合使用，搭载在数据采集车或飞行器等多种不同的平台上，记录的场景数据较为丰富，包含道路、城乡环境和其他交通参与者等多种交通信息，数据种类也涵盖了点云和图像等多种常见类型。

a) 数据采集车方案

自动驾驶数据采集车是一种专门用于采集自动驾驶数据的车辆。早期的数据采集系统多采用摄像头和全球导航卫星系统（GNSS）来记录和复现自然驾驶场景数据，并在道路车辆平台搭载。2016 年奔驰发布的 Cityscapes 数据集，采集系统仅采

用一对立体双目摄像机，包括左目和右目，基线距离 22cm，色彩深度为 16 位。立体双目可以看作一种特殊的激光雷达，能以物理测量级的精度获得 3D 信息，提供真值^[21]。为了丰富场景信息，昆士兰理工大学将车载摄像头拍摄的立体图片与 GNSS 提供的定位及地图数据结合，为后续处理提供原始数据^[22]。

随着自动驾驶技术的不断发展，其对场景数据的要求越来越高。现阶段的数据采集车多采用激光雷达、毫米波雷达、双目摄像头、广角摄像头、超声波传感器和 GNSS 等多种传感器融合方案，采集的场景数据较为丰富，包含道路、城乡环境和其他交通参与者等多种交通信息，数据种类涵盖了点云和图像等多种常见类型。KITTI 数据集的采集平台装配有 64 线激光雷达、惯性导航系统、灰度相机、彩色相机及光学镜头，激光雷达的坐标系遵循右手定则，相机坐标系遵循左手定则，见图 2.11。为了生成双目立体图像，相同类型的摄像头相距 54cm 安装。立体灰度摄像机和彩色摄像机相距 6cm 安装^[23]。nuScenes 数据集的传感器套件包含 6 个摄像头、1 个激光雷达、5 个毫米波雷达、GPS、IMU，见图 2.12。6 个摄像头的分辨率均为 1600×900 ，车辆后部相机 FOV 为 110 度，其他 5 个相机 FOV 为 70 度。激光雷达 32 线，速率 20Hz^[24]。Waymo Open Dataset 的数据采集车配备了 5 个激光雷达和 5 个相机，激光雷达在车辆前半段左、中、右位置各 1 个，车顶 1 个，车尾 1 个；5 个摄像头均安装车顶位置。此外，Waymo 第五代无人驾驶平台所有传感器的实时数据，输入进神经网络模型，以定位三维空间中的关键点^[25]。

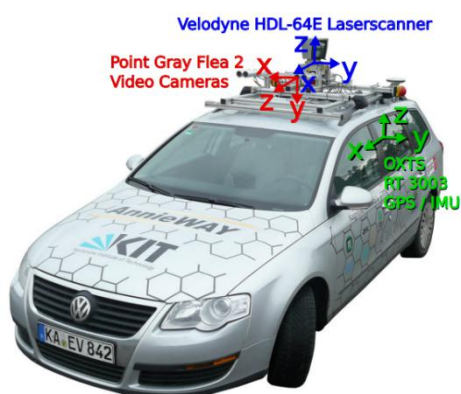


图 2.11 KITTI 采集车

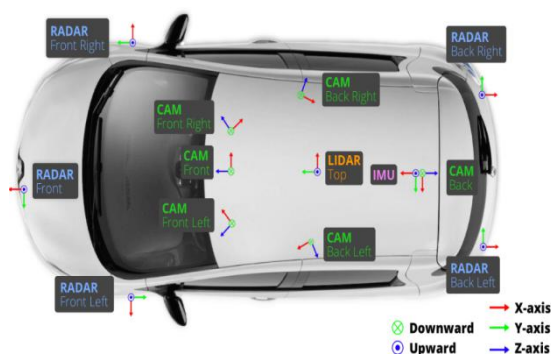


图 2.12 nuScenes 数据采集车

b) 路侧采集方案

路侧采集是将摄像机、激光雷达等高精度传感设备固定在道路上方，实现长距离、广视角的交通车流轨迹数据的大范围采集，最典型的是美国 NGSIM 数据集，通过在高速公路上空数十米高度处架设高清摄像机实现 640m 范围内的车流数据采集。

c) 无人机采集方案

无人机数据采集技术是指利用无人机搭载各种传感器和设备，对目标区域进行数据采集和处理的技术。随着传感器技术的不断进步，无人机可以搭载更先进的传感器，如高精度相机、多光谱相机等，可以实现更加精准的数据采集。

使用无人机进行数据采集时，传感器的安装位置和方式需要根据具体的数据采集需求和无人机的型号进行选择。无人机搭载的主要传感器为云台摄像机，通常安装在无人机的底部或侧面，通过无人机的飞行轨迹和姿态控制，可以实现相机的稳定工作。highD 数据集采用的无人机具有较好的稳定性和飞行控制能力，搭载的相机具有较高的图像质量和拍摄速度^[18]。

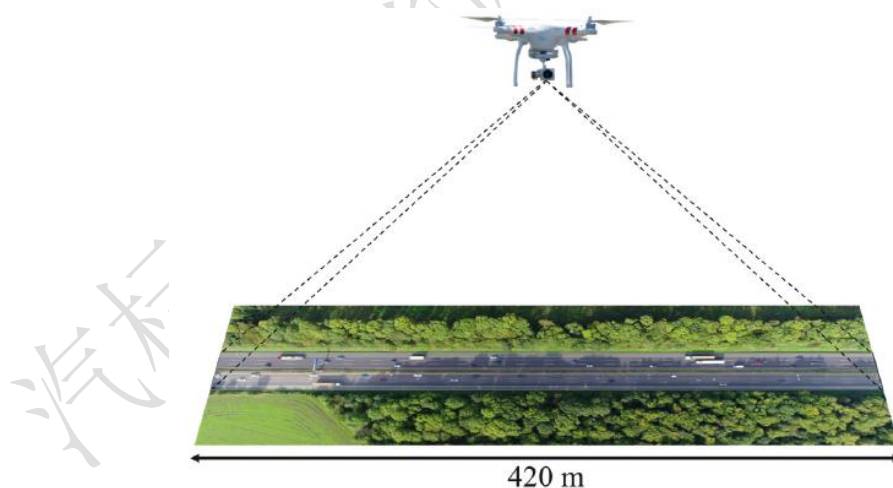


图 2.13 highD 无人机采集

综上所述，早期的数据采集系统所用传感器类型单一，所用平台以搭载摄像头和 GNSS 为主，采集场景的种类较少。随着技术的不断发展，现有的数据采集系统多使用多传感器融合的方案，这方面的研究工作将会推进交通场景库的建成和自动

驾驶的落地应用。但以上方案采集的数据类型和数据格式均有较大差异，采集到的场景很难在行业内共享。

2.2.2 国内发展现状

国内数据采集研究开展相对较晚，但随着自动驾驶功能升级，需测试与验证的场景数量呈几何级增加，场景数据的采集已成为产品开发至关重要的一环。现阶段，中汽数据、中国汽研、百度 Apollo 等都在自动驾驶领域开发了大量的数据采集系统，为测试场景库的构建提供强有力的技术手段。

a) 数据采集车方案

基于数据采集车进行自然驾驶场景采集仍是构建测试场景库最重要的数据来源。现阶段，众多厂商已经开展针对中国道路的智能网联汽车驾驶场景数据库研究及应用工作，深入研究驾驶场景采集技术，搭建了基于多传感器融合方案的采集平台，但不同厂商的部署方案略有差距。如中汽数据搭载 6 路周视高清摄像头、1 路智能摄像头、高线程激光雷达、5 路毫米波雷达、高精度组合惯导、激光雷达融合系统等。中国汽研以自研采集平台为基础，车辆前方安装激光雷达、视觉传感器，有效弥补毫米波雷达的感知弱势，见图 2.14。赛目科技现有数据采集方案中摄像头类型包括智能摄像头、高清摄像头及普通摄像头三种。

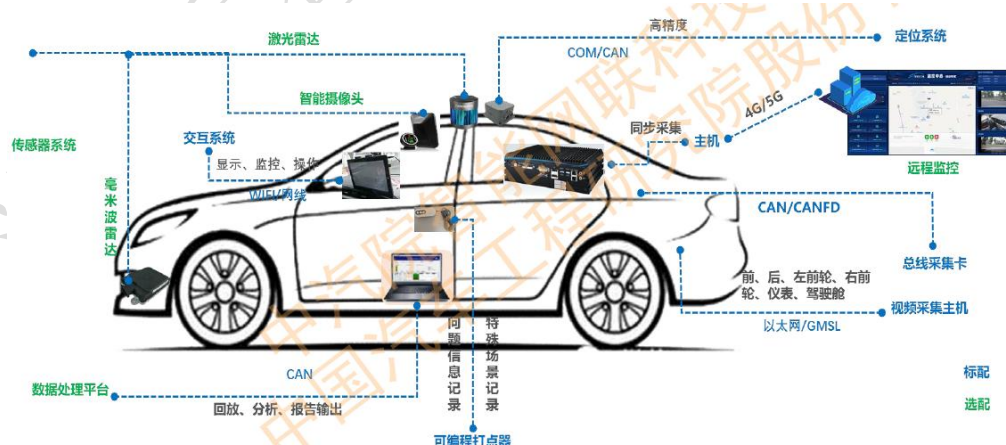


图 2.14 中国汽研采集方案

b) 路侧采集方案

路侧采集是指在道路旁边或道路上方安装传感器设备，采集道路上的交通信息、行人信息及车辆信息等数据，主要设备包括交通监控摄像头、车载摄像头、激光雷达等。清华苏研院的 Mirror-Traffic 数据集采集方案以视觉算法为基础，通过交通监控视频和无人机视频的信息提取技术所获取的车辆和行人真实轨迹数据，通过摄像机结合地图影像，在标定时无需特殊标定物即可完成摄像机标定，通过结合目标检测算法和视频抖动处理算法降低误判率。招商车研基于车端感知系统与路侧单元并行采集基础场景数据，其中路端采集感知系统布置方案见图 2.15。在交叉路口、直道、丁字路口三种关键道路典型地段路侧布置激光雷达、摄像头以及雨量传感器和光照传感器。利用多激光雷达拼接形成的连续感知能力，获取交通参与者的精确信息，如目标物类型、外观尺寸、经纬度、速度、加速度、方向角等，生成标准化的数据包并由消息中间件为上层应用提供信息，配合摄像头视频流实现融合感知，利用多雷达、多传感器感知融合算法，实现场景数据 24 小时无间断高效采集。

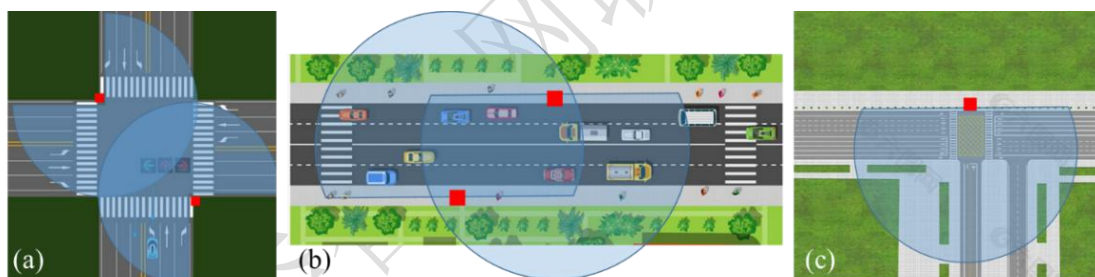


图 2.15 路端感知系统布置方案

c) 无人机采集方案

大疆车载采用无人机自带相机，在交通流上方 100m 处悬停，垂直向下进行拍摄，采集 4K 30 帧视频，采集包含 5-6 条车道线的路段，视野中路段长为 130m，见图 2.16。南京理工大学采用高分辨率的航拍数据进行道路场景的重建；西安电子科技大学使用无人机采集的实时视频图像作为场景数据进行分析处理^[22]。无人机采集方案具有视野广、灵活性强、透视变形低等优点，且随着无人机的广泛应用，飞行器平台的成本得到了进一步降低，但其目前所采集数据的种类仍以图像为主，与

其他采集方案相比较为单一。

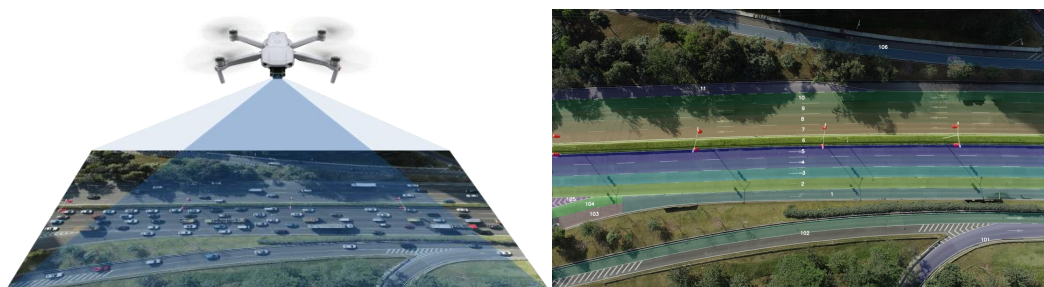


图 2.16 大疆无人机采集方案

综上所述，国内现有的数据采集系统以多传感器融合方案为主、由实车采集方案向多种采集方式结合转化，能对多种类型的场景进行采集。但目前自然驾驶场景数据采集规范不统一，传感器类型并不完全统一，传感器型号也不尽相同，导致采集的数据难于共享共用。

2.2.3 目前存在问题

当前数据采集技术在自动驾驶领域的发展是多方面的。一方面，传感器技术不断进步，激光雷达、摄像头和毫米波雷达等传感器能够提供更高分辨率、更准确的环境信息。另一方面，数据采集方式也由原先的单一车端采集方式发展为多种采集方式并行的多样化采集方式，并且采集技术也较为成熟，但现阶段数据采集在标准统一、业内共享等方面存在以下问题：

- a) 采集方案及采集设备要求不统一。自然驾驶数据采集多通过视觉传感器、激光雷达、毫米波雷达以及 GNSS 等传感器实现，采集设备的配置及采集方式缺少相应规范，不利于数据的收集与应用。此外，不同传感器产生的数据格式及类型存在较大差异，并且原始数据中存在大量无效数据、错误数据。
- b) 高精度采集车成本较高。目前自然驾驶数据多采用人工驾驶汽车或具有一定辅助驾驶功能的汽车采集，始终缺少高等级自动驾驶及网联协同驾驶的行车环境数据，难以满足高等级自动驾驶汽车测试需求。

2.3 测试场景数据分析技术发展现状

自动驾驶数据分析技术是实现自动驾驶的关键步骤之一，是将采集数据转化为测试场景的基础。数据分析技术大致可以分为数据预处理、场景标注、特征提取、场景聚类、场景生成等步骤，可以实现场景库的高效构建，围绕自然驾驶数据的分析应用国内外有关学者与机构开展了相关的研究工作。

2.3.1 国外发展现状

国外自动驾驶数据分析技术相对来说已经比较先进，通过对数据的提取与分析，寻找关注对象，利用聚类分析等方法建立场景数据库，为自动驾驶技术的发展和應用提供了重要的支持和保障。

数据预处理：自动驾驶数据预处理是指对采集到的原始数据进行清洗、筛选、转换等操作，以便于后续的数据分析和模型训练。PEGASUS 项目的数据预处理过程包括数据脱敏、数据上传、数据解析、数据模型转换、数据质检等，但这些过程的顺序并不固定，数据脱敏和数据上传会根据不同场景需求放在合适的位置^[26]。数据清洗主要包括清除冗余、删除缺失数据、数据修复等，其中数据修复可以进行关键信息的人工补全或者按照数据的统计学规律进行修复。目前数据清洗技术已经得到了不错的发展，不仅仅是通过人工手段，更加高效的是算法加人工辅助的方式^[27]。

场景标注：场景标注是指对图像或视频中的场景元素进行标注，以提取出场景中的各种信息，如目标、区域、属性等。目前，场景标注的方法包括自动标注和手动标注两种，多采用自动标注和手动标注相结合的方式。自动标注利用计算机视觉和机器学习技术，自动对图像或视频进行标注，速度快但准确率有限。自动标注常用的标注方法包括：基于语义分析的方式、基于半监督学习的方式、基于贝叶斯学习的方式^[28]。手工标注需要人工对图像或视频进行标注，耗时耗力，但标注结果准确可靠；Waymo Open Dataset 在数据标注方面，考虑到了人工标注费时费力，在人工标注真值的数据上训练模型，使用经过训练的模型来预测无标签数据，从而创建伪标签。此外，将标签数据和新生成的伪标签数据结合起来作为新的训练数据^[25]。

加拿大滑铁卢大学考虑到高精度语义地图及相关数据采集所需的成本较高，提出了一种基于接近度的由图像到地图的数据收集和注释框架，其使用神经网络和在线数据库中的可用标签进行学习，并实现了不同天气下道路场景的自动标注^[22]。

特征提取：特征提取是指从图像或视频中提取出与自动驾驶相关的特征，对场景中的静态和动态等元素进行参数化描述，自动驾驶场景特征可以包括车辆、行人、道路、交通标志、路标等各种目标和区域，以及它们的位置、速度、方向、大小等属性。特征提取的方法包括传统的特征提取方法和基于深度学习的特征提取方法。传统的特征提取方法包括 SIFT、HOG、SURF 等，这些方法可以提取出图像中的边缘、角点、纹理等特征；基于深度学习的特征提取方法包括卷积神经网络（CNN）、循环神经网络（RNN）等，这些方法可以自动学习图像中的特征。Elrofai 等基于自然驾驶数据，提出了一种场景自动识别方法，开展了对场景的提取和分类研究，并针对转弯和变道场景进行了分析验证。Waymo 使用了基于深度学习的语义分割算法，如 DeepLab、PSPNet 等，可以将图像中的每个像素分配到不同的类别中。

场景聚类：通过定义场景分类规则，将场景进行分类、聚类处理，将符合分类规则的场景聚类成为相应的逻辑场景，并明确场景要素的参数空间，按不同特征对场景进行分类。常用的聚类算法主要有 K-Means 聚类、层次聚类、混合高斯模型等。在 Waymo Open Dataset 中，K-Means 算法被用于对图像数据中的目标进行检测和识别。通过对聚类结果进行分析和处理，可以识别出图像中的不同目标，如车辆、行人、自行车等。

场景生成：基于各种动静态元素的关联关系或人工经验对场景参数进行重新组合，生成新的场景，提高场景的覆盖率。特斯拉等项目中根据基于规则的场景生成技术生成不同的驾驶场景，例如包含交通规则和道路标志等的场景。Rocklage 等在测试场景生成中考虑车辆运动轨迹规划，通过建立变道轨迹网络模型，利用网格间随机连线生成变道轨迹^[29]。

综上所述，随着深度学习等在自动驾驶领域的不断应用，测试场景的数据分析技术取得了明显的发展。目前在数据融合、场景理解、场景标注、场景分类等领域

已经相对成熟，但在场景提取、场景聚类以及场景生成等方面均未形成标准的方法，各家的实现方式也没有相应的规范，不利于测试场景库的构建。

2.3.2 国内发展现状

国内的众多研究机构、企业在吸收和借鉴国外一些大型数据采集与分析项目的基础上，也发展出了一套完整的数据分析方法。目前测试场景数据分析技术已经取得了不错的发展，不仅是通过人工手段，更加高效的算法加人工辅助的方式得到了很好的应用。数据分析技术流程大致包含数据清洗、场景理解、特征提取、场景聚类、场景生成、场景泛化等，但不同机构在技术细节上会有一些差异。

数据预处理：自动驾驶数据集数据预处理是指对自动驾驶车辆采集到的数据进行清洗、转换和标准化等处理，以便于后续的数据分析和建模。在数据采集之后，中汽数据在数据清洗阶段通过归一化、缺失值补充、冗余值剔除、异常值修正、格式转换等提高数据的质量。

场景标注：现阶段多采用自动标注和手动标注相结合的方式。其中，中国汽研支持典型场景自动标注，特殊场景人工标注。招商车研根据主车与目标物横纵向运动关系，设定关键参数，达到阈值条件时，标出相应类别场景。东软睿驰正在开发数据半自动标记工具，可对样本中的自车速度、道路类型、动作场景等进行初次标记，再通过人工进行二次复核，最终实现场景精细化分类。

特征提取：对场景的各种动、静态元素进行提取，从而完成数值化的描述，以便场景的进一步应用。目前多采用抽样、分类、聚类、非参数检验等方法，从原始数据中完成某一类或某几类场景的提取。中国汽研可支持根据切入、切出、换道、危险场景等进行场景提取，以测试不同的自动驾驶功能。清华苏研院支持场景切片，场景可离线自动识别、人工筛选，处理规则可灵活配置；支持离线自动场景提取与手动场景打标，场景标签类别包括但不限于本车驾驶行为、目标物、环境/交通信息、道路类别等。赛目科技基于场景规则，通过自编码、聚类算法、算法融合等机器学习、深度学习算法，实现逻辑场景的自动化提取。

场景聚类：对场景特征进行进一步分析，对场景库中的场景进行分类和聚类。ApolloScape 用算法对原始的数据进行挖掘，一种是静态场景，比如车道线、交通信号灯，甚至路边洒水车这种特殊障碍物。此外，还会去做一些仿真场景的挖掘，比如说闯红灯、交通并道、或者超车这样的行为，并将不同的仿真场景进行分类^[30]。招商车研使用机器学习算法、统计分析算法等对场景的特征参数进行分类、聚类、回归等处理，以得到场景的关键参数。刘颖等基于自然驾驶数据，利用聚类分析的多元统计学方法对样本进行研究，得到了 5 类具有典型特征的 AEB 危险场景，设计了典型交通工况的 AEB 行人测试场景^[31]。一些学者在研究中引入深度学习模型，可以根据数据内在特征实现场景自动聚类。中国汽车技术研究中心以采集的自然驾驶场景数据库为基础，提出一种基于轻量级梯度提升机的决策树模型，提取场景特征，实现场景的自动聚类及生成^[20]。

场景生成：场景生成目前以生成 OpenX 标准格式场景文件为主，包括.xodr 格式的道路文件和.xosc 场景文件，保证了数据处理的高可靠性、测试场景格式的强兼容性。中国汽研支持 1:1 还原的场景转换和基于场景分类要素的自动生成功能；百度、腾讯等利用数据驱动的方法，分析真实道路数据和车辆行驶数据，学习道路结构和车辆行为的模式，生成类似的测试场景。长安大学朱宇等为解决测试场景构建不完整的问题，提出一种基于蒙特卡洛模拟的变道虚拟测试场景自动生成算法，作为一种结合模型和真实数据的场景生成方法，该算法简单、生成轨迹数量多、生成过程解释性强^[32]。

综上所述，目前国内的数据分析技术也已经形成较为完善的流程，能够提供数据-场景-场景管理的全过程，尽管已经取得了一定发展，并在实际应用中发挥着重要作用，但仍然存在许多问题需要解决。这些问题包括数据清洗与预处理、自动标注与手动标注相结合、特征提取与描述精确性、聚类与分类准确性以及生成真实且多样化测试场景等方面。

