

自动驾驶开源数据体系：现状与未来

李弘扬^{1*Θ}, 李阳^{1Θ}, 王晖杰^{1Θ}, 曾嘉^{1Θ}, 徐慧琳¹, 蔡品隆¹,
陈立¹, 严骏驰², 徐丰³, 熊璐⁴, 王井东^{5†}, 朱福堂^{6†}, 颜凯^{7†},
许春景^{8†}, 汪天才^{9†}, 夏飞^{10†}, 穆北鹏^{10†}, 彭志辉^{11†}, 林达华¹, 乔宇^{1*}

1. 上海人工智能实验室, 上海 200232

2. 上海交通大学, 上海 200240

3. 复旦大学, 上海 200433

4. 同济大学, 上海 200092

5. 百度, 北京 100085

6. 比亚迪, 深圳 518118

7. 长安汽车, 重庆 400000

8. 华为, 深圳 518129

9. 旷视科技, 北京 100096

10. 美团, 北京 510030

11. 智元机器人, 上海 201315

* 通信作者. E-mail: hy@opendrivelab.com, qiaoyu@pjlab.org.cn

作者分工参见文末。Θ 共同第一作者。†同等合作贡献, 按机构拼音首字母排序。

本工作受科技创新 2030 “新一代人工智能”重大项目（2022ZD0160104）、国家自然科学基金青年项目（62206172）、国家自然科学基金重大研究计划重点项目（92370201）、国家自然科学基金优秀青年项目（62222607）、上海市启明星计划（22QA1412500）、中国博士后科学基金（2023M741848）、上海市青年科技英才扬帆计划（23YF1462000）资助。

摘要 随着自动驾驶技术的不断成熟与应用, 系统性梳理开源自动驾驶数据集有利于产业生态良性循环。现有自动驾驶数据集可大致分为两代, 第一代数据集的传感模态复杂度相对较低、数据集规模相对较小, 且大多局限于感知级任务, 以发布于 2012 年的 KITTI 为代表。相比于第一代数据集, 第二代数据集的特征为传感模态复杂度较高、数据集规模与多样性较丰富、所设置任务从感知扩展到预测、规控上, 以 2019 年前后提出的 nuScenes、Waymo 为代表。本评述联合学术界、产业界同仁, 首次系统性梳理了国内外七十余种开源自动驾驶数据集, 对如何构建高质量数据集、数据在算法闭环体系中发挥的核心作用、如何利用生成式大模型规模化生产数据等进行了总结。此外, 就未来第三代自动驾驶数据集应该具备的特质和数据规模, 以及需要解决的科学与技术问题, 我们进行了详细分析与讨论。希望本文的归纳与展望能促进新一代自动驾驶数据集与生态体系的建设、推动关键领域自主原创与科技自强的发展。更多内容请参见 <https://github.com/OpenDriveLab/DriveAGI>。

关键词 自动驾驶, 数据算法闭环, 大模型, 数据集与挑战赛

引用格式: 李弘扬, 李阳, 王晖杰, 曾嘉, 等. 自动驾驶开源数据体系: 现状与未来. 中国科学: 信息科学, 在审文章

H. Li, Y. Li, H. Wang, J. Zeng, et al. Open-sourced Data Ecosystem in Autonomous Driving: the Present and Future (in Chinese). Sci Sin Inform, for review

1 引言

当前发展人工智能已经成全球之势, 引发了新一轮国际竞争, 许多国家已将其上升为国家战略^[1~3]。习近平总书记曾多次强调^[4], 人工智能是新一轮科技革命和产业变革的重要驱动力量, 加快发展新一代人工智能是我们赢得全球科技竞争主动权的重要战略抓手, 是推动我国科技跨越发展、产业优化升级、生产力整体跃升的重要战略资源。国务院在《新一代人工智能发展规划》^[1]中明确, 以人工智能技术突破带动国家创新能力全面提升, 引领建设世界科技强国进程, 其中凸显了以自动驾驶为代表的智能运载工具的关键性作用, 自动驾驶技术的突破有利于加快人工智能关键技术转化应用和促进技术集成与商业模式创新。

作为人工智能重要应用领域之一, 自动驾驶有望重塑现有的交通和运输模式, 大大提高交通效率和安全性, 深刻影响着未来的城市和社会发展。中共中央、国务院印发《交通强国建设纲要》^[5]明确, 加强智能网联汽车包括智能汽车、自动驾驶、车路协同等在内的研发, 形成自主、可控、完整的产业链。现如今, 我国智能网联汽车产业迈入了商业化的试水和起步阶段, 道路测试和示范应用场景趋于成熟, 自动驾驶功能技术加速迭代, 车联网应用场景日益丰富, 各层面相关法规政策加速出台^[6], 共同推动市场进入高速发展期。自动驾驶技术需要大量数据来训练算法模型, 以识别和理解道路环境, 从而做出正确的决策和行动, 实现准确、稳定和安全的驾驶体验。自动驾驶落地当前有诸多难题亟待解决, 需要产业界与学术界共同努力, 数据集扮演着产业界与学术界之间的桥梁角色。数据为算法发展提供了基础, 算法的发展也在进一步挖掘数据的价值并促使数据本身的演进。**数据的建设对于自动驾驶技术的发展至关重要**。以美国特斯拉公司为例, 截至 2023 年第二季度, 其完全自动驾驶测试版系统 (Full Self-Driving Beta) 的总行驶里程已达到前所未有的 3 亿英里 (约 4.83 亿公里), 并将随着测试用户规模的扩大实现指数增长。特斯拉所积累的海量数据和驾驶场景是保持其算法优势的重要原因。另一方面, **自然语言处理和通用视觉领域大模型的出现, 更加印证了海量高质量数据的重要性, 给予自动驾驶的数据集建设以启发**。

图 1 展示了本文主要内容与组织架构。第二章从自动驾驶任务侧, 将数据集分为感知类、建图类、预测与规划类数据集, 分别介绍各种类型数据集现状及发展, 并围绕社区生态以及相关国际赛事展开数据集影响力方面的讨论。第三章面向数据算法闭环体系, 对比多家商业解决方案, 针对数据算法闭环中数据标注与质量把控、仿真技术、基于大模型的数据生成以及自动标注技术环节进行阐述。第四章就新一代数据集应有的特性进行展望。

本文系统地梳理了现有的自动驾驶数据集。图 2 展示了公开数据集按照时间发展的脉络与分类 (见第 2.1 节)。该图以数据集发布时间为横轴、数据集影响力估计值为纵轴, 其中数据集影响力估计值由数据集质量与数据生态决定, 具体定义在 2.2 节中说明。根据不同数据集的传感模态复杂度、数据规模与场景丰富度、任务多样性等综合因素, 我们把目前开源的近百种数据集分为两代: 第一代数据集以 KITTI^[7]为标志, 该数据集于 2012 年提出, 输入传感模态由前视摄像头与激光雷达构成, 并提出了一系列综合感知任务。第二代数据集由 nuScenes^[8] 及 Waymo^[9] 数据集为代表, 传感模态复杂度变高, 环视相机、激光雷达、定位信息以及高精地图成为常见组成部分, 下游任务面向感知、建图、预测与路径规划综合任务。随着对自动驾驶任务认知不断加深, 数据集建设呈现传感模态不断丰富、数据集规模和场景丰富度不断扩充、由单一感知任务演变为感知及决策综合性任务的发展趋势。

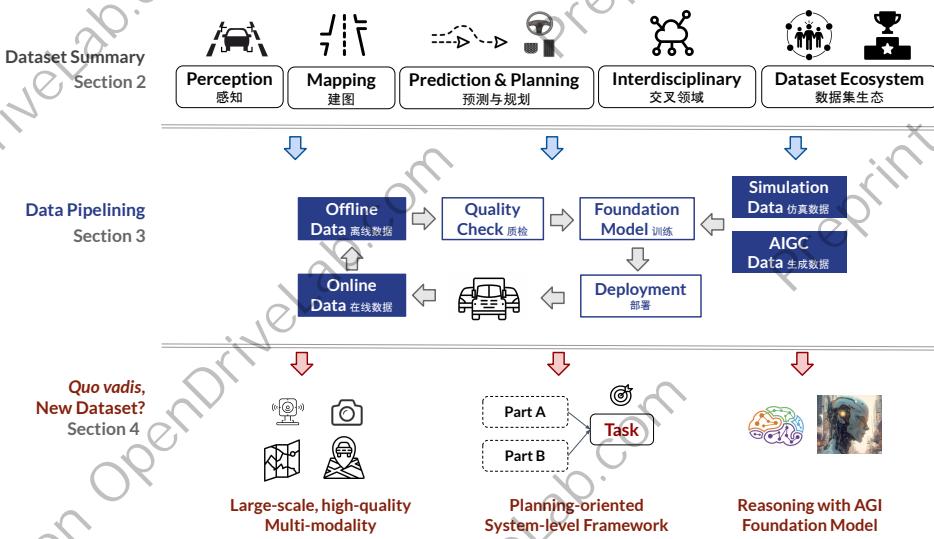


图 1 本文整体架构与内容分布。第 2 章总结现有自动驾驶数据集的情况与发展脉络；第 3 章介绍数据算法闭环体系以及数据的作用；第 4 章提出新一代数据集建设要点。

Figure 1 Layout of this survey. Section 2 describes the details of approximately 70 existing autonomous driving datasets in the wild. Section 3 depicts the data engine solution, and elaborates on how data plays a vital role. Section 4 delineates key essentials for building the next-generation dataset.

随着自然语言和视觉大模型的迅速发展，自动驾驶数据集的建设迎来了新的机遇和挑战。传统的数据集搭建需要经过多传感器标定、数据采集、数据标注与清洗等多个步骤。传统数据搭建方案存在采集成本过高、过度依赖人力标注、标注质量难以把控等问题，如何有效提高数据质量并降低数据标注成本逐渐成为数据侧的新研究热点。近年来出现新的数据集搭建方法，如人工智能生成内容（Artificial Intelligence Generated Content, AIGC），该技术可以生成逼真的城市道路环境、多类型的交通参与者、道路标志标线、天气状态等多种数据元素，构建出逼真的自动驾驶场景，用于面向多流程自动驾驶任务的算法测试与评估。通过 AIGC 技术获得的合成数据不受地点、时间和气候等条件的限制，可以覆盖更广泛的场景和环境，尤其是罕见但十分重要的危险场景。因此，通过 AIGC 生成自动驾驶数据可以大大降低数据获取成本和缩短自动驾驶研发流程。然而，AIGC 生成的虚拟环境与真实世界的差异性仍然较大，生成数据的质量和适用性需重点关注。

自动驾驶技术研究已经步入深水区，面向集成化多任务、复杂人机交互、多类型驾驶任务的应用，对新一代的自动驾驶数据在数据体量、质量、精细化加工等方面提出了更高的要求。通过第二章和第三章中对现有数据集与技术的总结分析，我们认为新一代自动驾驶数据集需要具备以下特点：(1) 传感器类别全覆盖，数据量充足并全面覆盖各种自动驾驶场景，并具有高质量的原始数据与标注；(2) 数据组织形式灵活多样，可以覆盖自动驾驶短中长期发展需求，并能够充分支持端到端框架、世界模型等任务新范式；(3) 面向智能化，能够赋能对自动驾驶系统可解释性的检验，并具有语言等支持逻辑推理的数据。

综上，本文贡献如下：(1) 全面分析现有自动驾驶数据集分布、侧重点等，提出数据集综合性影响力估计指标，并对现有数据集归纳总结；(2) 分析搭建高质量数据集关键要素与核心科学技术问题；(3) 联合产业界，分析实际自动驾驶诉求，展望与规划大模型背景下新一代自动驾驶数据集。

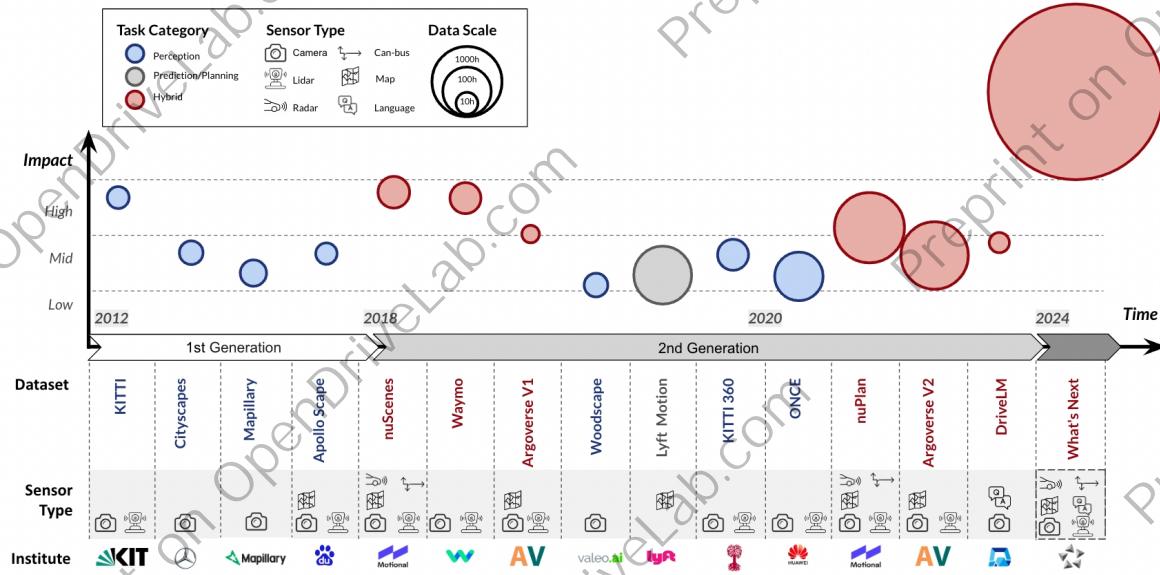


图 2 自动驾驶数据集影响力估计随时间发展的脉络图。随着人工智能技术的快速迭代，自动驾驶数据集趋于大规模、多模态、多任务、端到端的发展态势。希望本文能促进产业与学术的深度融合，建立跨机构间的“星空联盟”，打造大模型背景下的新一代数据体系，发挥海量高品质数据对自动驾驶技术在社会效益中的促进作用。

Figure 2 A chronological roadmap of popular datasets in terms of Impact since the 2010s. We define the Impact (y-axis) of a dataset based on sensor configuration, input modality, task category, data scale, ecosystem, etc.

2 自动驾驶数据集

2.1 整体概述

本文统计 2012 年至今近百种数据集，图 2 选取了其中影响力较高的 14 个数据集进行展示，其以数据集发布时间为横轴、数据集影响力估计值为纵轴，定性展示了主要公开数据集的情况。其中纵轴的影响力估计值，依据第 2.2 节定义的数据集影响力估计评价指标，由数据质量与行业生态两个维度决定。由图 2 可知，其中 KITTI、nuScenes、Waymo 这三个数据集影响力估计值处在第一梯队。

传感器模态复杂度逐渐提高。2012 年 A. Geiger 等人发布了 KITTI^[7] 数据集，该数据集模拟司机的视角，采用前视相机与激光雷达作为传感器，并定义了目标检测、深度估计以及目标跟踪等任务。随后陆续发布的一些自动驾驶数据集^[10~14]，尽管在数据量与任务定义上不同，但是都采用了前视相机并保持一致的任务范式。为了克服前视相机下因观测范围受限而导致驾车盲区问题，一些数据集^[8, 9, 15, 16] 在前视相机基础上新增了多个侧视与后视相机，实现环视场景理解，进一步增大了行车过程中的观测范围。这些数据集同样配备了激光雷达传感器，以赋能给予激光雷达的检测、分割等研究以及相机图像与激光雷达的跨模态融合研究。由于激光雷达传感器成本较高且采集频率低，成本低廉且采集频率高的声波雷达传感器受到了部分数据集青睐^[8, 17]。全球定位系统（Global Positioning System, GPS）、惯性测量单元（Inertial Measurement Unit, IMU）与高精地图（High-definition Map, HD Map）因能较好地实现定位功能与导航功能，能为车辆提供精确的位置参考和丰富的道路环境信息，逐渐成为数据集重要组成部分^[16, 18, 19]。在环视相机、激光雷达、高精地图、超

声波雷达传感器成为打造自动驾驶功能的基本组成部分的同时,也出现配备新传感器的数据集,如鱼眼相机^[10]、双目相机^[20]及场景文本信息^[21],这些数据集旨在降低量产成本,降低算法设计难度,提高算法可解释性。自动驾驶形成了以环视相机、激光雷达为主流传感器配置,并提供高精地图信息作为辅助的新格局。随着自动驾驶数据集涉及的传感器总类增多,而各个模态的数据采集频率通常不同,各传感模态间的统一对齐是个需要关注的工程问题。

数据集规模与多样性日益增长。不同于封闭场景下的机器人视角,自动驾驶场景下的感知是一个开放域问题,开放场景下自动驾驶系统的泛化表现依赖数据集规模与多样性。数据的丰富度是决定自动驾驶能力的关键因素之一,数据采集时间长度与数据集场景多样性是衡量数据集丰富度的两个重要维度。由图2可以看出,主流自动驾驶数据集的采集时长由最初的10小时左右逐渐提升至100小时,随着自动标注技术及标注工具的演进,近些年也出现了超过1000小时的数据集^[15, 18, 22]。除了数据丰富度,驾驶场景的多样性是自动驾驶系统表现的另一关键因素。为了提高算法在特定场景下的表现能力,Reasonable Crowd^[23]数据集采集了多个季节和极端天气下人车混行的复杂场景。为增强算法在不同地域的泛化能力,Mapillary^[11, 20]等数据集分别在多个大洲多个城市进行采集。

数据集任务从感知延伸至预测与规划。2016年推出的Cityscapes^[10]与Mapillary^[11]等数据集下游任务聚焦于动态物体检测。2019年推出的SemanticKITTI^[24]、DrivingStereo^[25]等数据集引入语义分割、深度估计、光流估计等任务。在传统预测与规划模块一般应用数值计算、优化、搜索等方法求解。在落地过程中,研究者们认识到传统方案存在性能上的瓶颈,开始将数据驱动的范式引入预测与规划模块。2019年前后提出的nuScenes^[8]、Waymo^[9]、Argoverse V2^[19]等数据集,不仅包括感知任务还涵盖预测与规划任务,实现了在同一数据集上进行多种任务研究,同时引领社区在传统多个模块范式下端到端自动驾驶研究的潮流。

总的来说,KITTI^[7]集合了多传感器并定义自动驾驶任务范式,实现了自动驾驶领域初步尝试,引领第一代自动驾驶数据集潮流。面向落地场景下出现的行车安全与可靠性问题,以nuScenes与Waymo为首的数据集,极大的提高了数据丰富度并确定了传感器范式。时至今日,随着自动驾驶领域的快速发展,场景丰富性极大提高,传感器配置趋向多样化,但在实际工程中依然存在一些问题:(1)因模块化设计框架,各模块优化方向不一致导致最终驾驶性能无法达到全局最优;(2)针对长尾场景,数据样例采集困难的问题亟待解决。

2.2 数据集影响力估计指标

我们提出自动驾驶数据集影响力估计值的评价指标,目的在于评估数据集的可用性、准确性和适用性,确保数据集能够有效地用于自动驾驶算法的开发、训练和评估。目前公开文献中,这类指标较少,而数据集众多,为方便相关从业者针对研究问题甄选目标数据集,本文提出数据集影响力估计评价指标。本文定义了一个数据集的影响力(Impact)指标,由数据质量评分(Q_s)与行业生态评分(C_s)共同构成,如式1所示,其中 Q_s 与 C_s 定义参见式2与式3:

$$Impact = Q_s + C_s \quad (1)$$

$$Q_s = \frac{NOC}{5} \cdot \frac{NST - NST_{min}}{NST_{max} - NST_{min}} \cdot \frac{\lg DSL}{\lg DSL_{max}} \cdot \frac{\lg NOS}{\lg NOS_{max}} \quad (2)$$

$$C_s = \mathbb{1}(NOCH) + \mathbb{1}(NOLB) \quad (3)$$

式 2 中 NOC (Number Of Continents) 为数据集分布的大洲数量, NST (Number of Sensor Types) 表示所使用的传感器类型数量, 例如相机、激光雷达、超声波传感器等, DSL 表示驾驶数据集的总时长, 以小时为单位。 NOS (Number of Senses) 表示场景的数量。式 3 中 $NOCH$ (Number of Challenges) 表示举办相关比赛次数。 $NOLB$ (Number of Leaderboards) 表示维护的任务榜单数量。函数 $\mathbb{1}(\cdot)$ 表示当 · 大于等于 1 时返回 1, 小于 1 时返回 0。

