



智能网联汽车感知训练数据集 标准化需求研究报告

全国汽车标准化技术委员会
智能网联汽车分技术委员会
自动驾驶标准工作组

2023年11月

前言

随着技术水平的不断提升和相关产业的全面融合，我国智能网联汽车产业进入新的发展阶段。高精度环境感知是智能网联汽车技术创新体系中的关键基础技术之一，具有中国特色感知数据训练集又是高精度环境感知技术的重要组成部分，其质量一定程度上决定着感知能力的水平。感知训练数据集是自动驾驶环境感知算法训练与实现的重要基础。为了实现高安全高可靠的自动驾驶，感知训练数据集要应对多重挑战：交通场景复杂性、天气多变性、光照挑战、传感器误差以及标注一致性等。标准化不仅是解决问题、提升感知算法鲁棒性的重要一环，更是引领自动驾驶技术蓬勃发展的关键步骤。

本研究报告集合行业主流技术供应商、车企和第三方检测机构，通过分析行业现状、技术发展、法律法规及国内外标准等对智能网联汽车感知训练数据集标准化需求及其可行性进行梳理与分析，为下一步国家标准制定提供参考。

在本研究报告编制过程中，各起草单位参阅了大量材料，并借鉴了行业的部分素材，鉴于篇幅有限，这里不一一列举，仅作诚挚的感谢！

在此，再次衷心感谢参与研究报告编写的各个单位和组织：中国汽车技术研究中心有限公司、上海交通大学、北京觉非科技有限公司、北京赛目科技有限公司、北京百度智行科技有限公司、宁波吉利汽车研究开发有限公司、一汽解放汽车有限公司、上海临港绝影智能科技有限公司、重庆长安汽车股份有限公司、中国软件评测中心（工业和信息化部

软件与集成电路促进中心)、北京智能车联产业创新中心有限公司、北京云测信息技术有限公司、东风汽车集团有限公司技术中心、长城汽车股份有限公司、泛亚汽车技术中心有限公司、上海机动车检测认证技术研究中心有限公司、小米汽车科技有限公司、招商局检测车辆技术研究院有限公司、福特汽车(中国)有限公司、上汽通用五菱汽车股份有限公司、高通无线通信技术(中国)技术有限公司、西华大学、武汉路特斯科技有限公司、江苏大学、天津大学、华为技术有限公司、东风悦享科技有限公司、江铃汽车股份有限公司、采埃孚商用系统有限公司、高新科技集团股份有限公司。

主要编写人：华一丁、袁悦、王亚飞、郝爽、白智敏、程周、金晨、吴爱文、宋方方、刘旭磊、于鹏、贾宇航、庾新娟、罗智、王荣、田生明、楼金彪、刘卫利、周德阳、曾杰、丁雪聪、杨洋、党利冈、柯任辰、蒋慧、陈艺、熊云杰、孙兆瑜、王静、张奥博、廖尉华、王海、李祎承、李绍鹏、石俊杰、马佳彬、戴娇、周文峰、姜腾龙、邓宗乾、李若尧、郭昊，孙卫萍、骆嫚、王科未、张明明、张进、李默涵、王百泉，周建武。

目录

1 感知训练数据集相关背景研究.....	6
1.1 研究背景与现状.....	6
1.1.1 车端感知训练数据集现状.....	6
1.1.2 路端感知训练数据集现状.....	11
1.1.3 数据集标注能力现状.....	15
1.2 相关概念.....	19
1.2.1 术语定义.....	19
1.2.2 缩略语.....	24
1.3 研究目的及意义.....	26
2 感知数据采集及传感器部署方案研究.....	28
2.1 车端感知数据采集.....	28
2.1.1 采集车的种类及型号.....	28
2.1.2 传感器的选型.....	30
2.1.3 传感器配置数量及部署安装方案.....	35
2.1.4 采集系统配置.....	39
2.1.5 其他硬件系统.....	40
2.2 路端感知数据采集.....	41
2.2.1 传感器选型.....	41
2.2.2 传感器配置数量及部署安装方案.....	42
2.2.3 采集软件配置.....	47
2.2.4 其他硬件系统.....	49
3 感知训练数据集应用场景和规模.....	50
3.1 国内外主流的感知算法模型及神经网络分类.....	50
3.1.1 图像感知算法.....	50
3.1.2 激光点云算法.....	70
3.1.3 融合算法.....	82
3.2 模型任务所应用的场景维度.....	87
3.2.1 道路场景类型.....	87
3.2.2 关键道路特征.....	89
3.2.3 路面情况.....	91
3.2.4 天气与光照.....	98
3.2.5 交通参与者与交通标志.....	99
3.2.6 其他.....	104
3.3 感知训练数据集子集所需量级建议.....	104
3.3.1 视觉图像算法所需数据量级.....	104
3.3.2 激光点云算法所需数据量级.....	105
3.3.3 融合算法所需数据量级.....	106
4 感知训练数据集数据属性和要求.....	107
4.1 视觉图像算法数据属性和要求.....	107
4.1.1 图像 2D 目标检测算法数据集.....	107

4.1.2 图像目标追踪算法数据集.....	114
4.1.3 图像语义分割算法数据集.....	114
4.1.4 图像车道线检测算法数据集.....	118
4.1.5 图像车位线检测算法数据集.....	118
4.2 激光点云算法数据属性和要求.....	119
4.2.1 点云 3D 目标检测算法数据集.....	119
4.2.2 点云目标追踪算法数据集.....	121
4.2.3 点云语义分割算法数据集.....	122
4.2.4 点云车道线检测算法数据集.....	123
4.3 融合算法数据属性和要求.....	124
4.3.1 图像点云融合检测算法数据集.....	124
4.3.2 图像点云融合追踪算法数据集.....	124
4.4 训练数据标注精确度行业普遍情况.....	125
4.4.1 抽检比例.....	125
4.4.2 准确率.....	125
4.4.3 贴合度 (cm/pixel)	125
4.4.4 航向角.....	126
4.4.5 连续帧的最短帧间隔时长.....	126
4.4.6 速度的标注以及最大误差.....	126
4.4.7 图像标注的像素误差.....	126
4.5 训练集数据标注格式.....	126
4.5.1 图像标注数据导出格式.....	126
4.5.2 点云标注数据导出格式.....	127
5 感知训练数据集测评方法.....	128
5.1 前言.....	128
5.2 感知数据质量评价指标.....	128
5.2.1 图像数据质量评价指标.....	128
5.2.2 点云数据质量评价指标.....	129
5.2.3 融合感知数据评价指标.....	134
5.3 感知训练数据集评价方法.....	137
5.3.1 感知数据集数据类型及分布评价.....	137
5.3.2 数据集划分.....	138
5.3.3 分类问题评估指标.....	139
5.3.4 回归问题评估指标.....	139
5.3.5 数据集划分方式.....	139
6 标准化建议.....	142
附录 A	143
A.1 感知数据集数据结构图.....	143
A.2 结构元数据说明.....	143

1 感知训练数据集相关背景研究

1.1 研究背景与现状

感知训练数据集是自动驾驶环境感知算法训练与实现的重要基础，是目标跟踪，语义分割等自动驾驶训练任务必不可少的关键部分。在感知训练数据集构建过程中，有两个关键环节：数据采集与数据标注。目前国内外主流的感知训练数据集以采集途径来分类共有两种：车端数据集和路端数据集。本节将分别介绍车端感知训练数据集、路端感知训练数据集现状以及数据集标注能力现状。

1.1.1 车端感知训练数据集现状

在所有种类的感知数据集中，车端感知训练数据集得到了各厂家和研究机构的广泛关注。近年来各个企业花费大量人力物力广泛收集车端数据集：从 2012 年全球第一个车端数据集 Kitti 公布开始，越来越多的企业开始着手研发自己的数据集：Google、Argo、Motional、百度等国内外公司先后公布了自采的数据集。车端数据集具有数据规模大，模态丰富等特点。绝大部分数据集由视频和图片组成，许多数据集也包含了激光雷达点云数据，部分数据集还包含了高精度地图信息。车端数据集不仅包含驾驶数据集，还有交通标志数据集，行人数据集，车道线检测数据集等。