2.3.3 目前存在问题

在自动驾驶汽车测试场景数据分析方面，各国机构开展了大量工作，并已经在辅助驾驶功能的测试方面发挥了重要的作用，但**尚未形成从数据分析挖掘到场景构建应用的完整方法链和工具链**，仍存在以下问题：

- a) 场景转换自动化程度较低。场景元素识别和分类等相关研究较为成熟，但数据标注、场景聚类、场景生成等技术的发展仍处于较为初步的阶段，无法实现完全的自动化，导致场景转换过程耗时又费力。
- b) 数据的准确性和真实性。自然驾驶数据需要准确地反映真实驾驶环境中的各种因素，包括道路、交通标志、交通信号灯、行人、其他车辆等。然而，采集到的数据可能存在噪声和误差，在进行场景构建时可能无法完全准确地反映真实驾驶环境。
- c) 数据的泛化能力。转化后的数据需要具有一定的泛化能力，即能够适应不同的驾驶场景和情况。然而，转化后的数据可能存在泛化能力不足的问题，无法完全适应所有可能的驾驶场景。此外，泛化过程中真实性损失也是一大问题。

第三章 自动驾驶系统测试场景政策及标准现状

3.1 自动驾驶系统测试政策及标准现状

3.1.1 我国自动驾驶系统测试地方政策

在全国政策文件的指导下，各省市制定了配套实施政策，指导开展自动驾驶系统测试，在促进智能网联汽车产业发展，提高智能网联汽车的安全性、效率和便利性，改善交通环境和出行体验，提升城市品质方面发挥了积极作用，并且不同地方的政策具有一定的开拓性，如：

- a) 2021 年 4 月，北京市发布《北京市智能网联汽车政策先行区总体实施方案》。随后 7 月份配套发布了《北京市智能网联汽车政策先行区高速公路及快速路道路测试及示范应用管理实施细则（试行）》。同年 11 月，**全国首个自动驾驶出行服务商业化试点**在北京开放，这标志着国内自动驾驶领域从测试示范迈入商业化试点的新阶段。
- b) 2022 年 8 月，深圳市发布《深圳经济特区智能网联汽车管理条例》，这是国内首部关于智能网联汽车管理的法规。也是深圳首次准许完全自动驾驶的智能网联汽车可以不配备驾驶人，在划定的区域、路段行驶。2023 年 2 月，上海市发布《上海市无驾驶（安全）员智能网联汽车测试技术方案》，除实际道路测试方法外，也包括模拟仿真测试、封闭场地测试方法。
- c) 2018 年 5 月，重庆市发布《自动驾驶开放道路准入测试方案》，对测试项目类型、环境要求等进行了较为详细的规定。如，规定了重庆市内自动驾驶测试区（城市道路测试区、高速测试区、乡道及山路测试区）内 20 种准入测试项目类型，包括自动紧急制动、环岛通行、夜间行驶等，并对各种类型的相关测试方法提出测试规范要求。

各省市政策法规在規定常見測試場景的前提下，根據地方特色，規定特殊場景的測試要求，如：

- a) 2018 年 2 月，北京市發布的《北京市自動駕駛車輛道路測試能力評估內容與方法（試行）》羅列了 20 多種測試場景，並給出了詳細的判定規則。
- b) 2022 年 10 月，吉林省印發的《吉林省智能网联汽車道路測試與示范應用管理實施細則（試行）》規定“在吉林省開展冬季低溫、冰雪路面等場景項目測試的，應取得省內第三方檢測機構出具的附加項目檢驗報告”，這是國內政府部門首次明確提出的關於冬季低溫、冰雪路面測試的要求。
- c) 2020 年 9 月，重慶市《重慶市自動駕駛道路測試管理辦法（試行）》（以下簡稱《辦法》），突出了重慶山地道路特色，首次明確了自動駕駛道路測試的道路類型：普通道路、山地道路和城市快速路，並把測試類型分為一般測試、載人測試、載物測試和編隊行駛測試四個類型，完善了測試類型，滿足了場景多樣性的需求。地方性政策文件所涉及的測試場景總結見表 3.1。

表 3.1 主要地方政策中測試場景總結

序號	政策名稱	測試場景或項目
1	《北京市自動駕駛車輛道路測試能力評估內容與方法（試行）》	綜合駕駛能力測試場景包括：起步、停車、跟車、變更車道、直行通過路口、通過人行橫道線、路口左轉彎、路口右轉彎、路口掉頭、靠邊停車、通過公共汽車站、會車、通過環島、主輔路行駛、通過模擬苜蓿葉式立交、通過學校區域、通過隧道、超車、倒車入庫、側方停車、通過雨區道路、通過霧區道路、通過濕滑路面、通過遺撒路面、避讓應急車輛、夜間行駛
2	《上海市智能网联汽車道路測試管理辦法》	（一）交通信號識別及響應；（二）道路交通基礎設施與障礙物識別及響應；（三）行人与非机动车識別及響應；（四）周邊車輛行駛狀態識別及相應；（五）車輛定位；（六）自動緊急避險；（七）動態駕駛任務干預及接管；（八）風險減緩策略；（九）网联協作式駕駛能力測試；（十）地下及內部道路測試；（十一）車輛信息傳輸安全測試；（十二）軟件升級測試

表 3.1 主要地方政策中测试场景总结（续）

序号	政策名称	测试场景或项目
3	《重庆市智能网联汽车道路测试与应用实施细则（试行）》	限速信息识别及响应、跟车行驶、车辆碰撞自动紧急制动、车辆变更车道检测及响应、障碍检测及响应、并道行驶、超车、靠路边停车、交通信号灯识别及响应、行人和非机动车识别及避让、交叉路口通行、环形路口通行、弱势群体碰撞自动紧急制动、车道保持控制、探测并避让对向来车、停车场通行、网联通信、人工操作接管、编队行驶、匝道通行、坡道行驶、弱基准环境巡航、人机交互、动态交通规则响应、特定场景作业
4	《吉林省智能网联汽车道路测试与示范应用管理实施细则（试行）》	交通信号识别及响应（包括交通信号灯、交通标志、交通标线等）、道路交通基础设施与障碍物识别及响应、行人与非机动车识别及响应（包括横穿道路和沿道路行驶）、周边车辆行驶状态识别及响应（包括影响本车行驶的周边车辆加减速、切入、切出及静止等状态）、动态驾驶任务干预及接管、风险减缓策略、自动紧急避险（包括自动驾驶系统开启及关闭状态）、车辆定位。

受限于客观条件，地方政策法规也存在一定局限性。各地方政策规定的测试场景范围不一致，对于测试场景的描述及测试通过条件描述详细程度均不一致，未形成全国统一的自动驾驶道路测试所应包含的测试场景；当前测试标准主要针对 ADAS 辅助驾驶系统测试，且对测试场景进行了规定，难以包含实际驾驶场景的复杂情况；对于 AD 自动驾驶系统的测试标准比较少，所包含的测试场景有限，尤其对需要进行测试的危险场景缺少系统梳理；总之，未形成统一的测试场景范围。

此外，各地开放道路测试的范围仍然较小，无法满足未来智能网联汽车量产上路的需求。因此急需制定统一的自动驾驶测试数据库，通过大量模拟仿真测试，加快自动驾驶测试进程，提高效率，降低成本，推动自动驾驶技术发展及应用。

3.1.2 国内外自动驾驶系统测试标准

联合国（即 UN/ECE）发布了 L3 级自动驾驶相关的法规 R157，旨在制定关于自动驾驶车道保持系统（ALKS）车辆认证的统一规定，包括关于型式认证、技术要求、审计和报告的规定，以及 ALKS 在测试场和公开道路下的测试。该标准的附件 3 做了关于交通干扰临界场景指导，对其他车辆切入、切出、减速的场景搭建车辆性能模型，确定了车辆应该避免碰撞的条件；给出的熟练的人类驾驶员性能模型为日本统

计的驾驶员模型，但与我国驾驶员驾驶行为习惯不同，需要根据我国自然驾驶场景数据统计获得符合我国驾驶员行为习惯的中国驾驶员模型。

国际标准化组织 ISO/TC22/SC32/WG8 功能安全工作组制定的 ISO 21448: 2022 《道路车辆 预期功能安全》，旨在为全球自动驾驶车辆的安全开发和测试评价提供技术指导。该标准立足对自动驾驶安全影响更广泛的非故障安全领域，重点关注自动驾驶汽车的行为安全，解决因自身设计不足或性能局限在遇到一定的触发条件（如环境干扰或人员误用）时导致的整车行为危害。该标准提供了用于确保预期功能安全（SOTIF）的通用论证框架和措施指南，但并未给出明确的预期功能安全场景和测试方法。

此外，国外针对主动安全测试的新车评估程序标准，如欧盟新车安全评鉴协会的 ENCAP、欧洲委员会发布的 R79/GSR、美国高速公路交通安全管理局（NHTSA）的 US-NCAP 等。此类标准规定的安全测试都需要在标准规定的封闭场地进行，测试场景数量有限，且纵向测试数量远大于横向。如，欧盟新车安全评鉴协会 ENCAP2022 涵盖的主动安全场景包括 AEB-车对车，AEB 车对两轮车，AEB 车对行人，LSS-HMI/LKA/ELK，HMI-人机界面等。

我国 2022 年发布了国标《智能网联汽车 自动驾驶功能场地试验方法及要求》，规定了智能网联汽车自动驾驶功能的道路试验条件、试验方法及要求，提出了交通信号识别及响应等 8 类测试项目。国内与 ENCAP 相对应的测试标准有 CNCAP（中国新车评价规程，中国汽车技术研究中心），i-VISTA（智能汽车集成试验区（评价规程）），CCRT（中国汽车消费者研究及测试中心（评价体系））等。其中，由中国汽车技术研究中心发布的 CNCAP 2021（2023），涵盖的主动安全场景包括 AEB-车对车，AEB 车对两轮车，AEB 车对行人，LKA-车道保持辅助，HMI-人机界面等五项试验项目。此外，还有 BSD-盲点检测-车对车，BSD-盲点检测-车对行人，SAS-限速标志识别，LDW-车道偏离预警等四项可选审核项，总分 56 分。

基于封闭场地场景的测试，可以在一定程度上反映行驶车辆在类似的场景中的

性能，而于公共道路上复杂多变的驾驶环境相比，这部分的测试远远不能代表车辆在公共道路上的实际性能表现。

全国汽标委智能网联汽车分标委(SAC/TC114/SC34)联合中国智能网联汽车产业创新联盟于 2018 年联合发布《智能网联汽车自动驾驶功能测试规程（试行）》，规定了网联汽车自动驾驶功能检测项目的测试场景、测试方法及通过标准等。

此外，汽车相关社会团体组织制定了多项关于自动驾驶测试的团体标准，但不同的团体标准在相关术语、功能要求、性能要求、检测范围等方面规定的内容存在差异。

此外，还有针对特定系统要求及测试的标准，但标准中测试方案虽然对测试场景的环境有所描述，但对测试场景都有明确规定，且只考虑了在能见度较好的情况，而车辆在现实中的行驶环境是及其复杂的，场地测试难以覆盖复杂的现实驾驶环境。驾驶自动化系统测试标准见表 3.2。

表 3.2 驾驶自动化系统测试标准

辅助系统	标准编号	标准名称	标准主要内容、测试项目或场景
车道保持 辅助系统 LKA	ISO 11270	智能交通系统 车道保持辅助(LKA) 系统性能要求与测试方法	本文主要对车道保持辅助系统的功能要求、性能要求以及测试要求进行介绍。主要的测试项目包括直线道路测试，弯曲道路测试
	(EU) 2021/646	机动车辆紧急车道保持系统	针对 LKA 车道保持功能要求
	GB/T 39323-2020	乘用车车道保持辅助（LKA）系统性能要求及试验方法	规定了乘用车车道保持辅助 (LKA) 系统的要求、试验条件和试验方法。 本标准适用于安装有车道保持辅助 (LKA) 系统的 M 类汽车，其他车辆可参照执行。
	GB/T 41796-2022	商用车车道保持辅助系统性能要求及试验方法	规定了直车道和弯车道的车道保持系统测试
车道偏离 预警 LDW	ISO 17361	智能交通系统 车道偏离报警（LDW） 系统性能要求与测试方法	针对车道偏离预警（LDW）功能要求和测试

表 3.2 驾驶自动化系统测试标准（续）

辅助系统	标准编号	标准名称	标准主要内容、测试项目或场景
车道偏离 预警 LDW	ECE R130	车道偏离 (LDW)	针对车道偏离预警 (LDW) 功能要求和测试
	GB/T 26773-2011	智能运输系统 车道偏离预警系统性能要求与检测方法	针对车道偏离预警 (LDW) 功能要求和测试
	JT/T 883-2014	营运车辆行驶危险预警系统技术要求和实验方法	针对两种危险预警系统, 级 LDW 车道偏离预警、FCW 前向碰撞预警功能要求及实验方法
盲点信息 系统 BSI	ECE R151	盲点信息系统 (BSI)	针对盲区预警 (BSD) 功能要求及试验方法
盲区监测 系统 BSD	GB/T 39265	道路车辆 盲区监测 (BSD) 系统性能要求及试验方法	规定了汽车盲区检测 (BSD) 系统的一般要求、性能要求和实验方法
自适应巡航系统 ACC	ISO 15622	智能交通系统 自适应巡航控制系统性能要求与测试方法	本文主要对自适应巡航系统的功能要求、性能要求以及测试要求进行介绍。主要的测试项目包括自动停止能力测试, 目标探测距离测试, 目标识别测试, 曲线能力测试
	(EU) 2021/1958	智能车速辅助	针对 ACC 自适应巡航级 TSR 交通标志识别功能要求
	GB/T 20608	智能运输系统 自适应巡航控制系统性能要求与检测方法	规定了 ACC 系统的基本控制策略、最低功能要求、基本人机交互、故障诊断及处理的最低要求以及性能检测规程。
前向碰撞 预警系统 FCW	ISO 15623	智能交通系统 前方车辆碰撞报警系统 性能要求与测试方法	本文主要对前碰撞预警系统的功能要求、性能要求以及测试要求进行介绍。主要的测试项目包括检测区域测试, 报警距离范围及精度测试, 目标辨识能力测试
	GB/T 33577	智能运输系统 车辆前向碰撞预警系统性能要求和测试规程	规定了车辆前向碰撞预警系统性能要求和测试方法

表 3.2 驾驶自动化系统测试标准（续）

辅助系统	标准编号	标准名称	标准主要内容、测试项目或场景
前方车辆碰撞缓解系统	ISO 22839	智能运输系统 前方车辆碰撞缓解系统 操作、性能和验证要求	本文主要对前方车辆碰撞缓解系统的功能要求、性能要求以及测试要求进行介绍。主要的测试项目包括头顶识别测试，弯道横向目标识别测试，直道横向偏移识别测试，直道横向识别测试，纵向识别测试
自动泊车系统 APS	ISO 16787	智能运输系统 自动泊车系统 (APS) 性能要求与测试方法	本文主要对辅助泊车系统的功能要求、性能要求以及测试要求进行介绍。主要的测试项目包括空间车位测试，线车位测试
	GB/T 41630	智能泊车辅助系统性能要求及试验方法	规定了智能泊车辅助系统的一般要求，性能要求和试验方法。适用于装备有智能泊车辅助系统的 M 类车辆，其他类型车辆可参照执行
车道变更决策辅助系统 LCDAS	ISO 17387	智能交通系统 车道变更决策辅助系统 (LCDAS) 性能要求与测试方法	本文主要对车道变更决策辅助系统的功能要求、性能要求以及测试要求进行介绍。主要的测试项目包括盲点警告测试，接近车辆警告测试，换道报警测试
行人及骑车人碰撞告警 MOIS	ECE R159	行人及骑车人碰撞告警 (MOIS)	针对 FCW 前碰撞预警功能要求及测试方法
智能车速辅助 ISA	(EU)2021/1958	机动车智能速度辅助系统 (ISA)	规定了关于机动车智能速度辅助系统型式批准的具体测试程序和技术要求
自动紧急制动 AEB	ECE R131	预先紧急制动系统 (AEBS)	针对 AEB 自动紧急制动功能要求及测试规程
	ECE R152	轻型车预先紧急制动系统	针对 AEB 自动紧急制动功能要求及测试规程
	GB/T 38186	商用车辆自动紧急制动系统 (AEBS) 性能要求及试验方法	规定了商用车辆自动紧急制动系统 (AEBS) 的技术要求、试验方法。适用于安装有 AEBS 的 M ₂ 、M ₃ 和 N 类车辆。

表 3.2 驾驶自动化系统测试标准（续）

辅助系统	标准编号	标准名称	标准主要内容、测试项目或场景
自动紧急制动 AEB	GB/T 39901	乘用车自动紧急制动系统（AEBS）性能要求及试验方法	规定了乘用车自动紧急制动系统（AEBS）的技术要求及测试方法；适用于安装有 AEBS 的 M1 类车辆。
	JT/T 1242	营运车辆自动紧急制动系统性能要求和测试规程	规定了营运车辆自动紧急制动系统性能要求和测试规程
倒车探测 RVC	ECE R158	倒车检测	针对倒车场景（倒车雷达、倒车影像、补盲后视镜）功能要求
L4 级自动驾驶系统	ISO 22737	智能运输系统-用于预定路线的低速自动驾驶（LSAD）系统-性能要求、系统要求和性能测试程序	针对 L4 级自动驾驶系统的标准，规定了速度不超过 32km/h 的低速自动驾驶系统的功能要求、最低安全要求、性能要求、ODD，以及给出针对 LSAD 系统的性能验证测试的参考，涉及行人、自行车危险情况测试、转弯测试、可驾驶区域测试、最小风险策略（MRM）测试等，从而确保搭载 LSAD 系统的汽车产品符合功能需求。
L4 级自动代客泊车系统子系统	ISO 23374	智能交通系统-自动代客泊车系统（AVPS）第一部分：系统框架、自动驾驶要求和通信接口	针对停车场内 L4 级自动代客泊车系统子系统，规定了 AVPS 的系统架构，AVPS 系统功能要求，停车场场地要求，系统操作员、设施管理员干预的系统管理要求和不同应用场景分类的测试规程。
自动驾驶功能	GB/T 41798-2022	智能网联汽车 自动驾驶功能场地试验方法及要求	本文件规定了智能网联汽车自动驾驶功能进行场地试验时的一般要求、试验过程及通过条件、试验方法。本文件适用于具备自动驾驶的 M 类、N 类车辆，其他车辆类型可参考执行。测试项目包括交通信号识别及响应、道路交通基础设施与障碍物识别及响应、行人与非机动车识别及响应、周边车辆行驶状态识别及响应、自动紧急避险、停车、动态驾驶任务干预及接管、最小风险策略等。

表 3.2 驾驶自动化系统测试标准（续）

辅助系统	标准编号	标准名称	标准主要内容、测试项目或场景
自动驾驶功能	国家标准	《智能网联汽车 自动驾驶功能仿真试验方法及要求》（制定中）	规定了 30 余个典型自动驾驶仿真测试场景，例如前方车辆紧急制动、前方车辆切入等，以及 20 余个危险仿真测试场景，例如应急车辆识别和响应、弯道超车等。
	国家标准	《智能网联汽车 自动驾驶功能道路试验方法及要求》（已报批）	规定了智能网联汽车自动驾驶功能的道路试验条件、试验方法及要求。测试内容包括：系统激活、执行动态驾驶任务、系统后援、试验车辆状态显示等。同时还规定了自动驾驶功能试验道路要素要求及要素采集方法。
封闭测试场地	国家标准	GB/T 43119-2023 自动驾驶封闭测试场地建设技术要求	规定了自动驾驶车辆封闭试验场地的一般要求，场地要素要求和配套设施要求，适用于汽车驾驶自动化分级 3 级及以上等级车辆封闭场地测试。

总体而言，仿真测试、封闭场地测试、道路测试是自动驾驶系统测试的三支柱，但目前标准未对仿真测试、封闭场地测试、道路测试所应包含的场景进行详细区分及明确验证方法。此外，还存在如下问题：

- a) 当前标准测试场景以地方法规规定或设计运行域（ODD）所包含的场景，未包含预期功能安全场景以及实车采集数据转换场景。测试场景应包含通过从整车层面、系统层面以及组件层面进行危害分析，识别产品的性能局限和功能不足，而构造的相应的预期功能安全测试场景以及通过实车采集数据转化而来的测试场景，保留真实交通场景的随机性、复杂性；
- b) 测试标准对于测试场景环境的描述不清晰，缺少定量描述。对于测试环境的要求缺少数据量化描述，影响测试场景数据采集。如，标准 GB/T 41630-2022《智能泊车辅助系统性能要求及试验方法》环境要求为试验路面为平整、干燥、具有良好附着能力的混凝或沥青路面；停车位的边界线应无破损，清晰可见等。GB/T 33577-2017《智能运输系统 车辆前向碰撞

预警系统性能要求和测试规程》中环境要求为测试应在干燥、平坦的沥青路面或者水泥混凝土路面上进行，测试可在日光条件下进行等，均只有定性说明而无量化指标。

- c) 国际国内实际情况及标准存在不一致。由于国际国内实际情况不一致，因此直接采标国际标准，可能不符合我国的实际情况。如，国内自动驾驶功能测试车速选取一般以 5km/h 递增，采标标准直接通过换算的车速规定不适应我国的实际情况，如：GB/T 26773-2011 《智能运输系统 车道偏离报警系统性能要求与检测方法》，该标准非等效采用 ISO 17361: 2007，但 4.3.2 操作要求中，规定 I 类系统应在车速大于等于 20m/s 时正常运行，20m/s 直接来源于国际标准规定的 72km/h，显然不符合我国实际情况，导致车速数据难以采集。

3.1.3 我国道路交通及基础设施标准

与自动驾驶系统测试相关的标准还包括测试场地建设和道路基础设施标准等，如道路建设（道路标志、标线、交通灯、停车位等）、路侧智能基础设施（路侧激光雷达、摄像头等）、封闭场地建设、封闭场地测试装备具要求、通信设施相关标准；自动驾驶导航地图、道路测试监管等相关标准，如下所述：

- a) 关于道路基础设施建设，交通运输部制定发布的国家标准 GB 5768 《道路交通标志和标线》是我国交通工程领域覆盖面最广的基础性标准，包含总则、交通标志、交通标线、作业区、限制区域、铁路道口、行人和非机动车、学校区域等 8 部分的完整体系。其中，一至三部分为通用规定，规定了基本原则及各种交通标志和标线的颜色、形状、图案和设置要求。交通运输部在 2020 年 4 月 26 日发布的行业标准《公路工程适应自动驾驶附属设施总体技术规范（征求意见稿）》针对高精度地图、定位设施、通信设施、交通感知设施、路侧计算设施、自动驾驶监测与服务中心等公路附属