行业生态评分与是否举办比赛以及是否维护相关任务榜单成正相关, 数据质量评分与数据集所涵盖大洲数量、传感器种类数量以及场景数量成正相关, 其评分由各自的归一化的乘积表示。在以数据驱动的感知任务范式下, 传感器种类越多, 场景越丰富, 场景数量越多, 一方面对于下游任务领域可以提供更加广阔的尝试空间, 另一方面也会提高相应算法鲁棒性的上限。

依据数据集影响力估计评分, 我们将现有公开数据集划分为低、中、高三个档位。其中低档位的影响力估计评分取值为 $[0, 1]$, 中档位取值为 $[1, 2]$, 高档位取值为 $[2, +\infty]$, 影响力信息在表 2 中呈现。例如, 虽然 KITTI-360 [26]、ONCE [15]、nuScenes [8] 等数据集都具有相似的传感器配置 (环视相机及激光雷达) 和公开的任务榜单, 它们对于社区的影响力却大不相同。nuScenes 数据集的影响力估计评分最高, KITTI-360 和 ONCE 数据集的估计评分较之逊色不少, KITTI-360 的影响力又要略高于 ONCE 数据集。此外, 我们还统计了各个数据集的 Github 代码库标星量与 Google Scholar 被引用量的情况, 这两个指标反应了相应社区研究人员关注度。截止论文撰写时, nuScenes [8]、KITTI-360 [26]、ONCE [15] 三个数据集 Github 代码库点赞量分别为 1900、314、73, 相应论文 Google Scholar 引用量分别为 3244、172、108。通过数值对比发现, 各个数据集的 Github 代码库点赞量和 Google Scholar 引用量与上述影响力估计评分大小关系是一致的, 这说明了本文所提出的影响力估计评分能够较为客观地反映出数据集的真实影响力。

2.3 国际挑战赛与榜单

表 1 列出了自动驾驶领域近几年主要的国际、国内挑战赛与榜单情况。从表中可以看出, 2012 年推出的 KITTI [7] 数据集包含了种类丰富的任务, 如深度估计、光流估计、目标追踪等, 覆盖了大部分自动驾驶相关的感知任务, 同时传感器类型也包含双目相机和激光雷达。KITTI [7] 数据集在推出至今的十几年时间里, 一直受到社区的广泛关注, 为自动驾驶数据集的发展夯实了基础。从第二代数据集开始, 如代表性的 nuScenes [8] 与 Waymo [9] 数据集, 环视相机成为标配, 愈来愈多的数据如高精地图等也被包含在数据集中。数据集相关任务也逐渐往预测方向发展, 如物体轨迹预测等。由于自动驾驶系统链路长, 复杂度高等因素, 相关比赛一直停留在感知或预测等模块中。但近年来, 随着不同自动驾驶挑战赛逐渐成为每年的热门赛事, 比赛任务的发展也开始往决策规划这一自动驾驶的最终任务发展, 赛题更加注重端到端系统与大模型方面。在 CVPR 2019 会议上, CARLA Team 在开源的自动驾驶模拟器上推出了闭环评测的端到端自动驾驶挑战赛 [29]。在今年由 OpenDriveLab 主办的 CVPR 2023 自动驾驶挑战赛中 [27], nuPlan Planning 赛道基于现实世界进行建模, 并在真实交通环境中对模型进行开环评测。在这两个比赛中, 模型评测方式也从单一面向的前景物体检测或背景地图描述, 变成更加全面的面向自动驾驶最终目标的评测, 如 CALRA 挑战赛中, 模型需要

表 1 自动驾驶国内外挑战赛与榜单(按时间排序)。可以看出,近几年的赛题不再局限于感知类任务,正逐步朝感知决策一体化、端到端、大模型等方向发展;模型评测的设置从单一的比较分数高低,到多元化、创新性的评价。

Table 1 Challenges in Autonomous Driving. “# Entry” denotes the number of public entries on the leaderboard by date of manuscript submission (September 2023). “Test Server” indicates whether the test server remains open or not.

Title	Host	Year	Task	Evaluation	# Entry	Test Server
Autonomous Driving Challenge [27]	OpenDriveLab	CVPR 2023	Perception / OpenLane Topology	OpenLane-V2 Score (OLS)		✓
			Perception / Online HD Map Construction	mAP	111	✗
			Perception / 3D Occupancy Prediction	mIoU		✓
		CVPR 2023	Prediction & Planning / nuPlan Planning	Mean Overall Score		✗
			Perception / 2D Video Panoptic Segmentation	weighted Segmentation and Tracking Quality		✓
			Perception / Pose Estimation	Pose Estimation Metric (PEM)	35	✓
			Prediction / Motion Prediction	Soft mAP		✓
			Prediction / Sim Agents	Realism Meta Metric		✓
		CVPR 2022	Prediction / Motion Prediction	Soft mAP		✓
			Prediction / Occupancy and Flow Prediction	AUC on Joint Occupancy and Flow Metric	128	✓
			Perception / 3D Semantic Segmentation	mIoU		✓
			Perception / 3D Camera-only Detection	Longitudinal Affinity Weighted LET-3D-AP		✓
Waymo Open Dataset Challenges [9, 28]	Waymo	CVPR 2021	Prediction / Motion Prediction	Soft mAP		✓
			Prediction / Interaction Prediction	mAP	115	✓
			Perception / Real-time 3D Detection	Mean Average Precision with Heading (APH)		✓
			Perception / Real-time 2D Detection	mAP		✓
			Prediction / Multi-agent Forecasting	Average Brier Minimum Final Displacement Error (avgBrierMinFDE)		✓
		CVPR 2023	Perception & Prediction / Unified Sensor-based Detection, Tracking, and Forecasting	Forecasting Average Precision	81	✓
			Perception / LiDAR Scene Flow	Three Way Average End Point Error		✓
			Prediction / 3D Occupancy Forecasting	L1 Error (L1), Absolute Relative L1 Error (AbsRel), Near-field Chamfer Distance (NFCDF)		✓
		CVPR 2022	Perception / 3D Object Detection	Composite Detection Score (CDS)		✓
			Prediction / Motion Forecasting	Average Brier Minimum Final Displacement Error (avgBrierMinFDE)	81	✓
			Perception / Stereo Depth Estimation	Number of Bad Pixels		✓
			Perception / Stereo Depth Estimation	Number of Bad Pixels		✗
			Prediction / Motion Forecasting	Brier Minimum Final Displacement Error (brier-MinFDE)	368	✓
Argoverse Challenges [16, 19]	Argoverse	CVPR 2021	Perception / Streaming 2D Detection	AP		✓
			Planning / CARLA AD Challenge 2.0	Driving Score, Route Completion, Infraction Penalty	-	✓
			Planning / CARLA AD Challenge 1.0	Driving Score, Route Completion, Infraction Penalty	19	✓
			Planning / CARLA AD Challenge 1.0	Driving Score, Route Completion, Infraction Penalty	-	✓
			感知 / 跨场景单目深度估计	算法精度、算法成本效率	-	✗
		NeurIPS 2022	感知 / 路侧毫米波雷达标定和目标跟踪	事件检测得分、车辆类别检测得分、雷达跟踪得分、雷达标定得分、算法实时性评测	-	✗
			感知 / 路侧三维感知算法	平均检测精度、平均地面中心相似度、平均朝向角相似度、平均面观相似度、平均底面四点相似度	-	✗
			感知 / 街景图像店面招牌文字识别	准确率、召回率	-	✗
		NeurIPS 2021	Perception / nuScenes Panoptic	Panoptic Quality (PQ)	11	✓
			Perception / nuScenes Detection	nuScenes Detection Score (NDS)		✓
			Perception / nuScenes Tracking	Average Multi Object Tracking Accuracy	456	✓
			Prediction / nuScenes Prediction	Minimum Average Displacement Error		✓
			Perception / nuScenes LiDAR Segmentation	mIoU		✓
计图 (Jittor) 人工智能算法挑 战赛 [32]	国家自然科 学基金委信 息科学部	2021	感知 / 交通标志检测	平均检测精度	37	✗
KITTI Vision Benchmark Suite [9, 33~35]	University of Tübingen	2012	Perception / Stereo, Flow, Scene Flow, Depth, Odometry, Object, Tracking, Road, Semantics	-	5,610	✓

平衡自动驾驶汽车的路线完成度与对驾驶场景中发生的各种事件的应对正确度。相比之下，国内自动驾驶比赛发展较晚，直到近年来才出现了一些自动驾驶相关比赛，而比赛题目均停留在感知领域，并未涉足预测与规划相关话题，仍有较大发展空间。从目前趋势看来，在未来自动驾驶比赛中，覆盖自动驾驶系统全链路的赛题会更加受研究人员青睐。例如在 2023 年举办的一些自动驾驶相关比赛仍然主要关注在感知方面，其参赛者数量相较于往年均有所下降，而 OpenDriveLab 主办的 CVPR 2023 自动驾驶挑战赛同时囊括感知、预测、规划三个方面，吸引了更多的研究人员参加比赛。

自动驾驶比赛与榜单在推动社区发展有至关重要的作用。测试服务器与榜单为广大研究者提供了一个公平比较模型结果的平台，同时让研究人员在榜单及时公开模型细节与代码。在一些较为活跃的榜单上，几乎每个月都能观察到榜首位置的迭代。而举办比赛与维护榜单需要主办方等多方面的共同努力，如提供数据下载途径、基准模型准备、测试服务器维护。特别是在近几年的发展中，大数据大模型的流行加大了举办比赛与维护榜单的难度。一方面，不断增长的数据量使下载训练数据变得并不简单。另一方面，大模型所需要的计算资源使得缺乏大量显卡的组织与个人，如高校学生，缺乏参与比赛的能力。因此，在未来的比赛与榜单中，广泛吸引各个方面的研究人员参与十分重要。数据的可用性、基线模型的易用性等都需要主办方和参与者的共同努力。

2.4 各类数据集介绍

下面我们按照不同任务划分，对各类自动驾驶数据集（感知、建图、预测、路径规划、学科交叉等）做详细的介绍与总结。

2.4.1 感知类数据集

自动驾驶感知类（Perception）任务包括二维与三维物体检测、追踪、可行驶区域检测、栅格检测（Occupancy Detection）等内容。表 2 给出了从 2012 年至 2023 年，感知类任务数据集的详细情况，包括传感器配置、任务、评价指标、影响力估计值等。从整体可以看出数据多样性在极大提高，主要体现在数据集时长、场景数量和数据地域分布上。前视视角（Front View）相机配置^[7, 10, 11]逐渐演变为环视视角（Multi-view, Surrounding View）^[20, 47]相机配置，激光雷达成为标配，多传感器融合成为主流趋势，定位信息以及高精地图成为新型数据集传感器数据。

感知任务类数据集主要由数据场景的丰富度，传感器种类以及标注类型几个要素构成。一般自动驾驶数据集以连续的几秒至几十分钟为一段进行切分，总时长从几小时^[24]到上千小时^[15]不等。所切分的场景数量越多，总时长越多，场景丰富度也越高。有的数据集给每一个片段打上场景标签，如晴天雨天^[8]、季节^[49]、白天黑夜^[9]以及乡村高速城市^[59]等标签，用以推进在多种场景下感知能力的研究。除了自然因素，人文因素也对自动驾驶场景丰富度至关重要。因为不同国家之间的语言、交通规则和交通标志往往不同，囊括多个地域的数据集也能为自动驾驶系统在不同场景的泛化能力提供保障，如 Cityscapes^[10, 20]等数据集，通过搜集网络数据，实现全球范围内数据搜集，nuScenes^[8]数据集通过分别在北美和亚洲布置数据采集车，获取两个大洲的数据。在传感器方面除了常见的前视觉相机与激光雷达传感器，如表中其他类型传感器所示，部分数据集^[8, 14, 20, 38]还配备了车身信号（Controller Area Network Bus, CAN-bus）、全球定位仪（Global Positioning System, GPS）、惯性测量单元（Inertial Measurement Unit, IMU）、红外相机（Infrared Camera）、超声波雷

表 2 感知类数据集一览。按照数据集支持的任务，整个表分为两部分，上半部为感知任务数据集，下半部为复合任务数据集。按照年份、数据多样性、传感器配置、标注数据以及影响力估计评分等维度，我们梳理了四十余种数据集。

Table 2 Perception datasets at a glance. The table is divided into two parts, with the upper half mainly on perception tasks and the lower half on hybrid tasks. For **Data Diversity**, **Scenes** indicates the number of video clips in a dataset. For **Region**, “AS” stands for Asia, “EU” for Europe, “NA” for North America, “Global” for all continents, and “Sim” for simulation scenarios. For **Annotation**, 2D/3D BBox indicates 2D or 3D bounding box, 2D/3D Seg means 2D/3D segmentation label, 3D Occ means 3D occupancy label. **Impact** implies the impact of the dataset defined in the context. ‘-’ indicates that a field is inapplicable.

Dataset	Year	Data Diversity			Sensor			Annotation	Impact
		Scenes	Hours	Region	Camera	LiDAR	Other		
KITTI [33]	2012	50	6	EU	Front-view	✓	GPS & IMU	2D BBox & 3D BBox	High
Cityscapes [10]	2016	-	-	EU	Front-view	✗		2D Seg	Mid
Lost and Found [12]	2016	112	-	-	Front-view	✗		2D Seg	
Mapillary [11]	2016	-	-	Global	Street-view	✗		2D Seg	Mid
DDD17 [36]	2017	36	12	EU	Front-view	✗	GPS & CAN-bus & Event Camera	-	
ApolloScape [14]	2016	103	2.5	AS	Front-view	✗	GPS & IMU	3D BBox & 2D Seg	Mid
BDD-X [37]	2018	6984	77	NA	Front-view	✗		Language	
HDD [38]	2018	-	104	NA	Front-view	✓	GPS & IMU & CAN-bus	2D BBox	Mid
IDD [39]	2018	182	-	AS	Front-view	✗		2D Seg	
SemanticKITTI [24]	2019	50	6	EU	✗	✓		3D Seg	
Woodscape [20]	2019	-	-	Global	360°	✓	GPS & IMU & CAN-bus	3D BBox & 2D Seg	Mid
DrivingStereo [25]	2019	42	-	AS	Front-view	✓		-	
Brno-Urban [40]	2019	67	10	EU	Front-view	✓	GPS & IMU & Infrared Camera	-	
A*3D [41]	2019	-	55	AS	Front-view	✓		3D BBox	Mid
Talk2Car [42]	2019	850	283.3	NA	Front-view	✓		Language & 3D BBox	
Talk2Nav [21]	2019	10714	-	Sim	360°	✗		Language	
PIE [43]	2019	-	6	NA	Front-view	✗		2D BBox	
UrbanLoco [44]	2019	13	-	AS & NA	360°	✓	IMU	-	
TITAN [45]	2019	700	-	AS	Front-view	✗		2D BBox	
H3D [46]	2019	160	0.77	NA	Front-view	✓	GPS & IMU	-	
A2D2 [47]	2020	-	5.6	EU	360°	✓	GPS & IMU & CAN-bus	3D BBox & 2D Seg	
CARRADA [17]	2020	30	0.3	NA	Front-view	✗	Radar	3D BBox	
DAWN [48]	2019	-	-	Global	Front-view	✗		2D BBox	
4Seasons [49]	2019	-	-	-	Front-view	✗	GPS & IMU	-	
UNDD [50]	2019	-	-	-	Front-view	✗		2D Seg	
SemanticPOSS [51]	2020	-	-	AS	✗	✓	GPS & IMU	3D Seg	
Toronto-3D [52]	2020	4	-	NA	✗	✓		3D Seg	
ROAD [53]	2021	22	-	EU	Front-view	✗		2D BBox & Topology	
Reasonable Crowd [23]	2021	-	-	Sim	Front-view	✗		Language	
METEOR [54]	2021	1250	20.9	AS	Front-view	✗	GPS	Language	
PandaSet [55]	2021	179	-	NA	360°	✓	GPS & IMU	3D BBox	
MUAD [56]	2022	-	-	Sim	360°	✓		2D Seg & 2D BBox	
TAS-NIR [57]	2022	-	-	-	Front-view	✗	Infrared Camera	2D Seg	
LiDAR-CS [58]	2022	6	-	Sim	✗	✓		3D BBox	
WildDash [59]	2022	-	-	-	Front-view	✗		2D Seg	
OpenScene [60]	2023	1000	5.5	AS & NA	360°	✗		3D Occ	
ZOD [61]	2023	1473	8.2	EU	360°	✓	GPS & IMU & CAN-bus	3D BBox & 2D Seg	Mid
nuScenes [8]	2019	1000	5.5	AS & NA	360°	✓	GPS & CAN-bus & Radar & HDMap	3D BBox & 3D Seg	High
Argoverse V1 [16]	2019	324k	320	NA	360°	✓	HDMap	3D BBox & 3D Seg	High
Waymo [9]	2019	1000	6.4	NA	360°	✓		2D BBox & 3D BBox	High
KITTI-360 [26]	2020	366	2.5	EU	360°	✓		3D BBox & 3D Seg	Mid
ONCE [15]	2021	-	144	AS	360°	✓		3D BBox	Mid
nuPlan [18]	2021	-	120	AS & NA	360°	✓		3D BBox	High
Argoverse V2 [19]	2022	1000	4	NA	360°	✓	HDMap	3D BBox	Mid
DriveLM [62]	2023	1000	5.5	AS & NA	360°	✗		Language	Mid

达传感器 (Radar)、事件相机传感器 (Event Camera) 等传感器数据, 用于探索不同类型传感器下自动驾驶算法研究以及多传感器融合研究 [63~65]。除了传感器类别不同, 同种传感器之间也存在着型号差异, 如自动驾驶系统中前视相机相较于后视相机, 前者需要较小的视场角以获得更远的视野, 使得对于远处小目标模型也能有较好感知能力, 提高车辆在高速行车过程中观测前方道路情况的能力, 拥有较大视场的后视相机能提供自车后方近处他车情况。表中传感器 (Sensor) 中其他 (Other) 一栏处空白表示数据集不包含其他传感器数据。

除了数据丰富度与传感器种类, 数据集配套的标注受到算法研发方案引导、技术路线影响也在不断迭代。针对动态物体的自动驾驶感知任务对于实现安全、智能、高效的自动驾驶至关重要。它涉及到碰撞避免、行为预测和决策、规划和路径规划、交互和协同行驶等关键方面。随着对问题理解不断加深, 感知任务本身在不断演进。基于相机成本低容易部署的特点, 相关工作 [66~68] 基于图像检测结果提取目标图像坐标系下位置, 结合一个额外的三维网络学习三维信息或生成一个伪点云用于输入基于点云的位置检测器。为避免级联误差, 后续工作 [66~68] 推出了端到端的检测方法, 这些方法可以在提高检测精度上限的同时降低计算消耗。针对单目相机感知范围小, 存在视野盲区问题, FCOS3D 算法 [69, 70] 推出了基于多相机的融合检测算法, 该算法有效解决了多相机融合问题, 在 nuScenes [8] 与 Waymo [9] 等知名榜单中摘得桂冠。随后出现基于鸟瞰图的工作 [71, 72] 首次将 Transformer 结构引入多传感器目标检测中, 同时较好的利用自动驾驶数据的时序特点。针对异形物体普通三维框无法有效表述, 实际驾驶过程中会导致磕碰等问题发生, 三维占据栅格数据集, 如 OpenOcc [60], 受到了社区广泛关注。三维占据数据通过三维体素表征空间物体的位置边界, 并将连续多帧点云通过位姿关系, 叠加到同一帧中, 实现了更加精密的三维外形表征。因现有数据标签简单, 只能做到对场景简单的感知, 无法实现全面的场景理解, 导致复杂场景下决策困难, 一些数据集 [21, 23, 42, 54] 提出了基于语言的标注, 在感知、决策阶段引入人的理解, 辅助系统模拟人脑思维过程。

参照 2.2 节数据集影响力估计值计算公式, 本文梳理了四十多个感知数据集影响力估计, 其中影响力为中高档的数据集在表中注明。有较高影响力的数据集在数据规模、数据多样性等方面较为突出, 如 Mapillary 等数据集 [11, 20] 横跨多个大洲, Cityscape [10] 数据集在欧洲五十个城市采集并涵盖了春、夏、秋三个季节的场景。影响力最高的几个数据集在创新性都有较大贡献, 如 KITTI [7] 数据集首次配备了激光雷达与相机传感器, 并定义了深度估计、目标检测、目标跟踪等 10 个任务与相应的评价指标。Waymo、nuScenes、Argoverse [8, 9, 16] 等数据集最先配备了环视相机, 并且场景规模更加丰富, 支持横跨感知、预测及决策规划等多个模块任务, 实现在同一个数据集上支持多个自动驾驶模块的研究, 给行业全链路量产团队完整的基准。其中 nuScenes 等 [8, 16, 18] 数据集首次配置了高精地图数据, 贴近实际量产环境下系统传感器输入, 吸引了更多量产行业的关注。

2.4.2 建图类数据集

自动驾驶建图任务是对于静态物体的感知和理解, 为下游任务构建出对静态环境的准确认知。静态场景中的交通元素感知不仅包含对路面元素如车道线、人行横道的识别, 还包括对交通灯、交通标志牌等与车辆行驶状态相关元素的识别 [73]。该任务涉及的目标处于不同的维度, 拥有不同的表达方式, 因此具有较大的挑战。

表 3 建图类数据集一览。在自动驾驶算法的发展历程中,建图类数据集真值表达形式也在不断的发展中发生改变。

Table 3 Review of Mapping Datasets. In the development of autonomous driving, the representation format of ground truth is continuously evolving. Numbers in “Frames” are the number of annotated frames over total frames respectively. “Inst.” indicates whether lanes are annotated instance-wise (c.f. semantic-wise). “Track.” implies whether a lane has a unique tracking ID.