1.1.1.1 车端数据集采集车部署方案

数据采集车常用传感器部署方式很多，不同厂商部署方案略有不同。如图 1-1 所示，高德的一种采集车顶部则是通过装配 2 个激光雷达（位于后方）和 4 个摄像机（两前两后）用于进行车辆周围的三维模型搭建。高德的另一种采集车搭载 6 台 CCD 相机，环形分布 5 个，顶部一个，方便采集全景画面，每个摄像机像素 500 万，总像素 3000 万。如图 1-2 所示，百度 ApolloScape 数据采集系统为 Rigel 的 VMX-1HA。VMX-1HA 配备 6 个摄像机，覆盖 360°全景。安装在车前端的两个为 917 万像素摄像机，其余四个是 500 万像素摄像机。Waymo 采集车布有 5 个激光雷达，车辆前半段左中右各一个，车顶一个，车尾一个。车顶有 5 个摄像机。Argoverse

采集车布有 2 个安装在车顶的激光雷达传感器，范围 200m。车顶装有七个高分辨率环形摄像机和两个前置立体摄像机，覆盖 360°全景视野。nuScenes 采集车顶部装有一个 20HZ 激光雷达，前、后、左、右、左前、左后方各装有一个摄像机。除后方摄像机视野为 120°外，其它摄像机视野均为 70°。



图 1-1 高德采集车示意图



图 1-2 百度 ApolloScape 数据采集车示意图

1.1.1.2 车端数据集的丰富度，质量，规模及应用

行业内主流的数据集主要有 nuScenes、Kitti、Argoverse、ApolloScape、Waymo 等。在现有阶段，nuScenes、Kitti、Argoverse 数据集得到了广泛的应用，这几个数据集标注类型比较完善，全球各地的科研团队基于这些数据集做出了很多优秀的成果，研发生态比较完善。ApolloScape 积累了大量的中国道路交通数据，也可以支持针对国内场景的模型开发。另外，IDD 数据集中包含了大量的印度道路交通数据，尤其是各种少见的交通场景，具有重要价值。当面向大规模的量产产品开发时，开源数据集的数据量、场景丰富程度均不能满足实际研发需求，企业需要通过各种手段建立更庞大更丰富的数据集。

1) nuScenes 数据集

nuScenes 数据集规模较大，包括在新加坡和美国波士顿拍摄的城市街道场景。数据集的采集使用了丰富的传感器套件，包含 6 个摄像机、1 个 20Hz 激光雷达、5 个毫米波雷达、GPS 及惯导系统，精确度较高。

nuScenes 数据集包含图像、点云、目标轨迹跟踪等多个子数据集，可用于 2D、3D 检测，目标跟踪等训练任务。图像数据集包含 1.4M 高清图像数据，标注帧数约 40k 帧。多目标跟踪数据集含 850 个标记序列，23 个类别，140 万个 3D 标注框。点云分割数据集含 850 个训练场景，150 个测试场景，14 亿个标注点，4 万个点云帧，32 个类别。

nuScenes 数据集数据采集场景包含晴天、雨天、阴天。数据采集区域为波士顿(港口及南波士顿)，以及新加坡(One North, Holland Village & Queenstown)，包含城区、住宅、郊区及工业区。数据采集时间包括白天和夜晚。数据采集交通场景包括 1000 多个场景，每个场景时长 20 秒。这些场景包括高交通密度场景(如十字路口、建筑工地)、特殊场景(如救护车、动物)、潜在危险的交通情况(如乱穿马路的行人、不正确的交通行为)以及对自动驾驶汽车来说可能存在困难的情况，还有一些其他场景，以尽量覆盖在不同空间、不同的场景类型以及不同的天气和照明条件方面的多样性。数据标注类别包括车辆、行人、自行车、卡车等 23 个类别。

2) Kitti 数据集

Kitti 数据集收集了长达 6 个小时的真实交通环境，数据集由经过校正和同步的图像、雷达点云、高精度的 GPS 信息和 IMU 加速信息等多种模态的信息组成。Kitti

数据集用一个 64 线 3D 激光雷达以 10HZ 的频率采样,包含从市区、乡村和高速公路等场景采集的点云数据 (29GB)、相机矫正数据 (16MB)、标签数据 (5MB)。除标签数据只包含训练集数据文件外,所有数据均包含训练集和测试集数据文件。标签数据包含汽车,行人,自行车等类别。

其中感知深度预测数据集包含 93000 张深度地图以及相应的激光雷达扫描数据和 RGB 图像数据。视觉里程计/SLAM 评估数据集包含 22 个立体序列,包含 11 个具有真值的序列和 11 个用于评估的序列。2D 目标物探测数据集包含 7481 个训练图像和 7518 个测试图像,共 80256 个标记对象。3D 目标物探测数据集包含 481 个训练图像和 7518 个测试图像以及相应的点云。目标跟踪数据集包含 21 个训练视频和 29 个测试视频。道路数据集包含 289 个训练图像和 290 个测试图像。语义和实例分割数据集包含 200 个训练图像和 200 个测试图像。

3) Argoverse 数据集

Argoverse 数据集是第一个蕴含高清地图数据的数据集,包含匹兹堡和迈阿密等美国六个城市 290km 带有几何形状和语义信息的高精地图数据,其中所有的点云数据都和地图数据进行过配准。Argoverse 数据集的采集平台配备 2 个 32 线激光雷达,平均 10Hz 产生一幅约含十万个点的点云地图。数据集包含 113 个场景的 3D 跟踪注释和用于运动预测的 324,557 条车辆轨迹,每条轨迹长度 5 秒,共含 15 种标注类别。

4) ONCE 数据集

ONCE 数据集由香港中文大学、华为诺亚方舟实验室、华为智能汽车解决方案事业部车辆云服务部门、中山大学和瑞士理工学院联合打造。数据包含 100 万个激光雷达场景和 700 万个相应的摄像机图像。源自 144 个小时的驾时时长。总标注量包含 1.6 万个场景,共 5 个类别(汽车、公共汽车、卡车、行人和自行车)的 3D 真值框,总共产生 41 万个 3D 框和 76 万个 2D 边框。主要应用于 3D 与 2D 目标检测。

数据采集时间包括早晨、中午、下午、晚上。采集天气包括晴天、阴天、雨天。采集场景包括市中心、郊区、隧道、高速公路、桥梁等。

5) ApolloScape 数据集

ApolloScape 是由百度研究院于 2018 年公布的数据集。这些数据是在中国的不同城市在不同的光照条件和交通密度下收集的，只被授权用于学术用途。ApolloScape 数据集使用高精度地图采集车，装有两个 250 线激光雷达，点云密度极高，具备稠密环境下的 3D 点云语义分割数据。

数据集包含采集自 1000km 行驶轨迹的 14 万张点云图像，点云环境包括 20 多个驾驶场地。数据集还包含密集的逐帧像素级语义标记，共 35 类，144K+张图像，35 类，160K+帧图像像素级车道标记，8 类，90K+帧图像语义 2D 实例分割，70K 帧 2D 汽车关键点和 3D 汽车实例标记，100 多个小时双目驾驶视频。

数据集采集时间包括早上，中午及晚上，采集地点包含中国的十个城市。数据内容包括四组共 24 个不同的标签（部分标签借鉴 Cityscapes 数据集）：汽车、公共汽车、桥梁、杆、交通标志等。数据可用于 3D 重建、自定位、语义解析、实例分割等训练任务。

6) Waymo 数据集

2019 年公布的 Waymo 数据集是有史以来最大、最丰富、最多样化的自动驾驶数据集之一，25 个城市收集了超过 1000 万英里的自动行驶里程数据，包含来自 1000 个片段（每个数据片段大小为 20s，采样频率为 10HZ）的激光雷达数据。数据集包含了 11 万个 3D 物体轨迹和地图信息，共划分成 1000 个训练集和 150 个测试集。

7) WoodScape 数据集

WoodScape 是一个自动驾驶鱼眼图像数据集，该数据集包括四台环绕视图相机和九项任务（包括分割、深度估计、3D 检测等），为 1 万多张图像提供了 40 类的语义标注，为超过 10 万张图像提供了其他任务的标注。

8) IDD 数据集

IDD 数据集由安装在汽车上的前置摄像头获得的图像组成，采集车主要在印度海德拉巴、班加罗尔等城市及其郊区进行采集。数据集包含 10004 张图像，包含道路种类，障碍物类别，路侧标识等 34 个类别。包含有印度公路上的 182 个驾驶序列。图像平均分辨率为 1678*968。