设施提出了总体技术要求,对于新基建和智慧公路建设将起到支撑性作用。此外,我国住房和城乡建设部发布的行业标准 JGJ 100-2015《车库建筑设计规范》。

- b) 全国道路交通安全管理标准化技术委员会(SAC/TC576),已发布公安首个车联网行业标准 GA/T 1743-2020《道路交通信号控制机信息发布接口规范》,规定了面向车联网服务应用类信息的交互接口与协议,规范了交通信号机与路侧网联设施的信息交互规范,对指导我国车联网交管路侧设施建设具有重大意义。此外,正在制定的国家标准包括《智能网联汽车运行安全测试环境技术条件 第1部分 公共道路》《智能网联汽车运行安全测试技术条件》,行业标准《智能网联汽车运行安全封闭场地测试场景要素及设置要求》《智能网联汽车运行安全半开放道路测试场景要素及设置要求》《智能网联汽车运行安全公共道路测试场景要素及设置要求》等。
- c) 全国智能运输系统标准化技术委员会(SAC/TC268)制定的国家标准《自动驾驶封闭测试场地建设技术要求》已发布,对封闭场地道路主体,交通设施等进行了较为详细的规定。此外,与智能驾驶电子道路图数据模型与表达相关的国家标准《智能运输系统 智能驾驶电子道路图数据模型与表达 第1部分:封闭道路》《智能运输系统 智能驾驶电子道路图数据模型与表达 第2部分:开放道路》已经报批。
- d) 智慧城市基础设施与智能网联汽车协同发展试点工作办公室(简称“双智试点办”),2022年2月11日在北京组织召开双智试点标准工作座谈会,会议提出由北京、上海、广州、武汉、长沙、无锡等第一批试点城市牵头六个标准导则编写工作,促进16个城市互认互用并不断完善,为其他城市提供可借鉴可复制的经验成果。

自动驾驶道路交通及基础设施相关标准的整体情况见表3.3。

表 3.3 自动驾驶道路交通及基础设施相关标准整体情况

标准编号	标准名称	归口组织
GB 5768.1-2009	《道路交通标志和标线 第1部分：总则》	交通运输部
GB 5768.2-2022	《道路交通标志和标线 第2部分：道路交通标志》	交通运输部
GB 5768.3-2009	《道路交通标志和标线 第3部分：道路交通标线》	交通运输部
GB 14886-2016	《道路交通信号灯设置与安装规范》	公安部
GB 14887-2011	《道路交通信号灯》	公安部
JGJ 100-2015	《车库建筑设计规范》	住房和城乡建设部
行业标准	《公路工程适应自动驾驶附属设施总体技术规范》（征求意见稿）	交通运输部
国家标准	《智能网联汽车运行安全测试环境技术条件 第1部分 公共道路》（制定中）	TC576
国家标准	《智能网联汽车运行安全测试技术条件》（制定中）	TC576
GA/T 1743-2020	《道路交通信号控制机信息发布接口规范》	TC576
/	《智慧灯杆网联化系统技术与工程建设规范》	双智试点办-北京
/	《充电设施网联化应用的技术接口要求》	双智试点办-上海
/	《自主代客泊车停车场建设规范》	双智试点办-广州
/	《智慧城市全息感知道路建设导则》	双智试点办-无锡
/	《车域网平台感知设备接入技术要求》	双智试点办-武汉
/	《智慧公交智能网联基础设施建设规范》	双智试点办-长沙
国家标准	《智能运输系统 智能驾驶电子道路图数据模型与表达 第1部分：封闭道路》（报批稿）	TC268
国家标准	《智能运输系统 智能驾驶电子道路图数据模型与表达 第2部分：开放道路》（报批稿）	TC268

随着自动驾驶技术发展理念的变化，尤其 L2 级以上自动驾驶功能的实现非单车智能所能全部承担，需要交通基础设施智能化、网联化的协同发展。作为融合了多种先进技术的产物，自动驾驶技术处在快速迭代进化进程中。对于我国来讲，标准在推动和规范行业健康发展的同时，也存在一定问题，主要是：

- a) 目前的基础设施建设标准存在无法满足自动驾驶功能实现和测试的情况。如，JGJ 100-2015《车库建筑设计规范》中，规定小型车平行式车位尺寸为 6.0m*2.1m（停车位毗邻时最小停车位宽度应为 2.4m）：垂直式车位尺寸为 2.4m*5.1m（停车位毗邻时最小停车位长度应为 5.3m）。而 GB/T

41630-2022《智能泊车辅助系统性能要求及试验方法》规定自动泊车功能测试所需的平行式车位尺寸为 6.0m*2.5m。我国交通运输部发布的 GB 5768.3-2009《道路交通标志和标线 第 3 部分：道路交通标线》中规定的平行式停车位标线尺寸为 6.0m*2.5m（宽度受限情况时宽度最小不应低于 2.0m）。其中，宽度受限时无法满足自动泊车功能的需要。

- b) 而且，标准中无法覆盖实际驾驶环境的复杂情况，如：基础设施标准在具体应用场景下，会经历不同定义、或前后标准混合出现的场景。例如，对于一段反复修路并更改车道线和箭头标识的路段，因为前后标识的叠加混合，就会给场景数据的采集和定义带来更多的不确定性。

3.1.4 政策和标准总结

我国幅员辽阔，国内复杂多变的自然环境、道路场景，无疑是对自动驾驶汽车性能严峻的挑战。对于自动驾驶车辆而言，相关的感知、通讯、规控算法等需要适应特殊气候和路面条件才能实现安全行驶。现阶段自动驾驶测试标准对自动驾驶车辆在地区适应性、环境适应性方面还未有明确的要求。

因此，需要结合地方政策、ODD 及实际交通场景，总结梳理出覆盖我国地貌及道路特色的测试场景、危险场景及扩充场景。扩充场景主要满足地区适应性需求，如冰雪环境、热带环境、山地环境、特殊光照环境等。

3.2 测试场景标准发展现状

3.2.1 国际自动驾驶系统测试场景标准发展现状

2018 年 4 月，我国向国际标准化组织道路车辆委员会 (ISO/TC22) 提出自动驾驶测试场景国际标准提案，获批组建自动驾驶测试场景工作组 (ISO/TC22/SC33/WG9) 并担任工作组召集人，联合德国、日本、英国、荷兰、美国等二十余个国家的专家，围绕自动驾驶测试场景共同规划了一系列国际标准项目。此外，自动化及测量系统

标准协会 ASAM 也制定了仿真测试相关的标准。

- a) ISO/IEC TC 22/SC 33 相关标准规定了自动驾驶系统测试的场景相关的《道路车辆 自动驾驶系统测试场景》系列标准。包括：
 - 1) ISO 34501:2022 “词汇”明确了场景、动态环境及实体要素之间的关系，与功能安全、预期功能安全等国际标准建立了配套关系。该标准形成了全球统一的自动驾驶测试场景相关术语“字典”。
 - 2) ISO 34502:2022 “基于场景的安全评估框架”给出了在产品开发过程中应用的基于场景的安全评估过程，侧重场景生成技术，为自动驾驶系统基于场景的安全评估框架提供指导。
 - 3) ISO 34503 “设计运行域规范”规定了用于指定自动驾驶系统(ADS) 设计运行域(ODD)定义的运行条件的层次分类法要求以及使用分类法定义 ODD 的定义格式要求，侧重 ODD 特征描述。
 - 4) ISO DIS 34504 “场景分类”给出对自动驾驶系统场景进行分类的方法，即通过提供带有场景信息的标签来定义场景，用于指导 L3 至 L4 自动驾驶系统。
 - 5) ISO WD 34505 “场景评估和测试用例生成”规定了自动驾驶系统测试场景的评价方法，提出测试场景关键度、复杂度、危险度等功能性评价指标的判定要求，并定义测试用例生成的一般性方法及其必要特征，包括测试目标、测试步骤、执行条件和预期结果等。
- b) 自动化及测量系统标准协会（Association for Standardization of Automation and Measuring Systems, ASAM）组织为解决仿真测试中各环节的数据、接口格式等不统一带来的问题，制定 ASAM OpenX 标准，创建一个工程、模拟、测试和自动化环境，其中：
 - 1) 数据管理和分析的标准有 ASAM CEA、ASAM ODS，是存储、检索和分析车辆模拟、测试、生产和运行过程中采集的大量数据的标准。

2) 仿真测试标准有 ASAM penDRIVE、OpenCRG、OpenLABEL、OpenODD、OSI，用于驾驶和交通模拟的道路网络描述及驾驶操作和测试场景规范。仿真测试场景道路网络部分（如道路拓扑结构、交通标志标线等）由 OpenDRIVE 文件描述，道路的表面细节（如坑洼、卵石路等）由 OpenCRG 文件描述，仿真测试场景的动态部分（如交通车的驾驶行为）由 OpenSCENARIO 文件描述。

综上，国际标准从宏观层面对自动驾驶测试场景的术语、基于场景的安全评估、ODD 特征描述、场景分类、场景分类方法的相关内容进行了介绍，但并未涉及具体的测试场景集。

3.2.2 国内自动驾驶系统测试场景标准发展现状

国内研究自动驾驶系统测试场景的相关标准组织包括 TC114/SC34、TC28/SC42 自动驾驶研究组、CCSA、中国汽车工程学会、中国汽车工业协会、中电标协、中关村智通智能交通产业联盟、中国道路交通安全协会、中国公路学会、上海市标准化协会、国家智能网联汽车产业创新联盟—智能网联汽车测试示范工作组等。制定的自动驾驶测试场景相关国家行业标准包括：通用自动驾驶测试场景术语、特定仿真测试场景、V2X 自动驾驶场景、道路基础地理数据规范等。具体见表 3.4。

表 3.4 国内自动驾驶测试场景数据采集和分析相关标准

序号	标准编号	标准名称	标准组织	主要内容
1	国家标准	道路车辆 自动驾驶系统测试场景术语	TC114 SC34	定义了自动驾驶系统测试场景中的术语。
2	国家标准	道路车辆 自动驾驶系统测试场景基于场景的安全评估框架	TC114 SC34	给出了在产品开发过程中应用的基于场景的安全评估过程，为自动驾驶系统基于场景的安全评估框架提供指导。
3	2022-1325T-SJ	人工智能 自动驾驶系统仿真测试场	TC28/SC42	规定了人工智能领域自动驾驶系统无人配送车部分的仿真测试场景描述要求、场景集分类和仿

		景要求 第 1 部分： 无人配送车		真测试场景数字格式要求。
--	--	----------------------	--	--------------

表 3.4 国内自动驾驶测试场景数据采集和分析相关标准（续）

序号	标准编号	标准名称	标准组织	主要内容
4	2022-1326T-SJ	人工智能 自动驾驶系统仿真测试场景要求 第 2 部分：小型客车	TC28/SC42	规定了人工智能领域自动驾驶系小型客车部分的仿真测试场景描述要求、场景集分类和仿真测试场景数字格式要求。
5	YD/T 3977-2021	面向 5G V2X 自动驾驶场景和数据集	CCSA TC10 WG5	规定了增强的 V2X 业务应用层交互数据要求，包括增强的 V2X 业务应用架构及其组成，增强场景及业务流，增强业务对通信系统要求，以及增强业务应用层交互数据要求。
6	2022-1348T-YD	基于移动互联网的车路协同应用场景及技术要求	CCSA TC10 WG5	面向 4G、5G 移动智能终端开展基于移动互联网的车路协同应用场景，提出基于移动互联网的车路协同系统架构、适用于移动互联网的车路协同应用数据交互内容。

此外，场景采集和分析还发布了一些团体标准，相关内容涉及采集或测试场景、采集工具、采集要求和要素、数据标注、数据格式和质量等。团体标准由社会团体组织成员单位制定，在本团体内协调一致即可，因此需要制定国家标准，以形成更高范围的协调一致。不同团体标准之间的不一致性如下举例说明：

- a) T/ZSA 40—2020 《自动驾驶仿真测试场景集要求》和《人工智能 自动配送车自动驾驶系统 仿真测试场景定义和要求 第 1 部分：城市道路》两个标准都规定了场景基本信息，但字段范围、字段名称、数据类型均存在不同，见表 3.5；
- b) 数据采集要求不统一，如《智能网联汽车道路测试与示范应用道路交通事故信息采集技术规范》中车辆驾驶模式、速度、加速度等采集频率是不低于 50Hz，而《自动驾驶测试车辆事故责任数据采集及分析技术要求（征求意见稿 2021 年 7 月）》中车辆驾驶模式采集频率大于 1Hz，速度和加速度采集频率大于 10Hz 等；

-
- c) 中国汽车工程学会发布的《智能网联汽车 自动驾驶地图采集要素模型与交换格式》中对测试场景的分类为：室外道路场景和公共停车场道路场景。
- 场景要素分为 5 个图层：场景基本信息层、道路交通标志层、道路交通标线层、智能路侧设备层、其他道路安全设施物层。

汽标委智能网联汽车分会

表 3.5 场景基本字段要求

字段		定义		类型		参考取值范围	
标准 A	标准 B	标准 A	标准 B	标准 A	标准 B	标准 A	标准 B
ADS__NAME	Name	场景名称	场景名称	string	string	——	——
ADS__TYPE	Type	场景类型	场景类型	cnum	cnum	0: 路侧开环数据;1: 路侧闭环数据 2: 人工创景场景;3: 危险驾驶场景 4: 自定义	Logsim、Wordsim
ADS__ID	ID	场景 ID	场景编号	int	int	0-99999999	——
/	Version	/	场景版本	/	string	/	——
PRIOPRITY	/	场景优先级	/	cnum	/	Low: 低优先级;High: 高优先级; Middle: 中优先级	/
STATUS	/	场景状态	/	cnum	/	Online: 已上线;Offline: 已下线 Pending: 处理中	/
/	MapName	/	地图名称	/	string	/	——
/	MapForm	/	地图格式	/	string	/	——
MAP__VERSION	MapVersion	地图版本信息	地图版本	string	string	——	——
MAP__REGION	/	场景所在地图的区域信息	/	string	/	/	/
/	CarName	/	车辆名称	/	string	/	——
/	Tag	/	标签	/	string	/	无
/	Level	/	级别		Cnum	/	Critical、Major、Minor
CREATE__TIME	/	场景创建时间	/	date	/	yyyy-MM-dd HH:mm:ss	/
CREATE__User	Creator	创建用户	制作人	string	string	——	——

表 3.5 场景基本字段要求（续）

字段		定义		类型		参考取值范围	
/	Decription	/	描述		string	/	——
CONTROL__ MODE	/	控制模式选择	/	cnum	/	0: 完美控制模式 1: 动力学控制模式 2: 自定义	——
PERCEPTION__ MODE	/	感知模式选择	/	cnum	/	0: 完美感知模式 1: 传感器模型模式 2: 自定义	——
SIMU__TIME	/	模拟仿真运行时间	/	int	/	0-9999s	——

注 1: 标准 A 代表 T/ZSA 40—2020 《自动驾驶仿真测试场景集要求》

注 2: 标准 B 代表《人工智能 自动配送车自动驾驶系统 仿真测试场景定义和要求 第 1 部分: 城市道路》

3.2.3 采集及分析标准总结

自动驾驶测试场景是开展自动驾驶汽车测试评价的重要前提，可以有效地复现实际情况下各种可能场景，从而测试自动驾驶系统的决策控制能力，发现问题，降低风险提高开发效率。然而现在自动驾驶场景采集和分析标准存在如下问题：

- a) 在国家标准层面，场景数据采集和分析标准缺失；
- b) 路端和车端采集的数据兼容问题未形成统一解决方案；
- c) 数据采集设备缺少统一要求。

由于存在上述问题，降低了自动驾驶测试的效率和质量，增加了测试成本和风险，限制了自动驾驶系统的性能和安全性的评价和验证，影响了用户的信任和接受度，阻碍了自动驾驶技术的创新和交流，造成了技术壁垒和信息孤岛。场景采集和分析国家标准的制定，可以形成统一的测试场景数据标准，指导自动驾驶测试场景数据采集，同时也进一步支撑自动驾驶测试的标准化，规范化。后续建议丰富自动驾驶系统测试场景数据标准，统一数据采集、数据标注、数据存储、数据质量要求。

3.2.4 术语定义

- a) 功能场景 functional scenario

驾驶场景在概念层面上用自然语言描述，通常没有特定的物理值。

[来源：ISO 34501，3.22，有修改]

- b) 逻辑场景 logical scenario

驾驶场景描述，包含参数，其中某些参数的值定义为范围。

[来源：ISO 34501，3.24，有修改]

- c) 具体场景 concrete scenario

驾驶场景用显式参数值描述，描述物理属性。构成一个具体场景的多个显示参数能够提取作为测试用例（3.5）的输入、执行条件、执行步骤和预期结果。

[来源：ISO 34501，3.25，有修改]

d) 测试场景 test scenario

用于测试和评估驾驶辅助或自动驾驶系统的场景，在描述上与功能场景均采用自然语言描述，通常指一类场景，每一类场景包含多个测试用例。

e) 测试用例 test case

为了实现特定的测试目标而开发的包含输入、执行条件、执行操作和预期结果（通过/失败标准）等内容的组合。这些测试目标是为了执行特定的程序/架构（软件单元、车辆级别等），来验证是否符合特定要求。一个具体的测试场景，各类要素应是可量化的、可执行的，并且具备拟真性。

f) 设计运行域 operational design domain; ODD

驾驶自动化系统设计时确定的适用于其功能运行的外部环境条件，典型的外部环境条件有道路、交通、天气、光照等。

[来源：GB/T 40429—2021，2.11，有修改]

3.3 测试场景采集及分析标准化的意义

测试场景数据库的建设是一项庞大的工作，测试场景数据贯穿自动驾驶产品开发、车型测试、品质评价、示范运营、商业化运营等产业发展全过程。但国内外场景采集设备要求和数据格式不统一，采集分析技术及所建场景等公开性及标准化远远不足，已制定的国内外标准尚不能支撑场景库高质量搭建，为了进一步推进场景库建设，必须开展相关标准化技术研究及建设。场景库相关标准化可带来以下积极作用：

- a) 技术层面，标新聚力，通过创新场景库标准建设，持续打破**国外关键技术垄断“卡脖子”**，攻克不同采集及分析技术方案之间的壁垒，实现国内外采集平台数据的互通和共享，并将共性要素进行提取、分析和统一，实现数据兼容，**避免形成数据孤岛**。另外，规范化的场景数据有利于促进行业快速完善测试场景数据库的标准化建设，从而提升自动驾驶产品研发测试

能效；

- b) 应用层面，准势启航，用标准引领催生更为科学高效的数据库产业化生态模式，快速推进场景库在产品工程目标、控制策略优化、仿真测试、验证与认证、产品准入及智能网联试验场建设等重大方面的应用，同时通过标准化建设可高效推进国内外相关行业共建共享，有效降低各家海量数据大范围采集带来的巨额成本，有效缩短各自开展大数据分析技术的研究周期，全面提高测试场景的利用率和单位价值，深度为智能网联汽车发展及标准体系建设赋能。

第四章 自动驾驶数据采集技术

4.1 概述

自动驾驶数据采集是指通过车辆搭载的传感器、摄像头、雷达等设备收集车辆周围环境、车辆状态和驾驶行为等数据。这些数据可以包括图像、视频、激光雷达扫描、惯性测量单元（IMU）数据、GPS 位置信息等。汽车数据采集的目的是提供实时、准确的车辆和环境数据作为自动驾驶系统的输入，用于感知、决策和控制。汽车制造商和技术公司对车辆数据的收集和分析日益重视，以提供更好的驾驶体验、车辆安全性和车辆维护。自动驾驶技术的兴起为数据采集和分析提供了更大的需求。此外，车辆与其他智能设备和系统的连接性增强，也促使了汽车数据采集需求的增长。当前数据采集行业面临一些新的痛点和挑战，包括以下方面：

- a) 数据标准化和互操作性：由于不同汽车制造商和技术供应商采用不同的数据格式和接口标准，导致数据采集的互操作性问题。这使得数据采集和整合变得复杂，限制了数据的有效利用和共享。
- b) 数据质量和准确性：汽车数据采集过程中可能面临数据质量和准确性的挑战。传感器的误差、噪声和故障可能导致数据不准确或不完整。数据采集系统需要具备高精度和可靠性，以确保采集到的数据质量达到要求。
- c) 处理和存储的挑战：随着数据量的增加，对数据的处理和存储也带来了挑战。大规模的数据处理需要高性能的计算和存储资源，以支持实时分析和决策。此外，长期存储和管理大量数据也需要解决数据存储和备份的问题。

解决这些痛点和挑战急需统一采集流程、制定更统一的数据标准、改进数据质量和准确性、提供高效的数据处理和存储解决方案。

4.2 采集流程

数据采集流程见图 4.1，如下所述：

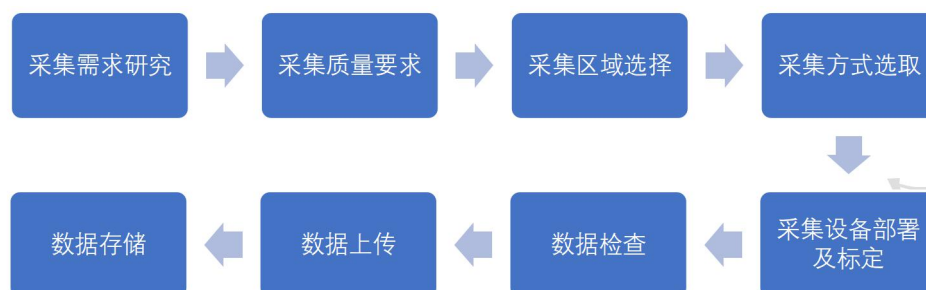


图 4.1 数据采集流程

- a) 采集需求研究：基于采集需求，确定环境、本车、道路及交通设施、其他交通参与者等参数确定的数据类型；
- b) 数据质量要求：主要包括空间配准质量要求和数据采集质量要求，数据采集质量要求涵盖数据完整性、数据可用性、数据精确性、数据稳定性、数据同步性、数据唯一性及数据安全性等要求；
- c) 采集区域选择：主要基于不同地理位置、道路类型、气候环境（季节/天气/时段）、交通复杂度及动态交通参与者等因素，分析不同目标区域内公共道路的类型与特征，同时结合道路测试经验、交通数据分析等方式来选择采集区域和路线，以实现更短的路线覆盖更多的场景；
- d) 采集方式选取：主要包含实车采集、无人机和路侧采集三种常见的数据采集方式；
- e) 采集设备部署及标定：基于实车采集、无人机和路侧采集等三种常见的数据采集方式进行采集设备部署，标定分为时间同步及空间同步；
- f) 数据检查：分为数据通用性检查及数据同步性检查等要求；
- g) 数据上传：涵盖车端传输要求及路侧设备传输要求；
- h) 数据存储：车辆采集端存储应需要综合考虑数据完整性与安全性、数据质量与访问控制等因素，以确保存储数据的完整性和安全性、硬件设备可持续使用性及隐私合规性。

4.3 采集需求研究

在自动驾驶数据采集技术中，采集需求研究主要是指对于自动驾驶所需数据的种类、数据来源和数据量的研究，采集需求研究是一个复杂且关键的环节，需要综合考虑各种因素来确保采集到的数据能够满足自动驾驶系统的需求。对于自动驾驶需要收集的数据类型，包括环境、本车、道路及交通设施、其他交通参与者等，部分数据详见表 4.1。