Dataset	Year	Data Diversity		Sensor		Annotation			
		Scenes	Frames	Camera	LiDAR	Type	Space	Inst.	Track.
Caltech Lanes [74]	2008	4	1224/1224	Front-view Image	x	Laneline	PV	✓	x
VPG [75]	2017	-	20K/20K		x		PV	x	-
TUsimple [76]	2017	6.4K	6.4K/128K		x		PV	✓	x
CULane [77]	2018	-	133K/133K		x		PV	✓	-
ApolloScape [14]	2018	235	115K/115K		✓		PV	x	x
LLAMAS [78]	2019	14	79K/100K		x		PV	✓	x
3D Synthetic [79]	2020	-	10K/10K		x		PV	✓	-
CurveLanes [80]	2020	-	150K/150K		x		PV	✓	-
VIL-100 [81]	2021	100	10K/10K		x		PV	✓	x
OpenLane-V1 [82]	2022	1K	200K/200K		x		3D	✓	✓
ONCE-3DLane [83]	2022	-	211K/211K		x		3D	✓	-
OpenLane-V2 [84]	2023	2K	72K/72K	Multi-view Image	x	Lane Centerline, Lane Segment	3D	✓	✓

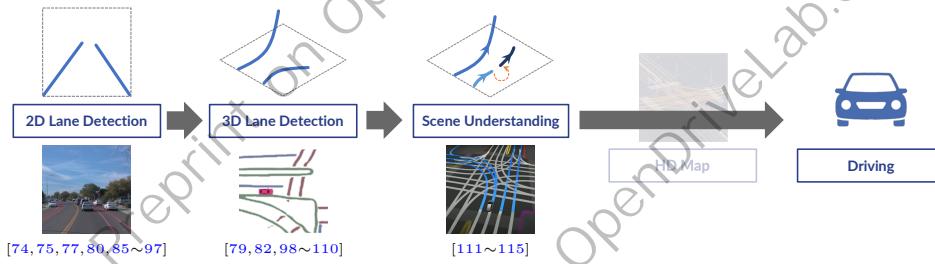


图 3 道路结构认知领域的发展脉络。相较于高精地图的使用,“重感知,轻地图”是未来潜在的赋能驾驶的方案。

Figure 3 The roadmap of road structure cognition. Instead of relying on high-definition maps, online mapping methods are more promising for autonomous driving in the future.

在表 3 总结的建图类工作中,数据集的真值表达形式即在线建图的目标表达形式在不断的发生变化。计算机视觉的发展从对图像的分类、检测、分割等工作开始,地图感知数据集的发展也始于二维图像的表达,如交通灯检测数据集 [116~126]。交通灯检测数据集可以被视为一种特定类别的图像检测数据集。初始的车道线检测数据集 [14, 74~81] 在二维图像坐标系中检测车道线,然后通过逆透视变换 (Inverse Perspective Mapping, IPM) 投影矩阵获得三维车道线。由于 IPM 算法基于路面符合平面假设的设定,而现实中大多数路面存在高度变化,导致在透视图中表示的车道线在投影到三维空间的过程中容易出现错误。为解决这个问题,近几年的车道线数据集 [82, 83] 提出直接进行三维车道线检测的任务。由于车道线并不是车道的完备表达,无法包含车道方向与车道之间的连接等关系,进一步地,OpenLane-V2 [84] 引入了车道的实例级表达方式,并且通过拓扑关系的构建赋予其连接性及其与交通标识的关联性。建图类数据集的发展使模型预测结果所包含的信息越来越接近高精地图。

图 3 描述了道路结构认知领域的表征方式及发展脉络。传统自动驾驶方案主要基于高精地图

表 4 自动驾驶预测与规划任务数据集。依据任务输入输出形式, 将预测与规划数据集分为不同类型。

Table 4 Review of prediction and planning datasets. The tasks are represented differently according to task formulations.

Subtask	Input	Output	Evaluation	Dataset	Reference
Motion Prediction	Surrounding Traffic States	Spatiotemporal Trajectories of Single/Multiple Vehicle(s)	Displacement Error	Argoverse [16]	[136], [137], [138]
		Trajectories of		nuScenes [8]	[139], [140], [141]
		Single/Multiple		Waymo [9]	[142], [143], [144]
		Vehicle(s)		Interaction [145]	[146], [147], [148]
Trajectory Planning	Motion States for Ego Vehicles, Scenario Cognition and Prediction	Trajectories for Ego Vehicles	Displacement Error, Safety, Compliance, Comfort	MONA [149]	
				nuPlan [18]	[150], [151], [152]
				CARLA [30]	[153], [154], [155]
				MetaDrive [156]	[157], [158], [159]
Path Planning	Maps for Road Network	Routes Connecting to Nodes and Links	Efficiency, Energy Conservation	Apollo [160]	[161], [162], [163]
				OpenStreetMap [164]	[165], [166], [167]
				Transportation Networks [168]	[169], [170], [171]
				DTAlite [172]	[173], [174], [175]
				PeMS [176]	[177], [178], [179]
				New York City Taxi Data [180]	[181], [182], [183]

(High-Definition Map, HD Map) 来获取道路结构信息, 通过感知结果与高精地图的匹配获得自行车在高精地图中的位置信息。虽然高精地图包含丰富的道路信息, 如道路形状、道路标记、交通标志、障碍物等, 精度可以到厘米级别, 但其获得成本高而且更新周期较长。相比基于高精地图的“重地图, 轻感知”方案, 近年来自动驾驶系统往“重感知, 轻地图”方向发展。通过感知模型在线建图, 摆脱了对离线高精地图的依赖, 有效地控制了成本, 避免了实时性与定位误差的问题, 也使得自动驾驶系统的扩展性更强。但由此也引申出另一问题, 神经网络的精度不足产生错误感知结果, 导致下游模块做出错误决定, 所以需要海量高质量数据来使模型性能尽可能高。目前地图感知结果所包含的信息仍然明显少于高精地图的信息, 地图感知结果需要具备何种特性是最能被下游任务有效利用的相关研究还有待开展。

2.4.3 预测与规划类数据集

传统的模块化规控技术路线会将规控任务具体拆分为不同维度的子任务, 具体的可以分为路网级别、道路级别与车辆级别。其中路网级别, 根据高清地图提供的静态道路拓扑数据和浮动车辆或卡口提供的实时道路流量数据 [127~131], 可以进行交通流预测和路线规划, 为车辆规划从起点到终点的行驶路线, 通常以最小化行驶路程或旅行时间为目。道路级别, 以完成路线引导的驾驶目的为导向, 根据车载相机、激光雷达、毫米波雷达等传感器所提供的自行车附近的小范围的道路场景感知信息, 预测周围车辆未来数秒时间内的驾驶行为和运动轨迹 [132~135], 进而规划出安全、高效、舒适行驶的自行车运动轨迹。车辆级别, 结合车辆运动学和动力学模型, 获取操纵车辆加减速和转向所对应的油门与刹车踏板行程、方向盘或转向轮转角, 以最小操纵代价的前提下驱使车辆完成给定的行驶轨迹。

表 4 列出了自动驾驶预测与规划任务的数据集。在运动预测方面, Argoverse 运动预测数据集 [16] 是从 1000 多个小时的驾驶数据中筛选出 324557 个场景, 每个场景长度为 5 秒, 场景中的车辆行为包括通行交叉口左右转及变道等多种行为, 并提供场景中每个跟踪对象的三维鸟瞰图的中心位置信

息。nuScenes [8] 数据集中也提供了相应的运动预测任务，预测时长为 6 秒，采样频率为 2Hz，目标输出即为对象的未来轨迹坐标。Waymo [9] 运动数据集提供了矢量地图和车辆运动信息，该数据集由 103354 个片段组成，每个片段包含时长 20 秒、频率 10Hz 的物体轨迹及其对应区域的地图数据，这些片段进一步被划分适配任务的窗口，即 1s 的历史数据作为输入和 8 秒的未来数据作为输出来训练和验证模型。Interaction [145] 是重点关注于道路或路口的车辆和行人之间交互过程的数据集，从中国、美国、德国、保加利亚采集了包括环岛、信号交叉口、无信号交叉口、合流与并道等 4 种场景，总共 12 个地点的数据，并提供了经过语义标注的 Lanelet2 格式地图。

在轨迹规划方面，nuPlan [18] 提供了包含来自美国和亚洲四个交通模式迥异的城市（波士顿、匹兹堡、拉斯维加斯和新加坡）的 1500 小时人类驾驶数据，并提供通过离线感知系统实现接近人类水平的自动标注结果。CARLA [30] 是支持自动驾驶系统研发的开源仿真器，提供了城市布局、建筑物、车辆等开放三维数字资产，平台支持三维地图生成、静态与动态参与者的完全控制、多种传感器信号模拟等功能。MetaDrive [156] 是面向自动驾驶决策规划应用的轻量级仿真器，具备低级别的三维渲染和场景可视化界面以支持高效运行，并提供大量过程生成的场景可便于进行强化学习算法的部署和训练。Apollo 仿真平台 [160] 可提供贯穿自动驾驶研发迭代过程的完整解决方案，内置高精地图的仿真场景包括各种路型、障碍物、交通信号等，支持多场景高速并行，拥有碰撞检测、违规检测、到达检测等多个场景判别指标，并提供测试过程算法模块的可视化信息。

在路径规划方面，OpenStreetMap [164] 是开放地图协作计划，目标是创造内容自由且能让所有人编辑的世界地图，用户可以通过提供 GPS 信息、航拍照片、卫星影响、政府开放资料等信息参与地图绘制，同时可以免费下载矢量地图进行使用。Transportation Networks [168] 是面向宏观交通研究数据集，提供包括路网及流量数据，可支持交通分配、路线规划等研究工作。DTAelite [172] 是轻量级的支持动态流量分配研究的开源软件，可用于大规模交通路网建模及其相关应用研究，包括路径规划、流量分配、瓶颈分析、收费机制等。PeMS [176] 是美国加州运输公司在加州的高速公路上收集的交通数据库，通过将 44681 个检测器每隔 30 秒数据合成粒度为 5 分钟的聚合数据，包括流量、速度、车辆行驶里程等多种类型，还包括巡检和监测的事故数据，可用于高速功能的性能评估、运营决策、路网瓶颈分析等研究。纽约出租车数据集 [180] 是由纽约市出租车管理委员会提供的 2009 年到 2016 年纽约出租车乘车记录数据，包括乘客上下车时间及地点、行程距离、票价及费率、乘客数量等相关字段信息，可用于面向城市范围的多车调度、共享出行等研究。

规控任务的评价方式可以划分为开环评价和闭环评价。开环评价是指以数据集所提供的传感器信息和高阶行车指令信息（前进、左转、右转等）作为输入，来获取出车辆未来轨迹的方法。开环评价通常会评估模型输出轨迹和数据集原始轨迹的差距，包括平均位移误差 (Average Displacement Error, ADE)、最终位移误差 (Final Displacement Error, FDE)、轨迹落在可行驶区域的比例 (Drivable Area Compliance, DAC) 等量化指标指标。开环评价的好处是可以充分利用数据集所生成的状态-动作对来模仿自动驾驶人的驾驶行为，缺点是模型可能过拟合于数据采集环境和有限的自动驾驶人行为模式而导致模型泛化性较差 [184~187]，遇到与数据集中差距较大的场景可能得到出乎意料的糟糕轨迹结果，这对于强调安全性的自动驾驶来说是难以接受的。闭环评价是指让所构建的模型能够在动态场景中进行校验，模型输出驱动自车进行行驶，周围车辆会根据自车行为做出交互式反应。闭环评价要求真实道路环境和实时车辆交互反馈，由于模型研发到部署周期长，研发阶段的车辆安全性无法保证，因此真正的闭环评价是不可能实现的。退而求其次，基于仿真的闭环评价

表 5 交叉领域数据集及其代表性工作。根据不同应用领域, 对数据集的要求也会有所不同。

Table 5 Interdisciplinary datasets and their representative works. Dataset requirements vary according to different fields.

Field	Description	Dataset	Reference
Robotics	Replacing People for Multiple Operations and Services	Roboturk [188], PaLi-X [189], Robonet [190]	[191], [192], [193]
V2X	Collaborative Sensing and Decision Making based on Wireless Communication	Dair-V2X [194], V2X-Sim [195], OPV2V [196]	[197], [198], [199]
UAV	Aircraft with Multiple Sensors for Aerial Operations	UAV123 [200], Blackbird [201], UAVid [202]	[203], [204], [205]
USV	Operation and Control of the Unmanned Surface Vehicles	MODS [206], MaSTR1325 [207], USVInland [208]	[209], [210]
Transportation	Allocation and Coordination of Products, Raw Materials, and Services in the Supply Chain	NEO Benchmark Datasets [211], LARA, Divvy data [212]	[213], [214], [215]

是当前研究的热点, 通过图形学或隐式表征方式可以构建虚拟道路环境或者构建真实道路场景的数字孪生, 再让模型驱动的车辆在道路上运行。在 nuPlan [18] 中, 闭环评价指标包括交通规则违反情况、与人类驾驶的相似性、车辆动力学代价、行驶目标完成度、特定场景的评价等。然而, 基于仿真的闭环评价的交通流在仿真开始之后就完全脱离数据集, 环境车辆通常是通过基于规则的偏保守风格的模型来控制的, 存在自车与环境车辆的交互的真实性不足的问题。

规控任务是自动驾驶技术的下游, 也是车辆运行效果的最终呈现。规控任务的完成质量是自动驾驶车辆能否上路的关键。尽管自动驾驶技术已经经历了数轮的技术迭代, 量产应用仍然处于需要驾驶员保持注意力、准备随时接管车辆的 L2 级别。其中重要的原因之一是规控算法在边缘场景 (Corner Case) 中的泛化能力依然一般。我们认为, 现阶段的规控任务技术挑战可以总结为以下几个方面:

- 如何通过数据集尽可能营造真实的仿真环境, 提升模型评价指标的可信度。
- 如何让仿真尽可能覆盖多样化的场景以更全面评估模型的规控性能。
- 如何利用基础模型强化自动驾驶模型的零样本学习能力, 在未见过的场景中也能做出正确决策。

2.4.4 交叉领域数据集

自动驾驶与其他应用领域的技术耦合同样值得关注, 表 5 按照分别按照机器人 (Robotics)、车联网 (Vehicle-to-everything, V2X)、无人机 (Unmanned Aerial Vehicle, UAV)、无人船 (Unmanned Surface Vehicle, USV)、物流 (Transportation) 等领域罗列相关数据集, 并介绍不同应用领域相关数据集对自动驾驶的促进作用。

机器人是由人类设计的能够半自主或全自主工作的机器, 通过所具备的感知、决策和执行等基本能力来辅助或替代人类完成各类复杂、危险、繁重的工作 [216]。机器人范畴较大, 包括工业机器人、服务机器人、特种机器人等, 针对不同的任务构建了相应的数据集。严格意义上说, 自动驾驶是机器人在应用领域的一个重要应用, 因为它的最终目的是由机器取代人类来完成操作车辆服务于人们的出行需求。因此, 机器人领域相关的技术和算法也能够迁移到自动驾驶领域进行应用 [217]。机器人领域有许多公开的数据集, 影响力较大的包括 Roboturk [188]、Cornell Grasp Dataset [218] 和 RoboNet [190]。Roboturk [188] 包括针对 3 个具有挑战性的操作任务的 111 个小时的机器人操作数据, 能够让机器人与环境交互来学习到最佳策略以完成任务; Cornell Grasp Dataset [218] 提供了 885 幅多视角拍摄的抓取图像, 可作为计算机视觉的抓取、规划、控制等方法的基准。RoboNet [190] 包

含 1500 万个视频帧,是在桌面环境中与不同对象进行交互的不同机器人收集的,包括图像、手臂姿势、力传感器读数、抓取器状态等信息。

车联网是新一代通信技术与自动驾驶驾驶领域的融合应用,其最早的设计方案根据美国和欧洲分别推动的专用短程通信(DSRC) [219] 和合作智能交通系统(C-ITS) [220] 来为车辆与其他设备和网络元素之间进行通信互联来传输数据。蜂窝无线电LTE 网络和新的5G 生态系统的不断发展,为支持V2X 设计了专门的标准规范,其中由国际标准组织3GPP 所发布的R16 标准是第一个5G NR V2X (5G New-Radio Vehicle-to-everything) 标准 [221],引入组播和广播等多种通信方式以支持车车、车路直连通信,并提出优化感知、调度、重传以及车车间连接质量控制等技术。车联网是车路协同自动驾驶方案的重要组成部分,通过协同感知、协同决策等技术手段,能够有效弥补单车智能的局限性,为迈向L5 完全自动驾驶提供有力支撑 [222~224]。在数据集建设方面,DAIR-V2X [194] 数据集是首个用于车路协同自动驾驶研究的大规模、多模态、多视角数据集,全部数据采集自真实场景,同时包含二维与三维标注;V2X-Sim [195] 是面向车路协作感知的数据集,支持检测、跟踪和分割等任务,并提供了开源测试平台;OPV2V [196] 也是多车协同感知数据集,包含相同时间截下的多辆自动驾驶汽车的三维点云与相机RGB 图像,涵盖了 73 个不同的场景,包括了 12K 帧 LiDAR 点云,48K 帧 RGB 图像,230K 帧三维标注框。近年来还有围绕着车联网数据存储 [225]、通信延迟 [226] 以及网络理论 [227] 等通信基础设施方面的工作与挑战也受到行业的关注 [228]。