9) PandaSet 数据集

PandaSet 是 2019 年由禾赛科技和 ScaleAI 发布的数据集。PandaSet 面向科研及商业应用公开。首次同时使用了机械旋转式和图像级前向激光雷达进行数据采集，输出点云分割结果，其中包含 16000 个激光雷达扫描点云图像(超过 100 个 8 秒场景)。数据集包含了汽车、自行车、交通灯、行人、建筑物等各种各样的交通信息，标注了 28 类物体。在大多数场景下还进行了语义分割，共 37 种语义标签。

表 1-1 总结了上述自动驾驶数据集相关信息。

表 1-1 常用国内外车端数据集对比

数据集	作者	发布时间	地点	使用设备	场景	标注类别	数据规模
nuScenes	Motional	2019	波士顿和新加坡	1 个 32 线雷达, 6 个摄像机	市区, 住宅, 郊区和工业区	图像 23 类, 点云 32 类	1.4M 高清图像, 4 万帧点云
Kitti	KIT& Toyota	2012	德国卡尔斯鲁厄	1 个 64 线雷达, 4 台摄像机	市区, 乡村和高速	9 类	29GB, 15000 个数据文件, 超过 200k 3D 标注物体图像
Argoverse	Argo, CMU	2019	皮茨堡和迈阿密	2 个 32 线雷达, 9 个摄像机	市区	15 类	跟踪 113 个场景, 预测 32 万条轨迹, 每个轨迹 5 秒
ONCE	华为	2021	中国多个城市	1 个激光雷达和 7 个摄像机	市中心、郊区、隧道、高速公路、桥梁等	9 类	1.6 万个场景, 41 万个 3D 框和 76 万个 2D 边框
Waymo	Waymo	2019	美国六个城市	5 个 64 线雷达, 5 个摄像头	市区、郊区等	23 类	检测 2030 个场景, 预测 10 万个场景, 113k 个 3D 轨迹和 160k 个 2D 轨迹
ApolloScape	百度	2018	中国十个城市	2 个 250 线雷达, 6 个摄像机	市区, 乡村和高速	35 类	144K+张图像, 70K 帧 3D 实例标记, 1000km 行驶轨迹

1.1.2 路端感知训练数据集现状

路端数据集相比车端数据集的区别在于其包含了采集自路端传感器的数据。车路协同技术感知范围大，可以实现人车路的有效互动，从而构建安全、高效的智能交通系统，但现有的路端感知数据集仍较少。

1.1.2.1 路端采集部署方案

路端数据采集主要部署场景为城市中的交通路口，城市交通路口人员密集，障碍物众多，对于路端辅助感知有迫切需求。另外在城市道路、高速路、其余封闭场景中，也可以部署路端设备，完成数据采集。

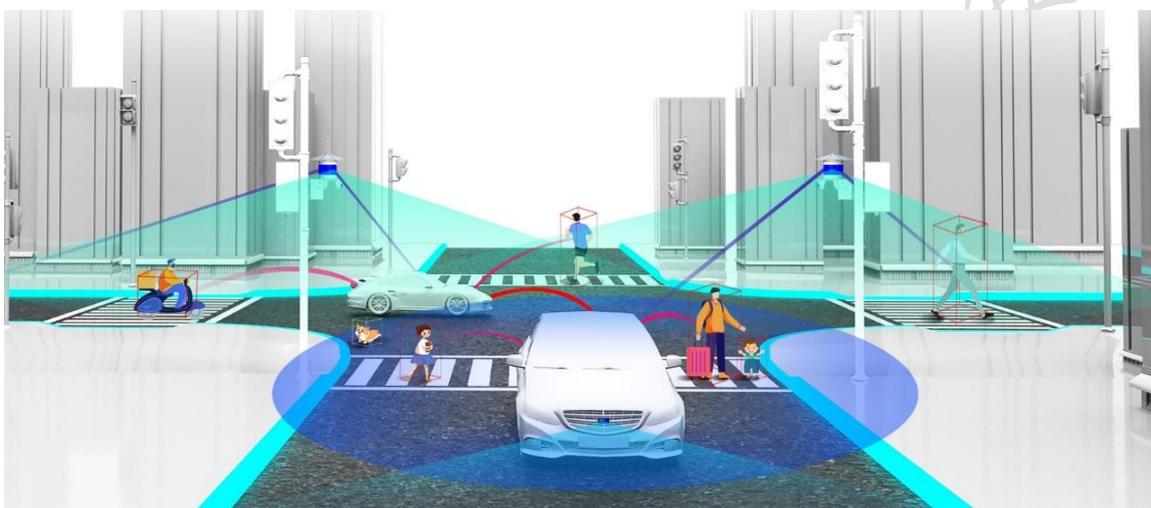


图 1-3 路端感知示意图

图 1-3 为路端感知示意图。路端感知设备主要包括激光雷达、摄像机、毫米波雷达等，为了实现区域的完全感知覆盖，感知设备大多安装在高 3-5m 的设备杆上，也有部分设备安装于现有交通设施上，如路灯杆等。部署的位置包括道路两侧、道路中央正上方、道路拐角处等，在虚拟数据集设备部署中，也会有将虚拟机械式激光雷达置于十字路口正中央的布设方式。大部分布设方案会将激光雷达与摄像机绑定位置，然后添加额外摄像机进行补盲，实现对于区域的无死角覆盖。

路端感知设备不断感知周边的道路交通信息，包括障碍物信息（行人、骑行者、机动车以及其他静态或动态物体）、交通设施（信号灯、交通标志）、路面状况（坑洼、道路维修或封闭等）、行驶环境（天气环境、交通状况等），所感知内容包括物体的位置信息、速度信息、物体大小、物体描述、历史轨迹并预估所感知物体的运动轨迹；路端感知设备将感知到的信息实时传送给自动驾驶智能路端计算控制单元，

自动驾驶智能路端计算控制单元实时处理接收到的感知信息，储存在本地或云端用于感知训练。

1.1.2.2 路端及车路协同感知训练数据集的丰富度、质量、规模及应用情况

本节介绍几个主要的路端及车路协同数据集，包括 DAIR-V2X、Rope3D、IPS300+、BAAI-VANJEE Roadside Dataset、V2XSim 等：

1) DAIR-V2X

DAIR-V2X 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 2D&3D 标注，总计 71254 帧图像数据和 71254 帧点云数据，共分为三个部分：

DAIR-V2X 协同数据集(DAIR-V2X-C)，包含 38845 帧图像数据和 38845 帧点云数据；DAIR-V2X 路端数据集(DAIR-V2X-I)，包含 10084 帧图像数据和 10084 帧点云数据；DAIR-V2X 车端数据集(DAIR-V2X-V)，包含 22325 帧图像数据和 22325 帧点云数据。

数据涵盖了脱敏后的原始图像和点云数据、标注数据、时间戳、标定文件等。障碍物目标共十类数据。同时标注了障碍物的遮挡、截断等信息。数据涵盖晴天/雨天/雾天等丰富场景。

2) Rope3D

Rope3D 数据也是清华大学智能产业研究院推出，相比于 DAIR-V2X，Rope3D 数据集更加聚焦于纯视觉的单目 3D 感知任务。包含 18166 帧 DAIR-V2X 中的带 3D 标注的路侧图像数据和额外的 31843 帧路侧数据。

障碍物类别：一共 4 大类，包括小车、大车、行人、非机动车。

3) IPS300+

IPS300+是清华 SmartLion 团队在 OpenMPD 数据集中发布的路侧感知数据集，是第一个可用于大规模城市交叉口场景中路边感知任务的多模态数据集，具有高标签密度和高挑战性。数据集包括图像数据和点云数据，共 14198 帧，每帧平均有 319.84 个标签。

点云和图像数据均包含 Car、Cyclist、Pedestrian、Tricycle、Bus、Track、EngineerCar

七个类别。

4) BAAI-VANJEE Roadside Dataset

BAAI-VANJEE Roadside Dataset 是 2021 年由北京智源人工智能研究院和北京万集科技股份有限公司联合发布“智源-万集”路侧数据集，智源-万集数据集采集自中国城市复杂路口和高速道路，涵盖了不同的天气、交通状况，支持 3D 点云检测、2D 图像检测以及激光雷达和图像融合检测任务。

该数据集包含 2500 帧 3D 点云数据和 5000 帧 RGB 图像数据(其中 20%为同时采集的数据)。3D 点云数据包含目标中心点三维坐标、目标类别、长宽高和航向角信息，2D 图像数据包含目标框位置和目标类别信息，提供了 74387 个 3D 标记实例和 105538 个 2D 标记实例。