表 4.1 场景数据采集内容

数据类型	数据名称
环境	天气
	气温
	PM2.5
	光照
	风速
本车	方向盘转角
	方向盘转角速度
	自车横向速度
	自车纵向速度
	发动机转速
	档位
	航向角
	
道路及 交通设施	道路名称
	道路等级
	路段路面
	路段曲率半径
	路段纵向坡度
	路口类型
	匝道形状
	桥涵类型
	车道宽度
	车道数量
	车道线类型
	

表 4.1 场景数据采集内容（续）

数据类型	数据名称
交通参与者	目标物类型
	目标物纵向距离
	目标物横向相对速度
	目标物纵向相对速度
	目标物横向相对加速度
	目标物纵向相对加速度
	目标物最近点横向距离
	目标物最近点纵向距离
	

4.4 数据质量要求

4.4.1 空间配准质量要求

多传感器场景数据采集，要求相同场景不同传感器采集的目标，在空间上完全配准，确保场景数据目标定位准确、标注正确。根据场景数据空间配准的方式，规定场景数据目标空间配准精度要求。

集成标定是依据不同传感器的成像特性，布设相应的靶标完成目标的空间配准，因此可以获得较高的目标定位精度，通过布设的控制点评定精度要求如下：

- a) 摄像头：使用标定场光学图像靶标评定，平面位置中误差不大于 10cm；
- b) 激光雷达：使用标定场激光雷达靶标评定，距离向中误差不大于 10cm，垂直距离向中误差不大于 15cm；
- c) 毫米波雷达：使用标定场毫米波雷达靶标评定，目标位置中误差不大于 30cm。

4.4.2 数据采集质量要求

4.4.2.1 数据完整性

场景数据采集完整性包括采集范围的完整性与场景数据描述的完整性。其中：

采集范围的完整性需要考虑不同地理位置、道路类型、气候环境、交通复杂度及参与者类型、季节效应、采集时段、采集人员及采集车辆类型等因素的影响；场景数据描述的完整性是指采集过程要完整记录影响测试场景的各个参数，从采集的参数类型、参数精度、采样频率、同步性等方面确保数据的完整性。

场景数据采集需要针对自动驾驶的设计功能及应用用例，完整还原测试场景元素及其演进的过程，包括道路条件、气候环境条件、交通参与者数量类型、驾驶员行为等。

采集的数据应包含视频、报文、时间戳、真值等文件，保证数据完整，不可缺项。在数据采集过程中，由于传感器故障、数据传输中断、存储设备故障等原因，数据可能会丢失或不完整，导致后续数据分析和应用时出现问题，因此要保证原始数据完整。

数据采集的内容需包含本车行为数据、本车传感器感知数据（其他参与者动态数据）、道路环境数据三部分，如下所述：

- a) 本车行为数据包括本车 GPS 信号、车速、加速度（x/y/z）、方向盘转角、发动机转速、档位、油门踏板位置、刹车踏板位置、转向灯、驾驶员状态和动作；
- b) 本车传感器感知数据包括原始数据和融合后的数据，其中原始数据是传感器的原始输出数据，例如激光雷达原始点云、相机原始图像等；融合后的数据包括其他交通参与者及车道线的信息，包含交通参与者的 ID、外形尺寸、速度、加速度（x/y/z）、与本车的距离等信息以及车道线的 ID、宽度、类别、曲率及车身与行车线的距离等信息；
- c) 道路环境参数包括交通信息、道路类型信息、环境参数。其中交通信息包括路面材质信息、车道数信息、限速标志、交通拥堵情况；道路类型信息包括高速道路、省干道、县道、乡村道路、园区道路、沙石路等；环境参数包括：天气信息（晴、雨、雾、雪、冰雹等）、气温、PM2.5 等。

4.4.2.2 数据可用性

采集的数据应有效可用，所有数据可打开、有内容、可运行，并且是本次训练或测试所需要的数据。实车采集时，司机和操作员的不规范操作是自动驾驶数据采集中的一个常见问题，而这些操作可能会导致数据损坏、遗漏或不准确，从而影响数据的质量。例如，司机可能会在行驶过程中不按照预设路线行驶，或者操作员可能会在采集过程中误操作，导致数据不准确，因此，需要对采集数据时人员的行为进行规范，确保数据可用。

此外，在采集数据时，应检查车辆集成的每个传感器是否采集到有效数据，前向、后向等主传感器必须连续记录有效数据，否则采集的数据是无效的；传感器在使用前也需要进行校准，包括内部参数校准和外部参数校准，如果传感器校准不准确，会影响到后续的数据采集和应用；如果有重叠视场的单个传感器数据异常，也应检查确定是否录用所采集的数据，并将详细问题进行记录。

4.4.2.3 数据精确性

数据采集时应根据测试功能与测试场景构建流程等确定数据采集的精度要求，确保精度满足测试需求。采集数据的精度可能影响到构建测试场景的精度，最终影响到测试场景的准确性。

在数据采集场景中，分别任意抽选每个工作日的起始段、中间段、结束段，包含易于提取静态和动态目标的场景，测量动态和静态目标的位置重合精度，精度应至少满足：

- a) 摄像头在成像距离 50m 范围内，平面位置中误差不大于 15cm；
- b) 激光雷达在成像距离 50m 范围内，距离向中误差不大于 15cm，垂直距离向中误差不大于 20cm；
- c) 毫米波雷达在成像距离 50m 范围内，目标位置中误差不大于 50cm。

为了数据精确性，通常也会针对不同性质、不同数据类型的传感器，选择合适的匹配和融合方式，对多源传感器的数据进行处理。

4.4.2.4 数据稳定性

采集数据应不受环境、光照、天气等因素的影响，质量、周期等要稳定无大幅波动。例如，在采集视频时，一般规范采集视频的帧率为 30fps，同时还要保证帧率稳定性大于等于 90%，只有足够稳定的数据才能用于后续模型训练与台架回放测试。

4.4.2.5 数据同步性

不同传感器采集的数据应有统一的对齐方式，达成同步。目前传感器、CAN 与采集工控机一般使用 PTP、NTP 或 PPS 等方式统一时钟源，进行时间同步，从而达到数据同步的效果。

在采集数据过程中，为了确保时间同步，避免时间戳不对齐、时间戳异常等问题，样本检查时应重点检查以下信息：

- a) REC 时间戳个数 = 视频帧数；
- b) REC 时长 = 视频时长；
- c) REC 文件 ID 与视频文件 ID 对应；
- d) CAN/激光雷达/RT-Range 与 REC 时间戳单位一致；
- e) $|\text{CAN/激光雷达/RT-Range 时长} - \text{REC 时长}| \leq 2\text{s}$ ；
- f) $|\text{CAN/激光雷达/RT-Range 起始时间戳} - \text{REC 起始时间戳}| \leq 2\text{s}$ ；
- g) CAN/激光雷达/RT-Range 时间戳周期稳定性 $\geq 90\%$ ；
- h) CAN/激光雷达/RT-Range/REC 时间戳：后一帧 - 前一帧 > 0 。

4.4.2.6 数据唯一性

数据唯一性是指各传感器采集的数据都有绝对唯一的格式与命名方式，不可混淆。例如，针对不同类型传感器采集的视频，可以赋予其不同的 ID，并在文件命名时以采集结束时间+视频 ID 进行区分，如表 4.2 所示。

表 4.2 视频数据存储唯一性

数据类型	视频
格式	avi（支持 YUV 格式）
文件 ID	前广：0 前窄：1 左侧前：2 右侧前：3 左侧后：4 右侧后：5 环视：6 后视：7
举例	2023 年 9 月 11 日 9 点 16 分 28 秒录制的前广视频：20230911091628_0.avi

对于环视而言，又分前、后、左、右四路，可再添加数字对视频文件加以区分：前路 6_1、后路 6_2、左路 6_3、右路 6_4。以此类推，对不同类型不同场景采集的视频文件 ID 进行对应命名，可以保证数据唯一性，清晰明了，避免后续应用时产生混淆。

4.5 采集区域选择

场景则是支撑智能网联汽车测试评价的核心要素与关键资源，测试场景数据的主要来源为道路自然驾驶场景，为满足高等级智能网联汽车测试对于场景的需求，采集自然驾驶场景数据时应遵循全面覆盖、重点采集的原则，而我国不同地区的道路条件、路网分布、地形分布、气候条件、交通环境以及驾驶员驾驶习惯等因素差异较大，对采集区域及道路有效选取提出了较高的要求。实车采集过程中，需要尽可能覆盖各种行驶场景。根据场景数据采集需求，分析不同目标区域内公共道路的类型与特征，同时结合道路测试经验、交通数据分析等方式来选择采集区域和路线，以实现更短的路线覆盖更多的场景。

自然驾驶场景采集需求应考虑不同地理位置、道路类型、气候环境（季节/天气/时段）、交通复杂度及交通参与者等因素，如图 4.2。

道路环境包括天气、光照、道路类型、路面状态以及道路拥堵等，如下：

- a) 天气：晴天、多云、雨天、雾天、雪天等；
- b) 光照：白天、夜晚、清晨、黄昏等；
- c) 道路类型：高速/高架、城区道路、乡镇道路、施工道路、隧道等；

- d) 路面状态：干、湿、雪等；
- e) 道路拥堵：正常、拥堵。

道路目标是指自动驾驶功能要求传感器探测的目标类型，包含静态目标和动态目标。

静态目标包括以下：

- a) 交通标志：限速标志、匝道标志、停止标志灯等；
- b) 交通灯：车辆交通灯、行人交通灯等；
- c) 地标：锥桶、圆筒、隧道入口、龙门架、升降门等；
- d) 车道线：实线、虚线、单线、双线、反光线等；
- e) 路边沿：路肩、护栏、草坪等；
- f) 路面标记：箭头、人行横道、停止线等。

动态目标则包含：车辆、行人、两轮车和动物等。

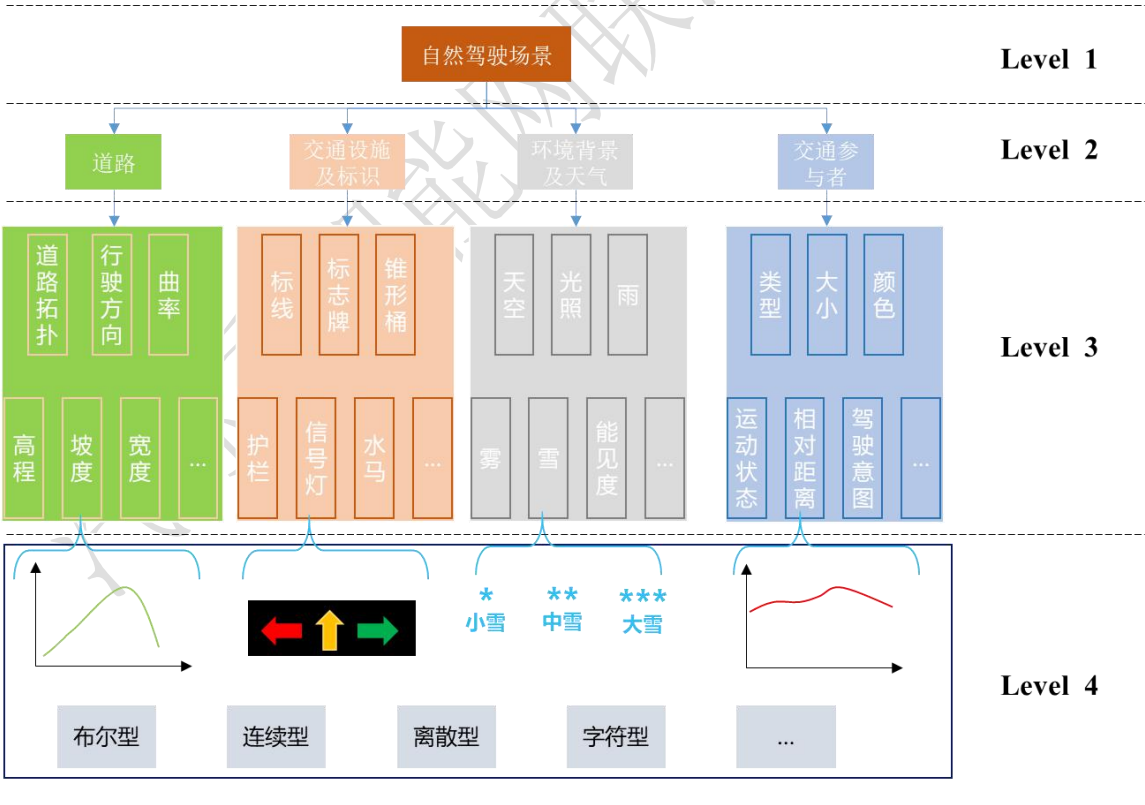


图 4.2 场景参数分级示意图

4.6 采集方式选取

4.6.1 实车采集方式

4.6.1.1 实车配置方案

基于实车的采集方式是指车辆搭载摄像头、激光雷达、毫米波雷达、定位设备、车辆网联设备以及车辆基础传感器数据采集设备等，由数据采集人员驾驶，在驾驶过程中采集道路环境信息，车辆行驶状态、交通参与者等信息的方式。

根据数据采集车车型尺寸，配置方案分为大型车辆配置方案和小型车辆配置方案，区别在于大型车辆配置方案加强了对车辆两侧盲区范围的探测，如图 4.3 所示。

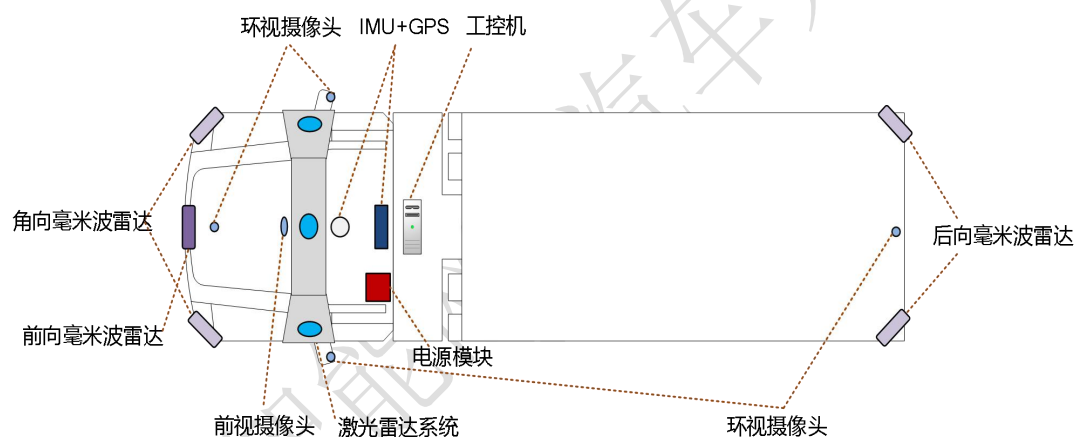


图 4.3 大型车辆传感器配置方案

大型车辆包括大型客车（载客人数 8 人以上）和大型货车，这类车辆的特点是，外形尺寸较大，车顶平坦宽阔，若使用单一的顶置激光雷达，会导致车辆周围有较大的盲区。因此，大型车辆选用 3 合 1 的激光雷达系统方案，增加两侧斜置盲区探测激光雷达，将点云数据拼接后输出，补全视野范围。

小型车辆配置方案如图 4.4 所示，其采用单一顶置激光雷达，并在部分车辆上加装智能摄像头传感器（如 Mobileye Q3），在原有的场景数据采集的基础上，增加目标级数据输出，可用于快速提取场景目标信息，构建场景模型。

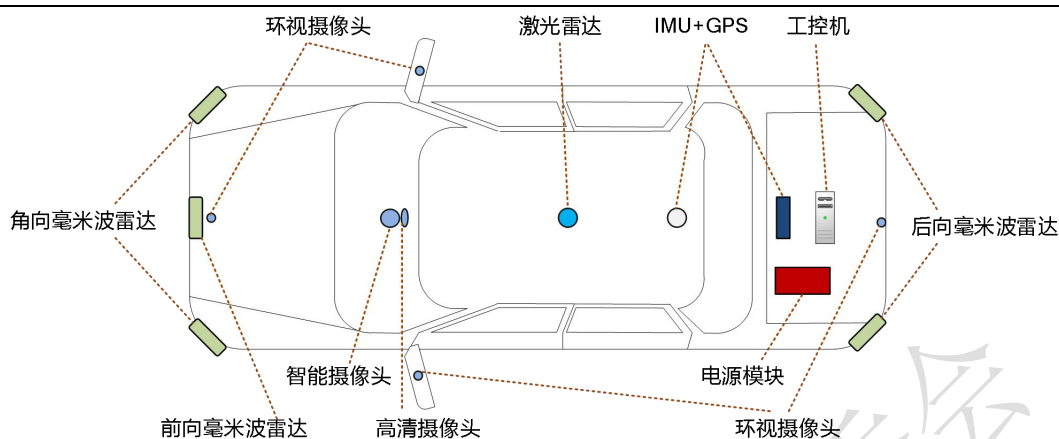


图 4.4 小型车辆传感器配置方案

4.6.1.2 实车采集系统

为了同时处理多种类型传感器采集到的数据，满足实车功能测试与数据同步采集的需求，多模块实车采集系统应运而生。

多模块实车采集系统分为“数据采集”与“功能测试”两大部分，通过数据分流的方式将多路传感器信息（视频、超声波雷达、毫米波雷达、IMU、RTK、车身 CAN）和真值系统信息（激光雷达、RT-Range）分别传送至“数据采集”与“功能测试”系统，实现原始数据采集与自动驾驶功能测试两个模块同步运行。当功能测试出现异常情况时，系统可同步将该场景对应的原始数据进行标记，作为重点场景进行后续数据分析。东软睿驰使用的多模块实车采集系统拓扑如图 4.5 所示。

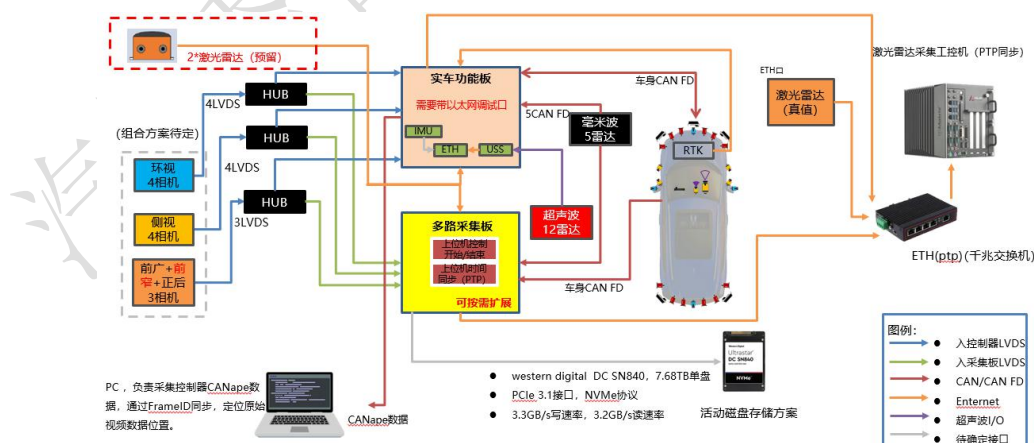


图 4.5 多模块采集系统拓扑图

4.6.1.3 驾驶员的选择

由于自动驾驶车端采集存在时间长、成本高、效率低等问题，会对采集人员的安全意识与业务能力有着极高的要求。因此本节将从安全意识与业务能力两个维度出发，对驾驶员的选取提出通用与专业两种要求。

为了保障采集过程的安全性，对驾驶员的通用要求如下：

- a) 有相应车型的驾驶证，并且驾驶经历在 3 年以上；
- b) 最近连续 3 个记分周期内没有满分记录，最近一年内没有超速 50%以上或闯红灯等严重违法行为；
- c) 从未有过酒驾记录、服用国家管制药品、麻醉药品记录，也从未有过致人死亡或重伤的交通事故责任记录。

只有经验丰富、驾驶稳健、身心健康的司机，才能有效应对采集难题。

为了提高采集效率，对驾驶员的专业要求如下：

- a) 经过自动驾驶系统培训，熟悉自动驾驶测试流程；
- b) 掌握自动驾驶测试设备操作方法，具备紧急状态下应急处置能力；
- c) 理解自动驾驶各功能定义，能够快速完成进入状态，完成功能场景采集；
- d) 对环境变化敏感，能够在复杂交通环境下筛选并记录特殊场景。

只有具备上述专业能力，才能提高采集效率，完成从“里程”采集到“场景”采集的蜕变。

此外，自动驾驶在很长一段时间内会存在人机共驾的情况，驾驶员的驾驶习惯在数据采集过程中尤为重要。在采集过程中，需要有不同驾驶风格的驾驶员参与。因此，驾驶员模型被用于构建自演绎仿真场景来测试自动驾驶汽车的连续决策能力。

人类的操纵决定了其驾驶车辆的行驶特征，不同驾驶风格会导致驾驶行为出现差异。面向不同道路的运行环境，按照人类驾驶员的经验、能力、驾驶习惯将驾驶风格分为以下三种并对其行为特征进行设定，总结如下：

- a) 保守型驾驶员：该模型代表自然驾驶环境中经验欠缺、能力较弱的一类驾驶员。此类驾驶员对交通环境信息和其他驾驶员行为的理解能力稍差，偏

向于做出保守的决策，以避免出现无法处理的事故和危险；

- b) 普通型驾驶员：该模型是自然驾驶数据的综合体现，代表了具有人类平均驾驶水平和经验的驾驶员。此类驾驶员最具典型性，对交通环境信息有较好的理解和分析能力，能够与其他同类型驾驶员进行正常交互并做出合理反应；
- c) 激进型驾驶员：该模型的经验和能力与普通型驾驶员模型相似，但由于驾驶习惯、性格、心情等原因倾向于做出激进的驾驶行为。

基于以上三种驾驶风格的模型，可以根据不同的特征进行分类，具体特征如下所述：

- a) 保守型：只进行纵向动作决策；
- b) 普通型：倾向于保持当前的行驶状态，会做出加减速或换道动作以保证行驶效率，在安全受到威胁时会采取急加减速动作；
- c) 激进型：倾向于做出有一定风险性的驾驶行为，较高的平均车速、平均加速度，较低的跟车车头间距，高频的急加/减速及换道动作。

在实车采集时，一方面可以根据以上三种驾驶员模型，综合考虑驾驶模型的分布情况，选择不同特征的驾驶员；另一方面还能根据车辆用户画像有侧重点地实现某种驾驶模型的数据采集。因此，在对招募到的驾驶员进行驾驶风格调查后，由测试工程师根据测试用例确定驾驶员模型，选用符合要求的驾驶员。

4.6.2 路侧设施采集方式

基于实车的采集方式采集设备均搭载于采集车上，视角低造成的视野遮挡以及传感器性能限制，无法得到除本车周围车辆以外的其他交通参与者的信息，这部分数据的缺失极大地限制了自动驾驶车辆对周围车辆的行为预测，以及对本车的路径规划，因此将数据采集设备固定至路侧高空位置进行数据采集可以有效解决上述数据缺失的问题。