无人机,又称无人飞行器、遥感飞航器或者空中机器人,是指一类没有飞行员操纵杆和控制舱的航空器,通常由地面遥控操作员使用遥控器进行控制或者按照程序自主飞行 [229~231]。无人机可以搭载各种传感器和载荷执行任务,如拍摄高清晰度图像、地形测量、环境监测、电力线路巡检、搜索救援等。除了军事用途,无人机在农业、林业、城市规划、建筑施工等领域也得到广泛应用。无人机与自动驾驶汽车的协同合作在城市交通领域也具有广泛的前景。无人机可以搭载高精度的激光雷达等传感器获取道路拓扑、交通路况、障碍物等信息,使得自动驾驶汽车合理规划路径以避免拥堵和事故 [232]。在数据集建设方面,UAV123 [200] 是由低空无人机捕获的视频跟踪数据集,包含总共 123 个视频序列,总图像帧超过 110K 帧,可以应用于长时空中跟踪任务;Blackbird [201] 通过 5 种不同的视觉环境以不同的最大速度追踪 18 条不同的轨迹,采集到时长超过 10 小时的数据,适合用于开发视觉惯性导航、三维重建和深度估计的算法;UAVid [202] 是高分辨率无人机语义分割数据集,在运动目标识别及时间融合保持等方面具有显著特点,包括 30 个视频序列,以倾斜的视角捕获 4k 高分辨率图像,并提供了 300 张涵盖 8 种类别的语义标签。

无人船,又称无人水上运载工具,是一种可自主导航、能够在水面上行驶并执行任务的机器人系统,它不需要人类在船上操作控制,并且具有高度的灵活性和可靠性,能够适应各种复杂的水域环境,被广泛应用于海洋勘探、环境保护、港口安全等领域 [233]。无人水上机器人与自动驾驶类似,包括 GPS 导航、多传感器融合感知、自主决策和控制、远程通信与协作等功能。由于无人水上机器人需要在各种复杂、动态的水上环境中工作,风浪大小、潮汐变化等因素都会对其造成影响,因此需要具有较强的环境适应性和安全保障 [234]。在数据集建设方面,MODS [206] 是多样化海事评估数据集,包含与船载惯导同步的约 81K 立体图像,其中提供了超过 60K 的标签,支持海上物体检测和障碍物分割任务;MaSTr1325 [207] 是面向语义分割训练的数据集,用真实的无人船采集了覆盖两年的 1325 幅图像,并提供了语义标签,可用于开发小型沿海无人船的障碍物检测方法。USVInland [208] 是在真实内河水域场景采集的多传感器、多天气条件下的无人船数据集,总行驶距离超过 26 公里。

交通运输领域通常以提高交通系统的效率、减少拥堵、提高安全性、降低成本为目的, 通过运筹优化等方法提出相应的解决方案。在保证交通系统可靠性和安全性的前提下, 优化交通流量组织、路网布局、信号灯配时、车辆调度等交通资源配置问题。随着自动驾驶技术的快速发展, 交通领域面临的优化问题也变得更加复杂和多样化, 车路联合优化和多车协同调度等方面受到越来越多的关注。在数据集方面, NEO [211] 提供了研究车辆路径问题的许多实例和最优解; LARA 是针对人类仓储活动的数据集, 通过惯导和相机捕捉 14 名被试者行为, 并提供了 758 分钟的录音标注, 包括了 8 个活动类别和 19 个二进制粗语义描述; Divvy Data [212] 涵盖了美国芝加哥、华盛顿和纽约等三个城市共计 90 万条骑行数据, 并提供了城市、年龄、性别、用户类型等信息。

3 数据算法闭环体系

数据对于现有人工智能算法与神经网络是不可或缺的。模块化自动驾驶系统包括感知、决策、规划、控制等组件, 其中大部分功能是通过数据驱动的神经网络模型实现的。对于这些模块来说, 海量和高质量的数据是确保模块性能的必要条件。长尾问题是自动驾驶工程中一个长期存在的问题, 尤其是在应对极端天气条件、少见但复杂的交通场景以及地理上不常见的道路结构方面, 在这些场景下保证系统的安全和可靠性至关重要。研究人员一直尝试设计针对长尾问题的算法以解决在复杂和罕见场景下模型性能的下降或错判问题 [235~237]。但究其根本原因在于训练模型的数据量不足而导致存在少量情况未被模型学习, 而在模型推理阶段, 模型并不能对这些边缘场景给出正确的结果。另外, 对于基于规则的模块, 现有的方式是通过人工设计各种规则来使模块输出符合人为设计逻辑的结果。这个方法耗时耗力, 并且难以覆盖所有情况, 有可能导致自动驾驶系统在某些未见场景下失效。而使用数据驱动的神经网络代替这些模块是一个可能的解决方案。由此, 海量数据的引入对于解决现存自动驾驶系统中的各种问题都很有必要。同时, 在神经网络学习过程中, 数据噪声的引入会不可避免地对优化过程产生负面影响, 并降低模型性能。数据质量不仅包括传感器数据的分辨率和同步性等, 还包括标签的准确性。在这两个方面中, 任意一个方面存在质量问题都直接影响着自动驾驶系统的性能和安全性。由此, 海量和高质量的数据成为构建自动驾驶系统必不可少的一个环节。

现阶段, 如何高效地构建海量高质量的数据仍是一个开放性的问题。学术界与工业界对于构建自动驾驶数据集所采用的方案不尽相同, 数据采集、质量把控、标注技术等方面都根据各自拥有的平台与技术有所变化。在本章节中, 我们将对比不同厂商构建数据闭环体系的方案, 提取其共性与精华。同时, 我们还将对其中各个关键技术进行分析, 为下一代自动驾驶数据集的构建夯实基础。

3.1 商业解决方案对比

数据在自动驾驶汽车发展周期中具有极大作用。如表 6 所示, 相关厂商通常都会构建自己的数据算法闭环体系来挖掘其价值。数据算法闭环体系通常包括对数据的采集、预处理、存储、可视化、检索、标注、使用等, 其目的是高效地利用大量数据提升自动驾驶系统可靠性。

值得注意的是, 绝大部分厂商都将数据算法闭环解决方案视作核心资产, 不公开技术细节, 仅通过技术发布会或宣发稿等渠道公开部分信息。但各个解决方案提供商都会为自动驾驶汽车厂商等客户提供定制化的方案, 这也使得不同方案的技术实现和功能各有参差。但对于在自动驾驶研究领

表 6 商业化数据算法闭环解决方案对比。不同厂商对于特定功能的实现方式存在差别，但整体方案趋于相同。

Table 6 Comparison of data engine solutions. A consensus on implementation among different entities has yet to be reached.

Solution	Owner	Data Catalog	Data Retrieval	Auto-labeling	Model Training	Simulation	Open-Source
Autonomous Driving Data Framework (ADDFF) [238]		Scene Description via Scene Metadata	OpenSearch	Object and Lane Detection via Open-source Models	✓	✓	✓
Full Self Driving (FSD) [239]		Misprediction Identification, Label Correction, and Selection on Most Valuable Examples		Static Scene Annotation via Multi-trip Reconstruction	Dojo Supercomputer	Scene Generation in Minutes	✗
MagLev [240]		Generating Dataset via Searching, Collection, Labeling, and Export	Elastic Search and Categorization via Active Learning	-	Multi-node Training and Parallel Evaluation	DRIVE Sim	✗
OpenTrek [241]			Multi-modal Retrieval based on Semantics, Images, Labels, Similarity, etc.	-	✓	✓	✗

域的数据集来说，严格遵守固定的数据算法闭环体系方案有可能会导致投入资源过大、链路过长等问题。与之相反，灵活运用数据算法闭环体系中的某些环节能够节省资源提高数据集质量。在本章节余下部分，我们将关注某个特定环节中的细节，并通过具体案例详细分析自动驾驶数据算法闭环体系在构建新一代数据集中的作用。

3.2 数据质量把控

数据质量的高低直接影响到业务决策的质量和效率，通常通过以下维度进行分析。(1) 完整性：确保所有必需的数据都被收集并记录在系统中，可建立数据标准和质量控制流程来监控缺失或不完整的数据；(2) 有效性：保证数据可用且准确，可使用验证规则和逻辑检查来识别无效或错误的数据，并设立修正的机制；(3) 及时性：确保数据在需要时能够及时获取，可设置适当的存储策略和处理过程，以便快速访问和提取数据；(4) 一致性：确保数据在整个组织内的一致性，可建立标准化的数据定义和格式，以及一致的命名约定；(5) 准确性：确保数据的真实性和可靠性，可采用多方验证和数据清洗技术来发现和纠正错误和异常值；(6) 唯一性：确保数据的唯一性和避免重复，可实施唯一标识符和主键约束，以避免数据冗余和不一致性。

数据标注对于构建面向监督学习的数据集是非常必要的环节。传统的数据标注过程需要人类介入，通常是将特定的标注任务指派给若干名标注人员进行判断和选择，生成可用于监督模型训练的数据标签。数据标注质量包括准确性、一致性、置信度等方面的评价指标。

数据标注准确性需要事先了解基线方法的效果，常用的基线方法包括随机标注、最大频率标注和数据频率三种。以类别标注为例，随机标注是指从待选标签中以相同概率赋予数据标签，最大频率标注是指以数据频率最高的类别来标注目标数据，数据频率是指按照标签分布概率来标注目标数据。假设已知目标数据集标注真值，通过以上三种基线方案可以计算出相应的数据标注准确度下限，其中最大频率标签法是准确度最高的基线方法。当标注员完全熟悉数据集之后的标注准确度低于基线方法，则说明数据标注质量较差，应及时修改标注任务和指南或更换标注人员。

数据标注一致性体现的是群体智慧的思想，如果多个标注员同时标注一批数据时，则以选择类别最多的标签作为最终结果。通过数据标注一致性，可以判断数据集和标注人员的可靠性，并能够甄别标注人员间的抄袭行为。当数据集与标注任务适配度不高或者标注难度较大时，标注一致性也

会变差, 此时需要及时调整方案。一致性的客观衡量指标可以使用 Krippendorff's alpha 统计量来确定, 需确定期望一致性和实际一致性, 计算式为 $(pe - pa)/(1 - pe)$ 。其中, 期望一致性 pe 由每个标签的数据频率的平方和来确定, 实际一致性 pa 是每个标注任务中标注员间一致性的平均值, 根据调整因子 ϵ 进行微调, ϵ 是每个标签计数的倒数和。

数据标注结果的置信度也需要重点关注。我们通常希望标注过程中标注员之间相互独立以提升标注质量。标注结果不确定度的计算公式为 $1 - np/(n + 1)$, 其中 p 是最高置信标签所对应的概率, n 是标注结果数。当某个数据的标注结果不确定度较高时, 可以安排更多标注人员或由相关专家判定, 也可以认为该数据质量较差, 从数据集中剔除。对于复杂的标注任务应该尽量拆分成分阶段的简单任务, 再交由标注人员标注, 以提升标注准确性。

3.3 生成与仿真技术

在以数据驱动的自动驾驶技术中, 提高数据集的多样性, 对提高系统应对复杂场景的能力至关重要, 如极端天气、复杂施工道路、非结构化道路等长尾场景 [242, 243]。获取海量丰富自动驾驶数据, 让自动驾驶算法达到充分的测试与验证, 对行车安全至关重要。

生成与仿真技术可以快速、低成本地获得海量自动驾驶数据。虽然获取海量数据最为直观的方法是采用大量采集车在道路中不断采集数据, 但这个方法的问题在于采集海量数据所需要的时间和成本是不可估量的。特斯拉通过大量用户返回的道路数据来优化它的完全自动驾驶 (Full Self-Driving, FSD) 系统。但该收集数据方法的前提是已经有大量用户购买并使用带有自动驾驶辅助系统的特斯拉汽车。这对于规模不大的车厂和没有采集车的学术机构来说是十分困难的。再者, 在处理用户返回数据时, 隐私、长尾场景等方面仍然不可被忽视。为此, 研究人员转向人为合成自动驾驶数据。相较于采集数据, 合成数据的优势在于: (1) 可根据需求产生对应数据, 对于现实场景中难以出现的边缘场景, 经过一定的场景设计就能产生大量数据; (2) 采集数据中可能涉及个人隐私、地理位置等敏感信息, 而合成数据不需要考虑隐私问题; (3) 虽然开发一套数据合成工具需要一定成本, 但合成海量数据的花费相比于实车采集大量数据十分低廉, 且边际成本低。

随着 AIGC (AI-Generated Content) 技术的发展, 愈来愈多的工作关注于使用人工智能模型生成自动驾驶场景。目前, 我们能够看到各个 AIGC 工作不尽相同, 如基于 VQ-VAE (Vector Quantized-Variational AutoEncoder) 的 comma ML Simulator [244] 与 BEVGen [245], 基于 NeRF (Neural Radiance Fields) 的 UniSim [246] 与 Drive-3DAug [247], 和基于 Diffusion 的 Tesla 世界模型 [248]、GAIA-1 [249]、VideoLDM [250]、GeoDiffusion [251] 等。相比于较为成熟的仿真系统, 这些方案都存在一定程度上的不足。首先, 相关工作没有对源代码进行开源, 研究人员无法很好的使用与探索相关模型。其次, 由于神经网络的参数由优化算法决定, 无法对生成的数据进行显式的几何或时序约束。最后, 传感器类型较为局限, 无法根据需求自定义采集所需模态的数据。

仿真软件的发展也使得在仿真环境中获取逼真数据成为可能。早期的自动驾驶仿真工具通常使用专业的游戏引擎, 如 Grand Theft Auto (GTA) 引擎, 来还原足够真实的场景。近年来, 数量丰富的自动驾驶仿真系统被开发出来, 如 CARLA [30]、SUMMIT [252]、SVL [253]、PGDrive [254]、AirSim [255] 等。这些仿真系统通常包含自定义景观、路网、天气等场景信息, 配置不同类型的传感器, 使得通过其采集的数据种类丰富。另一个特点则是行为者 (如交通和行人) 的行为可以根据上一

时刻的行为进行改变;为闭环评测提供了可能性。另外,在仿真系统中,由于车辆行人等所有物体都存在物理模型,相应的检测框等标签均可以自动生成,减少了大量标注成本。需要注意的是,由于现实环境与仿真环境存在着显著的域间差异,通过仿真系统生成数据训练的模型不能应用到现实环境中。如何结合现实与仿真环境,仍是一个值得关注的领域。

3.4 自动标注技术

自动驾驶系统需要大量的标注数据来训练和验证其算法。数据标注是将原始传感器数据(如摄像头、雷达和激光雷达数据)与目标(如车辆、行人、道路标志等)进行关联的过程。准确的数据标注是训练和验证自动驾驶算法的基础,对于系统的性能和安全性至关重要。针对传统的数据标注方法过度依赖标注员水平、成本居高不下以及标注效率低等问题,自动标注技术通过算法和模型自动地对数据进行标注,在一定程度上减少对人工标注的依赖。现有的市场上标注工具部分已经将该算法集成至工具箱中。

表 7 从自动标注技术、价格以及功能多个方面对比了亚马逊、百度、微软、数据堂、Segments.ai 以及 Scale AI 多家标注工具。这些标注工具都提供了一整套工具和相应功能,用于自动化和简化数据标注过程,并支持多种标注任务,如图像分类、目标检测、语义分割等。其中部分工具集成了半自动标注算法,如针对点云数据提供的三维目标标注任务,在亚马逊 Sagemaker Ground Truth 与数据堂 Data Pro 以及 Scale AI 的 Data Engine 标注平台上,标注员仅需要标注部分数据,或仅需要通过鼠标点击的方式给予算法提示,即可快速生成全量高质量的三维边界框标,实现降本增效。其相应的标注费用都在表 7 中说明,其中 Multi-sensor Labeling Platform, Sagemaker Ground Truth 及 Data Engine 需要对平台单独收费, VOTT (Visual Object Tagging Tool) 是开源工具对外免费,其他均不对外单独授予权。

以知名的 nuScenes [8] 数据集为例,其中数据量为 4 万帧图像与点云对,共包含 1,166,187 个实例目标,如果用表 7 中工具对其进行标注,标注任务为二维检测框。亚马逊、百度与数据堂的标注人民币价格分别为 62.4 万、9.3 万、9 万。以上价格通过官方渠道对外公示价格计算而得,在此仅供参考,并且实际标注价格还受到地区物价水平、标注质量以及标注效率影响。

3.5 应用案例

在这一章节中,我们选择了一个案例来阐释本章节中提及的方案。该案例的大致流程如图 4 所示。在这个案例中,研究人员的目标为开发一个基于语言的 AIGC 模型,用于根据文本内容生成与自动驾驶相关的图片。面临的一个挑战是,由于语言和自动驾驶的结合并不紧密,现有自动驾驶数据集基本上不含有语言相关标签。为了解决这个问题,研究人员利用 ChatGPT 语言模型,通过自动标注的方式为 nuScenes 数据集中的图片生成相应描述作为文本标签。由于 ChatGPT 并不是为自动驾驶专门设计的,研究人员为一张图片使用不同的模板生成文本,并且经过多次人工校验与挑选,以生成切实可用的数据标签对。在数据集准备完成后,研究人员使用标注好的数据对基于文本的生成模型 text2image 进行训练。但由于 nuScenes 数据集图片数量有限、覆盖场景较少, text2image 模型并不能根据文本生成某些特定场景的图片,如黑夜、雨天等。对此,研究人员引入外部驾驶数据集,如从 YouTube 上获取的海量驾驶图像,对现有 img2text 模型进行微调,使之成为一个专门

表 7 自动驾驶标注工具对比。以点云与图像标注任务为例, 本文搜集了国内外若干家数据公司的信息及其报价, 并以 nuScenes 数据集为例计算出所需标注价格。

Table 7 Comparison of annotation toolkits. **2D/3D Seg** means the 2D or 3D segmentation label. **2D/3D BBox** is for 2D or 3D bounding box. **2D BBox-c/ 3D BBox-c** indicates 3D bounding box label task for continuous frame data. **MA** represents manual annotation mode, while **SA** represents semi-automatic annotation mode. Under “Point Cloud” and “Image”, “Price” means the price for annotating a single instance. “Toolkit Price” refers to the cost of using annotation tools. “-” implies that related information is inaccessible to the public. “X” denotes that the target tool does not support this feature.

Company	Owner	Point Cloud				Image				Toolkit Price	Price of Sample Task
		3D Seg	Price	3D BBox	Price	3D BBox-c	2D Seg	Price	2D BBox	Price	
Sagemaker Ground Truth [256]	amazon	MA	-	SA	-	SA	SA	5.360	SA	0.255	SA
智能云 [257]	Baidu 百度	MA	0.3	MA	0.2-0.25	MA	MA	0.2	MA	0.06-1.1	MA
VOTT [258]	Microsoft	X	-	X	-	X	MA	-	SA	-	SA
Data Pro [259]	数据堂 DataTang	SA	0.25	SA	0.3	SA	SA	0.25	SA	0.07	SA
Multi-sensor Labeling Platform [260]	Segments.ai	SA	-	SA	-	SA	SA	-	SA	-	SA
Data Engine [261]	scale	SA	-	SA	-	SA	SA	18.2	SA	5.8	SA

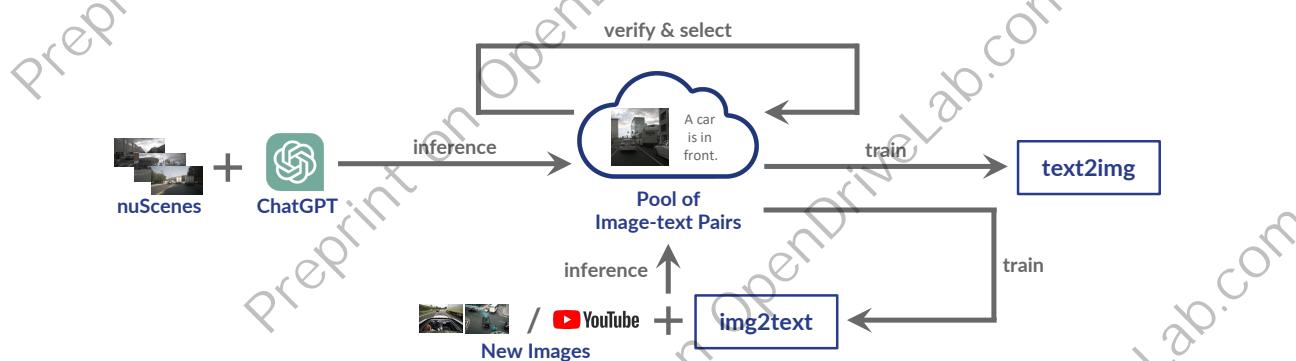


图 4 应用案例: 大语言模型通过数据算法闭环体系, 生成海量高质量数据。

Figure 4 Large Language Models (LLMs) are applied in the data-algorithm pipeline.