5) V2XSim

V2XSim 是由纽约大学 AI4CE 实验室和上海交通大学 MediaBrain 小组开发的第一个自动驾驶 V2X 辅助协同感知数据集。V2XSim 数据包含三个城镇多个路口的数据，数据总量达到 500G 以上。该数据集提供了各种标签，例如 3D 边界框、逐像素和逐点语义标签。每个框由其中心在 x、y、z 坐标中的位置及其宽度、长度和高度定义，共有行人、建筑物、地面等二十三类，还提供了精确的深度值用于深度估计。

6) A9-Dataset

A9 数据集是基于德国 Providentia++ 试验场地进行采集得到，包含高速公路、十字路口、桥梁等场景。数据集分为 R0 和 R1 两个版本，R0 版本包含了从原始数据中随机选择的数据段，并进行标注，最终得到的数据结果约 3.85GB，共 1098 帧数据。标注类别包括九类。R1 版本包含了 3 种特殊场景，包括大雪、大雾和交通事故发生的场景。

表 1-2 总结了路端及车路协同数据集的相关信息。

表 1-2 路端及车路协同数据集的情况

数据集	作者	发布时间	数据量	数据类别	其他
DAIR-V2 X	清华	2022	71254 帧图像数据 71254 帧点云数据	10 类目标物	包含障碍物遮挡、截断等信息
Rope3D	清华	2022	50009 帧图像数据	9 类目标物	
IPS300+	清华	2021	14198 帧图像数据 14198 帧点云数据	7 类目标物	
BAAI-VA NJEE Roadside Dataset	北京智 源、北京 万科	2021	5000 帧图像数据 2500 帧点云数据	12 类目标物	
V2XSim	上海交 通大学	2021	三个 Carla 小镇 500G 以上数据	23 类目标物	大型虚拟数据集
A9-Datase t	慕尼黑 工业大 学	2022	1098 帧数据	9 类目标物	包含大雪、大雾、交通事故特殊场景

1.1.3 数据集标注能力现状

数据集标注是指将数据集中的原始数据（图片、点云、文本等）进行人工或自动处理，将其转换成计算机能够理解、能够被算法处理的标准形式。数据标注通常是针对特定任务进行，例如图像识别、点云分割、语音识别等等。数据集的质量不仅取决于所采集的数据丰富度、数据量等因素，也与数据标注质量有密切联系，因此数据标注的质量与能力直接影响到机器学习算法的性能。本节具体介绍数据集标注方法及标注能力。

1.1.3.1 通用标注方式

数据集标注主要包含人工标注，半自动标注，自动标注三种类型。

1) 人工标注

人工标注指由人工对文本、图像、视频、3D 点云等原始数据进行清洗，标注，质检等操作，以提供满足机器学习训练要求的数据标注结果。这种方法通常需要大量的人工劳动，但它可以提供高质量的标注数据。人工标注的标注能力可以覆盖 90% 以上的场景，在一定的标注规则下，在一定的标注规则下，通常认为人工标注的交并比（Intersection over union, IOU）为 100%。人工标注应具备的能力及流程包括：

a. 标注需求理解：明确标注数据量级，标注方法，数据安全，数据准确度要求，标注数据格式，工期约定，以及标注规则的细节说明文档，针对人工标注的培训文档。

b. 标注准备：包括数据获取、数据预处理、操作规程、质检方案、工具/平台、人员能力和试标注等环节。

c. 标注过程：按照操作规程对指定数据集进行标注，并按照质检方案进行质量控制和反馈。

d. 标注结果：生成满足需求方模型训练要求的训练样本，并按照数据交付格式进行验收交付。

2) 半自动标注

这种技术结合了人工智能和人工审阅。使用预训练的模型，可以自动为数据生成初步的标签，然后由人工审阅员进行修改和确认。现有辅助标注模型的场景覆盖率可以达到 50% 以上，通过特定自动标注和筛选策略结合，可以实现自动化标注 99% 以上的数据，并且达到与人工标注近似的标注效果，召回率达到 95% 以上。感知大模型助力标注提效可以将标注速度加速 600 倍，标注成本减少 500 倍，迭代优化速度加速 18 倍。半自动标注应具备的能力及流程包括：

a. 算法预处理和预标注：利用已有的模型或算法对原始数据进行筛选、分类等操作，以提高数据的有效性和质量，并对数据进行初步的标注结果生成。

b. 人工修正和补充：对算法预处理和预标注的结果进行检查和修改，并添加一些算法难以识别的信息或特征，以提高数据的准确性和完整性。

- c. 标注结果：同人工标注相同。

3) 自动标注

指由算法对原始数据进行完全的处理和标注，无需人工干预的操作。自动标注应具备的能力及流程包括：

- a. 算法处理和标注：利用高效且成熟的模型或算法对原始数据进行处理和标注，以达到人工水平或超过人工水平的效果。
- b. 标注结果：同人工标注相同。

1.1.3.2 传感器融合标注

自动驾驶中常用多传感器融合实现感知范围增大与感知准确度增加，针对多传感器融合数据，主要包括数据同步、数据对齐、标注物体、标注关联、质量检查几个步骤：

- a. 数据同步：首先，需要将来自不同传感器的数据进行时间同步。由于每个传感器可能有自己的采样率和时间戳，因此需要通过插值、重采样等方法，使得所有数据在时间上对齐。
- b. 数据对齐：对于空间数据（如图像和点云），需要将它们对齐到一个公共的坐标系。这通常需要知道每个传感器的位置和方向（即外参），并进行坐标变换。
- c. 标注物体：在同步和对齐的数据上进行物体标注。这包括 2D 图像标注（如边界框和分割），也包括 3D 标注（如 3D 边界框和点云分割）。在进行 3D 标注时，通常需要考虑物体的位置、方向和尺寸。
- d. 标注关联：对于同一个物体，需要将其在不同传感器数据中的标注关联起来。例如，一个车辆在图像中的边界框和在点云中的 3D 边界框应该关联到同一个物体。
- e. 质量检查：完成标注后，通常需要进行质量检查，确保标注的准确性和一致性。这包括自动化的错误检测，以及人工的标注审查。

1.1.3.3 路径与轨迹标注

针对目标跟踪、轨迹预测等任务，需要对数据序列进行路径与轨迹标注，标注

步骤包括数据收集、检测物体标注、跟踪物体标注、轨迹生成、轨迹标注与质量检查：

- a. 数据收集：首先需要收集连续的图像序列或视频数据。这些数据可以来自摄像机、LiDAR 或者其他适合的传感器。
- b. 检测物体标注：在每一帧图像或数据中，需要通过物体检测技术（如边界框、分割等）识别出感兴趣的物体。
- c. 跟踪物体标注：识别出物体后，需要进行物体跟踪，即在连续的帧中追踪同一物体的位置。这通常可以通过物体跟踪算法实现。
- d. 轨迹生成：基于物体跟踪的结果，可以生成物体的轨迹。每个轨迹通常表示为物体在连续帧中的位置序列。
- e. 轨迹标注：在生成的轨迹上进行标注，这可以包括物体的类别（如车辆、行人等）、运动状态（如速度、方向等），以及特定的行为（如转弯、停止、超车等）。
- f. 质量检查：完成轨迹标注后，需要进行质量检查。

1.1.3.4 场景理解标注

数据标注除了对目标物的状态、位置信息进行标注外，还需要标注一些场景相关信息，主要包括道路属性、环境属性、上下文信息、场景分类等：

- a. 道路属性标注：道路属性包括车道线、交通标志、交通灯、人行横道、停车位、路口等。这些属性通常需要在图像或视频帧中进行标注，例如，为车道线和交通标志画出精确的轮廓，或为交通灯添加点标注。
- b. 环境属性标注：环境属性包括天气条件（如晴天、雨天、雾天等）、照明条件（如白天、夜晚、黄昏等）、道路条件（如干燥、湿滑、积雪等）等。这些属性通常需要通过文本或标签的形式进行标注。
- c. 上下文信息标注：上下文信息包括交通流量、路面拥堵情况、道路工程、突发事件（如事故、路障等）等。这些信息通常需要通过文本描述或者特定的标签进行标注。
- d. 场景分类标注：场景分类是对整个场景的全局标注，例如，将场景分类为城市街道、高速公路、乡村道路、停车场等。这通常需要通过文本或标签的形式进行