按照目前行业的研究进展，车路协同路侧感知系统主要由路侧感知单元、数据

传输单元、路侧计算单元、附属配套设施等组成，系统组成架构见图 4.6 所示。

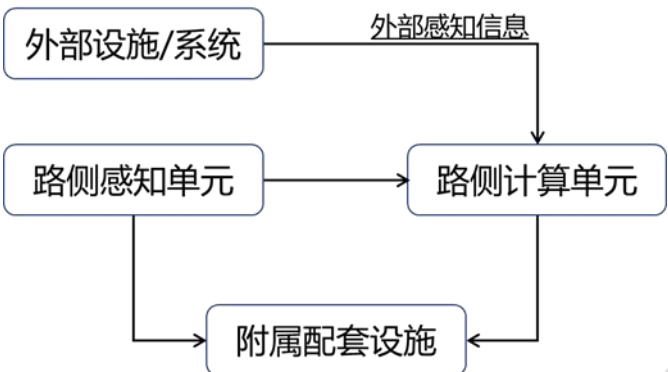


图 4.6 车路协同路侧感知系统组成架构

- a) 路侧感知单元：用于提取道路交通状态的各类要素，如交通参与者的运动学信息、判定交通事件触发的信息、计算交通流相关指标的支撑信息等，通常包括摄像机、毫米波雷达、激光雷达等交通检测器，以及包括用于道路状态监测等场景的传统交通检测器。考虑到路侧感知单元的选择与系统支持的应用场景有关，在实际应用中对路侧感知单元的具体组成不作限定；
- b) 外部设施/系统：用于为特定应用提供来自外部的感知信息，通常包括信号机、RSU、云平台、交通管控系统等外部设施或系统；
- c) 数据传输单元：用于系统组成设备之间以及系统与外部设备/系统进行通信，通常包括各类有线或无线通讯设施；
- d) 路侧计算单元：用于对路侧感知单元的原始数据或结果数据进行存储、融合处理分析，生成较高精度的感知结果信息；
- e) 附属配套设施：用于为系统提供部署、供电、时间同步、信息安全、气象感知等支撑服务的相关设备。系统授时源可来自上层网络授时，或本地卫星授时。

目前已广泛应用的传统路侧单元，一般仅仅包含摄像头，摄像头与毫米波雷达（测速）的组合，并不能准确且充分地获取形成自动驾驶场景所需的全部运动信息。因此，可利用多激光雷达拼接形成的连续感知能力，获取交通参与者的精确信息，如图 4.7 所示。激光雷达采集的原始数据为点云数据，点云数据经过 MQTT 感知算

法处理之后，形成目标列表数据。目标列表数据是不同时刻激光雷达检测到的目标的运动参数的列表。包含了所有进入探测范围的目标物的经纬度位置、局部坐标系位置、速度、加速度、航向角等运动信息，它是复现场景动态要素、判断场景是否有价值的关键数据。摄像头采集视频流文件，为目标列表提供必要的补充校核信息。雨量光照传感器采集雨量和环境光照数据，为场景提供自然环境信息，也是组成场景的重要要素之一。激光雷达点云数据经过感知算法处理后生成的目标数据以及雨量光照传感器获取的雨量光照值，封装为标准化的 RSM 消息按主题发布于 MQTT 中转站中。应用层通过订阅主题的方式获取相应的数据。对于摄像头获取的视频流，因为文件体量较大，采取更高效的 RTSP 取流方式，实现数据的远程传输。

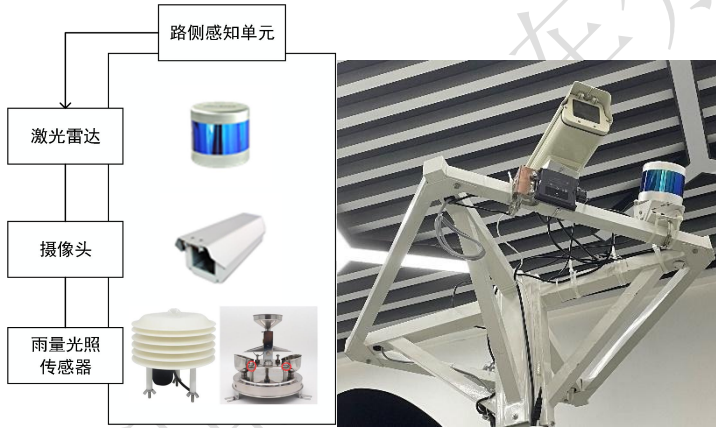


图 4.7 路侧感知单元

4.6.3 无人飞行器采集方式

相比于基于路侧设备的采集方式，基于无人机的采集方式所具有的-90°的俯视视野可以获取更广泛的交通参与者的行为，加之无人机悬停位置灵活、悬停高度更高的优势，使用基于无人机的采集方式更有利于自然交通流的数据采集。

基于无人飞行器的采集方式是指对某些特定的关键场景包括十字路口、环岛、高速公路、城市拥挤路段等，使用带有相机的无人飞行器在关键场景上方悬停进行数据采集。相关设备包括：无人飞行器、遥控器、RTK 模块、RTK 移动站等。

采集流程：根据研究目的选择采集地点后，使用遥控器将无人机飞至合适位置悬停进行数据采集；RTK 模块可使用网络 RTK 服务实现时间同步功能，也可结合

RTK 移动站使用 RTK 自定义服务，实现厘米级定位进行相关测绘工作；数据采集完成后，将储存在无人机内的相关文件导出，进行后续数据处理。

4.6.4 采集方式差异化分析

数据采集车、无人机和路侧采集是三种常见的数据采集方式，它们在应用场景、数据收集范围和操作灵活性等方面有所不同。

a) 实车采集

- 1) 应用场景：主要适用于道路上的数据收集，如道路状况、交通流量等。
- 2) 数据收集范围：由于车辆行驶在道路上，可以获取到相对全面的环境信息。同时，通过安装多种传感器（如摄像头、激光雷达等），可以获取高精度的地图和环境感知数据。
- 3) 操作灵活性：需要遵守交通规则，并受到道路条件和其他车辆的限制。操作相对复杂且受限制。

b) 无人机采集

- 1) 应用场景：主要适用于空中或难以到达的区域进行数据收集，如城市建筑物外观检查、农田监测等。
- 2) 数据收集范围：能够提供俯瞰式视角，并覆盖大范围区域。通过搭载摄像头或其他传感器，可以获取高分辨率影像或特定环境参数。
- 3) 操作灵活性：具有较高的机动性和自由度，在合法飞行区域内可自由移动并进行任务执行。

c) 路侧采集

- 1) 应用场景：主要适用于特定位置或设施进行局部化数据收集，如交通信号灯状态监测、停车位检测等。
- 2) 数据收集范围：局部化且针对特定目标进行观察与记录。通过安装在道路旁边或设施上的传感器（如摄像头、雷达），可以实时监控并记录相关信息。
- 3) 操作灵活性：需要选择合适位置并安装设备，在一定程度上受限于布

置位置以及周围环境条件。

4.7 采集设备部署及标定

4.7.1 采集设备部署

4.7.1.1 实车采集设备部署选取

数据采集设备实车部署通常需要遵循以下步骤：

- a) 选择合适的采集设备：根据数据采集需求选择合适的采集设备，例如 GPS 定位仪、Lidar、摄像头等需要有专用设备进行采集，同时需要考虑各设备时间对齐；
- b) 部署采集设备：根据车辆的结构和尺寸，选择合适的部署位置和方式，例如安装在车辆外部或内部。通常传感器安装在车辆外部，采集设备集中放置在车辆内部。安装时需考虑与车身硬联接，且要考虑缓冲减震，避免车辆行驶震动造成设备故障；
- c) 连接采集设备与数据处理系统：将采集设备与数据处理系统连接，以便实时传输数据并进行处理。连接方式一般采用专用线材进行连接，同时对于部分信号也可以考虑无线形式。需结合实际情况选取；
- d) 测试和调试：在实车部署前进行测试和调试，确保采集设备能够正常工作，并且数据传输和处理没有问题。安装完成后需要进行试运行，以验证设备稳定性；
- e) 维护和保养：定期对采集设备进行维护和保养，以确保其长期稳定的工作状态。

数据采集系统组成结构示意图如图 4.8 所示。按照自动驾驶汽车场景数据采集需求及当前传感器探测视场、探测距离的能力，传感器采集数据应覆盖车体周围 360 度，前方最远探测距离不小于 150 米，后方最远探测距离不小于 80 米，左右侧向探测距离不小于 30 米。数据采集传感器包括摄像头、激光雷达、毫米波雷达、红

外夜视摄像头、超声波雷达设备，以及 GPS/北斗和惯性导航的位姿测量设备。

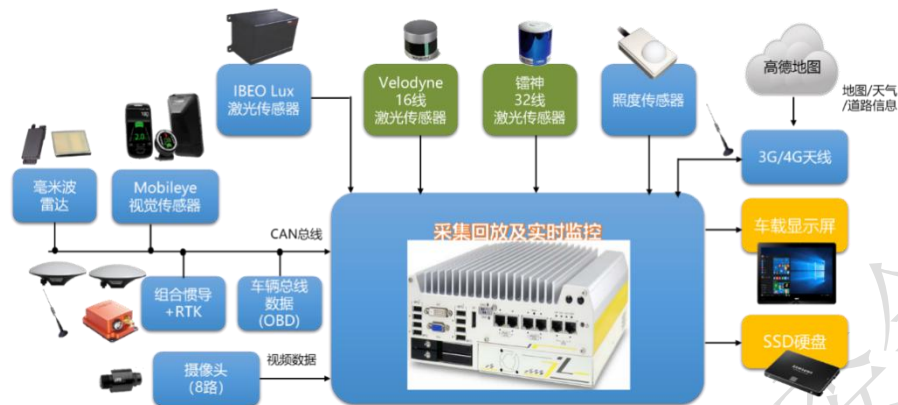


图 4.8 数据采集系统结构组成示意图

a) 摄像头

摄像头用于采集车辆周围环境有利于可见光识别的目标、驾驶员的疲劳驾驶监视信息等。摄像头对行人、车辆、交通标志、路面等目标具有较好的识别能力，根据车辆传感器的配置不同，包括前视和环视摄像头。前视摄像头成像距离较远，应配置高分辨面阵和窄视场长焦镜头，环视摄像头成像距离较近，使用中分辨面阵和宽视场短焦镜头。

驾驶员信息采集包括两个监控画面，其一是与车辆/环境/背景相关的驾驶员眼睛/头部运动；其二是驾驶员与车辆的相互作用，以及驾驶员动作（踏板，换挡，方向盘处理等）。数据采集项主要包括事件发生时刻、驾驶员危险驾驶（分心、疲劳、不良情绪）图像、驾驶员视线方向图像、对事件预先意识及发生事件后的反应图像等。

场景数据采集车辆内部摄像头包括驾驶员眼睛/头部运动监控、驾驶员动作监控和车内全景监控摄像头，驾驶员监控拟选用成熟的商用 DMS 系统，实现驾驶员行为监控与数据采集。

b) 毫米波雷达

毫米波雷达用于采集车辆前方、后方和侧向运动目标的位置和运动速度，以及毫米波易于识别的静态目标。毫米波雷达通过向外发送无线电波探测目标，是一种可全天时工作的主动传感器，毫米波雷达具有探测范围广、环境干扰小、成本低等

特点。目前车载毫米波雷达主要有 24 GHz 和 77 GHz 两种产品，采集数据项包括：车辆前向、后向和侧向障碍物体的位置和速度等信息，根据场景数据采集的不同要求，规范提出了两种不同需求的毫米波雷达性能指标要求。

c) 激光雷达

激光雷达用于采集车辆周围的三维目标信息，包括车辆前后方的运动目标信息、路面障碍物，以及车辆周围的环境要素信息。

激光雷达具有高精度采集车辆周围环境目标 3D 数据的能力，且有利于障碍物的检测。激光雷达通过发射红外激光探测物体的距离，是一种主动传感器，可以全天时工作。目前，自动驾驶汽车测试配置的激光雷达传感器以 360 度环扫为主，其次是固定视场半固态激光雷达。

d) 超声波雷达

超声波雷达用于采集车辆周边障碍物信息，获取智能汽车自动泊车所需避让的目标距离、方位等数据和信息。超声波雷达是一种小型汽车标配的主动传感器，主要配置在车辆前方、后方、前侧和后侧位置，近距离探测车辆运动方向的障碍物，避免车辆擦碰。

e) 定位系统

定位系统用于获取汽车的实时位置和运动轨迹，通过定位系统可以计算得到不同传感器采集目标的全局空间坐标或地图坐标。目前通用的定位系统以 GPS 为主，以及 GLONASS、北斗定位系统形成的组合全球导航定位系统(GNSS)。汽车定位系统应兼顾北斗导航定位系统，并保证接收不少于两个通道。

场景数据采集要求达到厘米级的空间定位精度，建议选用全国 GPS 网的 CORS 站提供的差分数据处理服务，特殊情况时采用布设基站方案实现动态差分数据处理。

f) 惯性导航系统

惯性导航系统用于获取汽车的运动姿态变化参数，通过车辆的姿态参数，可以确定不同传感器采集目标相对于车辆的方位。

惯性导航系统是一种不依赖于外部信息、也不向外部辐射能量的自主式导航系

统，能够采集车辆在导航坐标系中的速度、偏航角和位置等信息。目前，高性能 MEMS 惯导性价比更好，百度无人车使用一种基于紧耦合技术的一体式 MEMS 组合导航系统（SPAN-IGM-A1），作为场景数据采集惯性导航系统参考。

g) 上位机功能要求

上位机主要用于对车辆集成传感器的集成控制、收发指令、数据采集和存储等方面。上位机采用工控机，并与车辆的多传感器集成应用，多传感器集成对工控机的数据处理、传输与存储有较高要求。

上位机性能指标要求为系统最低要求，建议使用车规级的工控机进行传感器的集成控制、收发指令、数据采集和存储。

4.7.1.2 路侧设施采集设备部署选取

路端采集感知系统布置方案如图 4.9 所示，在交叉路口、直道、丁字路口三种关键道路典型地段路侧布置激光雷达、摄像头以及雨量传感器和光照传感器。

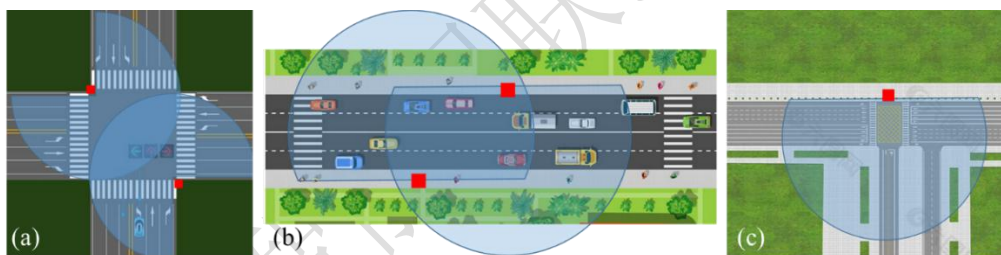


图 4.9 路端感知系统布置方案

在十字路口中进行路侧设备布局，如图 4.10 所示，整个部署方案中包含 12 个相机，4 个毫米波雷达以及 2 个激光雷达。在该道路的路口对侧，设置 1 个毫米波雷达，可以安装在红绿灯杆上朝向监控灯杆，负责感知到达停止线前的高速运动物体。另外还可以在路口对角线的端点处分别设置 1 个激光雷达，激光雷达安装在路口的两个对角的灯杆立杆上，负责感知与其相邻的两个方向的车道。

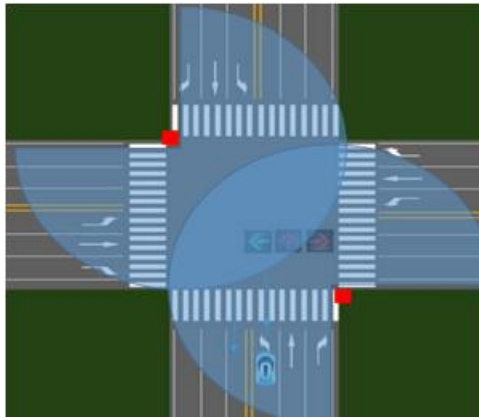


图 4.10 十字路口点位激光雷达部署方案

利用多激光雷达拼接形成的连续感知能力，获取交通参与者的精确信息，如目标物类型、外观尺寸、经纬度、速度、加速度、方向角等，生成标准化的数据包并由消息中间件为上层应用提供信息，配合摄像头视频流实现融合感知，利用多雷达、多传感器感知融合算法，实现场景数据二十四小时无间断高效采集。

4.7.1.3 无人飞行器采集设备部署选取

基于无人飞行器的采集方式对无人机相关设备的参数要求可参考目前已有的航测数据库视频帧数与时间，以及所研究方向要求进行选择。

大部分已有航测数据集无人机采集视频帧数都是 25fps 左右，根据时间同步的需要选择是否搭载 RTK 模块，根据测绘功能的需要选择是否搭配 RTK 移动站。

悬停高度的选择如图 4.11，根据公式 $L = \frac{P}{F} \times H$ ，可得出无人机 100m 高度下 1pixel = 3.75cm，该公式中：

P：视野内目标物所占像素值；

F：相机镜头焦距；

H：悬停高度；

L：目标物实际长度。

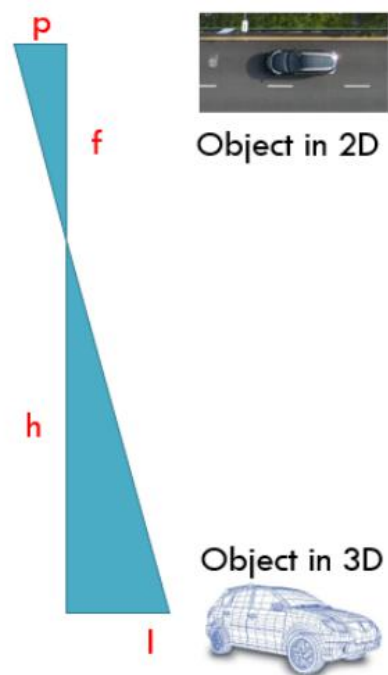


图 4.11 无人飞行器采集悬停高度计算示意图

无人机采集平台相关配置参数参考如下：

- a) 相机：无人机搭载一台相机，位于无人机前端或底部，至少满足录制 4k 25fps 视频功能（参考已有数据集视频参数）；
- b) 电池：电池容量 2250mAh—5000mAh，满足无人机在空悬停时间 30min 以上；
- c) 储存：支持外插 SD 卡储存；
- d) 云台：保持相机稳定，并支持相机-90° 的俯视拍摄；
- e) 卫星导航系统：采用的卫星导航系统有 GPS、GLONASS、Galileo、BeiDou 几种，大多为 GPS + GLONASS + Galileo，高端机型上存在 GPS + Galileo + BeiDou 的组合方式；
- f) 遥控器：工作频段 2.400 GHz 至 2.4835 GHz、5.725 GHz 至 5.850 GHz；
- g) RTK 模块与 RTK 移动站：使用网络 RTK 服务或是自定义 RTK 服务，满足时间同步需求或是测绘需要。

4.7.2 时间同步

对于智能网联汽车搭载有摄像头、激光雷达、毫米波雷达等传感器，以及 GNSS 和惯性测量单元（IMU）等位置姿态传感器。GNSS、IMU 和成像传感器对应不同的时间系统、存在定时偏移，各自的时间起点和基准不一致，因此需要场景数据采集多种传感器的时间同步方案，以保证多传感器所采集的数据对齐。

基于 GPS UTC 时间的多传感器时间同步，是搭载有 GPS 全球定位系统的多传感器集成应用的一种通用时间同步方法。通过公用的时间同步模块设计可以实现多传感器的时间同步，将 1PPS 和 GPS 时间信息都输入到公用的时间同步模块中来完成时间同步。如图 4.12 所示，时间同步系统由三部分组成：能够输出 1PPS 的 GPS 接收机、时间同步数据采集模块和数据采集软件。GPS 接收机输出 1PPS 作为时间参考并触发整个系统，时间同步数据采集模块由模拟信号输入/输出、计数器/计时器和数字 I/O 等组成，采集软件可将多传感器数据打上同步时间戳，该方法可以获得毫秒级的时间同步精度。

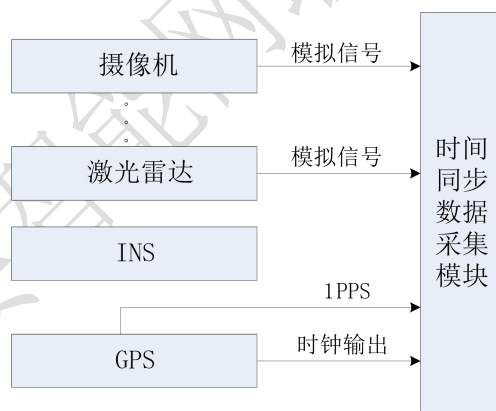


图 4.12 时间同步模块的同步原理

GPS 接收机的数据更新率通常为 1Hz，INS 能够达到 200Hz 或更高的数据更新率，视频的采集频率 20-30Hz，激光雷达数据采集帧频/扫描周期为 10-20Hz，将多传感器的采样时间以 GPS 的秒脉冲信号对齐，进一步对各传感器按照帧频和时间进行线性内插，得到各传感器数据采集的时间戳。

4.7.3 空间同步

智能汽车多传感器融合的目的是实现汽车周边环境目标的有效识别和定位，由于各传感器成像和测量数据坐标系不同，在场景数据的空间位置配准前首先要统一各传感器测量坐标系。在统一的空间坐标系下，通过多传感器集成标定或同名目标特征，实现场景数据的空间配准。

智能汽车各传感器涉及多种坐标系统，包括摄像头成像平面坐标系、激光雷达设备扫描三维坐标系、毫米波雷达三维成像坐标系、INS 惯导系统北天东坐标系、GPS 的 WGS84 坐标系以及汽车载体坐标系等。必须将所有传感器的坐标系进行统一转化，才能实现场景数据的空间配准。集成 GPS 的智能汽车的坐标系转换要求如下：

集成 GPS 的多传感器系统的场景数据空间配准在 WGS84 空间直角坐标系/国家 CGCS2000 坐标系下进行，将各传感器测量的目标坐标分别归化到该坐标系，再进行场景数据的空间位置配准，如图 4.13 所示。