为自动驾驶相关图片标注模型。同时, 大量新的自动驾驶图片被采集, 并通过 img2text 模型进行推理。在大量图像文本对中, 符合要求的数据对被挑选出来并再次用于微调模型。如图 4 中的图像文本对池子所示, 上下两个循环, 即人工干预与数据生成, 交替进行, 持续地提高池子中的数据质量。在多次迭代后, 数据规模和场景丰富度都相较于 nuScenes 数据集有很大提升。使用基于这些数据训练得出的 text2image 模型也因此可以生成高质量的驾驶场景数据。可以看出, 在数据算法闭环体系中, 不同模块的使用并不需要严格按照它们原有的顺序进行, 而可以根据项目的特定需求自由组合, 以达到最佳效果。

在图 5 中, 我们可以看到该目标模型生成的样例结果。这个模型能够根据输入的文本内容生成与自动驾驶相关的高质量图片, 并且能够通过改变文字含义有效地生成不同的自动驾驶场景图片。如子图 (a) 所示, 文本中只有 “Tokyo” (东京) 一词暗示自动驾驶汽车所在地域, 模型能够生成在肉眼上看风格十分接近东京街景的图片。而子图 (b) 的文本与之相比只有一个单词的差别, 生产的图片虽然不能看出所在地点, 但在风格上能够明显区别于子图 (a)。由子图 (c) 与 (d) 可以看出, 模

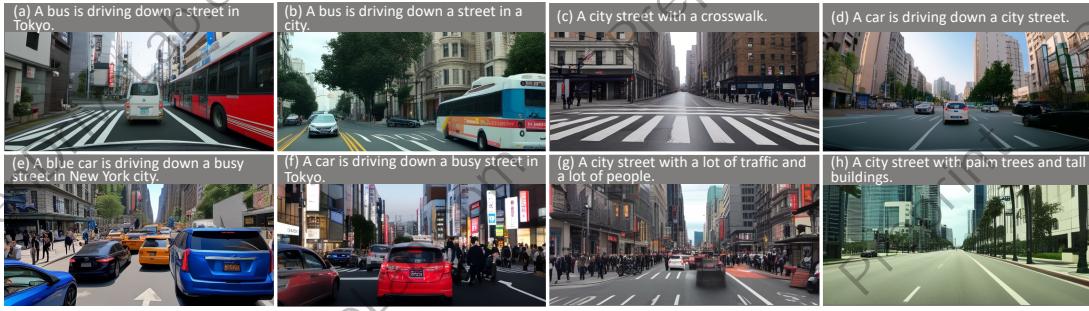


图 5 案例模型生成结果。通过对语言大模型的运用，研究人员达成通过文本输入生成高质量自动驾驶图片的目标。

Figure 5 Examples of generated images. By utilizing Large Language Models (LLMs), researchers achieve the goal of generating high-quality autonomous driving images based on text.

表 8 大模型对数据与模型参数的要求。与其他行业大模型对比，我们估算出自动驾驶大模型所需数据量与参数量。

Table 8 Comparison of data and model parameters in the development of large models. “# Parameter” denotes the number of parameters in the neural network. Under “# Parameter”, “B” denotes billion, “V” means vision models, “VL” indicates vision-language models, and “Gen” represents generative models.

Field	Method	Organization/Date	Data Scale	# Parameter	Open-source	Feature
自然语言	GPT-4 [262]	OpenAI/2023.3	13T Tokens	1800B (估)	✗	综合表现最强的 LLM
	LLaMA 2 [263]	Meta/2023.7	2T Tokens	7-70B	✓	开源、全栈 LLM
通用视觉	ViT-22B [264]	Google/2023.2	4B Images	22B	✗	目前最大的纯视觉模型 (单模态)
	BLIP-2 [265] (多模态)	Salesforce/2023.1	129M Images	12.1B	✓	LLM + Vision → VLM
机器人	PaLM-E [266]	Google/2023.3	Unknown	562B	✗	LLM + Robotics
	RT-2 [267]	DeepMind/2023.7	1B Image-Text Pairs, 13 Robots, 17 Months	5/12/55B	✗	LLM + Robotics
医疗	Med-PaLM 2 [268]	Google/2023.4	1M Samples	562B	✗	LLM + Med
自动驾驶	World Model Demo [269]	Tesla/2023.9.30	Unknown	1.1B (估)	✗	自驾领域世界模型
	DriveAGI [270]	OpenDriveLab /Scheduled for 2024	4M Images, LiDAR, etc. + 32M Images	1.2B (V/VL) + 2.6B (Gen)	✓	强推理, 多模态, 闭环: 感知 + 生成

型对道路场景与单词的对应关系有正确的认识，能够根据“crosswalk”（人行横道）与“city street”（城市街道）生成对应的道路场景。在第二行的四幅图中，模型也能对颜色、物件、事物概率等有充分认知。总的来说，生成图像的整体风格符合文本描述，如果要对图像中细节进一步调整，如车道线、路标等交通元素的位置与属性等，需要更加精细的文本描述用以生成定制化的场景数据。由此得出，经过数据算法闭环体系训练出来的模型拥有生成任意自动驾驶场景图片的能力。

4 新一代自动驾驶数据集

4.1 大模型的启示

通用大模型 vs 自动驾驶大模型。如表 8 所示，当前基础大模型在自然语言处理、计算机视觉等领域取得了举世瞩目的成果，掀起了新一波的人工智能研究热潮。但目前市面上还没有面向自动驾

驶垂直领域的大模型。在数据规模与模型参数量方面,自动驾驶与其它领域相比仍有较大差距。受到数据采集合规性以及成本等方面的限制,数据规模明显小于其他领域。而较小的数据规模也不足以支撑大模型的优化,限制了模型参数增长的可能性。虽然在自动驾驶领域,较小的模型有利于在算力较小的车辆平台上达到更好的实时性,但模型性能也会受到限制,也不利于达到在离线平台上的性能要求。由此可见,在数据量与模型规模上,自动驾驶大模型仍有较大的发展空间。以其他领域的大模型作为参照,新一代数据集至少应将数据量提升至与其他领域相近才能够赋能自动驾驶大模型。值得注意的是,在大多数领域大模型的引入也意味着自然语言的引入,也更偏向于通用大模型。由于自动驾驶系统的输入为传感器数据而输出为车辆路径规划,与自然语言的结合并不直观。已有部分工作尝试将自然语言引入自动驾驶领域中,如 NuInstruct^[271] 将基于驾驶视频的问答文本作为数据标签对外公开,旨在鼓励业界探索语言在自动驾驶中的作用, GAIA-1^[272] 将生成模型与语言结合起来生成定制化的视频场景,还有工作^[273] 初步证明了引入语言能够提高自动驾驶系统在长尾场景下的表现能力。

数据规模 vs 算法性能。尽管无限的数据可以使神经网络达到接近百分之百的水平,但获得如此大量的成本也是不可估量的。通过之前的一些工作可以看出^[274~278],数据规模的增大能够显著提升模型性能,但当数据量达到一定程度之后,模型性能增长趋于平缓。并且,自动驾驶车辆在真实世界中会不可避免地遇到训练数据之外的场景。因此,大规模地应用自动驾驶技术必然要求模型能够在罕见场景中做出正确行为,避免发生危险或功能失效的情况。由此看来,在自动驾驶领域我们并不需要一味地去扩充数据量。对于绝大多数交通场景来说,并不需要十分大量的数据就能够支撑自动驾驶可靠的运行。而更需要关注的是长尾场景,由于某些交通场景十分罕见,如撞车等,相关数据的缺失会对自动驾驶系统的性能影响巨大。总的来说,在保证数据数量的前提下,场景丰富度对算法性能更为重要。

4.2 新数据集建设要点

第一、二代自动驾驶数据集已经不能够继续满足自动驾驶系统的发展需求,新一代数据集的建设亟待提上日程。在大模型蓬勃发展的背景下,大数据成为新一代数据集不可缺少的一个特点。同时,模块化设计的自动驾驶系统在落地过程遇到迭代成本高、性能上界受限等问题,端到端自动驾驶架构逐步受到业界的青睐。除此之外,多模态传感器、高质量标注、模型逻辑推理能力等方面也需要得到重视。在这一章节中,我们总结归纳了新一代自动驾驶数据集的发展目标。

4.2.1 面向多模态、保质保量

作为自动驾驶系统了解外部环境的第一步,传感器数据的获取为后续系统的可靠运行提供了保证。尽管相机个数在现有不同公开数据集上仍然存在差异,环视视角已经成为现有开源自动驾驶数据集的共识。除此之外,激光雷达也是一个重要的传感器数据来源。虽然现有量产自动驾驶系统通常依赖这两类传感器,其它类别传感器对自动驾驶系统的作用仍然值得探索。在新一代数据集中,更多的传感器类型数据,如鱼眼相机、双目相机、毫米波雷达、惯性测量单元等,应该被包含在其中。单一类型的传感器也应具有多样性,如在相机中需要包含长焦、鱼眼、广角、窄角等镜头。在保证传感器种类丰富的同时,不同类型传感器数据之间的时间、空间、覆盖范围的对齐是一个需要特别

重视的问题。同时,卫星导航、高精地图、导航地图等定位数据也需要被考虑在多模态数据中。针对现有传感器各自优缺点,如何有效的融合提升自动驾驶系统整体性能成为多模态数据集需要回答的问题。

除了传感器的多样性,数据场景多样性也同样重要,包括结构化和非结构化道路、城市和郊区、不同国家与地区、不同天气条件等场景,以及对自动驾驶系统行为有不同要求的场景,如城区、高速、泊车等。对于静态场景建图数据,需要考虑是否应包含在同一地点多次采集的数据,以赋能联合建图等任务。与数据数量相似,标注的质量与多样性对自动驾驶技术发展同样重要。标注规则的合理统一、标注人员的技能水平等因素决定了一个数据集的标注质量。而多样的数据标签则有利于研究人员在同一数据集上进行多种自动驾驶任务,如检测框适用于目标检测任务、环境深度信息适用于深度估计任务等。

4.2.2 面向端到端、决策导向

现有自动驾驶系统通常以模块化形式存在,整个系统主要被拆分成感知、预测、规划与控制模块。而每个模块会被进一步的拆分成不同的任务,如感知模块中含有深度估计、物体检测、道路结构认知等任务。模块化设计使得单个任务流程十分简短,有利于研究人员对特定任务进行针对性设计。但对于自动驾驶系统来说,模块化设计不可避免地造成模块内误差在系统不同模块间不断积累,导致级联误差。端到端技术将自动驾驶各个子模块与任务整合到统一的网络模型中,面向最终规控任务的优化范式可有效避免级联误差。端到端自动驾驶技术早在1988年被Alvinn^[279]提出,伴随着神经网络、高性能计算、车载传感器等技术的持续演进,在近两年得到迅猛发展。新一代的端到端自动驾驶技术以UniAD^[280]、MP3^[281]、ST-P3^[282]为代表,采用深度学习框架主干网络整合和优化自动驾驶全流程的关键任务,从而在感知、预测和规划等方面达到更高的准确性和效率。可以看出,第一与第二代自动驾驶数据集以感知任务为主,辅以少部分预测和规划任务,不能直接用于端到端自动驾驶模型的训练与评测。CARLA^[30]提供的仿真环境使得端到端闭环评测成为可能,但真实世界与仿真环境的虚实差距不可忽视。在新一代自动驾驶数据集中,赋能端到端模型是一个必备特点,同时也需要考虑如何将闭环评测、现实环境、仿真环境进行有机统一。同时,对于小型研究团队来说,过于冗长的端到端开发流程可能会限制其参与自动驾驶相关研究。新一代数据集中也需要包含一些尚未被深入研究的自动驾驶任务新范式,如使用占据栅格统一动态物体和静态场景、使用世界模型获得更好的预测能力、构建语言相关的DriveGPT模型等。对于端到端系统来说,较为简短的任务能够对自动驾驶技术中短期发展起到有效的带动作用。

4.2.3 面向智能化、逻辑推理

如何根据外部环境的变化作出正确的决策,是自动驾驶系统必须要回答的问题。这一过程并不止步于神经网络具有的通过大量数据训练得出的模仿功能,还需要确保做出决策的逻辑推理过程的正确性。神经网络常作为人工智能的实现形式,但其黑箱的特性使得完全的正确性不能够得到保证。尽管根据神经网络在其他任务上的表现,其可以模仿几乎所有的函数,但在实际落地过程中并非如此,任何一瞬的系统失效对于车辆及乘客来说仍可能是致命性的打击。基于隐式模型的自动驾驶系统不可避免的存在如上问题,因此提高自动驾驶系统可解释性,提高车辆安全性至关重要。大语言

模型具有出色的推理能力和知识理解能力，面对复杂场景可以模拟人的决策过程。我们希望新一代自动驾驶数据集能结合大语言模型的推理能力，提高自动驾驶系统可解释性与安全性。

4.3 产业推广与应用

除了技术难题需要攻克，搭建新一代的自动驾驶数据集需要面临诸多其他挑战，也需要产业界、学术界、相关组织与行业协同配合。在这一章节中，我们列举了部分需要考虑的方面。

数据隐私与合规。自动驾驶数据包括自动驾驶汽车设计、生产、销售、使用、运维等过程中所涉及个人数据、重要数据、国家核心数据、地理信息数据、企业数据。本文主要关注通过摄像头、雷达等传感器从汽车外部环境采集的道路、建筑、地形、交通参与者等数据，以及对其进行加工后产生的数据（相关国家标准 GB/T 41871-2022《信息安全技术汽车数据处理安全要求》）。多家单位在采集数据的同时，相关部门也在围绕数据隐私与合规加快立法进程。例如，2021 年五部委联合发布了《汽车数据安全管理若干规定（试行）》，规范汽车数据处理活动，保护个人、组织的合法权益，维护国家安全和社会公共利益，促进汽车数据合理开发利用。

数据地域与政策。有限的开放数据获取权限使得自动驾驶涉及的模型研发难以通过大规模的数据及场景来加强算法的性能设计及验证。尤为突出的是数据缺乏代表性的问题，目前可开放获取的自动驾驶数据大多采集于国外的道路环境。这些数据集通常包括多个城市中的不同区域，涵盖丰富的交通事件、气候天气、光照条件等因素。而国内鲜有成规模、质量高的数据集能够开放获取权限。目前，随着智能网联汽车技术创新和产业化进程不断加快，国家及各地方省市支持建设了 17 个国家级智能网联汽车测试示范区，推动全国 26 个省（市）出台管理细则，开放 3200 多公里测试道路，发放 700 余张测试牌照，在公路客货运输、城市出行与物流、园区内运输、特定场景作业等领域布局试点项目持续推进，并在北京、上海等地支持开展载人载物示范应用。数据集建设从面向科研的小规模数据集向大规模城市及场景采集的海量数据集迈进，离不开政策和产业的持续支持和投入。

数据共建与联盟。数据的获取成本十分昂贵，对于大多数数据使用者来说都难以承担。近些年来，由多个企业或组织自发组建数据联盟促进数据资源共享和合作，整合各方数据资源，减少重复采集、重复标注及处理等各自数据冗余和浪费，提高数据的利用效率和质量。例如，中国汽车工程学会、中国测绘学会、国家智能网联汽车创新中心、中国地图出版社集团有限公司作为发起、联合汽车、地信、交通、示范区等 30 余家企事业单位共同成立的智能网联汽车自动驾驶地图基础平台产业联盟，致力于面向智能网联汽车产业的平台共建、数据共享、产业跨域合作，在符合国家数据安全前提下推动自动驾驶地图的有序建设和安全应用。数据资源的整合需要在产业共建联盟的推动下，充分发挥资源整合的优势，打破数据壁垒，加强企业和机构之间的合作，不断推进技术创新。

5 总结

本文详细评述了自动驾驶公开数据集的现状与挑战。针对数据算法闭环体系，结合当前大模型发展趋势，提出了下一代自动驾驶数据集的愿景与规划。我们系统性地总结了自动驾驶发展历程中所使用的数据集，并展示了通过挑战赛与榜单促进社区发展的重要性。紧接着，我们概括性地分析了自动驾驶数据算法闭环体系，并总结其中各个重要环节的作用，最后通过应用案例展现对数据算

法闭环体系的使用方法。基于以上内容,我们总结并展望未来自动驾驶数据集的方向,借此希望能够进一步推动自动驾驶在学术界与产业界中的发展。

作者贡献 李弘扬、乔宇负责整体项目方向与定位;李弘扬撰写引言部分,李阳负责撰写全文(第二至第四章节),并整体修改与制图,王晖杰负责撰写挑战赛、数据闭环部分,曾嘉负责全文内容、格式、风格校验与统一,徐慧琳负责文章内容的修改与格式统一,蔡品隆负责撰写法律法规、交叉数据集等内容;林达华、陈立,严骏驰、徐丰和熊璐从各自的角度对文章做出了修改。王井东、许春景对全文架构、整体传达思想做了修改;朱福堂、颜凯、汪天才、夏飞、穆北鹏和彭志辉(即稚晖君)对新一代数据建设提出了具体建议和修改。

致谢 感谢OpenDriveLab浦驾成员在提纲讨论、素材提供、文章校对等方面给予的诸多协助。特别地,感谢杨佳智、王嘉豪、司马崇昊、李天羽、汪邦骏、谢城恩、贾萧松、吴鹏浩、曹晓旭等对本工作提供的素材校对与文稿整理。感谢产业界同事对本工作提出的建议与指导,包括张骞、张祥雨、孙诚骁、王昊奋等。

参考文献

- 1 The State Council of the People's Republic of China. Next Generation Artificial Intelligence Development Plan, 2017. [中华人民共和国国务院. 新一代人工智能发展规划. 中国政府网. 2017-7-20].
- 2 European Commission. White Paper on Artificial Intelligence: a European approach to excellence and trust, 2020.
- 3 National Science and Technology Council of USA. NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN 2023 UPDATE, 2023.
- 4 Jinping Xi. *On Self-reliance and Strength in Science and Technology*. Central Party Literature Press, 2023. [习近平. 论科技自立自强. 中央文献出版社. 2023].
- 5 The Central Committee of the Communist Party of China and The State Council of the People's Republic of China. Outline on Building China's Strength in Transportation, 2019. [中国共产党中央委员会, 国务院中华人民共和国国务院. 交通强国建设纲要. 中央文献出版社. 2019-09-19].
- 6 The Ministry of Industry and Information Technology, the Ministry of Public Security, and the Ministry of Transport of the People's Republic of China. Management Specifications for Road Testing and Demonstration Application of Intelligent Connected Vehicles (Trial), 2021. [中华人民共和国国务院工业和信息化部公安部交通运输部. 关于印发《智能网联汽车道路测试与示范应用管理规范(试行)》的通知. 中国政府网. 2021-07-27].
- 7 A Geiger, P Lenz, C Stiller, et al. Vision meets robotics: The KITTI dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013.
- 8 Holger Caesar, Varun Bankiti, Alex H Lang, et al. NuScenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11621–11631, 2020.
- 9 Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2446–2454, 2020.
- 10 Marius Cordts, Mohamed Omran, Sebastian Ramos, et al. The Cityscapes Dataset for Semantic Urban Scene Understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016.
- 11 Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulo, et al. The Mapillary Vistas Dataset for Semantic Understanding of Street Scenes. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4990–4999, 2017.
- 12 Peter Pinggera, Sebastian Ramos, Stefan Gehrig, et al. Lost and found: detecting small road hazards for self-driving vehicles. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1099–1106, 2016.
- 13 Harald Schafer, Eder Santana, Andrew Haden, and Riccardo Biasini. A commute in data: The comma2k19 dataset. *arXiv preprint arXiv:1812.05752*, 2018.
- 14 Xinyu Huang, Peng Wang, Xinjing Cheng, et al. The apolloscape open dataset for autonomous driving and its application. *IEEE transactions on pattern analysis and machine intelligence*, 42(10):2702–2719, 2019.