标注。

1.2 相关概念

1.2.1 术语定义

1.2.1.1 数据集相关术语定义

数据集相关术语定义如表 1-3 所示，主要介绍了数据集、训练集、测试集、验证集、批量的定义。

表 1-3 数据集相关术语定义

术语名称	术语解释
数据集	数据的集合，本文中特指服务于感知算法的数据集合，根据需要可以划分为训练集、测试集、验证集。
训练集	数据集的子集，用于训练模型。
测试集	数据集的子集，用于在模型经过验证集的初步验证后，进行测试模型。
验证集	数据集的一个子集（与训练集不同），用于调整超参数。
批量	一次输入供模型计算用的数据量。

1.2.1.2 数据集标注相关术语

数据集标注相关术语定义如表 1-4 所示，主要介绍了数据标注、标注框、其他标注相关概念的定义。

表 1-4 数据集标注相关术语定义

术语名称	术语解释
数据标注	对数据集的图像/点云进行标记以训练机器学习模型的过程。
标注框	以 2D 框、3D 框、多边形框等形式标注出图像/点云中的指定目标对象。
遮挡	被背景遮挡的标注目标。
截断	位于图像/点云边缘且被图像/点云截断的标注目标。
航行线	目标在道路上运行的路线方向。

点云	激光获取物体表面每个采样点的空间坐标，所得到的一系列表达目标空间分布和目标表面特性的海量点的集合。
语义分割	根据物体的属性，对图像/点云进行区域划分，区分出图像/点云中的物体类型。
目标追踪	从连续帧图像/点云数据中按帧捕捉某一对象，并进行画框标注。
WGS84 坐标系	国际上采用的地心坐标系，坐标原点为地球质心。
GCS-02 坐标系	由中国国家测绘局（G 表示 Guojia 国家，C 表示 Cehui 测绘，J 表示 Ju 局）制订的地理信息系统的坐标系统。
时间戳	使用数字签名技术产生的数据，对签名对象进行数字签名产生时间戳，签名的对象包括了原始文件信息、签名参数、签名时间等信息。

1.2.1.3 感知设备相关术语定义

1) 摄像机相关术语定义

摄像机相关术语定义如表 1-5 所示，主要介绍了摄像机各类参数的定义。

表 1-5 摄像机相关术语定义

术语名称	术语解释
探测距离	摄像机能探测到物体的最远距离。
视场角	以镜头为顶点，以被测目标的物像可通过镜头的最大范围的两条边缘构成的夹角，可分为垂直视场角(HFOV)和水平视场角(VFOV)。
分辨率	当摄像机摄取等间隔排列的黑白相间条纹时，在监视器（比摄像机的分辨率要高）上能够看到的最多线数。
最低照度	即图像传感器对环境光线的敏感程度，或者说是图像传感器正常成像时所需要的最暗光线。
动态范围	摄像机拍摄的同一个人画面内，能正常显示细节的最亮和最暗物体的亮度值所包含的区间。
内参矩阵	代表相机的内部结构参数。

畸变参数	分为径向畸变系数和切向畸变系数。径向畸变发生在相机坐标系转图像物理坐标系的过程中，切向畸变发生在相机制作过程。
旋转矩阵	描述世界坐标系到相机坐标系的旋转变换。
平移矩阵	描述世界坐标系到相机坐标系的平移变换。
四元数	四元数在姿态解算中用来表示旋转。四元数的乘积可表示三维空间上任意的伸缩旋转变换。
欧拉角	构件在三维空间中的有限运动，可以用三个相对转角（每次的旋转轴是被转动坐标系的某一坐标轴）表示，即俯仰角，横滚角，航向角。

2) 激光雷达相关术语定义

激光雷达相关术语定义如表 1-6 所示，主要介绍了激光雷达各类参数的定义。

表 1-6 激光雷达相关术语定义

术语名称	术语解释
探测距离	激光雷达能探测到物体的最远距离。探测精度是指探测距离的精确度，探测精度越高，3D 景深刻画的越准。
视场角	激光雷达传感器覆盖的角度或者是激光雷达信号发射的角度分为水平视场角和垂直视场角。
线束	激光雷达分为单线束和多线束。单线束激光雷达只有一个激光发射器，随着雷达转动形成一条水平扫描线，所以只能检测前方有没有障碍物。多线激光雷达在垂直方向有多个激光发射器。
角分辨率	角分辨率分为水平和垂直两个方向，水平分辨率指左右两个扫描的激光点形成的夹角，垂直分辨率是指上下两个线束形成的激光点的夹角。
出点数	也叫周期采集点数，出点数越多，扫描的效果越好。
旋转平移变换矩阵	激光雷达测量坐标系相对于其他传感器测量坐标系的相对变换关系。
角度盲区	激光雷达视场角范围没有覆盖的区域，无法获取这些区域内的目标物信息。

1.2.1.4 目标检测跟踪相关术语定义

1) 目标检测跟踪相关术语定义

目标检测跟踪相关术语定义如表 1-7 所示，主要介绍了目标检测跟踪算法与网络相关的各类参数的定义。

表 1-7 目标检测跟踪相关术语定义

术语名称	术语解释
检测	通过算法得到目标在图像中的位置的过程称为检测。（包含 2D\3D 检测）
跟踪	不同帧中，物理意义下的同一目标相关联的过程。目标跟踪根据每帧跟踪目标个数不同又划分为单目标跟踪和多目标跟踪。
非极大值抑制	通过计算交并比，从较多候选框中去掉重复的候选框。
真值框	数据集中手动标记的边界框。
正样本	与真值对应的目标类别来说该样本为正样本。
负样本	与真值不对应的其他所有目标类别来说该样本为负样本。
简单样本	预测时与真值标签误差较小的样本。
困难样本	预测时与真值标签误差较大的样本。
区域建议	Region Proposal Network(RPN)产生的候选簇，最终检测的结果从其中产生。
边框回归	将产生的区域建议框调整到和真值框位置尽可能重叠。
特征融合	目标检测中普遍存在的低层特征分辨率高语义信息较少，高层特征分辨率低语义信息较高的问题，将两者进行高效的融合从而改善检测器的性能。
上采样	放大图像或称为上采样或图像插值的主要目的是放大原图像,从而可以显示在更高分辨率的显示设备上。
下采样	缩小图像或称为下采样或降采样的主要目的有两个：1、使得图像符合显示区域的大小；2、生成对应图像的缩略图。
卷积层	卷积层的作用是提取输入图片中的信息，这些信息被称为图像特征，这些特征是由图像中的每个像素通过组合或者独立的方式所体现，比如图片的纹理特征，颜色特征。
池化层	使用降采样的方法从原图中选择一部分像素生成新的特征图。
全连接层	全连接层将池化层的所有特征矩阵转化成一维的特征大向量，全连接层一般放

	在卷积神经网络结构中的最后，用于对图片进行分类。
特征图	特征图是卷积过滤器的输出结果。
激活函数	激活函数给神经元引入了非线性因素，使得神经网络可以任意逼近任何非线性函数。
归一化	归一化数据分布，提高网络收敛速度，提高网络泛化能力。
损失/目标函数	定义单个批次/整个训练集样本与真实值之间的误差，网络训练目标是使损失函数最小化。
优化器	网络训练反向传播过程中,指引损失函数的参数往正确的方向更新合适的大小,使得更新后的各个参数让目标函数不断逼近全局最小点。
学习率	控制优化器参数更新的系数。

2) 2D 目标检测跟踪评价指标术语定义

2D 目标检测跟踪评价相关术语定义如表 1-8 所示，主要介绍了 2D 目标检测跟踪评价标注相关的参数的定义。

表 1-8 2D 目标检测跟踪评价相关术语定义

术语名称	术语解释
真值	ground truth, GT 框，表示标注真值。
交并比	计算过两个边界框交集和并集之比。
准确率	所有预测正确的（包括正类和负类）占总的比例。
精确率	正确预测为正类的占全部预测为正类的比例。
召回率	即正确预测为正类的占全部实际为正类的比例。
PR 曲线	描述精确率-召回率关系的曲线，该曲线与坐标轴围成面积越大越好。
平均像素误差	预测目标中心位置与真实位置的像素距离。
平均重叠率	预测的 b-box 与真值的交并比取帧平均。
平均定位误差	预测框与真实框中心位置的欧式距离取帧平均。
平均精确度	对 P-R 曲线上的精确率值求平均值。
多目标跟踪精度	除了误报(FP)、丢失目标(FN)、ID 异常切换(ID Sw.)情况以外的正确预测样本占有所有样本的比率。