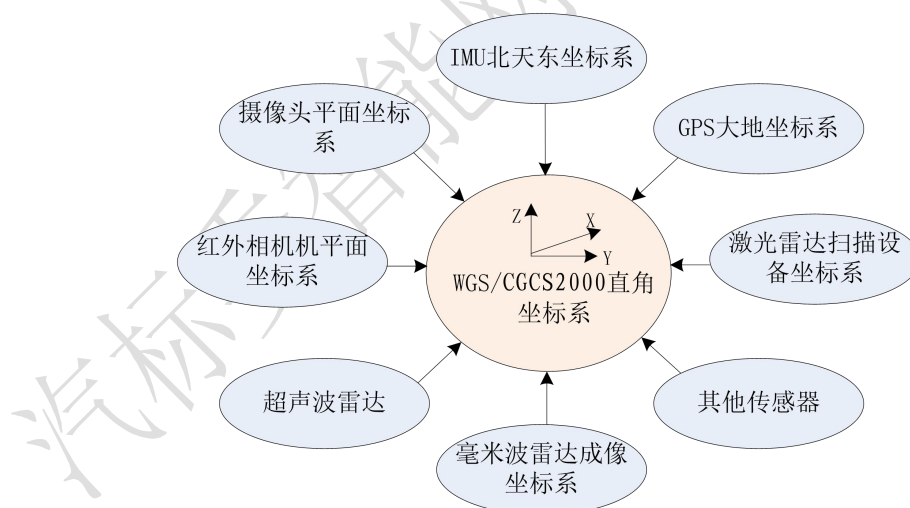


图 4.13 CGCS2000 直角坐标基准

激光雷达、毫米波雷达和摄像头等多传感器集成标定实验环境选择室外约 30m×15m 较为空旷的场地，路面为平坦道路，道路周围安装多组车载传感器支撑架，在支撑架上和道路周围地面传感器视场内，分别均匀布设激光雷达、毫米波雷达和摄像头用靶标，组成车用多传感器集成定标的实验环境，如图 4.14 所示。

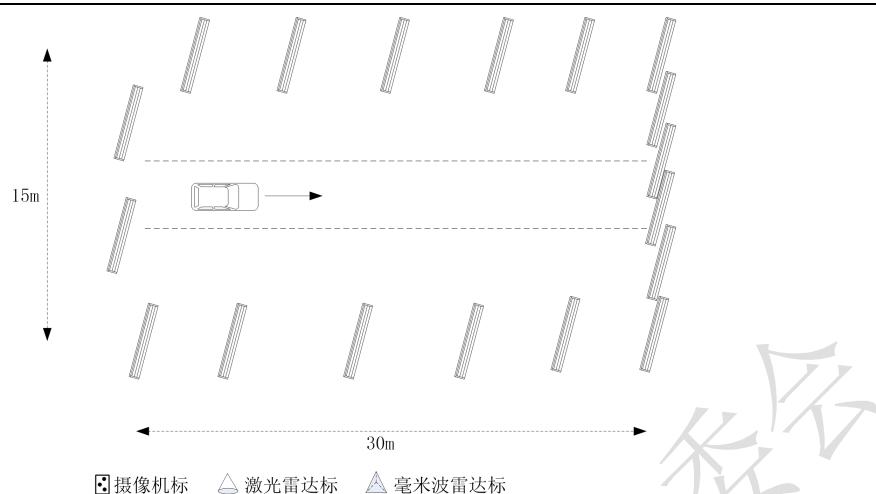


图 4.14 多传感器集成标定实验环境

对于没有进行集成标定的多传感器数据采集系统，可以通过激光雷达、毫米波雷达和摄像头等各传感器在静态或慢速平稳运动同时时刻的成像和测量数据，提取同名目标实现空间配准。在场景数据配准时，选择场景中的静态目标作为特征提取对象，采用 SIFT、SURF 等特征提取算法，提取目标有效的同名特征。

4.8 数据检查要求

4.8.1 通用要求

为保证数据同步的质量，对同步后的数据有如下指标要求：

- 数据同步后的数据刷新频率应不低于 10Hz；
- 数据同步后，对于同一目标，不同固态激光雷达探测到的目标数据应相同，即输出同一个目标物的数据；
- 数据同步后，对应同一目标，多个机械激光雷达之间输出的数据偏差不超过 10ms，对同一目标上的最近点，各机械激光雷达之间探测值的差的绝对值应小于 5cm；
- 数据同步后，对于同一目标，在 X 方向-25m~50m、Y 方向-5m~5m 范围内固态激光雷达与智能摄像头的探测值，在 X 轴方向的差的绝对值应小于 1m，在 Y 轴方向的差的绝对值应小于 50cm。

4.8.2 常规数据检查

场景数据检查流程主要包括数据完整性检查、数据连续性检查、数据一致性检查、数据准确性检查、检查结果记录及抽样回放，如图 4.15 所示。



图 4.15 场景数据检查流程

- a) 数据完整性检查旨在确认采集到的数据文件是否完整，数据完整性主要检查内容如下：
 - 1) 采集的整个数据文件种类及数量是否缺失：检查采集的数据文件中各种类型的文件是否齐全，确保没有丢失任何关键数据文件。
 - 2) 整个数据文件集及每种数据文件的占用空间大小是否满足给定要求：检查每种数据文件的占用空间大小，确保其满足预定的要求，以防止数据过大或过小导致问题。
 - 3) 数据文件中各字段的信息是否缺失：检查数据文件中各字段的信息是否缺失，确保每个字段都有相应的数据。
- b) 数据连续性检查主要关注数据之间的时间关系，以确保数据连续性和一致性，数据连续性检查要求如下：
 - 1) 时间戳间隔检查：确保连续的数据集之间的时间戳间隔不应超过 1 秒，以维护数据的时序性。
 - 2) 数据集内时间戳间隔检查：数据集内时间戳之间的间隔不应大于 0.1 秒，以确保数据的高频采集。
- c) 数据一致性检查旨在验证采集的数据与实际情况的一致性，包括以下方面：
 - 1) 合理性检查：采集的数据是否与实际情况相反，确保采集的数据在合

理的取值范围内，例如，各个场景数据变量是否在合理的范围内，以避免数据异常情况的出现。

- 2) 语义一致性检查：对于不同类型传感器采集的同一场景数据，需要进行语义一致性检查。例如，通过 Kappa 检验约束来检查场景数据中的定性参数的一致性，如天气信息。对于定量参数，可以使用 Kendall 协调系数来进行一致性检查。
- d) 数据准确性检查旨在检查数据中可能存在的错误或异常情况，包括以下内容：
- 1) 数量级错误检查：检查整个数据集中某个字段的数据是否存在数量级的记录错误。通过统计分析的方法，如计算平均数和中位数，可以发现这类问题。
 - 2) 异常值检查：检查数据集中是否存在个别的异常值。这可以通过计算最大值和最小值的统计量，或使用箱线图或 3σ 原则等方法进行检查。
 - 3) 计算结果错误检查：确保各项数据的计算方法和计算结果是否存在错误，以避免数据处理过程中的问题。
- e) 检查结果记录是每次数据检查后，输出数据检查结果，包括检查时间、数据来源、数据集名称、数据类别、数据大小及检查是否通过，用于评估数据采集质量。
- f) 抽样回放是一种人工检查方法，用于从每天的数据集中抽取 20% 进行回放检查，以验证数据的可用性和一致性。回放检查要求包括以下内容：
- 1) 数据可回放性检查：检查数据集内各种类型的数据是否可以正常回放，以确保数据的实际可用性。
 - 2) 数据同步性检查：确保数据集内各种类型的数据能够同步回放，避免出现明显的错位现象。
 - 3) 卡顿现象检查：检查数据集内各种类型的数据回放是否出现卡顿现象，以评估系统的性能。

4.8.3 数据同步性检查

数据同步性检查是场景数据检查流程中的重要环节，旨在确保不同类型的数据能够在时间上同步回放，以维持数据的时序关系和一致性。数据同步性检查主要关注数据之间的时间同步性，以保证各种类型的数据在回放时能够按照实际时间发生的顺序进行同步。这对于分析自动驾驶系统在真实场景中的行为至关重要。数据同步性检查包括以下方面：

- a) 时间戳比对：针对不同类型的数据（如图像、激光雷达、GPS 等），检查它们的时间戳是否一致。时间戳通常用于标识数据采集的时间点。通过比对时间戳，可以确保不同数据来源的数据在时间上是一致的。以下是时间戳比对的主要方法和公式：

- 1) 绝对时间戳比对：比较不同数据源的绝对时间戳，通常以秒为单位。

如果两个数据源的时间戳在给定的时间差范围内相等，就认为它们是时间同步的。公式如下：

$$|t_1 - t_2| \leq \Delta t$$

t_1 是第一个数据源的时间戳。

t_2 是第二个数据源的时间戳。

Δt 是允许的时间差阈值。如果 $|t_1 - t_2|$ 小于或等于 Δt ，则认为两个时间戳是同步的。

- 2) 相对时间戳比对：相对时间戳比对将时间戳转换为相对时间，通常以毫秒或微秒为单位。比较两个数据源的相对时间戳，确保它们之间的时间差在可接受范围内，公式如下：

$$|t_{1r} - t_{2r}| \leq \Delta t_r$$

t_{1r} 是第一个数据源的相对时间戳。

t_{2r} 是第二个数据源的相对时间戳。

Δt_r 是允许的相对时间差阈值。如果 $|t_{1r} - t_{2r}|$ 小于或等于 Δt_r ，则认为两个相对时间戳是同步的。

时间戳比对公式的选择取决于具体的数据采集和处理情况，以及同步要求的精

确度。在实际应用中，通常会根据系统需求和实验条件来确定适当的时间同步方法。

- b) 时序关系验证：验证各种类型的数据在回放时是否保持正确的时序关系。
例如，车辆位置数据应该与相机拍摄的图像在时间上一一对应。
- c) 错位检查：检查数据回放过程中是否出现了明显的错位现象。这可能导致数据不同步，需要及时纠正。
- d) 数据采集频率匹配：确保不同数据源的采集频率相匹配。如果某个传感器以不同的频率采集数据，需要对其进行插值或降采样，以保持同步性。
- e) 数据对齐标定：在一些情况下，可能需要进行数据对齐标定，以将数据同步到相同的坐标系中。例如，将激光雷达数据与相机图像进行校准，确保它们在相同坐标系下表现一致。
- f) 数据播放测试：在数据同步性检查的过程中，进行数据回放测试，模拟自动驾驶系统在实际场景中的行为。这有助于发现数据同步性问题，并验证系统是否正确处理了同步数据。

4.9 数据上传要求

4.9.1 车端传输要求

4.9.1.1 通用要求

车端数据在传输过程中主要考虑数据的完整性、安全性和可用性。车端传输车辆数据的关键要求如下：

- a) 数据完整性：确保上传的数据是完整的，没有丢失或损坏的部分。这可以通过数据校验和错误检测机制来实现，例如使用校验和或哈希值来验证数据完整性。
- b) 数据压缩：考虑到车辆数据通常具有大量的冗余信息，采用数据压缩技术可以降低数据传输的带宽要求，减少传输时间和成本。
- c) 数据安全性：车辆数据传输需要采用安全的通信协议，以确保数据不会在传输过程中被篡改或窃取。常见的做法包括使用加密和身份验证机制。

-
- d) 数据格式标准化：车辆数据的格式应该遵循标准化的格式，以便接收端能够正确解析和处理数据。通常使用行业标准的数据格式，如 JSON、XML 或 Protobuf。
 - e) 网络带宽管理：考虑到车辆数据可能需要上传到云端或远程服务器，需要在车辆端管理带宽，以确保数据上传不会影响车辆的正常操作，同时也要优化上传策略，确保高优先级数据能够及时传输。
 - f) 重传机制：为了应对车辆突然下电、网络不稳定或丢包的情况，需要实施数据重传机制，以确保数据的可靠性。这可以通过 TCP 协议或自定义的重传策略来实现。
 - g) 数据权限控制：确保只有授权的用户或系统可以上传和访问车辆数据。这可以通过访问控制列表（ACL）或身份验证来实现。
 - h) 上传日志和诊断：在上传过程中记录日志和诊断信息，以便监控数据上传状态和及时解决问题。
 - i) 数据隐私：车辆数据上传需要遵守相关的数据隐私法规和政策，确保用户的个人数据得到充分保护。

数据传输应符合下列技术要求：

- a) 采用大端模式的网络字节序来传递无符号整型（WORD、DWORD）、时间戳（TIMESTAMP、TS_MIN）等数据类型；
- b) 经纬度等坐标信息采用符合国家要求的坐标系；
- c) 视频信息符合 GB/T 28181-2022《公共安全视频监控联网系统信息传输、交换、控制技术要求》；
- d) 字符串采用 UTF-8 编码格式；
- e) 时间戳为东八区 UTC 时间。

4.9.1.2 网络连接及规则

场景数据采集端与云端的网络连接指车载采集终端通过运营商 4G 或 5G 网络与云存储端建立的网络连接，如图 4.16 所示。

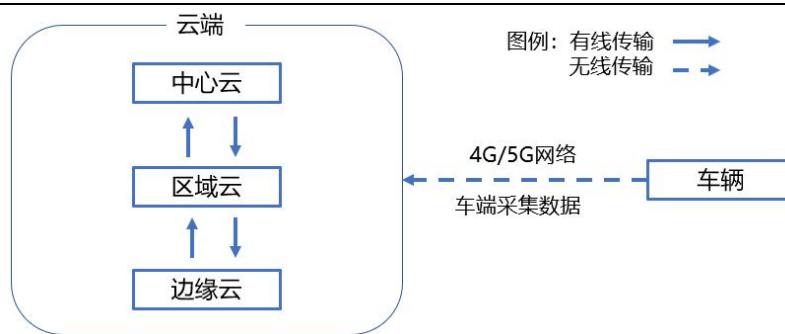


图 4.16 场景数据采集端与云端传输示意图

车载采集端与云端车云网关连接时，建立基于 TLS1.2 及以上版本的双向认证的 TCP 连接，并使用云端签发的客户端证书。

车端设备启动后，发起请求和云端建立 TCP 连接，用于高频实时数据上传。

TCP 连接维持规则：连接建立后，车端应按固定频率向云端发送心跳消息，云端收到后向车端发送心跳回应消息。

TCP 断开规则：

- a) 云端和车端都可以根据协议主动断开连接；
- b) 云端判断 TCP 断开方法：
 - 1) 根据协议断开车端连接；
 - 2) 相同身份车端建立连接，断开原连接；
 - 3) 在一定的时间内未收到车端发出的消息或者心跳；
- c) 车端判断 TCP 断开方法：
 - 1) 根据协议断开云端连接；
 - 2) 达到重传次数后仍未收到应答。
- d) 车端和云端发现任意 TCP 连接断开后，由车端主动请求重新建立连接，车端处于非工作状态时，保持连接断开状态。

4.9.1.3. 数据类型

a) 二进制类数据类型

协议约定的数据传输格式为二进制格式，二进制类数据类型定义应符合表 4.3 的要求。

表 4.3 二进制类数据类型

序号	英文名称	长度	中文名称	取值范围
1	BIT	1 bit	比特	[0..1]
2	BYTE	1 B	无符号单字节整型	[0..255]
3	BYTE[N]	N B	无符号整型数组	[0..256 ^N -1]
4	WORD	2 B	无符号双字节整型	[0..65535]
5	DWORD	4 B	无符号四字节整型	[0..2 ³² -1]
6	TIMESTAMP	8 B	UTC+8 时间（单位：ms），当前时刻距 1970 年 1 月 1 日 0 时整的毫秒数	[0..2 ⁶⁴ -1]
7	TS_MIN	4 B	当前时刻距 1970 年 1 月 1 日 0 时整的分钟数，TS_MIN = TIMESTAMP/60000	[0..2 ³² -1]
8	STRING[N]	N B	字符串类型（UTF-8）	-

注：长度列中“B”表示“字节”，“bit”表示比特，对于长度为 N 个字节的数据，BIT0 表示最低位，BIT(8N-1)表示最高位。

b) 结构体类数据类型

部分数据传输类型采用结构体进行描述，其定义及成员数据类型定义应符合表 4.4 的要求。

表 4.4 结构体类数据类型

序号	英文名称	中文名称	长度	成员名	成员中文名称	数据类型	取值范围及单位
1	POSITION	位置	12B	longitude	经度	DWORD	[0..3600000000], 经度, 单位: 1e-7°, offset=-1800000000, 表示实际值: [-180.0000000..180.0000000], 大于 0 表示东经, 不可缺省, 0xFFFFFFFF 表示异常
				latitude	纬度	DWORD	[0..1800000000], 纬度, 单位: 1e-7°, offset=-900000000, 表示实际值: [-90.0000000..90.0000000], 大于 0 表示北纬, 不可缺省, 0xFFFFFFFF 表示异常
				elevation	高程	DWORD	[0..200000], 海拔高度, 单位: dm, offset=-100000, 表示实际值: [-100000..100000], 0xFFFFFFFF 表示缺省

表 4.4 结构体类数据类型（续）

序号	英文名称	中文名称	长度	成员名	成员中文名称	数据类型	取值范围及单位
2	POSITION2D	二维位置	8B	longitude	经度	DWORD	[0..3600000000]，经度，单位： $1e-7^{\circ}$ ，offset=-1800000000，表示实际值：[-180.0000000..180.0000000]，大于0表示东经，不可缺省，0xFFFFFFFF表示异常
				latitude	纬度	DWORD	[0..9000000000]，纬度，单位： $1e-7^{\circ}$ ，offset=-9000000000，表示实际值：[-90.0000000..90.0000000]，大于0表示北纬，不可缺省，0xFFFFFFFF表示异常
3	EQUATION	五次曲线方程	28B	factor5	五次项系数	DWORD	[0..4280000000]，单位：0.0000001，offset=-214，表示实际值：[-214.0000000..214.0000000]，0xFFFFFFFF表示缺省

4.9.2 路侧设备传输要求

4.9.2.1 通用要求

传输应符合下列要求：

- a) 经纬度等坐标信息采用符合国家要求的坐标系；
- b) 字符串采用 UTF-8 编码格式；
- c) 时间戳为东八区 UTC 时间。

4.9.2.2 云端与 RSU 传输规则

可根据应用选择MQTT、HTTPS等传输协议，宜采用MQTT传输的数据协议。对于采用MQTT传输协议的，宜使用JSON格式，也可以根据消息集定义，采用protobuf等格式。云端与路侧设备传输如下图4.17所示。

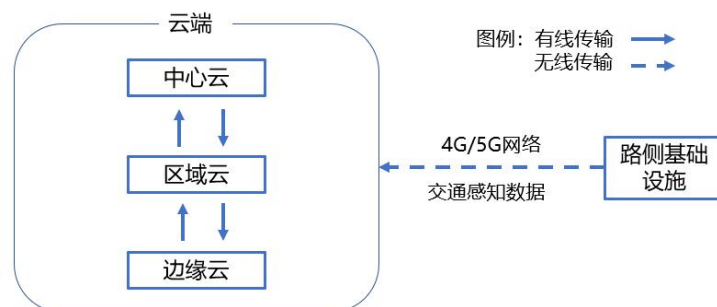


图 4.17 云端与 RSU 传输示意图

4.9.2.3 数据类型

a) 字符串数据类型

根据消息集定义，采用JSON，protobuf等传输格式，宜采用JSON格式。JSON字符串数据类型应符合表4.5的要求。

表 4.5 JSON 数据类型

数据类型	长度	数据描述
JSON_INT	4 BYTE	整型，-2147483648~2147483647
JSON_DOUBLE	8 BYTE	双精度浮点型或数据长度大于4byte的内容，-1.79E+308~1.79E+308
JSON_STRING	不固定	字符串（UTF-8）
JSON_ITEM	不固定	JSON字符串基本单元，格式为：键/值对key:value，用半角冒号分割
JSON_OBJECT	不固定	JSON字符串的一个对象，对象的内容本身是一个json字符串
JSON_ARRAY	不固定	JSON字符串中的一个数组，一个数组中可能包含一个或多个ITEM、OBJECT、ARRAY
JSON_TIMESTAMP_UTC	8 BYTE	无符号整型，自1970年1月1日00:00:00以来的毫秒数

b) 二进制数据类型

对于采用二进制格式传输的消息集，二进制数据类型应符合表 4.6 的要求。

表 4.6 二进制数据类型

数据类型	长度	描述
BYTE	1 BYTE	无符号整型
BYTE[n]	n BYTE	无符号整型数组
WORD	2 BYTE	无符号整型
DWORD	4 BYTE	无符号四字节整型（双字，32位）
STRING[n]	n BYTE	字符串（UTF-8）
TIMESTAMP	8 BYTE	UTC时间，当前时刻距1970年1月1日 0时整的毫秒数，经过东八区转换后的数值

4.10 数据存储要求

4.10.1 通用要求

车辆数据采集后，需要在采集端进行临时存储，采集端的存储要求如下：

- 数据完整性：确保数据不丢失、不损坏，可以使用校验和、哈希值或数据签名等技术来验证数据的完整性。
- 擦写次数管理：特别对于 NAND Flash 存储介质（如 eMMC），需要关注

擦写次数的管理，采取措施减少擦写操作，例如数据缓存、数据压缩、均衡擦写策略等，以延长存储介质的寿命。

- c) 数据备份与恢复：定期备份存储重要数据，并实施数据恢复策略，以防止数据丢失。定期上传存储介质中的重要数据，以防止数据丢失。
- d) 数据质量与访问控制：在数据存储之前进行质量控制，排除错误或损坏的数据，以确保存储的数据质量。另外，仅允许授权的用户或系统访问车端存储的数据，并实施访问控制策略，以确保数据的隐私和安全性。
- e) 数据容量管理：根据数据保留政策和法规要求，制定数据清理策略，有效管理存储容量，确保足够的空间用于数据存储，数据量达到一定阈值后，定期清理或循环覆盖不再需要的数据，以降低存储成本和风险。根据需要，使用数据压缩技术来减少存储空间的占用，并优化数据存储结构以提高访问效率。
- f) 合规性和法规遵守：确保存储系统符合适用的法规和合规性要求，特别是涉及个人数据的情况下，要进行脱敏，隐私数据要遵守隐私法规。

因此，车辆采集端存储需要综合考虑各方面因素，以确保存储数据的完整性和安全性、硬件设备可持续使用性及隐私合规性。

4.10.2 采集端存储要求

采集端存储采用固态硬盘，存储的文件主要是由数采设备采集得到的 ROSbag 数据包等原始文件。

采集时按照需求修改数据分包要求，默认按时间进行分包（或按照里程计距离、bag 包大小进行分包）。开始录制时，在 Panels 选项下打开 Topicsview 界面，对需要一同录制的 topic 进行勾选以确定录制内容。

bag包为传感器产生的数据，使用ROS系统的rosvbag文件定义，以该数据包开始录制的起始时间作为命名。bag内包含的数据一般有：

- a) 带 rslidar_packets 字样的数据为激光雷达相关数据，图中分左中右三个激光雷达的数据，每个数据有 rslidar_packets 和 rslidar_packets_difop 两种数据，分别为激光雷达的网络数据包（即激光产生的点云数据）和激光雷达设备

信息数据（即内参等）；

- b) 带 camera 字样的数据为相机相关数据，图中为 6 个相机产生数据。其内为直接的图片数据，采用 ROS 系统自带图像消息格式 `sensor_msgs/CompressedImage` 进行的定义；
- c) 带 radar 字样的为毫米波雷达相关数据，图中为前后两个毫米波雷达的数据，其采用自定义的消息格式体 `miliwave_msg/miliwave_list` 进行记录，内含的数据为毫米波直接输出的目标数据；
- d) `/rs/gps`、`/rs/imu`、`/rs/odom` 为 RTK 惯导产生的数据，分别代表 GPS 位置（经纬度）、IMU 和融合输出的里程计结果。其分别采用 ROS 系统自带的 `sensor_msgs/NavSatFix`、`sensor_msgs/Imu`、`nav_msgs/Odometry` 三种消息格式定义；
- e) 以上毫米波雷达，激光雷达的数据皆为自定义消息格式，如需解析 bag 对应数据，需要专用消息体 msg 文件。