- 15 John Houston, Guido Zuidhof, Luca Bergamini, et al. One thousand and one hours: Self-driving motion prediction dataset. In *Conference on Robot Learning*, pages 409–418, 2021.
- 16 Ming-Fang Chang, John W Lambert, Patsorn Sangkloy, et al. Argoverse: 3D Tracking and Forecasting with Rich Maps. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8748–8757, 2019.
- 17 A. Ouaknine, A. Newson, J. Rebut, et al. CARRADA dataset: Camera and automotive radar with range-angle-doppler annotations. In *2020 25th International Conference on Pattern Recognition*, pages 5068–5075, 2021.
- 18 Holger Caesar, Juraj Kabzan, Kok Seang Tan, Whye Kit Fong, Eric Wolff, Alex Lang, Luke Fletcher, Oscar Beijbom, and Sammy Omari. NuPlan: A closed-loop ML-based planning benchmark for autonomous vehicles. *arXiv preprint arXiv:2106.11810*, 2022.
- 19 Benjamin Wilson, William Qi, Tanmay Agarwal, et al. Argoverse 2: Next Generation Datasets for Self-driving Perception and Forecasting. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021.
- 20 Senthil Yogamani, Ciaran Hughes, Jonathan Horgan, et al. WoodScape: A multi-task, multi-camera fisheye dataset for autonomous driving. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9308–9318, 2019.
- 21 Arun Balajee Vasudevan, Dengxin Dai, and Luc Van Gool. Talk2Nav: Long-range vision-and-language navigation with dual attention and spatial memory. *International Journal of Computer Vision*, 129:246–266, 2021.
- 22 J. Houston, G. Zuidhof, L. Bergamini, Y. Ye, A. Jain, S. Omari, V. Iglovikov, and P. Ondruska. One Thousand and One Hours: Self-driving Motion Prediction Dataset. <https://level5.lyft.com/dataset>, 2020.
- 23 Bassam Helou, Aditya Dusi, Anne Collin, et al. The Reasonable Crowd: Towards evidence-based and interpretable models of driving behavior. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 6708–6715, 2021.
- 24 J. Behley, M. Garbade, A. Milioto, et al. SemanticKITTI: A dataset for semantic scene understanding of lidar sequences. In *2019 IEEE/CVF International Conference on Computer Vision*, pages 9296–9306, 2019.
- 25 Guorun Yang, Xiao Song, Chaocin Huang, et al. DrivingStereo: A large-scale dataset for stereo matching in autonomous driving scenarios. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 899–908, 2019.
- 26 Yiyi Liao, Jun Xie, and Andreas Geiger. KITTI-360: A novel dataset and benchmarks for urban scene understanding in 2d and 3d. *IEEE transactions on pattern analysis and machine intelligence*, 2022.
- 27 OpenDriveLab. Autonomous Driving Challenge. <https://opendrivelab.com/AD23Challenge.html>, 2023.
- 28 Waymo. Waymo Open Dataset Challenges. <https://waymo.com/open/challenges>, 2023.
- 29 CARLA. CARLA Autonomous Driving Challenge. <https://leaderboard.carla.org>, 2023.
- 30 Alexey Dosovitskiy, German Ros, Felipe Codevilla, et al. CARLA: An open urban driving simulator. In *Conference on robot learning*, pages 1–16, 2017.
- 31 番洲实验室. 粤港澳大湾区（黄埔）国际算法算例大赛. <https://iacc.pazhoulab-huangpu.com>, 2023.
- 32 计图. 计图人工智能算法挑战赛. <https://cg.cs.tsinghua.edu.cn/jittor/news/2021-1-22-13-14-comp>, 2021.
- 33 Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE conference on computer vision and pattern recognition*, pages 3354–3361, 2012.
- 34 Jannik Fritsch, Tobias Kuehnl, and Andreas Geiger. A New Performance Measure and Evaluation Benchmark for Road Detection Algorithms. In *16th International IEEE Conference on Intelligent Transportation Systems*, pages 1693–1700, 2013.
- 35 Moritz Menze and Andreas Geiger. Object Scene Flow for Autonomous Vehicles. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3061–3070, 2015.
- 36 Jonathan Binas, Daniel Neil, Shih-Chii Liu, et al. DDD17: End-to-end davis driving dataset. *arXiv preprint arXiv:1711.01458*, 2017.
- 37 Jinkyu Kim, Anna Rohrbach, Trevor Darrell, et al. Textual explanations for self-driving vehicles. In *European Conference on Computer Vision*, pages 563–578, 2018.
- 38 Vasili Ramanishka, Yi-Ting Chen, Teruhisa Misu, et al. Toward Driving Scene Understanding: A Dataset for Learning Driver Behavior and Causal Reasoning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7699–7707, 2018.
- 39 Girish Varma, Anbumani Subramanian, Anoop Namboodiri, et al. IDD: A dataset for exploring problems of autonomous navigation in unconstrained environments. In *2019 IEEE Winter Conference on Applications of Computer Vision*, pages 1743–1751, 2019.

- 40 Adam Ligocki, Ales Jelinek, and Ludek Zalud. Brno Urban Dataset – the new data for self-driving agents and mapping tasks. In *2020 IEEE International Conference on Robotics and Automation*, pages 3284–3290, 2020.
- 41 Quang-Hieu Pham, Pierre Sevestre, Ramanpreet Singh Pahwa, et al. A*3D Dataset: Towards Autonomous Driving in Challenging Environments. In *2020 IEEE International conference on Robotics and Automation*, pages 2267–2273, 2020.
- 42 Thierry Deruyttere, Simon Vandenhende, Dusan Grujicic, et al. Talk2Car: Taking control of your self-driving car. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 2088–2098, 2019.
- 43 Amir Rasouli, Iuliia Kotseruba, Toni Kunic, et al. PIE: A large-scale dataset and models for pedestrian intention estimation and trajectory prediction. In *Proceedings of the IEEE international conference on computer vision*, pages 6262–6271, 2019.
- 44 Weisong Wen, Yiyang Zhou, Guohao Zhang, et al. UrbanLoco: A full sensor suite dataset for mapping and localization in urban scenes. In *2020 IEEE international conference on robotics and automation*, pages 2310–2316, 2020.
- 45 Srikanth Malla, Behzad Dariush, and Chiho Choi. TITAN: Future forecast using action priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11186–11196, 2020.
- 46 Abhishek Patil, Srikanth Malla, Haiming Gang, et al. The H3D Dataset for Full-Surround 3D Multi-Object Detection and Tracking in Crowded Urban Scenes. In *2019 International Conference on Robotics and Automation*, pages 9552–9557, 2019.
- 47 Jakob Geyer, Yohannes Kassahun, Mentar Mahmudi, et al. A2D2: Audi autonomous driving dataset. *arXiv preprint arXiv:2004.06320*, 2020.
- 48 Mourad A Kenk and Mahmoud Hassaballah. DAWN: Vehicle detection in adverse weather nature dataset. *arXiv preprint arXiv:2008.05402*, 2020.
- 49 Patrick Wenzel, Rui Wang, Nan Yang, et al. 4Seasons: A cross-season dataset for multi-weather slam in autonomous driving. *arXiv preprint arXiv:2009.06364*, 2020.
- 50 Sauradip Nag, Saptakatha Adak, and Sukhendu Das. What's There in the Dark. In *2019 IEEE International Conference on Image Processing*, pages 2996–3000, 2019.
- 51 Yancheng Pan, Biao Gao, Jilin Mei, et al. Semanticpos: A point cloud dataset with large quantity of dynamic instances. In *2020 IEEE Intelligent Vehicles Symposium*, pages 687–693, 2020.
- 52 Weikai Tan, Nannan Qin, Lingfei Ma, et al. Toronto-3D: A large-scale mobile lidar dataset for semantic segmentation of urban roadways. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 202–203, 2020.
- 53 Gurkirt Singh, Stephen Akriigg, Manuele Di Maio, et al. ROAD: The road event awareness dataset for autonomous driving. *IEEE transactions on pattern analysis and machine intelligence*, 45(1):1036–1054, 2022.
- 54 Rohan Chandra, Xijun Wang, Mridul Mahajan, et al. METEOR: A dense, heterogeneous, and unstructured traffic dataset with rare behaviors. In *2023 IEEE International Conference on Robotics and Automation*, pages 9169–9175, 2023.
- 55 Pengchuan Xiao, Zhenlei Shao, Steven Hao, et al. Pandaset: Advanced sensor suite dataset for autonomous driving. In *2021 IEEE International Intelligent Transportation Systems Conference*, pages 3095–3101, 2021.
- 56 Gianni Franchi, Xuanlong Yu, Andrei Bursuc, et al. MUAD: Multiple uncertainties for autonomous driving, a benchmark for multiple uncertainty types and tasks. *arXiv preprint arXiv:2203.01437*, 2022.
- 57 Peter Mortimer and Hans-Joachim Wuensche. TAS-NIR: A VIS+NIR dataset for fine-grained semantic segmentation in unstructured outdoor environments. *arXiv preprint arXiv:2212.09368*, 2022.
- 58 Jin Fang, Dingfu Zhou, Jingjing Zhao, et al. LiDAR-CS Dataset: Lidar point cloud dataset with cross-sensors for 3d object detection. *arXiv preprint arXiv:2301.12515*, 2023.
- 59 Oliver Zendel, Katrin Honauer, Markus Murschitz, et al. WildDash - creating hazard-aware benchmarks. In *European Conference on Computer Vision*, pages 402–416, 2018.
- 60 Chonghao Sima, Wenwen Tong, Tai Wang, et al. Scene as Occupancy. In *Proceedings of the IEEE international conference on computer vision*, pages 8406–8415, 2023.
- 61 Mina Alibeigi, William Ljungbergh, Adam Tonderski, et al. Zenseact open dataset: A large-scale and diverse multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 20178–20188, 2023.
- 62 Chonghao Sima, Katrin Renz, Kashyap Chitta, et al. DriveLM: Driving with graph visual question answering. *arXiv preprint arXiv:2312.14150*, 2023.
- 63 Zhenyuan Zhang, Xiaojie Wang, Darong Huang, et al. idt: An integration of detection and tracking toward low-observable multipedestrian for urban autonomous driving. *IEEE Transactions on Industrial Informatics*, 2023.
- 64 Zhenyuan Zhang, Xiaojie Wang, Darong Huang, et al. Mrpt: Millimeter-wave radar-based pedestrian trajectory tracking

- for autonomous urban driving. *IEEE Transactions on Instrumentation and Measurement*, 71:1–17, 2021.
- 65 Taoying Li, Xutong Li, Bo Ren, et al. An effective multi-scale framework for driver behavior recognition with incomplete skeletons. *IEEE Transactions on Vehicular Technology*, pages 1–15, 2023.
- 66 G.P. Stein, O. Mano, and A. Shashua. Vision-based ACC with a single camera: bounds on range and range rate accuracy. In *IEEE IV2003 intelligent vehicles symposium.*, pages 120–125, 2003.
- 67 Young-Woo Seo and Ragunathan Rajkumar. Use of a monocular camera to analyze a ground vehicle’s lateral movements for reliable autonomous city driving. In *Proceedings of IEEE IROS Workshop on Planning, Perception and Navigation for Intelligent Vehicles*, pages 197–203, 2013.
- 68 Mahdi Rezaei, Mutsuhiro Terauchi, and Reinhard Klette. Robust Vehicle Detection and Distance Estimation Under Challenging Lighting Conditions. *IEEE Transactions on Intelligent Transportation Systems*, 16(5):2723–2743, 2015.
- 69 Tai Wang, Xinge Zhu, Jiangmiao Pang, et al. FCOS3D: Fully Convolutional One-Stage Monocular 3D Object Detection. In *Proceedings of the IEEE international conference on computer vision*, pages 913–922, 2021.
- 70 Tai Wang, Qing Lian, Chemping Zhu, et al. Mv-fcos3d++: Multi-view camera-only 4d object detection with pretrained monocular backbones. *arXiv preprint arXiv:2207.12716*, 2022.
- 71 Zhiqi Li, Wenhui Wang, Hongyang Li, et al. BEVFormer: Learning bird’s-eye-view representation from multi-camera images via spatiotemporal transformers. In *European Conference on Computer Vision*, pages 1–18, 2022.
- 72 Tingting Liang, Hongwei Xie, Kaicheng Yu, et al. BEVFusion: A Simple and Robust LiDAR-Camera Fusion Framework. In *2023 IEEE international conference on robotics and automation*, pages 2774–2781, 2023.
- 73 Tianjiao Huo, Jiaqi Fan, Xin Li, et al. Traffic sign recognition based on resnet-20 and deep mutual learning. In *2020 Chinese Automation Congress*, pages 4770–4774, 2020.
- 74 Mohamed Aly. Real time detection of lane markers in urban streets. In *2008 IEEE intelligent vehicles symposium*, pages 7–12, 2008.
- 75 Seokju Lee, Junsik Kim, Jae Shin Yoon, et al. Vpgnet: Vanishing point guided network for lane and road marking detection and recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 1947–1955, 2017.
- 76 TuSimple. <https://github.com/TuSimple/tusimple-benchmark>, 2017.
- 77 Xingang Pan, Jianping Shi, Ping Luo, et al. Spatial As Deep: Spatial CNN for Traffic Scene Understanding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.
- 78 Karsten Behrendt and Ryan Soussan. Unsupervised Labeled Lane Markers Using Maps. In *Proceedings of the IEEE/CVF international conference on computer vision workshops*, pages 0–0, 2019.
- 79 Yuliang Guo, Guang Chen, Peitao Zhao, et al. Gen-lanenet: A generalized and scalable approach for 3d lane detection. In *European Conference on Computer Vision*, pages 666–681, 2020.
- 80 Hang Xu, Shaoju Wang, Xinyue Cai, et al. Curvelane-nas: Unifying lane-sensitive architecture search and adaptive point blending. In *European Conference on Computer Vision*, pages 689–704, 2020.
- 81 Yujun Zhang, Lei Zhu, Wei Feng, et al. VIL-100: a new dataset and a baseline model for video instance lane detection. In *Proceedings of the IEEE international conference on computer vision*, pages 15681–15690, 2021.
- 82 Li Chen, Chonghao Sima, Yang Li, et al. Persformer: 3D Lane Detection via Perspective Transformer and the OpenLane Benchmark. In *European Conference on Computer Vision*, pages 550–567, 2022.
- 83 Fan Yan, Ming Nie, Xinyue Cai, et al. ONCE-3DLanes: Building Monocular 3D Lane Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17143–17152, 2022.
- 84 Huijie Wang, Tianyu Li, Yang Li, et al. OpenLane-V2: A Topology Reasoning Benchmark for Scene Understanding in Autonomous Driving. *arXiv preprint arXiv:2304.10440*, 2023.
- 85 Jun Wang, Tao Mei, Bin Kong, et al. An approach of lane detection based on inverse perspective mapping. In *17th International IEEE Conference on Intelligent Transportation Systems*, pages 35–38. IEEE, 2014.
- 86 Zuo-Quan Li, Hui-Min Ma, and Zheng-Yu Liu. Road lane detection with Gabor filters. In *2016 International conference on information system and artificial intelligence*, pages 436–440, 2016.
- 87 Davy Neven, Bert De Brabandere, Stamatis Georgoulis, et al. Towards End-to-End Lane Detection: an Instance Segmentation Approach. In *2018 IEEE intelligent vehicles symposium*, pages 286–291, 2018.
- 88 Yuenan Hou, Zheng Ma, Chunxiao Liu, et al. Learning lightweight lane detection cnns by self attention distillation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1013–1021, 2019.
- 89 Zhenpeng Chen, Qianfei Liu, and Chenfan Lian. Pointlanenet: Efficient end-to-end cnns for accurate real-time lane detection. In *2019 IEEE intelligent vehicles symposium*, pages 2563–2568, 2019.
- 90 Xiang Li, Jun Li, Xiaolin Hu, et al. Line-CNN: End-to-End Traffic line detection with line proposal unit. *IEEE Transactions*

- on Intelligent Transportation Systems, 21(1):248–258, 2019.
- 91 Zequn Qin, Huanyu Wang, and Xi Li. Ultra fast structure-aware deep lane detection. In *European Conference on Computer Vision*, pages 276–291, 2020.
- 92 Hala Abualsaad, Sean Liu, David B Lu, et al. Laneaf: Robust multi-lane detection with affinity fields. *IEEE Robotics and Automation Letters*, 6(4):7477–7484, 2021.
- 93 Lizhe Liu, Xiaohao Chen, Siyu Zhu, et al. Condlanenet: a top-to-down lane detection framework based on conditional convolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3773–3782, 2021.
- 94 Oshada Jayasinghe, Damith Anhettigama, Sahan Hemachandra, et al. SwiftLane: Towards Fast and Efficient Lane Detection. In *2021 20th IEEE International Conference on Machine Learning and Applications*, pages 859–864, 2021.
- 95 Zhan Qu, Huan Jin, Yang Zhou, et al. Focus on local: Detecting lane marker from bottom up via key point. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.
- 96 Lucas Tabelini, Rodrigo Berriel, Thiago M Paixao, et al. Keep your eyes on the lane: Real-time attention-guided lane detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 294–302, 2021.
- 97 Jimming Su, Chao Chen, Ke Zhang, et al. Structure guided lane detection. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, pages 997–1003, 2021.
- 98 Sergiu Nedevschi, Rolf Schmidt, Thorsten Graf, et al. 3D lane detection system based on stereovision. In *Proceedings. The 7th International IEEE Conference on Intelligent Transportation Systems*, pages 161–166, 2004.
- 99 Nabil Benmansour, Raphaël Labayrade, Didier Aubert, et al. Stereovision-based 3D lane detection system: a model driven approach. In *2008 11th International IEEE Conference on Intelligent Transportation Systems*, pages 182–188, 2008.
- 100 Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. In *Advances in neural information processing systems*, pages 2017–2025, 2015.
- 101 Ruijin Liu, Dapeng Chen, Tie Liu, et al. Learning to Predict 3D Lane Shape and Camera Pose from a Single Image via Geometry Constraints. *arXiv preprint arXiv:2112.15351*, 2021.
- 102 Min Bai, Gellert Mattyus, Namdar Homayounfar, et al. Deep multi-sensor lane detection. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3102–3109, 2018.
- 103 Noa Garnett, Rafi Cohen, Tomer Pe'er, et al. 3D-LaneNet: End-to-End 3D Multiple Lane Detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2921–2930, 2019.
- 104 Yujie Jin, Xiangxuan Ren, Fengxiang Chen, et al. Robust Monocular 3D Lane Detection With Dual Attention. In *2021 IEEE International Conference on Image Processing*, pages 3348–3352, 2021.
- 105 Bencheng Liao, Shaoyu Chen, Xinggang Wang, et al. MapTR: Structured modeling and learning for online vectorized HD map construction. In *The Eleventh International Conference on Learning Representations*, 2022.
- 106 Bencheng Liao, Shaoyu Chen, Yunchi Zhang, et al. MapTRv2: An end-to-end framework for online vectorized hd map construction. *arXiv preprint arXiv:2308.05736*, 2023.
- 107 Limeng Qiao, Wenjie Ding, Xi Qiu, et al. End-to-End Vectorized HD-Map Construction With Piecewise Bezier Curve. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13218–13228, 2023.
- 108 Wenjie Ding, Limeng Qiao, Xi Qiu, et al. PivotNet: Vectorized Pivot Learning for End-to-end HD Map Construction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3672–3682, 2023.
- 109 Gongjie Zhang, Jiahao Lin, Shuang Wu, et al. Online Map Vectorization for Autonomous Driving: A Rasterization Perspective. *arXiv preprint arXiv:2306.10502*, 2023.
- 110 Siyu Li, Kailun Yang, Hao Shi, et al. Bi-Mapper: Holistic BEV Semantic Mapping for Autonomous Driving. *arXiv preprint arXiv:2305.04205*, 2023.
- 111 Yigit Baran Can, Alexander Liniger, Danda Pani Paudel, et al. Structured bird's-eye-view traffic scene understanding from onboard images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15661–15670, 2021.
- 112 Zhenhua Xu, Yuxuan Liu, Yuxiang Sun, et al. CenterLineDet: Road Lane CenterLine Graph Detection With Vehicle-Mounted Sensors by Transformer for High-definition Map Creation. In *2023 IEEE International Conference on Robotics and Automation*, pages 3553–3559, 2023.
- 113 Tianyu Li, Li Chen, Xiangwei Geng, et al. Topology Reasoning for Driving Scenes. *arXiv preprint arXiv:2304.05277*, 2023.
- 114 Juyeb Shin, Francois Rameau, Hyeonjun Jeong, et al. InstaGraM: Instance-level Graph Modeling for Vectorized HD Map Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.
- 115 Bencheng Liao, Shaoyu Chen, Bo Jiang, et al. Lane graph as path: Continuity-preserving path-wise modeling for online