多目标跟踪 准确度	默认检测框重合度阈值下正确预测的目标与预测目标总数之比。
--------------	------------------------------

3) 3D 目标检测跟踪评价指标术语定义

3D 目标检测跟踪评价指标与 2D 目标检测跟踪评价指标有很大重合，特针对 3D 目标检测补充相关术语定义如表 1-9 所示。

表 1-9 3D 目标检测跟踪指标相关术语定义

术语名称	术语解释
绝对轨迹误差	绝对轨迹误差是估计位姿和真实位姿的直接差值，直观反应算法精度和轨迹全局一致性。
平均方向相似度	所有预测样本与真值间余弦距离的归一化，衡量检测目标与真值之间的方向相似程度。
平均平移误差 ATE	中心点的 2D 欧式距离。
平均方向误差 AOE	目标和 GT 框之间的最小偏航角。

1.2.2 缩略语

表 1-10 列举了感知训练数据集常用缩略语。

表 1-10 感知训练数据集常用缩略语

缩略语	英文全称	中文全称	含义
AP	Average Precision	平均精度	对于一个类别的平均精确率。
APE	Average Pixel Error	平均像素误差	一般指预测框与真实框中心位置的像素距离误差的平均值。
DR	Detection Rate	检测率	检测到真实目标上所选点的比例。
DOF	Depth of Field	景深	摄影机镜头或其他成像器前沿能够取得清晰图像的成像所测定的被摄物体前后距离范围。
FN	False Negative	假阴性	被模型预测为正类的负样本。

FP	False Positive	假阳性	被模型预测为负类的正样本。
FOV	Field of View	视场角	以光学仪器的镜头为顶点，以被测目标的物像可通过镜头的最大范围的两条边缘构成的夹角，称为视场角。
GT	Ground Truth	真值	指打标签的真值数据。
HFOV	Horizontal Field of View	水平视场角	水平方向的视场角。
IoU	Intersection over union	交并比	检测结果的矩形框与样本标注的矩形框的交集与并集的比值。
mAP	Mean Average Precision	平均精度均值	即各类别 AP 的平均值。
ML	Most Lost	丢失轨迹占比	丢失的轨迹占总轨迹的比值。
MOTA	Multiple Object Tracking Accuracy	多目标跟踪准确度	多目标跟踪综合漏检率、误检率以及 ID 跳变率来评估跟踪准确度。
MOTP	Multiple Object Tracking Precision	多目标跟踪精度	算法输出与目标真值度量距离的平均偏差。
MT	Mostly Tracked	命中轨迹占比	命中的轨迹占总轨迹的占比。
NMS	Non-Maximum Suppression	非极大值抑制	非极大值抑制，通过计算交并比，从较多候选框中去掉重复的候选框。
ROI	Region of Interest	感兴趣区域	有很大可能性包含检测目标的区域。
TN	True Negative	真阴性	被模型预测为负类的负样本。
TP	True Positive	真阳性	被模型预测为正类的正样本。
VFOV	Vertical Field of View	垂直视场角	垂直方向的视场角。

1.3 研究目的及意义

随着技术水平的不断提升和相关产业的全面融合，我国智能网联汽车产业进入新的发展阶段。智能网联汽车标准体系建设第一阶段目标任务已圆满完成，初步构建起支撑驾驶辅助及低级别自动驾驶的标准体系。习近平总书记在 2020 年在吉林省考察时来到中国一汽集团发表的讲话中提到，“现在，国际上汽车行业竞争很激烈，信息化、智能化等趋势不断发展，对我们来讲有危有机，危中有机，一定要把关键核心技术掌握在自己手里。”《智能汽车创新发展战略》中提到要突破智能汽车创新发展的关键基础技术。开展复杂系统体系架构、复杂环境感知、智能决策控制、人机交互及人机共驾、车路交互、网络安全等基础前瞻技术研发。重点突破新型电子电气架构、多源传感信息融合感知、新型智能终端、智能计算平台、车用无线网络、高精度时空基准服务和智能汽车基础地图、云控基础平台等共性交叉技术。

高精度环境感知是智能网联汽车技术创新体系中的关键基础技术之一，具有中国特色感知训练数据集又是高精度环境感知技术的重要组成部分，其质量一定程度上决定着感知能力的水平。感知训练数据集是感知算法训练的基础，数据集数据量大、覆盖场景多、数据标注精度高，都是感知算法发展提升的必须要求，制定中国训练数据集标准是中国感知算法发展、中国智能网联汽车发展的必经之路。

构建具有中国特色感知训练数据集，并兼顾现有国外成熟感知训练数据集，将有效促进我国智能网联汽车感知算法能力提升，并为中国自主品牌“走出去”提供感知算法基础支撑。因此，针对感知训练数据集标准化的调研势在必行，本报告针对感知训练数据集相关背景、数据采集及部署方案、感知训练数据集应用场景和规模、感知训练数据集的属性和需求与感知训练数据集测评方法进行了深度调查与广泛研究，并构建符合标准的感知训练数据集。

1) 标准化感知训练数据集为行业相关技术研发提供基础支撑

ADAS/ADS 技术方案不断成熟，产品驾驶自动化等级逐步提高，企业对更高精度的感知能力的需求增加，目前广泛采用的数据集往往是国外研究机构提供的，与中国的道路交通参与者的特征相差加大，存在较大的不适应性。制定中国特有感知

训练数据集标准，构建中国特色标准数据集，可以适应中国环境，提升中国企业的技术实力与经济效益。

2) 标准化感知训练数据集为行业管理提供间接保障

政府对 ICV 的管理要求逐渐明晰，技术要求和测试方法相关标准加快发布实施，目前对于数据集测评手段及测试方法还存在较大缺失，标准化的测试方法是支撑标准实施的重要保障。

3) 标准化感知训练数据集为行业提供标准、高效、全面的感知数据训练平台

目前国内及国外存在诸多感知训练数据集，数据体量几千到几万帧不等，同时质量也难以保证，误标率、错标率难以支撑车企量产车算法，同时存在数据集覆盖度较为片面，缺乏统一及全面的统筹。构建统一的、全面的、高精度的感知训练数据集，是提升感知算法精度、促进自动驾驶技术发展完善的重要保障。

汽标委智能网联汽车

2 感知数据采集及传感器部署方案研究

2.1 车端感知数据采集

2.1.1 采集车的种类及型号

采集车是感知数据采集的平台和基础，不同的采集车可以得到不同类型和质量的数据集。一般选择项目目标量产车辆、企业自身开发平台车辆，经各部门评估选取合适的采集车辆。图 2-1 为 Kitti 数据集采集车辆，选择大众帕萨特 B6，搭载激光雷达、摄像机等感知设备。



图 2-1 Kitti 数据集采集车

用来采集地图数据的车辆，一般安装有一台或多台相机、激光、GNSS、IMU 等多种传感器设备并集成有电源系统、采集控制单元、前置运算的计算机等。通过这些设备可获得图像、点云、GNSS 的空间位置信息数据。从市场对于汽车需求的方面来看，涉及到了传感器扫描角度、供电能力、作业环境、续航能力与维修要求等方面。

对于机械式激光雷达和全景相机，水平视场 360 度，垂直视场较大，视线遮挡

在所难免，则需要尽可能减少遮挡比例，其中一个重要的方法就是选车身短小的 2 厢车或 SUV，如大众 Polo Cross，其长*宽*高（mm）为 3987*1705*1486。与三厢轿车相比长度方向要减小很多。

另外考虑车辆通过性，加装设备的整体高度一般不高于 2.5 米。很多地方都有限高，如桥梁、涵洞、市内环线等，2.5 米以内车高能确保大部分限高可安全通过。

为最大化满足业务需求，选车需要关注的参数较多，既要考虑通用需求，也要根据项目不同，关注特殊需要。对车的要求一般包括车身外形与尺寸、车内空间、车顶行李架、能源类型及油标号等等，一共有十多项参数。参考如下表 2-1：