4.10.3 云端存储要求

驾驶场景云端存储同样采用固态硬盘，储存的文件是由 ROSbag 原始文件后处理解析而来的场景文件，其储存结构如图 4.18 所示。



图 4.18 采集场景数据结构

车辆数据上传至云端后，云端存储以确保数据的可用性和有效性，有以下要求：

a) 高可用性：

- 1) 冗余和备份：使用冗余存储和自动备份策略，以应对硬件故障或其他中断，确保数据随时可用；
- 2) 负载均衡：采用负载均衡技术，以分散访问压力，提高系统的可用性；
- 3) 容量和扩展性：设计存储系统，以便在需要时能够方便地扩展存储容量，以满足不断增长的数据需求。

b) 数据备份和恢复：

- 1) 定期备份：建立定期备份策略，以确保数据不会因意外事件而丢失；
- 2) 快速恢复：实施快速数据恢复策略，以降低数据恢复时间。

c) 数据存储结构：

- 1) 索引和元数据：建立有效的索引和元数据管理系统，以支持数据查询、检索和管理；
- 2) 分区和分层存储：根据数据访问模式将数据分区并采用适当的存储层次结构，以提高访问效率。

d) 数据质量控制：

- 1) 数据验证：在存储之前进行数据验证，以排除错误或损坏的数据；
- 2) 数据清理策略：制定数据清理策略，删除不再需要的数据，以降低存储成本。

e) 访问控制：

- 1) 权限管理：建立权限模型，确保只有授权用户或系统可以访问和操作存储的数据；
- 2) 多因素认证：实施多因素认证以提高访问的安全性。

f) 性能优化：

- 1) 缓存和加速：使用缓存和数据加速技术，以提高数据查询和分析的性能；
- 2) 负载监控：实施实时负载监控，以及时调整系统资源以满足性能需求。

g) 长期存储策略：

- 1) 数据保留政策：制定合适的数据保留政策，以满足法规和合规性需求，同时管理存储成本；
- 2) 存档策略：将不再活跃的数据归档到更便宜的存储层次结构中，以减少成本。

云端存储是对每天采集的原始数据、场景标注的数据，清洗后的数据、数值计算及分析结果进行归档、整理和共享的过程，应满足以下要求：

- a) 应提供独立逻辑存储空间；
- b) 应具备可弹性伸缩、高可用的分布式对象存储的能力；
- c) 能为结构化、半结构化和非结构化类型数据提供不同级别的存放策略；
- d) 应创建不同类型的字段，根据用途和需求变化对字段进行增加、删除和修改操作，并提供数据导入/导出和数据迁移功能；
- e) 单车每天采集的数据（3t 左右量级），通过场景类型、标签检索场景检索的时间不超过 5min；
- f) 计算分析 10 万公里场景数据，时间应能达到小时级。

为满足原始数据的海量存储、快速分析及快速查询结果的需求，驾驶场景数据云端存储采用对象存储、关系型数据库及分布式存储联合的方式，其中采集的原始数据采用对象存储的方式存储，例如：Amazon S3（Amazon Simple Storage Service）；web 应用管理平台数据存储在关系型数据库中，例如 Mysql；场景标注及数值计算结果、数据分析结果及提取的测试用例采用分布式存储方式，例如：场景标注及数值计算用 ApacheDruid 存储，数据分析结果用 Hive 存储，提取的测试用例用 HBase 存储。

4.11 总结

由于不同汽车制造商和技术供应商采用不同的数据格式和接口标准，导致数据采集的互操作性问题。这使得数据采集和整合变得复杂，限制了数据的有效利用和共享。基于以上的需求，行业应开展更广泛的数据标准化活动，以促进数据的互通和互操作性。对数据质量、采集区域、采集方式、采集设备部署及标定、数据检查、数据上传及数据存储进行标准化，同时，执行自动驾驶测试场景数据采集流程时对于采集分析技术提出了要求，对数据清洗和处理、场景标注、数据统计及测试用例提取等关键技术将发挥关键作用。

第五章 自动驾驶系统测试场景数据分析技术标准化研究

5.1 场景数据分析流程

测试场景数据的分析流程主要包括数据清洗、场景标注、场景统计及测试用例生成，如图 5.1 所示，其中：



图 5.1 测试场景数据分析总体流程

- a) 数据清洗是指通过筛选清除、插值补充、滤波去噪等方法发现并纠正无效数据、错误数据、缺失数据、重复记录、不一致数据、传感器漏检、前期处理精度不足的数据等；
- b) 场景标注是依据场景标签定义，采用自动标注加人工校核的方法对道路、交通设施、临时变化、交通参与者及环境等场景信息进行标注；
- c) 场景统计是从采集的真实交通数据中检测并提取测试场景的参数值，分类汇总生成测试场景各参数的总体分布，并计算均值及方差等；
- d) 测试用例提取流程主要包括明确目标、测试场景选择、要素确认与值设定、测试用例优化、期望或评价指标设定、激活条件设定、测试用例优化。

5.2 场景数据清洗技术

5.2.1 场景数据清洗流程

数据清洗的目的主要是纠正无效数据、修正错误数据、消除重复记录、确保数据一致性、处理传感器漏检及改善前期处理精度，为后续的场景统计及测试用例生成提供更加可靠和准确的数据支持，清洗流程主要包括制定清洗规则、数据清洗、结果验证及结果记录，如图 5.2 所示，其中：



图 5.2 场景数据清洗流程

- a) 制定清洗规则主要是针对数据的异常类别，制定相应的数据清洗规则，包括数据有效性的检查和处理、数据精准性的检查和处理、数据完善性的检查和处理、数据一致性的检查和处理。
- b) 数据清洗是以自然驾驶场景原始数据为输入，根据清洗规则，通过数据修正、数据平滑滤波等一系列步骤剔除无效数据，修正错误数据，最终以期望的格式输出有效数据的过程。
- c) 结果验证主要是人工从每天的数据集中抽取 20% 进行回放检查，检查不同类型数据是否能够同步回放，是否有明显偏离设定范围的现象，不同类型数据对同一个目标值的探测值是否存在明显偏差。

5.2.2 场景数据清洗方法

5.2.2.1 摄像头视频数据的清洗

摄像头视频数据的清洗流程主要包括格式标准化、畸变校正、时间同步、空间同步及视频分割，如图 5.3 所示，其中：



图 5.3 场景数据清洗流程

- a) 格式标准化：所有视频数据采用相同的格式和编码，以简化后续处理步骤；
- b) 畸变校正：校正由于双目、环视和全景摄像头的镜头畸变引起的图像失真；
- c) 时间同步：考虑不同摄像头的时钟偏差，依据时间戳，通过同步算法使不同摄像头采集的视频数据在处理过程中保持时间上的一致性；
- d) 空间同步：在同一摄像头视频的每一帧中寻找和匹配公共特征，以确保同一视频数据不同帧之间的对准，在不同摄像头同一时间戳对应的每一帧中寻找和匹配公共特征，确保不同相机视角之间的准确对齐；

- e) 视频分割：将长时间录制的视频分割成更短的片段或帧，以方便后续处理和分析。

5.2.2.2 毫米波雷达点云数据的清洗

毫米波雷达数据清洗主要是去除无效信号、虚点信号及漂移点，如图 5.4 所示。



图 5.4 毫米波雷达点云数据清洗流程

a) 无效信号的去除

毫米波雷达的无效信号通常是由于各种原因而产生的噪声或干扰，去除方法主要包括信噪比过滤、距离过滤及数据截断，如下所述：

- 1) 信噪比过滤：通过计算信号的信噪比（Signal-to-Noise Ratio, SNR），判断信号是否有效，低于一定阈值的信号被认为是无效的，并将其进行滤除；
- 2) 距离过滤：根据目标与雷达的距离来过滤远距离无效信号；
- 3) 数据截断：将信号强度不在于设定阈值范围内的数据截断或替换为特定的值，以去除无效信号，例如当该轨迹点的径向距离、切向距离、瞬时速度跟相邻轨迹点均值相差超过 30% 时，将判定为错误的轨迹点进行删除。

b) 虚点信号的去除

毫米波雷达的虚点信号主要是由于道路环境中存在大量护栏、隧道壁等反射镜面引起的，去除方法主要包括多普勒滤波和多路径校正，如下所述：

- 1) 多普勒滤波：采用多普勒频率分析及滤波算法清洗多普勒频率异常的信号，去除毫米波雷达的虚点；
- 2) 多路径校正：分析信号的反射路径，检测是否存在多路径效应，通过校正或补偿多路径效应恢复真实的目标位置。

c) 漂移点去除

毫米波雷达的漂移点主要是由于目标位置信息发生短时间内剧烈变化引起，可

通过目标跟踪算法进行校正，以维持目标的连续性。

5.2.2.3 激光雷达点云数据清洗

激光雷达数据清洗主要是去除点云数据的噪声，主要包括噪声滤波、抖动校正、去除离群点及冗余点，如下所述：

- a) 激光雷达噪声滤波主要包括滑动窗口、中值滤波及高斯滤波等技术，以去除小尺度噪声；
- b) 抖动校正则可以通过运动补偿或插值方法来修正点云中的抖动问题，以提高点云的平滑度；
- c) 离群点去除是采用高斯滤波或半径滤波的方法，对点云数据进行平滑处理；
- d) 冗余点去除是通过体素滤波或最近点去重消除激光雷达扫描过程中由于重叠或多次采样引起点云中多个近似相同位置的点。

5.2.2.4 目标列表数据清洗

目标列表数据清洗主要是对目标检测传感器产生的目标列表中的缺失数据和噪声数据进行处理，如下所述：

- a) 缺失数据：当缺失率较低时，可以采用均值、中位数、众数统计模型补全缺失数据；当缺失率很高时，针对重要性低的数据可以采用忽略元组的方法，直接删除缺少属性值的场景数据，而针对重要性较高的数据可以采用模型驱动的修复方法、多源数据融合法及其他渠道补全法补全缺失数据；
- b) 噪声数据：清洗方法主要包括滤波法、移动平均法、回归法、异常值检测及处理。

5.3 场景标注技术

5.3.1 场景标注流程

场景标注流程主要包括标注标签定义、自动标注、人工校验及补充，如图 5.5

所示，其中：



图 5.5 场景标注流程

- a) 标签定义用于确定此次标注的场景所需标注的内容，包括但不限于交通参与者、道路几何、交通设施和自然环境类别的标签描述，应采用含义清晰且互斥的名词，不应使用含义相近的多个标签标记同一要素；
- b) 驾驶场景自动标注是基于机器学习算法，以人工搭建的仿真场景文件作为数据训练集，通过全监管学习的方式，训练算法自动化识别场景文件中所包含场景元素，并在场景文件中添加标签的形式，完成场景数据的标注；
- c) 人工校验及补充标注由数据回放、动态目标物补充标注、静态目标物补充标注和标注结果记录组成，实现自动驾驶测试场景采集数据回放，同时可参照场景标注标签定义校对修改自动标注错误的内容，并对自动标注无法标注的内容进行补充标注。

5.3.2 标注标签定义

5.3.2.1 交通参与者标签

交通参与者标签应描述包括主车在内的交通参与者的状态信息，应至少包含以下信息类别：

- a) 交通参与者的类别：包含车辆（乘用车、货车、客车、自行车、摩托车等）、行人（成人、儿童等）、动物；
- b) 交通参与者的位置：可表征交通参与者在道路中的位置信息。主车位置信息包括但不限于所在车道编号、相对道路边缘、车道中心线的距离等，其他交通参与者信息包括但不限于所在车道编号、相对道路边缘、车道中心线的距离、相对于主车的空间位置等；
- c) 交通参与者的行为：可表征交通参与者的运动状态，一般通过横向行为和

纵向行为描述。

5.3.2.2 道路几何标签

道路几何标签主要描述与道路形态有关的信息，应至少包含以下信息类别：

- a) 道路等级：道路的等级划分信息。道路的等级划分应该以国家或行业标准为基础，并结合相关使用途径和数据来源制定可选值，如：高速、城市快速路等；
- b) 道路结构：道路的形状特点信息的定性描述。包括但不限于：普通行车道、匝道、高速入口等。而对于高速入口这种道路连接处，则可以进一步进行相关结构标签等级的制定；
- c) 曲率：道路平面线形的具体信息。对于场景片段，为了使用的便捷，可以将曲率划分为不同的等级，并作为可选值；
- d) 坡度：道路垂向线形的具体信息。对于场景片段，为了使用的便捷，可以将坡度划分为不同的等级，或者区分为：上坡、下坡等，并作为可选值。

5.3.2.3 交通设施标签

交通设施标签主要描述道路上对交通参与者行为产生影响的相关设施的信息，应至少包含以下信息类别：

- a) 车道线：车道线定义了道路上车道的分布情况与交通参与者的运动限制条件。车道线标签一般进一步可以包含：车道线类型与车道线清晰度信息等；
- b) 交通标志：用文字或符号传递引导、限制、警告或指示信息的道路设施。交通标志信息一般进一步可以包含：交通标志类别、交通标志位置等。

5.3.2.5 自然环境标签

自然环境标签应该主要描述对交通参与者行为或者智能网联汽车产生影响的环境信息，应至少包含以下信息类别：

- a) 时间：描述场景发生的时刻或者时段。为了使用的便捷，可以根据太阳的

光照角度等信息区分为:黎明、白天、黄昏和夜晚等时间段;

- b) 天气: 描述场景发生时的气象状态。天气标签的可选值应该以国家标准或行业标准为基础, 并结合相关使用途径制定可选值, 如:晴天、雨天、雪天等;
- c) 能见度: 描述了交通参与者的最大视野范围。能见度的可选值应该以国家标准或行业标准为基础, 并结合相关使用途径制定可选值, 如: 1 级 (能见度 $>10\text{km}$)、2 级 ($2\text{km}\leq\text{能见度}<10\text{km}$) 等;
- d) 降水量: 描述了雨天、雪天等天气时的降水情况。降水量可选值应该以国家标准或行业标准为基础, 并结合相关使用途径制定可选值, 如根据具体降水量对道路交通的影响情况划分: 1 级 ($10\text{mm/h}\leq\text{降水量}\leq14.9\text{mm/h}$)、2 级 ($15\text{mm/h}\leq\text{降水量}\leq29.9\text{mm/h}$) 等。

5.3.3 自动标注方法

5.3.3.1 标注流程

自动标注流程主要包括车道线检测、目标检测、目标跟踪、目标位置标注、本车及目标动作标注, 如图 5.6 所示。



图 5.6 自动标注流程

5.3.3.2 车道线检测

车道线的检测主要包括以下 2 步: 一是提取几何或物理特征, 二是利用离散数据拟合成车道线, 目前使用的方法主要有基于图像的检测方法和基于激光雷达的检测方法:

- a) 基于图像的车道线检测方法主要包括 LaneNet 和 H-Net 两个网络模型, 其中 LaneNet 输出实例分割的结果, 为每个车道线像素分配一个车道线 ID,

H-Net 输出一个转换矩阵 H ，使用转换矩阵 H 对车道线像素进行修正，并对修正的结果拟合出一个三阶的多项式作为预测得到的车道线；

- b) 激光雷达检测车道线主要有四种方法：①基于激光雷达回波宽度；②基于激光雷达反射强度；③激光雷达 SLAM 与高精度地图配合，不仅检测车道线还进行自车定位；④利用激光雷达获取路沿高度信息或物理反射信息，再根据道路宽度推算出车道线位置。第一种方法适用于 4 线或单线激光雷达，后三种方法需要使用多线激光雷达，最少也是 16 线激光雷达。

5.3.3.3 目标检测

目标检测主要是获取交通参与者的类别、目标框的尺寸、目标框的位置及目标框的航向角度等信息，目前使用的方法主要有基于图像的目标检测法、基于激光雷达的目标检测法及图像与激光雷达联合目标检测法：

- a) 基于图像的目标检测法分为两种：单阶段法（YOLO、SSD、RetinaNet 等）和多阶段法（Fast RCNN、Faster RCNN、Cascade RCNN 等）。单阶段目标检测法只需一次提取特征即可实现目标检测，其速度相比多阶段的算法快，但一般精度稍微低一些；
- b) 基于激光雷达的 3D 目标检测法：根据其数据表示方法可以分为基于体素的表示法和基于原始点云的表示法。基于体素的表示方法主要有 PointPillar、VoxelNet 及 SECOND 等，基于原始点云的表示法主要有 PointNet、PointNet++、PV-RCNN 及 SAPV-RCNN 等；
- c) 图像与激光雷达联合目标检测主要分为 detection-level 和 point-level 融合方法。detection-level 即后期融合，融合形式简单，主要有 CLOCS；point-level 又称早期融合，利用相机功能增强了 LiDAR 点云，主要有 3D-CVF、MV3D、PointPainting 及 EPnet 等。

5.3.3.4 目标跟踪

目标跟踪主要是获取目标物的速度、跟踪 ID、运动轨迹、平滑物体的运动状态

等，分为三维目标跟踪和二维目标跟踪，跟踪流程主要如下：

- a) 目标检测预处理：根据得分高于某个阈值选择用于跟踪的边界框并删除多余的边界框；
- b) 运动预测：通常使用卡尔曼滤波（KF）或等速模型（CV）预测目标的运动轨迹及状态。前者例如 AB3DMOT，后者例如 CenterPoint；
- c) 关联匹配：将预测的轨迹和边框关联匹配起来，首先采用基于面积交并比（IoU）或基于距离进行相似度度量，然后使用匈牙利算法或通过贪婪算法迭代进行关联；
- d) 生命周期管理：通常采用基于计数或基于置信度（CBMOT）控制检测的边界框是否初始化为新的轨迹、边界框的轨迹是否应被删除及是否输出边界框的运动状态。

5.3.3.5 目标位置标注

目标相对本车位置标注原理如图 5.7 所示，逻辑标注规则如下：

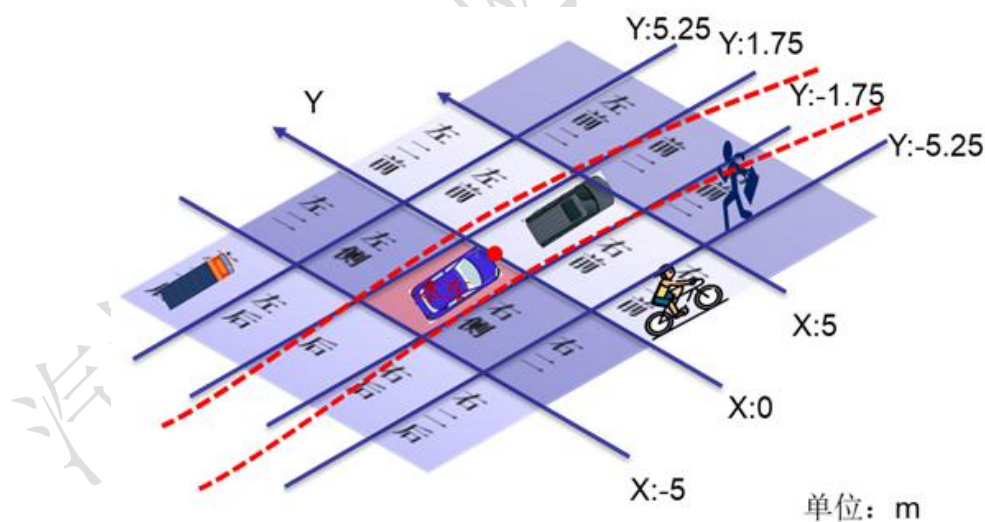


图 5.7 目标位置标注原理图

- a) 以本车车头中心为原点，建立坐标系，前进方向为 X 轴，正左侧为 Y 轴；
- b) 当车速大于 50km/h，方向盘转角 $>5^\circ$ ，则 X 方向边界线随本车轨迹半径进

行偏转;

- c) 当目标车辆中心点坐标 y 在相应坐标宫格范围内时, 则对车辆位置进行左右标注;
- d) 正前: 目标车中心点 $x_1 > 0$, $-1.75 < y < 1.75$;
- e) 正后: 目标车中心点 $x_1 < -5$, $-1.75 < y < 1.75$;
- f) 左侧: 目标车中心点 $-5 < x_1 < 0$, $1.75 < y$;
- g) 右侧: 目标车中心点 $-5 < x_1 < 0$, $y < -1.75$;
- h) 左前: $0 < x_1$, $1.75 < y < 5.25$;
- i) 右前: $0 < x_1$, $-5.25 < y < -1.75$ 。

5.3.3.6 本车及目标动作标注

通过识别本车及目标动作的特征参数变化, 选取特征参数的阈值变化范围制定标注规则确定本车及目标动作的起点和终点信息, 打上相关标签。本车及目标动作的特征参数主要包含方向盘转角、车道线曲率、航向角信息、横摆角速度、车速信息、车辆和车道线距离等, 自车掉头、自车转向及自车减速标注规则及阈值范围详见表 5.1。

表 5.1 本车动作标注规则示例

本车动作 •	标注规则	阈值范围
自车掉头	片段始末自车航向角变化值(单位: $^{\circ}$) 片段内自车横摆角速度(单位: $^{\circ}/s$)	横摆角 > 6 (或 < -6) 掉头航向角变化范围: $[170, 190]$
自车转向	片段始末自车航向角变化值(单位: $^{\circ}$) 片段内自车横摆角速度(单位: $^{\circ}/s$)	横摆角 > 6 (或 < -6) 转向航向角变化范围: $[80, 100]$
自车减速	自车纵向加速度(单位: m/s^2)	匀速: $[-1, 1]$ 加速: $[1, 10]$ 减速: $[-10, -1]$

基于以上规则, 可使用 Python 及 Spark 框架开发本车及目标动作自动化标注工具, 如图 5.8 所示。以自车右转向场景标注为例, 从采集的某批次自然驾驶场景中标注出 70 个自车右转场景, 经过人工审核, 其中 65 个为正确的自车右转场景, 准确率为 92.3%。