- lane graph construction. *arXiv preprint arXiv:2303.08815*, 2023.
- 116 Raoul De Charette and Fawzi Nashashibi. Real time visual traffic lights recognition based on spot light detection and adaptive traffic lights templates. In *2009 IEEE Intelligent Vehicles Symposium*, pages 358–363, 2009.
- 117 Rachid Belaroussi, Philippe Foucher, Jean-Philippe Tarel, et al. Road sign detection in images: A case study. In *2010 20th International Conference on Pattern Recognition*, pages 484–488, 2010.
- 118 Johannes Stallkamp, Marc Schlipsing, Jan Salmen, et al. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. *Neural networks*, 32:323–332, 2012.
- 119 Andreas Mogelmose, Mohan Manubhai Trivedi, and Thomas B Moeslund. Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey. *IEEE Transactions on Intelligent Transportation Systems*, 13(4):1484–1497, 2012.
- 120 Sebastian Houben, Johannes Stallkamp, Jan Salmen, et al. Detection of traffic signs in real-world images: The German Traffic Sign Detection Benchmark. In *The 2013 international joint conference on neural networks*, pages 1–8, 2013.
- 121 Radu Timofte, Karel Zimmermann, and Luc Van Gool. Multi-view traffic sign detection, recognition, and 3D localisation. *Machine vision and applications*, 25:633–647, 2014.
- 122 Vladislav Igorevich Shakhuro and AS Konouchine. Russian traffic sign images dataset. *Computer optics*, 40(2):294–300, 2016.
- 123 Zhe Zhu, Dun Liang, Songhai Zhang, et al. Traffic-sign detection and classification in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2110–2118, 2016.
- 124 Karsten Behrendt, Libor Novak, and Rami Botros. A deep learning approach to traffic lights: Detection, tracking, and classification. In *2017 IEEE International Conference on Robotics and Automation*, pages 1370–1377, 2017.
- 125 Andreas Fregin, Julian Muller, Ulrich Krebel, et al. The DriveU traffic light dataset: Introduction and comparison with existing datasets. In *2018 IEEE international conference on robotics and automation*, pages 3376–3383, 2018.
- 126 Christian Ertler, Jerneja Mislej, Tobias Ollmann, et al. The mapillary traffic sign dataset for detection and classification on a global scale. In *European Conference on Computer Vision*, pages 68–84, 2020.
- 127 Haipeng Cui, Qiang Meng, Teck-Hou Teng, et al. Spatiotemporal correlation modelling for machine learning-based traffic state predictions: state-of-the-art and beyond. *Transport Reviews*, 43(4):780–804, 2023.
- 128 Zhiyuan Liu, Cheng Lyu, Zelin Wang, et al. A gaussian-process-based data-driven traffic flow model and its application in road capacity analysis. *IEEE Transactions on Intelligent Transportation Systems*, 24(2):1544–1563, 2023.
- 129 Shukai Chen, Hua Wang, and Qiang Meng. An optimal dynamic lane reversal and traffic control strategy for autonomous vehicles. *IEEE transactions on intelligent transportation systems*, 23(4):3804–3815, 2021.
- 130 Lin Zhang, Jiawei Yuan, Kang Yuan, et al. Automated braking decision and control for pedestrian collision avoidance based on risk assessment. *IEEE Intelligent Transportation Systems Magazine*, 14(3):41–61, 2022.
- 131 Ge Guo and Shijie Zhao. 3d multi-object tracking with adaptive cubature kalman filter for autonomous driving. *IEEE Transactions on Intelligent Vehicles*, 8(1):512–519, 2023.
- 132 Ping Wu, Feng Gao, Xiaolin Tang, et al. An integrated decision and motion planning framework for automated driving on highway. *IEEE Transactions on Vehicular Technology*, pages 1–11, 2023.
- 133 Jiawei Wang, Yang Zheng, Keqiang Li, et al. Deep-lcc: Data-enabled predictive leading cruise control in mixed traffic flow. *IEEE Transactions on Control Systems Technology*, 2023.
- 134 Jinxin Liu, Yugong Luo, Zihua Zhong, et al. A probabilistic architecture of long-term vehicle trajectory prediction for autonomous driving. *Engineering*, 19:228–239, 2022.
- 135 Jian Zhao, Dongjian Song, Bing Zhu, et al. A human-like trajectory planning method on a curve based on the driver preview mechanism. *IEEE Transactions on Intelligent Transportation Systems*, 24(11):11682–11698, 2023.
- 136 Zikang Zhou, Zihao Wen, Jianping Wang, et al. QCNeXt: A Next-Generation Framework For Joint Multi-Agent Trajectory Prediction. *arXiv preprint arXiv:2306.10508*, 2023.
- 137 Chen Feng, Hangning Zhou, Huadong Lin, et al. MacFormer: Map-Agent Coupled Transformer for Real-Time and Robust Trajectory Prediction. *IEEE Robotics and Automation Letters*, 8(10):6795–6802, 2023.
- 138 Luke Rowe, Martin Ethier, Eli-Henry Dykhne, et al. FJMP: Factorized Joint Multi-Agent Motion Prediction over Learned Directed Acyclic Interaction Graphs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13745–13755, 2023.
- 139 Fei Wu, Luoyu Chen, and Hao Lu. ParallelNet: Multi-mode trajectory prediction by multi-mode trajectory fusion. *arXiv preprint arXiv:2212.10203*, 2022.
- 140 Yunpeng Zhang, Zheng Zhu, Wenzhao Zheng, Junjie Huang, Guan Huang, Jie Zhou, and Jiwen Lu. BEVerse: Unified

- Perception and Prediction in Birds-Eye-View for Vision-Centric Autonomous Driving. *arXiv preprint arXiv:2205.09743*, 2022.
- 141 Daehee Park, Hobin Ryu, Yunseo Yang, et al. Leveraging Future Relationship Reasoning for Vehicle Trajectory Prediction. In *The Eleventh International Conference on Learning Representations*, 2022.
- 142 Shaoshuai Shi, Li Jiang, Dengxin Dai, et al. MTR++: Multi-agent motion prediction with symmetric scene modeling and guided intention querying. *arXiv preprint arXiv:2306.17770*, 2023.
- 143 Qiao Sun, Xin Huang, Junru Gu, et al. M2I: From factored marginal trajectory prediction to interactive prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6543–6552, 2022.
- 144 Xiaosong Jia, Li Chen, Penghao Wu, et al. Towards capturing the temporal dynamics for trajectory prediction: a coarse-to-fine approach. In *Conference on Robot Learning*, pages 910–920, 2023.
- 145 Wei Zhan, Liting Sun, Di Wang, et al. Interaction dataset: An international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps. *arXiv preprint arXiv:1910.03088*, 2019.
- 146 Xiaosong Jia, Penghao Wu, Li Chen, et al. Hdgt: Heterogeneous driving graph transformer for multi-agent trajectory prediction via scene encoding. *IEEE transactions on pattern analysis and machine intelligence*, 45(11):13860–13875, 2023.
- 147 Chalavadi Vishnu, Vineel Abhinav, Debaditya Roy, et al. Improving Multi-Agent Trajectory Prediction Using Traffic States on Interactive Driving Scenarios. *IEEE Robotics and Automation Letters*, 8(5):2708–2715, 2023.
- 148 Xiaosong Jia, Liting Sun, Hang Zhao, et al. Multi-agent trajectory prediction by combining egocentric and allocentric views. In *Conference on Robot Learning*, pages 1434–1443, 2022.
- 149 Luis Gressenbuch, Klemens Esterle, Tobias Kessler, et al. Mona: The munich motion dataset of natural driving. In *2022 IEEE 25th International Conference on Intelligent Transportation Systems*, pages 2093–2100, 2022.
- 150 Daniel Dauner, Marcel Hallgarten, Andreas Geiger, et al. Parting with Misconceptions about Learning-based Vehicle Motion Planning. *arXiv preprint arXiv:2306.07962*, 2023.
- 151 Tung Phan-Minh, Forbes Howington, Ting-Sheng Chu, et al. DriveIRL: Drive in real life with inverse reinforcement learning. In *2023 IEEE International Conference on Robotics and Automation*, pages 1544–1550, 2023.
- 152 Weitao Xi, Liangchao Shi, and Guangzhi Cao. An Imitation Learning Method with Data Augmentation and Post Processing for Planning in Autonomous Driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023.
- 153 Zhejun Zhang, Alexander Liniger, Dengxin Dai, et al. End-to-end urban driving by imitating a reinforcement learning coach. In *Proceedings of the IEEE international conference on computer vision*, pages 15222–15232, 2021.
- 154 Kashyap Chitta, Aditya Prakash, and Andreas Geiger. NEAT: Neural attention fields for end-to-end autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15793–15803, 2021.
- 155 Hao Shao, Letian Wang, Ruobing Chen, et al. Reasonnet: End-to-end driving with temporal and global reasoning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13723–13733, 2023.
- 156 Quanyi Li, Zhenghao Peng, Lan Feng, et al. Metadrive: Composing diverse driving scenarios for generalizable reinforcement learning. *IEEE transactions on pattern analysis and machine intelligence*, 45(3):3461–3475, 2022.
- 157 Letian Wang, Jie Liu, Hao Shao, Wenshuo Wang, Ruobing Chen, Yu Liu, and Steven L Waslander. Efficient Reinforcement Learning for Autonomous Driving with Parameterized Skills and Priors. *arXiv preprint arXiv:2305.04412*, 2023.
- 158 Xuesong Wang, Jiazhi Zhang, Diyuan Hou, et al. Autonomous Driving Based on Approximate Safe Action. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–9, 2023.
- 159 Xu-Hui Liu, Feng Xu, Xinyu Zhang, et al. How To Guide Your Learner: Imitation Learning with Active Adaptive Expert Involvement. *arXiv preprint arXiv:2303.02073*, 2023.
- 160 Xinyu Huang, Xinjing Cheng, Qichuan Geng, et al. The ApolloScape dataset for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition workshops*, pages 954–960, 2018.
- 161 Wei Li, CW Pan, Rong Zhang, et al. AADS: Augmented autonomous driving simulation using data-driven algorithms. *Science robotics*, 4(28):e863–e863, 2019.
- 162 Andrea Piazzoni, Jim Cherian, Martin Slavik, et al. Modeling perception errors towards robust decision making in autonomous vehicles. *arXiv preprint arXiv:2001.11695*, 2020.
- 163 Yuanwei Bi, Chuanbiao Li, Xiangrong Tong, et al. An application of stereo matching algorithm based on transfer learning on robots in multiple scenarios. *Scientific Reports*, 13(1):12739, 2023.
- 164 Mordechai Haklay and Patrick Weber. Openstreetmap: User-generated street maps. *IEEE Pervasive computing*, 7(4):12–18, 2008.
- 165 Matthias Hentschel and Bernardo Wagner. Autonomous robot navigation based on openstreetmap geodata. In *13th*

- 166 Dennis Luxen and Christian Vetter. Real-time routing with OpenStreetMap data. In *Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems*, pages 513–516, 2011.
- 167 Allan M De Souza, Roberto S Yokoyama, Guilherme Maia, et al. Real-time path planning to prevent traffic jam through an intelligent transportation system. In *2016 IEEE symposium on computers and communication*, pages 726–731, 2016.
- 168 Transportation Networks for Research Core Team. Transportation Networks for Research. <https://github.com/bstablert/TransportationNetworks>, 2023.
- 169 Hongliang Guo, Xuejie Hou, Zhiguang Cao, et al. GP3: Gaussian process path planning for reliable shortest path in transportation networks. *IEEE Transactions on Intelligent Transportation Systems*, 23(8):11575–11590, 2021.
- 170 Jingxing Wang, Shu Lu, Hongsheng Liu, et al. Transportation origin-destination demand estimation with quasi-sparsity. *Transportation science*, 57(2):289–312, 2023.
- 171 Priyadarshan N Patil, Katherine C Ross, and Stephen D Boyles. Convergence behavior for traffic assignment characterization metrics. *Transportmetrica A: Transport Science*, 17(4):1244–1271, 2021.
- 172 Xuesong Zhou and Jeffrey Taylor. DTALite: A queue-based mesoscopic traffic simulator for fast model evaluation and calibration. *Cogent Engineering*, 1(1):961345, 2014.
- 173 Daocheng Fu, Pinlong Cai, Yilun Lin, et al. Incremental path planning: Reservation system in V2X environment. *Physica A: Statistical Mechanics and its Applications*, 623:128914, 2023.
- 174 Chenfeng Xiong, Xuesong Zhou, and Lei Zhang. AgBM-DTALite: An integrated modelling system of agent-based travel behaviour and transportation network dynamics. *Travel Behaviour and Society*, 12:141–150, 2018.
- 175 Lu Tong, Yuyan Pan, Pan Shang, et al. Open-source public transportation mobility simulation engine dtalite-s: A discretized space-time network-based modeling framework for bridging multi-agent simulation and optimization. *Urban Rail Transit*, 5:1–16, 2019.
- 176 Caltrans: California Department of Transportation. Caltrans Performance Measurement System (PeMS). <https://pems.dot.ca.gov>, 2017.
- 177 Tengchan Zeng, Jianlin Guo, Kyeong Jin Kim, et al. Multi-task federated learning for traffic prediction and its application to route planning. In *2021 IEEE Intelligent Vehicles Symposium*, pages 451–457, 2021.
- 178 Sakib Mahmud Khan, Kakan C Dey, and Mashrur Chowdhury. Real-time traffic state estimation with connected vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 18(7):1687–1699, 2017.
- 179 Zhensong Wei, Peng Hao, Matthew Barth, et al. Evaluating contraflow high-occupancy vehicle lane designs for mitigating high-occupancy vehicle lane performance degradation. *Transportation research record*, 2677(5):707–719, 2023.
- 180 Brian Donovan and DB Work. New York City Taxi Data (2010-2013). <http://dx.doi.org/10.13012/J8PN93H8>, 2014.
- 181 Nivan Ferreira, Jorge Poco, Huy T Vo, et al. Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips. *IEEE transactions on visualization and computer graphics*, 19(12):2149–2158, 2013.
- 182 Ming Zhu, Xiao-Yang Liu, and Xiaodong Wang. An online ride-sharing path-planning strategy for public vehicle systems. *IEEE Transactions on Intelligent Transportation Systems*, 20(2):616–627, 2018.
- 183 Duy Hoang Tran, Pieter Leyman, and Patrick De Causmaecker. Adaptive passenger-finding recommendation system for taxi drivers with load balancing problem. *Computers & Industrial Engineering*, 169:108187, 2022.
- 184 Hong Chen, Kang Yuan, Yanjun Huang, et al. Feedback is all you need: from chatgpt to autonomous driving. *Science China Information Sciences*, 66(6):1–3, 2023.
- 185 Yuning Wang, Junkai Jiang, Shangyi Li, et al. Decision-making driven by driver intelligence and environment reasoning for high-level autonomous vehicles: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 24(10):10362–10381, 2023.
- 186 Fang Zhang, Jian Lu, Xiaojian Hu, et al. A stochastic dynamic network loading model for mixed traffic with autonomous and human-driven vehicles. *Transportation Research Part B: Methodological*, 178:102850, 2023.
- 187 Xiangdong Chen, Xi Lin, Qiang Meng, et al. Coordinated traffic control of urban networks with dynamic entrance holding for mixed cav traffic. *Transportation research part E: logistics and transportation review*, 178:103264, 2023.
- 188 Ajay Mandlekar, Yuke Zhu, Animesh Garg, et al. Roboturk: A crowdsourcing platform for robotic skill learning through imitation. In *Conference on Robot Learning*, pages 879–893, 2018.
- 189 Xi Chen, Josip Djolonga, Piotr Padlewski, Basil Mustafa, Soravit Changpinyo, Jialin Wu, Carlos Riquelme Ruiz, Sebastian Goodman, Xiao Wang, Yi Tay, et al. PaLI-X: On Scaling up a Multilingual Vision and Language Model. *arXiv preprint arXiv:2305.18565*, 2023.
- 190 Sudeep Dasari, Frederik Ebert, Stephen Tian, et al. Robonet: Large-scale multi-robot learning. *arXiv preprint*

- arXiv:1910.11215*, 2019.
- 191 Danny Driess, Fei Xia, Mehdi S. M. Sajjadi, et al. PaLM-E: An Embodied Multimodal Language Model. *arXiv preprint arXiv:2303.03378*, 2023.
- 192 Sulabh Kumra and Christopher Kanan. Robotic grasp detection using deep convolutional neural networks. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 769–776, 2017.
- 193 Minting Pan, Xiangming Zhu, Yunbo Wang, et al. Iso-dream: Isolating and leveraging noncontrollable visual dynamics in world models. In *Advances in neural information processing systems*, volume 35, pages 23178–23191, 2022.
- 194 Haibao Yu, Yizhen Luo, Mao Shu, et al. Dair-V2X: A large-scale dataset for vehicle-infrastructure cooperative 3D object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21361–21370, 2022.
- 195 Yiming Li, Dekun Ma, Ziyuan An, et al. V2X-Sim: Multi-agent collaborative perception dataset and benchmark for autonomous driving. *IEEE Robotics and Automation Letters*, 7(4):10914–10921, 2022.
- 196 Runsheng Xu, Hao Xiang, Xin Xia, et al. Ovp2v: An open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication. In *2022 International Conference on Robotics and Automation*, pages 2583–2589, 2022.
- 197 Haolin Zhang, Mohammad Shareef Mekala, Dongfang Yang, et al. 3D Harmonic Loss: Towards Task-consistent and Time-friendly 3D Object Detection on Edge for V2X Orchestration. *IEEE transactions on vehicular technology*, pages 1–12, 2023.
- 198 Shunli Ren, Zixing Lei, Zi Wang, et al. Interruption-Aware Cooperative Perception for V2X Communication-Aided Autonomous Driving. *arXiv preprint arXiv:2304.11821*, 2023.
- 199 Yue Hu, Shaoheng Fang, Zixing Lei, et al. Where2comm: Communication-efficient collaborative perception via spatial confidence maps. In *Advances in neural information processing systems*, volume 35, pages 4874–4886, 2022.
- 200 Matthias Mueller, Neil Smith, and Bernard Ghanem. A benchmark and simulator for UAV tracking. In *European Conference on Computer Vision*, pages 445–461, 2016.
- 201 Amado Antonini, Winter Guerra, Varun Murali, et al. The blackbird UAV dataset. *The International Journal of Robotics Research*, 39(10-11):1346–1364, 2020.
- 202 Ye Lyu, George Vosselman, Gui-Song Xia, et al. UAVid: A semantic segmentation dataset for UAV imagery. *ISPRS journal of photogrammetry and remote sensing*, 165:108–119, 2020.
- 203 Zhiyong An, Xumin Wang, Bo Li, et al. Learning Spatial Regularization Correlation Filters with the Hilbert-Schmidt Independence Criterion in RKHS for UAV Tracking. *IEEE Transactions on Instrumentation and Measurement*, 72:1–12, 2023.
- 204 Giovanni Cioffi, Leonard Bauersfeld, and Davide Scaramuzza. HDVIO: Improving Localization and Disturbance Estimation with Hybrid Dynamics VIO. *arXiv preprint arXiv:2306.11429*, 2023.
- 205 Xinyu Li, Yu Cheng, Yi Fang, et al. 2DSegFormer: 2-D Transformer Model for Semantic Segmentation on Aerial Images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2022.
- 206 Borja Bovcon, Jon Muhovič, Duško Vranac, et al. MODS—a USV-oriented object detection and obstacle segmentation benchmark. *IEEE Transactions on Intelligent Transportation Systems*, 23(8):13403–13418, 2021.
- 207 Borja Bovcon, Jon Muhovič, Janez Perš, et al. The MaSTr1325 dataset for training deep USV obstacle detection models. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3431–3438, 2019.
- 208 Yuwei Cheng, Mengxin Jiang, Jiannan Zhu, et al. Are we ready for unmanned surface vehicles in inland waterways? the USVInland multisensor dataset and benchmark. *IEEE Robotics and Automation Letters*, 6(2):3964–3970, 2021.
- 209 Linghong Yao, Dimitrios Kanoulas, Ze Ji, et al. ShorelineNet: An efficient deep learning approach for shoreline semantic segmentation for unmanned surface vehicles. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5403–5409, 2021.
- 210 Yuheng Yin, Yangang Guo, Liwei Deng, et al. Improved PSPNet-based water shoreline detection in complex inland river scenarios. *Complex & Intelligent Systems*, 9(1):233–245, 2023.
- 211 NEO Research Group. NEO benchmark datasets. <https://neo.lcc.uma.es/vrp/vehicle-routingproblem>, 2013.
- 212 Divvy Bike. Chicago Divvy bicycle sharing data. <https://www.divvybikes.com/system-data>, 2015.
- 213 Nora Niazy, Ahmed El-Sawy, and Mahmoud Gadallah. Solving Capacitated Vehicle Routing Problem Using Chicken Swarm Optimization with Genetic Algorithm. *International Journal of Intelligent Engineering Systems*, 13(5), 2020.
- 214 Michael Kirchhof, Lena Schmid, Christopher Reining, et al. Chances of interpretable transfer learning for human activity recognition in warehousing. In *Computational Logistics: 12th International Conference*, pages 163–177, 2021.
- 215 Hongtai Yang, Yibei Zhang, Lizhi Zhong, et al. Exploring spatial variation of bike sharing trip production and attraction:

- A study based on Chicago's Divvy system. *Applied geography*, 115:102130, 2020.
- 216 Ruchi Goel and Pooja Gupta. Robotics and industry 4.0. *A Roadmap to Industry 4.0: Smart Production, Sharp Business and Sustainable Development*, pages 157–169, 2020.
- 217 Pan Zhao, Jiajia Chen, Yan Song, et al. Design of a control system for an autonomous vehicle based on adaptive-pid. *International Journal of Advanced Robotic Systems*, 9(2):44, 2012.
- 218 Yun Jiang, Stephen Mosenon, and Ashutosh Saxena. Efficient grasping from rgbd images: Learning using a new rectangle representation. In *2011 IEEE International conference on robotics and automation*, pages 3304–3311, 2011.
- 219 John B Kenney. Dedicated short-range communications (DSRC) standards in the United States. *Proceedings of the IEEE*, 99(7):1162–1182, 2011.
- 220 Andreas Festag. Cooperative intelligent transport systems standards in Europe. *IEEE communications magazine*, 52(12):166–172, 2014.
- 221 Mario H Castañeda Garcia, Alcjandro Molina-Galan, Mate Boban, et al. A tutorial on 5G NR V2X communications. *IEEE Communications Surveys & Tutorials*, 23(3):1972–2026, 2021.
- 222 Shixi Wen and Ge Guo. Control of connected vehicles in road network via traffic flow information feedback. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–14, 2023.
- 223 Kangning Hou, Fangfang Zheng, Xiaobo Liu, et al. Cooperative on-ramp merging control model for mixed traffic on multi-lane freeways. *IEEE Transactions on Intelligent Transportation Systems*, 24(10):10774–10790, 2023.
- 224 Ge Guo, Ziwei Zhao, and Renyongkang Zhang. Distributed trajectory optimization and fixed-time tracking control of a group of connected vehicles. *IEEE Transactions on Vehicular Technology*, 72(2):1478–1487, 2023.
- 225 Darong Huang, Jinfu Zhou, Bo Mi, et al. Key-based data deduplication via homomorphic ntru for internet of vehicles. *IEEE Transactions on Vehicular Technology*, 72(1):239–252, 2022.
- 226 Darong Huang, Shaoqian Li, Zhenyuan Zhang, et al. Design and analysis of longitudinal controller for the platoon with time-varying delay. *IEEE Transactions on Intelligent Transportation Systems*, 23(12):23628–23639, 2022.
- 227 Zhenping Deng, Darong Huang, Jinyu Liu, et al. An assessment method for traffic state vulnerability based on a cloud model for urban road network traffic systems. *IEEE Transactions on Intelligent Transportation Systems*, 22(11):7155–7168, 2020.
- 228 Qing Xu, Keqiang Li, Jianqiang Wang, et al. The status, challenges, and trends: an interpretation of technology roadmap of intelligent and connected vehicles in china (2020). *Journal of Intelligent and Connected Vehicles*, 5(3):1–7, 2022.
- 229 Rohit Chaurasia and Vandana Mohindru. Unmanned aerial vehicle (uav): A comprehensive survey. *Unmanned Aerial Vehicles for Internet of Things (IoT) Concepts, Techniques, and Applications*, pages 1–27, 2021.
- 230 Xin Fang, Jing Zhu, Darong Huang, et al. E 2 dtf: An end-to-end detection and tracking framework for multiple micro-uavs with fmcw-mimo radar. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–16, 2023.
- 231 Darong Huang, Zhenyuan Zhang, Xin Fang, et al. Stif: A spatial-temporal integrated framework for end-to-end micro-uav trajectory tracking and prediction with 4d mimo radar. *IEEE Internet of Things Journal*, 2023.
- 232 Shannon Hood, Kelly Benson, Patrick Hamod, et al. Bird's eye view: Cooperative exploration by UGV and UAV. In *2017 International Conference on Unmanned Aircraft Systems*, pages 247–255, 2017.
- 233 Ru-jian Yan, Shuo Pang, Han-bing Sun, et al. Development and missions of unmanned surface vehicle. *Journal of Marine Science and Application*, 9:451–457, 2010.
- 234 Edoardo I Sarda, Huajin Qu, Ivan R Bertaska, et al. Station-keeping control of an unmanned surface vehicle exposed to current and wind disturbances. *Ocean Engineering*, 127:305–324, 2016.
- 235 Guorun Yang, Yu Qiao, Jianping Shi, and Zhe Wang. Long-tailed object mining based on clip model for autonomous driving. In *2022 4th International Conference on Control and Robotics (ICCR)*, pages 348–352, 2022.
- 236 Weitao Zhou, Zhong Cao, Yunkang Xu, et al. Long-tail prediction uncertainty aware trajectory planning for self-driving vehicles. In *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, pages 1275–1282, IEEE, 2022.
- 237 Weitao Zhou, Zhong Cao, Nanshan Deng, Xiaoyu Liu, Kun Jiang, and Diange Yang. Dynamically conservative self-driving planner for long-tail cases. *IEEE Transactions on Intelligent Transportation Systems*, 24(3):3476–3488, 2022.
- 238 Amazon. Autonomous driving data framework. <https://docs.aws.amazon.com/prescriptive-guidance/latest/addf-security-and-operations/welcome.html>, 2022.
- 239 Tesla. Full self-driving. <https://www.tesla.com/support/autopilot>.
- 240 Nividia. How maglev speeds autonomous vehicles to superhuman levels of safety. <https://blogs.nvidia.com/blog/2018/09/13/how-maglev-speeds-autonomous-vehicles-to-superhuman-levels-of-safety>, 2018.

- 241 阿里巴巴. 自动驾驶云开发平台. <https://www.aliyun.com/product/iovcc>, 2022.
- 242 Jonah Philion. Fastdraw: Addressing the long tail of lane detection by adapting a sequential prediction network. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11582–11591, 2019.
- 243 Haoyi Niu, Yizhou Xu, Xingjian Jiang, and HU Jianming. Continual driving policy optimization with closed-loop individualized curricula. In *Second Agent Learning in Open-Endedness Workshop*, 2023.
- 244 comma. https://twitter.com/comma_ai/status/1681491118536691712, 2023.
- 245 Alexander Swerdlow, Runsheng Xu, and Bolei Zhou. Street-View Image Generation from a Bird’s-Eye View Layout. *arXiv preprint arXiv:2301.04634*, 2023.
- 246 Ze Yang, Yun Chen, Jingkang Wang, et al. UniSim: A Neural Closed-Loop Sensor Simulator. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1389–1399, 2023.
- 247 Wenwen Tong, Jiangwei Xie, Tianyu Li, et al. 3D Data Augmentation for Driving Scenes on Camera. *arXiv preprint arXiv:2303.10340*, 2023.
- 248 Tesla. https://www.youtube.com/watch?v=6x-Xb_uT7ts, 2023.
- 249 Wayve. <https://wayve.ai/thinking/introducing-gaia1>, 2023.
- 250 Andreas Blattmann, Robin Rombach, Huan Ling, et al. Align your latents: High-resolution video synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22563–22575, 2023.
- 251 Kai Chen, Enze Xie, Zhe Chen, et al. Integrating Geometric Control into Text-to-Image Diffusion Models for High-Quality Detection Data Generation via Text Prompt. *arXiv preprint arXiv:2306.04607*, 2023.
- 252 Panpan Cai, Yiyuan Lee, Yuanfu Luo, et al. SUMMIT: A Simulator for Urban Driving in Massive Mixed Traffic. In *2020 IEEE International Conference on Robotics and Automation*, pages 4023–4029, 2020.
- 253 Guodong Rong, Byung Hyun Shin, Hadi Tabatabaei, et al. LGSVL Simulator: A high fidelity simulator for autonomous driving. In *2020 IEEE 23rd International conference on intelligent transportation systems*, pages 1–6, 2020.
- 254 Quanyi Li, Zhenghao Peng, Qihang Zhang, et al. Improving the Generalization of End-to-End Driving through Procedural Generation. *arXiv preprint arXiv:2012.13681*, 2020.
- 255 Shital Shah, Debadeepta Dey, Chris Lovett, et al. AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles. In *Field and Service Robotics: Results of the 11th International Conference*, pages 621–635, 2018.
- 256 Amazon. Amazon sagemaker data labeling. <https://aws.amazon.com/sagemaker/data-labeling>.
- 257 百度. 百度智能云. <https://cloud.baidu.com>.
- 258 Microsoft. Visual object tagging tool: An electron app for building end to end object detection models from images and videos. <https://github.com/microsoft/VoTT>, 2017.
- 259 数据堂. 数据堂, 专业的人工智能数据服务提供商. <https://www.datatang.com>, 2011.
- 260 Segments.ai. Multi-sensor Labeling Platform. <https://segments.ai/>.
- 261 Scale AI. Scale data engine. <https://scale.com/data-engine>.
- 262 OpenAI. GPT-4 Technical Report. *arXiv.org*, 2303.08774, 2023.
- 263 Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, 2023.
- 264 Mostafa Dehghani, Josip Djolonga, Basil Mustafa, et al. Scaling vision transformers to 22 billion parameters. In *International Conference on Machine Learning*, pages 7480–7512, 2023.
- 265 Junnan Li, Dongxu Li, Silvio Savarese, et al. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. *arXiv preprint arXiv:2301.12597*, 2023.
- 266 Danny Driess, Fei Xia, Mehdi SM Sajjadi, et al. Palm-e: An embodied multimodal language model. *arXiv preprint arXiv:2303.03378*, 2023.
- 267 Anthony Brohan, Noah Brown, Justice Carbajal, et al. Rt-2: Vision-language-action models transfer web knowledge to robotic control. *arXiv preprint arXiv:2307.15818*, 2023.
- 268 Karan Singhal, Tao Tu, Juraj Gottweis, et al. Towards expert-level medical question answering with large language models. *arXiv preprint arXiv:2305.09617*, 2023.
- 269 Tesla. Building foundation models for autonomy. https://www.youtube.com/watch?v=6x-Xb_uT7ts, 2023.
- 270 OpenDriveLab. DriveAGI: Embracing foundation models into autonomous agent and system. <https://github.com/OpenDriveLab/DriveAGI>, 2023.
- 271 Xinpeng Ding, Jinahua Han, Hang Xu, et al. Holistic Autonomous Driving Understanding by Bird’s-Eye-View Injected

- Multi-Modal Large Models. *arXiv preprint arXiv:2401.00988*, 2024.
- 272 Anthony Hu, Lloyd Russell, Hudson Yeo, et al. GAIA-1: A Generative World Model for Autonomous Driving. *arXiv preprint arXiv:2309.17080*, 2023.
- 273 Chenbin Pan, Burhaneddin Yaman, Tommaso Nesi, et al. VLP: Vision Language Planning for Autonomous Driving. *arXiv preprint arXiv:2401.05577*, 2024.
- 274 Priya Goyal, Dhruv Mahajan, Abhinav Gupta, et al. Scaling and benchmarking self-supervised visual representation learning. In *Proceedings of the ieee/cvf International Conference on computer vision*, pages 6391–6400, 2019.
- 275 Tom Brown, Benjamin Mann, Nick Ryder, et al. Language models are few-shot learners. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, volume 33, pages 1877–1901, 2020.
- 276 Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- 277 Mannat Singh, Quentin Duval, Kalyan Vasudev Alwala, et al. The effectiveness of mae pre-training for billion-scale pretraining. *arXiv preprint arXiv:2303.13496*, 2023.
- 278 Hongqing Chu, Hejian Zhuang, Wenshuo Wang, et al. A review of driving style recognition methods from short-term and long-term perspectives. *IEEE Transactions on Intelligent Vehicles*, pages 1–15, 2023.
- 279 Dean A Pomerleau. Alvinn: An autonomous land vehicle in a neural network. In *Advances in neural information processing systems*, volume 1, 1988.
- 280 Yihan Hu, Jiazhi Yang, Li Chen, et al. Planning-oriented autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17853–17862, 2023.
- 281 Sergio Casas, Abbas Sadat, and Raquel Urtasun. MP3: A Unified Model To Map, Perceive, Predict and Plan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14403–14412, 2021.
- 282 Shengchao Hu, Li Chen, Penghao Wu, et al. ST-P3: End-to-end Vision-based Autonomous Driving via Spatial-Temporal Feature Learning. In *European Conference on Computer Vision*, pages 533–549, 2022.

Open-sourced Data Ecosystem in Autonomous Driving: the Present and Future

Hongyang Li^{1*}, Yang Li¹, Huijie Wang¹, Jia Zeng¹, Huilin Xu¹, Pinlong Cai¹, Li Chen¹, Junchi Yan², Feng Xu³, Lu Xiong⁴, Jingdong Wang^{5†}, Futang Zhu^{6†}, Kai Yan^{7†}, Chunjing Xu^{8†}, Tiancai Wang^{9†}, Fei Xia^{10†}, Beipeng Mu^{10†}, Zhihui Peng^{11†}, Dahua Lin¹ & Yu Qiao^{1*}

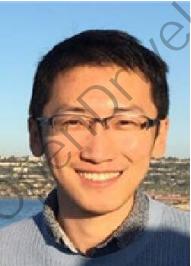
1. *Shanghai AI Lab, Shanghai 200232, China;*
2. *Shanghai Jiao Tong University, Shanghai 200240, China;*
3. *Fudan University, Shanghai 200433, China;*
4. *Tongji University, Shanghai 200092, China;*
5. *Baidu, Shanghai 100085, China;*
6. *BYD Auto, Shenzhen 518118, China;*
7. *Changan Automobile, Chongqing 400000, China;*
8. *Huawei, Shenzhen 518129, China;*
9. *MEGVII Technology, Beijing 100096, China;*
10. *Meituan, Guangzhou 510030, China;*
11. *AGIBOT, Shanghai 201315, China*

* Corresponding author(s). E-mail: hy@opendrivelab.com, qiaoyu@pjlab.org.cn

† Equal consulting contribution. These authors are listed in alphabetic order.

Abstract With the continuous maturation and application of autonomous driving technology, a systematic examination of open-source autonomous driving datasets becomes instrumental in fostering the robust evolution of the industry ecosystem. Current autonomous driving datasets can broadly be categorized into two generations. The first-generation autonomous driving dataset is characterized by relatively simpler sensor modalities, smaller dataset scale, and is limited to perception-level tasks. KITTI, introduced in 2012, serves as a prominent representative of this initial wave. In contrast, the second-generation datasets exhibit heightened complexity in sensor modalities, greater dataset scale and diversity, and an expansion of tasks from perception to encompass prediction and control. Leading examples of the second generation include nuScenes and Waymo, introduced around 2019. This comprehensive review, conducted in collaboration with esteemed colleagues from both academia and industry, systematically assesses over seventy open-source autonomous driving datasets from domestic and international sources. It offers insights into various aspects, such as the principles underlying the creation of high-quality datasets, the pivotal role of data within algorithmic closed-loop systems, and the utilization of generative foundation models to facilitate scalable data generation. Furthermore, this review undertakes an exhaustive analysis and discourse regarding the characteristics and data scales that future third-generation autonomous driving datasets should possess. It also delves into the scientific and technical challenges that warrant resolution. The synthesis and perspectives presented in this article provide valuable guidance for the development of a novel generation of autonomous driving datasets and ecosystems. These endeavors are pivotal in advancing autonomous innovation and fostering technological enhancement in critical domains. For further details, please refer to <https://github.com/OpenDriveLab/DriveAGI>.

Keywords Autonomous Driving, Data Pipelining, Foundation Model, Dataset and Benchmark



Hongyang Li is the Research Scientist at OpenDriveLab, Shanghai AI Lab. He obtained the Ph.D. from Chinese University of Hong Kong in 2019. His research focus is on autonomous driving and foundation model. As PI, he proposed the end-to-end planning-oriented framework for autonomous driving and won the CVPR 2023 Best Paper Award. He serves as Area Chair for CVPR 2023, 2024 and NeurIPS 2023. He is a Senior Member of IEEE.



Dahua Lin received the B.Eng. degree from the University of Science and Technology of China, Hefei, China, in 2004, the M. Phil. degree from The Chinese University of Hong Kong, Hong Kong, in 2006, and the Ph.D. degree from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2012. From 2012 to 2014, he was a Research Assistant Professor with Toyota Technological Institute at Chicago, Chicago, IL, USA. He is currently an Associate Professor with the Department of Information Engineering, The Chinese University of Hong Kong (CUHK), and the leading scientist at Shanghai AI Laboratory.



Jia Zeng received the B.E. from Central South University in 2017 and the Ph.D. degree from Robot Institute, Shanghai Jiao Tong University in 2023. He is currently a researcher at OpenDriveLab, Shanghai AI Laboratory. His research interests lie in computer vision, autonomous driving and embodied AI, with first-authored or co-authored papers published in CVPR, CoRL, IEEE JBHI, etc.



Yu Qiao is a professor with Shanghai AI Laboratory and the Shenzhen Institutes of Advanced Technology (SIAT), Chinese Academy of Science. His research interests include computer vision, deep learning, and bioinformation. He has published more than 300 papers in IEEE TPAMI, IJCV, IEEE TIP, CVPR, ICCV, etc. His H-index is 72, with 42,700 citations in Google scholar. He is a recipient of the distinguished paper award in AAAI 2021. His group achieved the first runner-up at the ImageNet Large Scale Visual Recognition Challenge 2015 in scene recognition, and the winner at the ActivityNet Large Scale Activity Recognition Challenge 2016 in video classification. He is a Distinguished Member of CCF.