表 2-1 汽车通用需求

项目	参数	重要度	对应需求	说明
1	车身外形与尺寸	特别重要	传感器扫描角度	涉及传感器遮挡，要求车头、尾短，且后备箱门与车顶棚夹角近 90 度。
2	车内空间	特别重要	车内空间	要求满足司机乘坐，可放得下机柜、电池、备胎并能方便操作显示器。
3	车顶行李架	特别重要	车顶设备安全	车顶行李架承载能力大于设备重量，一般要求承载能力达 50kg 以上。承载能力不够，长时间容易出现行李架疲劳断裂，造成行车风险。若评估承载能力不够，需设计加固。
4	发电机	特别重要	供电能力	需要有较大冗余，一般要求发电机 14V、110A 及以上；若冗余不够，在夏天开空调情况下，容易出现亏电，满足不了每天连续 9 小时作业要求。选车需注意，很多日系车发电机小于 100A。
5	车型级别	重要	车内空间	需关注车整体高度及通过性，一般选用紧凑型 2 厢轿车或小型 SUV。
6	能源类型及油标号	重要	作业环境、续航	考虑车辆全国作业，一般选用 92 号汽油车。
7	维修性	重要	能力与维	选择保有量较大而且售后维修点多的车型。

8	环保标准	一般	修	符合当前国家环保要求即可，如国 VI。
9	变速箱	一般		考虑驾驶安全，一般推荐自动挡。
10	发动机	一般		考虑承载及作业路况，一般选择 1.6L 或以上。
11	油箱容量	一般		尽可能大一些，保证续航里程。

可用的采集车型号包括但不限于：大众帕萨特 B6 (Kitti)、雷诺佐伊 Zoe (NuScenes)、商用车一汽解放 J6V 国六 L2+级智能重型半挂牵引车 (一汽解放)、蒙迪欧、福克斯、电马、探险者、F-150, Volo XC60 和吉利领克 09 等。

2.1.2 传感器的选型

智能驾驶汽车环境感知传感器主要有激光雷达、单/双/三目摄像机、鱼眼摄像机、环视摄像机以及毫米波雷达等。

2.1.2.1 激光雷达

激光雷达也称光学雷达，是以发射激光束探测目标的位置、速度等特征量的雷达系统。其工作原理是向目标发射探测信号（激光束），然后将接收到的从目标反射回来的信号与发射信号进行比较，作适当处理后，显示器就可获得目标的有关信息，如目标距离、方位、高度、速度、姿态、甚至形状等参数。

激光雷达可分为单线激光雷达和多线激光雷达，另外也可分为机械式激光雷达、混合固态激光雷达和固态激光雷达。机械式激光雷达具有 360°视角，多线激光雷达可以获得极高的速度、距离和角度分辨率，形成精确的 3D 地图，抗干扰能力强，车端应用机械式多线激光雷达较多。但是激光雷达目前也存在挑战：一是车规级产品难度高，二是成本高，三是寿命短。

常用传感器水平指标如下：

1) 通用水平：

- a. 时钟同步要求：支持基于 GNSS、PTP 或 NTP 的时钟同步功能，从外部时钟同步系统获得授时；
- b. 接口要求：具备至少 1 个 RJ45 1000M 以太网接口；
- c. 运维功能要求：具有远程运维及在线升级功能；

- d. 防护要求：防护等级符合 GB/T 4208 的 IP67 级；
- e. 安规要求：符合 IEC 60825-1:2014 的第 6 章、第 7 章以及第 9 章对 Class 1 类激光产品的规定；
- f. 激光安全等级：class1。

2)性能水平：

表 2-2 汇总了相关数据集及相关采集机构常用车端激光雷达性能水平。

表 2-2 车端激光雷达性能水平调研

雷达型号	Velodyne HDL-64E	Velodyne HDL32E	/
应用数据集	KITTI 数据集、H3D 数据集	nuScenes 数据集	Waymo 数据集
雷达线束	64 线	32 线	64 线
有效点频	220 万点/秒	70 万点/秒	/
帧率	5-15Hz	20Hz	/
测距能力	100m	100m	75m (中距) 25m (短距)
距离精度	2cm	±2cm	/
垂直视场角	+2°至 -24.8°	+10.67°至 -30.67°	-17.6°至 2.4°(中距);-90°至 30°(短距)
垂直角分辨率	0.4°	1.33°	/
水平角分辨率	0.08°-0.35°	0.1°-0.4°	/
工作温度	-10°C-50°C	-10°C-60°C	/

雷达型号	VLP-32	国产某型号	国产某型号
应用数据集	Argoverse 数据集	/	/
雷达线束	32 线	80 线	128 线
有效点频	120 万点/秒	288 万点/秒	153 万点/秒
帧率	5Hz-20Hz	10Hz/20Hz	10Hz/20Hz
测距能力	200m	最大 230m	180m(10%反射率)
距离精度	±3cm	±3cm	±5cm
垂直视场角	-25°至+15°	40°	25.4°
垂直角分辨率	0.33°	0.1° (-2.99°-+1.01°)	0.2°
水平角分辨率	0.1°-0.4°	0.2°/0.4°	0.1°
工作温度	-10°C-60°C	-40°C-60°C	-40°C-+85°C

2.1.2.2 摄像机

常用的摄像机包括单目摄像机，双目摄像机以及鱼眼摄像机等，主要应用于中远距离场景，能识别清晰车道线、交通标识、障碍物、行人。但是摄像机对光照、天气等条件很敏感，而且需要复杂的算法支持，对处理器的要求也比较高。

在车端感知上使用最多的是 RGB 相机，还有红外相机、黑白相机等，在功能上包括普通相机、人脸识别相机、动作触发相机等等。高清摄像机的有效探测距离可以达到 80m-150m，在分辨率上可以达到 800 万像素以上，在车道线检测、车位线检测、交通标识检测等方面应用较多。

常用摄像机传感器水平指标如下：

1) 通用水平：

- a. 基本要求：建议满足 GA/T 1127—2013 中 5.1、5.2.1 规定的要求；
- b. 编码格式：视频编码支持 H.265、H.264、MJPEG 中的一种或多种格式，图像应支持 JPEG、PNG 等格式；
- c. 曝光时间：最大曝光时间支持设置（即在达到最大曝光时间时，强制曝光，以应对不同光照强度的场景），每一帧图像在成像时，在曝光时打印出准确的时间戳；
- d. 多码流要求：支持至少 3 路及以上的并发请求，摄像机主码流、辅码流均应支持设置不同的分辨率、帧率及压缩比；
- e. 时钟同步要求：支持基于 GNSS、PTP 或 NTP 的时钟同步功能，从外部时钟同步系统获得授时；
- f. 接口要求：具备 RJ45 以太网接口、BNC 接口、USB 接口等；
- g. 防护要求：IP66 级。

2) 性能水平：

表 2-3、2-4、2-5 汇总了相关数据集及相关采集机构常用摄像机性能水平。

表 2-3 前视摄像机性能水平调研

摄像机布设	前视摄像机			
摄像机型号	FL2-14S3C-C	Basler acA1600-60gc	/	Grasshopper 3

应用数据集	KITTI 数据集	nuScenes 数据集	Waymo 数据集	H3D 数据集
像素	1392*1032	1600*1200	1920*1280	1920x1200
帧率	15fps	60Hz	10fps	/
工作温度	/	0-50°C	/	/
水平视场角	/	110°	/	90°/80°
垂直视场角	/	/	/	/

摄像机布设	前视摄像机					
摄像机型号	/	/	/	国产某品牌	国产某品牌	/
像素	3480*2160	1920*1080	1080P	2560*1440	3840*2160	3480*2160
帧率	30Hz	25Hz	15Hz	25fps	20fps	30fps
工作温度	-40°C至85°C	-20°C至85°C	-30°C至55°C	-30°C至60°C	-30°C至60°C	-40°C至85°C
水平视场角	120°	97°至120°	远距 28°， 中距 61°， 近距 121°	2.8mm 焦距 105°，4mm 焦距 88°，6mm 焦距 54°，8mm 焦距 39°	2.8mm 焦距 101.6°， 4mm 焦距 88.2°，6mm 焦距 54°	30±5°
垂直视场角	63°	43°	远距 16°中 距 32°近距 66°	2.8mm57°4mm4 4° 6mm29° 8mm21°	2.8mm 52.4°，4mm 48.7°，6mm 30.2°	17±5°

表 2-4 周视/侧视摄像机性能水平调研

摄像机布设	周视/侧视摄像机			
摄像机型号	/	/	/	/
应用数据集	/	/	/	/
像素	1920*1080	1920*1535	1920*1280	200 万像素
帧率	30fps	30fps	20fps	/
工作温度	-40°C-85°C	-40°C-86°C	/	-30°C-60°C
水平视场角	91°	100±5°	93.6°	/
垂直视场角	46°	70±5°	59.4°	/