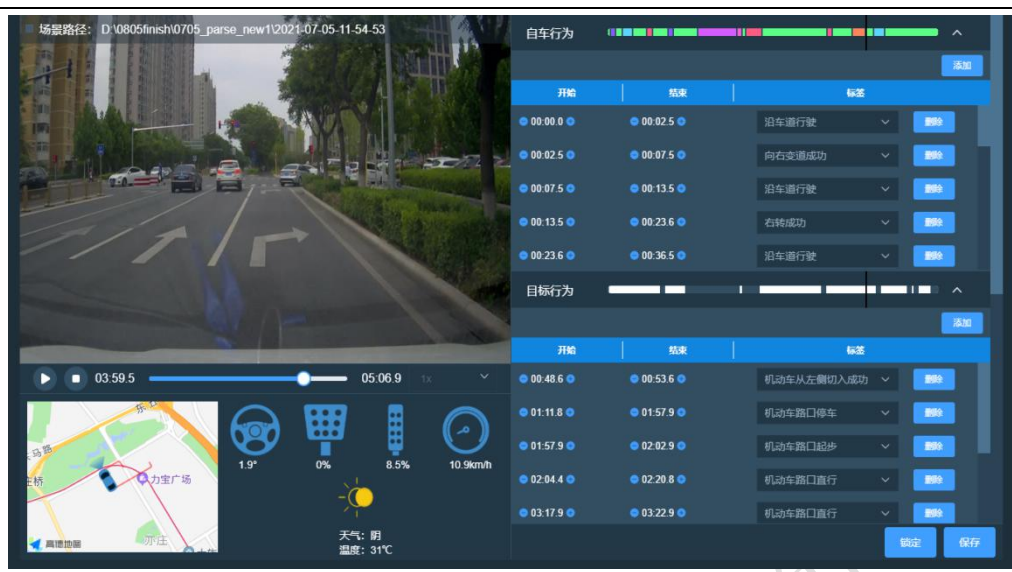


图 5.8 本车及目标动作标注结果

5.3.4 人工校验及补充标注要求

5.3.4.1 数据回放要求

人工校验及补充标注应能同步回放采集的测试场景数据，应满足以下要求：

a) 回放数据及回放形式详见表 5.2；

表 5.2 回放数据类型及形式

数据来源	数据类别	回放形式
车辆数据	CAN总线数据	图表
激光雷达	目标物数据	鸟瞰图
摄像头	视频数据	视频
麦克风	音频数据	声音
毫米波雷达	CAN总线数据	鸟瞰图
智能摄像头	CAN总线数据	鸟瞰图
组合导航	总线数据	图表

b) 每一回放数据应有多个独立窗口进行回放，回放界面可在显示器中任何区域进行拖拽、放大、缩小；

c) 每一回放数据的回放可实现暂停、快退、快进等基本播放功能；

5.4.4.2 动态目标物补充标注要求

动态目标物补充标注应建立初始索引表，以动态标注数据类别作为表头，为每一个动态标注数据类别在初始索引表中建立统一的时间轴，并与自然驾驶数据采集的时间轴保持一致。

动态目标物补充标注应有独立的可视化界面可对动态目标物进行标注，应包含以下几个区域：

- a) 标记修正区：该区域将周围车辆的行驶空间进行划分，与初始索引表建立映射关系，与数据回放同步时间轴；
- b) 位置选择区：该区域用于进行交通参与者对应位置的选择；
- c) 标注数据属性选择区：标注数据属性选择；
- d) 操作区：该区域可对标记框进行删除、切断等操作。

5.4.4.3 静态目标物补充标注要求

静态目标物补充标注应建立静态标注数据类别初始索引表，以静态标注数据类别作为表头，为每一个静态标注数据类别在初始索引表中建立统一的时间轴，并与自然驾驶数据采集的时间轴保持一致。

静态目标物补充标注应有独立的可视化界面，该界面应包含以下几个区域：

- a) 标记修正区：该区域将周围车辆按照静态标注数据类别进行划分，与初始索引表建立映射关系，与数据回放同步时间轴；
- b) 标注数据属性选择区：标注数据属性选择；
- c) 操作区：该区域可对标记框进行删除、切断等操作。

5.4 场景统计技术

5.4.1 场景统计流程

场景统计流程主要包括确定统计内容、场景数据提取、指标计算及结果描述，如图 5.9 所示。



图 5.9 场景统计流程图

5.4.2 确定统计内容

功能场景、逻辑场景及具体场景统计内容主要包括：

- a) 功能场景描述了最抽象的场景表示级别，针对该场景的统计内容主要为测试场景关键维度或他们之间关系的描述性统计，例如提取不同类型道路上跟随不同类型目标的概率分布；
- b) 逻辑场景是使用不同的参数范围来表示目标和目标之间的关系，例如跟随不同类型前方目标的停车距离分布。在该类分析中，对于连续性数据应使用 GB/T 3358.1 中分布的标准名称明确其分布类型，宜采用 5%分位、95%分位来提取边界，作为实际采集场景的参数范围；
- c) 具体场景是使用不同的参数值描述目标和目标之间的关系，例如前方目标为货车，本车为乘用车，跟随停车距离为 3.4m。针对该场景的分析，应在逻辑场景分析完成之后进行，其目的是基于边界值、高频值等要求来提取具有代表性的场景及其参数，作为测试用例提取的输入。

5.4.3 场景数据提取

5.4.3.1 总体要求

场景数据提取应遵循以下要求：

- a) 每个提取模型都应有对应的唯一编号，编号可参考“模型名称_大版本号_小版本号”的方式。当提取模型的条件发生改变时定义为大版本号更迭，当参数发生改变时定义为小版本号更迭；
- b) 被提取的每一段场景数据应有唯一索引，该索引可以由一个或多个字段组合而成；
- c) 被提取的每一段场景数据应含有原始数据的索引字段，用于溯源原始数据；

-
- d) 被提取的每一段场景数据应含有提取人名称或提取模型的编号，用于溯源场景生成方式；
 - e) 场景数据片段应保留完整动作过程，但非特征数据段不宜过长，允许在数据片段前后各有不超过 2s 的缓冲数据段。

5.4.3.2 提取方法

场景数据提取方法主要包括手动提取和自动提取，如下所述：

a) 手动提取方法

- 1) 分析原始数据，制定可量化的条件和参数，依此可确定所要提取场景的时间起点和时间终点；
- 2) 对照视频以及数据确认所要提取场景的时间起点和时间终点；
- 3) 按照 GB/T 2828.1 的方法对提取的场景片段进行二次抽样检验；
- 4) 当检验结果为接受时完成提取，并将提取人的唯一识别编号作为字段写入到对应的场景提取数据表中；
- 5) 当检验结果为拒绝时，应将该批场景全部重新标注。

b) 自动提取方法

- 1) 分析原始数据，制定可量化的条件和参数，依此可确定所要提取场景的时间起点和时间终点；
- 2) 编写场景自动化提取脚本或工具，对原始数据进行提取。在编写代码调试过程中，宜使用小批量数据。根据提取的场景时间起点和终点，对照视频确认特征是否符合预期，过程是否完整。若效果不佳则应重新调整其提取条件和参数；
- 3) 若处理全量数据的时间较短，可忽略此步骤。否则选择非调试阶段的数据源检验提取的场景片段是否符合预期，过程是否完整。当效果可接受时可进行下一步，否则应重新调整其提取条件和参数；
- 4) 使用提取模型对全量数据进行处理，并按照 GB/T 2828.1 的方法进行检验。当检验结果为接受时，完成提取，否则应重新调整其提取条件和

参数或提取模型；

- 5) 完成提取后，应将提取模型的编号作为字段写入到对应的场景提取数据表中。

5.4.4 场景指标计算

场景统计指标主要包括描述类和评估类。如下所述：

- a) 描述类指标主要是揭示场景各层级要素属性的取值范围、出现频次及占比等，如道路类型、驾驶员类型、本车动作等；
- b) 评估类指标主要是为了评估主车与单一目标物或场景元素发生碰撞的可能性，可以划分为时间指标、加速度指标和距离指标 3 类，如下所述：
 - 1) 时间指标通过主车到碰撞点的时间来表征风险，主要包括碰撞时间、修正碰撞时间及模型预测碰撞时间等；
 - 2) 加速度指标一般由车辆的极限加减速性能以及所在场景中车辆为避免碰撞所需的加减速性能确定，主要包括避撞减速率、制动威胁系数和转向威胁系数等；
 - 3) 距离指标通过主车到碰撞点的距离来表征风险，用于风险评估的距离指标本质上类似于时间度量，主要包括最小安全距离、碰撞距离等。

5.4.5 统计结果描述

5.4.5.1 文字类

使用文字描述统计结果时，应满足以下要求：

- a) 主要包括数据源描述、统计指标、趋势分析和关键数据点等；
- b) 由文字引出表格或图片时，应清晰标出表索引和图索引。

5.4.5.2 表类

使用表格描述数据统计结果时，应满足以下要求：

- a) 使用数据表格来呈现结构化数据。表格应包括列标题，列项类表格宜将首

列作为序号列，同列小数位数应统一，结果为“0”时，不应按照小数位要求补齐，应写为“0”；

- b) 对于多行项目相同的表格，应采取合并单元格的方式，减少页面重复内容，增加表格可读性；
- c) 对于大型数据集，应使用数据透视表进行数据分析和总结，包含聚合数据、汇总数据和计算字段；
- d) 表格数据宜与图表和图形结合使用，以便更直观地展示数据趋势和关联性。

5.4.5.3 图类

使用图描述数据统计结果时，应满足以下要求：

- a) 图像分析、地图数据和视觉数据应直接呈现图片或照片；
- b) 如果需要解释图片中的特定元素或区域，应使用图像标注工具来添加文本或箭头，以指示关键信息，应注意标记框或箭头应与图片底色呈高对比度；
- c) 不宜在图片上展示超过三维的信息；
- d) 基于表格数据绘制包括柱状图、折线图、散点图、环形图等。图应有图题，并且标记清楚各个轴所代表数据的含义以及单位，单位应与相关的表、结论一致。

5.5 测试用例提取技术

5.5.1 测试用例提取要求

5.5.1.1 测试用例提取模型要求

测试用例提取模型可以为代码、相关工程、配置文件等，应在代码内部或单独使用文件记录以下信息：

- a) 模型名称按照提取模型的主要任务，进行简要描述；
- b) 模型索引按照“智能驾驶功能缩写-功能场景编号-提取模型版本”的方式编写；

-
- c) 模型索引中功能场景编号，需要固定位数，自然序列递增，不足位用 0 补齐；
 - d) 对输入数据的格式、名称、频率等要求的描述；
 - e) 编写日期；
 - f) 改版日期及变更说明。

5.5.1.2 测试用例输出要求

输出的测试用例应满足以下要求：

- a) 可量化：测试用例中包含的各类要素所体现的特征应是可被量化的，例如道路要素的几何尺寸、线性特征，设施要素、临时要素的大小、位置，交通参与者要素的初始状态、行为，环境要素的温度、光照度、光照角度；
- b) 可执行：本车需要采取的动作、以及执行动作的时间点应清晰，对可量化的要素特征，应提供合理的误差范围或使用区间；
- c) 可溯源：应确认溯源信息是否能追溯到原始数据和数据分析模型；
- d) 可实现：提取的测试用例在真实驾驶环境中存在对应场景；
- e) 可评价：自然驾驶场景数据包含了驾驶员对不同驾驶场景的处理反馈，例如在不同弯道下的行驶速度、在不同车道与不同类型边界或目标的横向距离等，该量化结果可以作为评价智能驾驶产品控制的标准输入；
- f) 结构化：输出文件应使用结构化或半结构化的格式存储，例如二维表、可扩展标记语言。

5.5.1.3 测试用例条目要求

测试用例应包含以下条目：

- a) 输入数据的索引，可用于溯源输入数据的索引，例如库名称和表名称的组合信息、文件路径信息、参考文献索引等；
- b) 提取模型的索引；
- c) 测试场景名称；

-
- d) 测试用例编号，在模型索引的基础上，忽略其版本，提取的测试用例应编号递增；
 - e) 要素名称及其值；
 - f) 激活条件；
 - g) 满足激活条件时需要执行的操作；
 - h) 结束条件；
 - i) 评价指标或期望结果。

5.5.2 测试用例提取流程

5.5.2.1 明确目的

根据验证方向的不同，具体要素值的取值要求如下：

- a) 功能的用户边界场景验证：以自然驾驶场景覆盖范围的限值为基准，按照分布边界为目标，确定要素值，边界场景包含了复杂场景和危险场景；
- b) 功能的用户常见场景验证：以自然驾驶场景覆盖范围的限值为基准，按照分布最高频为目标，确定要素值；
- c) 功能的极限性能验证：以功能设计运行域为基础，确定提取要素及值，其要素值应在功能设计运行域范围内接近边界；
- d) 功能的边界逻辑验证：以功能设计运行域为基础，确定提取要素及值，其要素值应设置为跨越设计运行域边界值的一个范围，并标注出初始状态和结束状态，使单条测试用例中有且仅有一个要素值按此设计；该类测试用例应 2 个为 1 组进行设计，分别对应由设计运行域内到设计运行域外和由设计运行域外到设计运行域内。

5.5.2.2 测试场景选择

目前采集的驾驶场景中行车包含跟车场景、换道场景、超车场景、并入场景及驶离场景，泊车包含搜索车位场景、泊入车位场景、驶出车位场景。将自动驾驶功

能与驾驶场景相结合形成测试场景，例如自适应巡航跟车场景、智能泊车辅助功能泊入车位场景。

5.5.2.3 要素确认与值设定

确认选择构成测试用例的要素，例如弯道曲率半径、光照度、车道线、目标车速等。该环节中需要结合智能驾驶功能或自动驾驶系统的设计运行域进行。

5.5.2.4 测试用例优化

实车测试中不应使用所有要素值全交组合形成测试用例，需要基于危险度和复杂度，按照危险兼容安全、复杂兼容简单的原则对测试用例进行删减。

5.5.2.5 激活条件设定

部分需要人工操作来激活的应明确进行操作的时间点，例如“本车距离前车纵向距离 20 m 时”。

5.5.2.6 期望或评价指标设定

在具体场景下，相同条件由人类驾驶员控制车辆获得的数据可以作为期望或评价指标设定的参考。

第六章 自动驾驶系统测试场景数据采集和分析标准化建议

6.1 自动驾驶系统测试场景数据采集和分析标准化框架建议

本文着重研究了自动驾驶系统测试场景数据采集和分析技术，旨在建立具有科学性、全面性和系统性的标准。根据本文第四、五章的研究成果，提出标准框架，如图 6.1，主要包括驾驶场景大数据采集和分析两个部分。

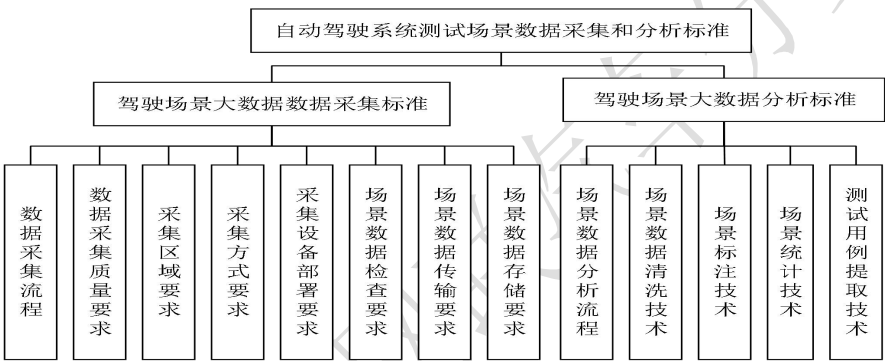


图 6.1 自动驾驶系统测试场景数据采集和分析标准框架

按照该框架，本标准基于数据驱动自动驾驶研发的主体路线，针对测试场景采集及分析，重点提出场景需求定义、采集方案设计、采集实施、大数据处理及测试用例提取五步法，如图 6.2，并结合 ISO3450X 系列标准，给出了场景采集、分析的要求及方法，这将对自动驾驶系统测试场景标准化工作开展起到基础性作用。



图 6.2 自动驾驶测试场景采集及分析标准化技术路线

6.2 自动驾驶系统测试场景数据采集和分析标准化需求研究总结

本文调研了自动驾驶技术与产品发展现状和发展趋势，分析了驾驶场景库建设对自动驾驶技术及产品开发、测试的重要意义。研究表明，测试场景数据贯穿自动驾驶技术及产品开发、测试、品质评价、示范及商业化运营等整个产业发展全过程，因此测试场景数据采集及分析标准化可快速推动自动驾驶汽车产业发展。

在此基础上，本文分析了自动驾驶系统测试场景数据采集和分析技术已开展的研究实践及应用情况，阐明了目前驾驶场景数据采集、理解、标注、提取及生成等已相对成熟，但缺少标准方法及规范，不利于场景库的国家发展及行业建设。本文还从我国自动驾驶测试地方政策、国内外测试相关标准及采集分析标准，进一步说明了数据采集分析国家标准缺失，应形成统一的标准，指导数据采集及分析，以此更好的支撑自动驾驶测试过程及结果的标准化及规范化。

本文重点研究了自动驾驶系统测试场景数据采集及分析的流程、要求及方法等，分别从采集和分析两个维度给出了标准化建议。关于测试场景大数据采集技术，首次提出采集范围、方式、区域选取、设备标定、数据质量、数据检查、数据上传及数据存储等要求及方法；针对测试场景大数据分析技术，以数据分析总体流程为依托，深度构建场景大数据清洗、场景标注、场景统计及测试用例提取等关键技术方法；最后，建立了测试场景数据采集和分析标准框架及技术路线，为进一步开展自动驾驶系统测试场景数据采集和分析标准化工作奠定了基础。

参考文献

- [1] Representatives of China, European Union, Japan and the United States of America. (2019). Proposal for Amendments to ECE/TRANS/WP.29/2019/34 Framework Document on Automated/Autonomous Vehicles (Levels 3 and Higher)[Informal Document], Geneva: United Nations.
- [2] United Nations Economic Commission for Europe. (2018). UN Regulation No.157-Automated Lane Keeping Systems (ALKS). Geneva: United Nations.
- [3] ChinaEV100. 德国自动驾驶法律法规借鉴 [EB/OL]. (2020-08-15). <https://www.yoojia.com/article/9402725631857198215.html>.
- [4] 张韬略, 钱榕. 迈入无人驾驶时代的德国道路交通法——德国《自动驾驶法》的探索与启示[J]. 德国研究, 2022, 37(1): 85-101.
- [5] 腾讯研究院. 瞄准 2025 年正式商用, 英国最新自动驾驶汽车政策意味着什么[EB/OL]. (2022-09-06). <https://36kr.com/p/1903608670890369>.
- [6] 赫荣亮, 关兵. 美国《自动驾驶汽车综合计划》三点启示[J]. 智能网联汽车, 2021(3): 61-63.
- [7] 智车派新闻. 修订交通法规 日本拟 2023 年允许 L4 级自动驾驶车上路[EB/OL]. (2022-10-28). https://www.sohu.com/a/600474346_120030907.
- [8] 环球网. 韩国发布自动驾驶汽车安全标准 预计 2020 年 7 月开始实施[EB/OL]. (2020-01-17). <https://www.861718.com/fagui/show-1295.html>.
- [9] 国家发展改革委. 关于印发《智能汽车创新发展战略》的通知[EB/OL]. (2020-02-10). https://www.ndrc.gov.cn/xxgk/zcfb/tz/202002/t20200224_1221077.html.
- [10] 工业和信息化部. 公开征求对《关于开展智能网联汽车准入和上路通行试点工作的通知(征求意见稿)》的意见[EB/OL]. (2022-11-02). https://www.miit.gov.cn/gzcy/yjzj/art/2022/art_4ae46de7edee4a72adb611b3c67b9d6e.html.
- [11] 盖世汽车. 美国首家 奔驰脱手驾驶获批在内华达州上路. [EB/OL]. (2023-01-08). https://auto.sina.com.cn/zz/hy/2023-01-08/detail-imxzisvm5568738.shtml?oaid=00101&c=spr_automato_trackid_4d7d188c9b7a61c4.
- [12] 曾德松, 高琛, 谭北海, 等. 自动驾驶仿真测试场景库构建研究[J]. 现代计算机, 2021, 27(31):13.
- [13] 九章智驾. 一文读懂自动驾驶仿真测试场景和场景库[EB/OL]. (2021-10-10). https://zhuanlan.zhihu.com/p/399032822?utm_id=0.
- [14] 数据集处理. KITTY 数据集[EB/OL]. (2022-6-07). https://blog.csdn.net/weixin_36670529/article/details/103774700.
- [15] yeshusheng. 一文解析 Waymo 的自动驾驶训练数据集 WOD[EB/OL]. (2022-10-26). http://www.360doc.com/content/22/1026/21/30375878_1053410890.shtml.
- [16] 整数智能 AI 研究院. 最全自动驾驶数据集分享系列一[EB/OL]. (2022-3-10). <https://blog.csdn.net/molardata/article/details/123406244?spm=1001.2014.3001.5506>.
- [17] 我爱研发网. Cityscapes 数据集的深度完整解析[EB/OL]. (2022-10-7). <http://www.taodudu.cc/news/show-4911652.html?action=onlick>.

- [18] Krajewski R, Bock J, Kloecker L, et al. The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems[C]//2018 21st International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2018.
- [19] 太平洋汽车. 百度 Apollo 加入 DeepDrive 自动驾驶联盟 [EB/OL]. (2022-10-7). <https://www.pcauto.com.cn/jxwd/2839/28391925.html>.
- [20] 邓伟文, 李江坤, 任秉韬等. 面向自动驾驶的仿真场景自动生成方法综述[J]. 中国公路学报, 2022, 35(01): 316-333.
- [21] 佐思汽车研究. 奔驰的 Cityscape[EB/OL]. (2022-10-19). <https://www.dongchedi.com/article/7156056479827624486>.
- [22] 高振海, 于桐, 孙天骏等. 面向无人驾驶的数据采集与分析系统研究综述[J]. 汽车技术, 2021(06): 1-11.
- [23] 格物钛. 详解 KITTY 数据集 [EB/OL]. (2020-11-30). <https://www.graviti.cn/article/xiang-jie-kittishu-ju-ji>.
- [24] Caesar H, Bankiti V, Lang A H, et al. nuScenes: A multimodal dataset for autonomous driving[J]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
- [25] 解析 Waymo 的自动驾驶训练数据集 WOD[EB/OL]. (2022-2-28). <https://new.qq.com/rain/a/20221026A010G800>.
- [26] 王润民, 朱宇, 赵祥模等. 自动驾驶测试场景研究进展[J]. 交通运输工程学报, 2021, 21(02): 21-37.
- [27] 曾德松, 高琛, 谭北海等. 自动驾驶仿真测试场景库构建研究[J]. 现代计算机, 2021, 27(31): 1-12+32.
- [28] 朱冰, 张培兴, 赵健等. 基于场景的自动驾驶汽车虚拟测试研究进展[J]. 中国公路学报, 2019, 32(06): 1-19.
- [29] Rocklage, Elias, Karatas, et al. Automated Scenario Generation for Regression Testing of Autonomous Vehicles[C]//IEEE. Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems. New York: IEEE, 2016: 476-483.
- [30] Apollo 智能驾驶. 轻成本、重体验的地图对自动驾驶至关重要[EB/OL]. (2022-12-08). <https://baijiahao.baidu.com/s?id=1751631679612469559&wfr=spider&for=pc>.
- [31] 刘颖, 贺锦鹏, 刘卫国等. 自动紧急制动系统行人测试场景的研究[J]. 汽车技术, 2014(03): 35-39.
- [32] 朱宇, 赵祥模, 徐志刚等. 基于蒙特卡洛模拟的无人车高速公路变道虚拟测试场景自动生成算法[J]. 中国公路学报, 2022, 35(03): 89-100.