表 2-5 环视摄像机性能水平调研

摄像机布设	环视摄像机					
摄像机型号	Basler acA1600-60gc	/	/	/	/	/
应用数据集	nuScenes 数据集	ApolloScape 数据集	/	/	/	/
像素	1600*1200	500 万像素	200 万 像素	130 万 像素	1920*1080	1920*1536
帧率	60Hz	/	/	30fps	20fps	30fps
工作温度	0 至 50°C	/	/	/	-40°C至 105°C	-40°C至 85°C
水平视场角	70°	/	100°	200°	93.6°	60±5°
垂直视场角	/	/	/	130°	59.4°	47±5°

2.1.2.3 毫米波雷达

毫米波实际上就是电磁波，波段使用范围为 30-300GHz 频域，波长为 1-10mm，介于微波和厘米波之间，因此毫米波雷达兼有微波雷达和光电雷达的一些优点。

毫米波雷达的测距原理是把无线电波（雷达波/电磁波）发射出去，根据接收回波与发送之间的时间差测得目标位置距离数据。一般的探测距离在 0-200 米之间，穿透能力强，具有全天候全天的特点，但在雨、雾和湿雪等高潮湿环境下，会降低探测距离。

常用毫米波雷达传感器水平指标如下：

1) 通用水平：

- a. 目标分类要求：支持对机动车、非机动车和行人进行分类；
- b. 接口要求：支持 CAN/CANFD/ETH 等接口；
- c. 防护要求：IP66 级。

2) 性能水平

- a. 扫描频率：≥14Hz；
- b. 测距能力：前向≥150m，角雷达 ≥100m；
- c. 测速范围：-400~200km/h
- d. 帧率：≥20Hz；

- e. 距离精度: $\pm 0.4\text{m}$ (长距)/ $\pm 0.1\text{m}$ (短距);
- f. 速度精度: $\leq 0.2\text{m/s}$;
- g. 角度精度: $\leq 0.5^\circ$;
- h. 距离分辨率: $\leq 0.5\text{m}$;
- i. 速度分辨率: $\leq 0.1\text{m/s}$;
- j. 水平视场角: 前向 $\geq 90^\circ$, 角雷达 $\geq 150^\circ$;
- k. 角度分辨率: $\leq 5^\circ$;
- l. 最大输出目标数: ≥ 32 ;
- m. 工作温度: $-20^\circ\text{C}\sim 60^\circ\text{C}$ 。

2.1.3 传感器配置数量及部署安装方案

数据采集车最主要的传感器是摄像机和激光雷达, 分别对图像数据和点云数据进行全面感知识别, 实现对交通标志、车辆、行人信息的综合感知。考虑到车辆对传感器的遮挡, 大部分传感器布置在车顶及前盖处。以下为主要传感器配置数量及部署安装方案:

2.1.3.1 摄像机配置数量及部署安装方案

大部分采集车辆使用 5-8 个摄像机, 包括前视摄像机、周视/侧视摄像机、环视摄像机和后视摄像机等。前视摄像机大多布置在车顶的正前方, 用于实现远距离的道路标志感知, 主要采用高清工业相机, 分为广角摄像机和长焦摄像机, 广角摄像机视场角可达 120° 至 150° , 检测距离较短, 主要应用于低速工况下的目标物检测。长焦摄像机视场角一般小于 60° , 检测距离可达 150m 以上, 常用于远距离目标检测、红绿灯检测等。

侧视摄像机一般安装四个, 分为两个侧前视摄像机和两个侧后视摄像机, 侧前视摄像机可以安装在车辆后视镜下方, 侧后视摄像机可以安装在车辆前侧翼子板上或 B 柱处, 视场角在 90° 左右, 检测距离在 80m 左右, 用于检测侧向的目标物, 并应用于车辆变道等场景。

环视摄像机主要安装在车身四周, 例如前后车标处与左右后视镜处, 常采用鱼

眼摄像机，是可以独立实现大范围无死角监控的全景摄像机。车载鱼眼相机感知水平视场角大多可达 180 度。摄像机拍摄得到的图像需要进行去畸变处理，转化为适合人眼的正常平面视图。车载鱼眼摄像机技术成熟，价格低廉，实现近距离高精度的感知，常用于自动泊车、交通堵塞辅助和低速紧急制动等场景。

后视镜摄像机一般安装在后备箱或后挡风玻璃上，主要用于辅助泊车，视场角较宽但检测距离较短。

2.1.3.2 激光雷达配置数量及部署安装方案

大部分数据采集车选择单个激光雷达，通常为 360°机械式激光雷达，布置在采集车正上方离地 2m-3m 处，根据车辆的结构，雷达的垂直视场角进行具体选取。

也有车辆选择使用两个机械式激光雷达，在车顶两侧对称布置，每个激光雷达与地面成 30-45 度角。另外，也可以采用多个机械式激光雷达，在车顶四周对称布置用于补盲，常采用比如 16 线或者 32 线等低线束激光雷达，在车顶正前方布置高线束激光雷达，比如 64 线或者 128 线，从而扩大整体感知范围。

除此之外，随着固态激光雷达的技术不断成熟，采用固态激光雷达采集数据的方案也不断增加。由于固态激光雷达的横向 FOV 普遍在 80-150 度，常用布置方案为车辆头尾各布置一台以上固态激光雷达，实现前向后向的感知覆盖。

2.1.3.3 其他传感器配置数量及部署安装方案

正向毫米波雷达主要布置在车顶部车辆中轴线上，使雷达波束中心平面基本平行于地面，考虑安装误差，需保证与路面夹角最大偏差不超过 5°。特殊情况下，允许正 Y 向最大偏执距离为 300mm。

侧向毫米波雷达在车辆四角呈左右对称布置，前侧向毫米波雷达与车辆行驶方向成 45°夹角，后侧向毫米波雷达与车辆行驶方向成 30°夹角，雷达波束的中心平面与路面基本平行，角度最大偏差仍需控制在 5°以内。

毫米波雷达的纵向 FOV 较小，因此需要选取合适的安装高度，根据车辆结构和毫米波雷达型号，一般选择安装在 1.5m-2.5m 高处。

2.1.3.4 目前较为成熟的数据采集车传感器布置方案

如图 2-2 所示，Kitti 数据集使用 1 个 GPS/IMU，1 台 Velodyne HDL-64E 激光雷达，2 台 Point Grey Flea 2(FL2-14S3M-C) 灰度相机，2 台 Point Grey Flea 2(FL2-14S3C-C) RGB 相机，4 个 Edmund Optics NT59-917 变焦镜头。车顶前方部署四个摄像机，相同类型的摄像机相距 54cm 安装，不同类型的摄像机相距 6cm 安装。

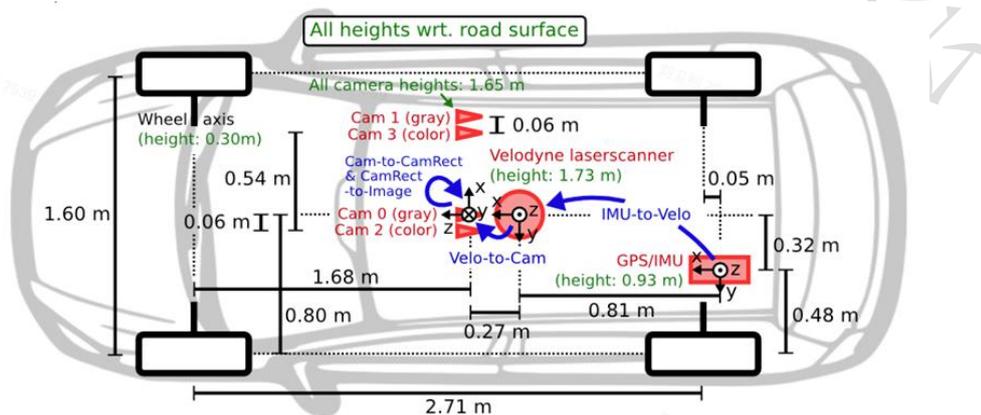


图 2-2 Kitti 数据集采集车传感器布置

NuScenes 数据集配置 6 台 RGB 相机，1 台激光雷达，5 台毫米波雷达 1 台 GPS/IMU，如图 2-3 所示，六个摄像机分别部署于数据采集车车顶的前端左右两边、前端中央、后端左右两边和后部中央，其中前置和侧置摄像机具有 70° 视场角并偏移 55°，后置摄像机的视场角为 110°；五个雷达分别布设于车辆的前端和后端；激光雷达部署于车顶中央，具有 360° 水平视场角和 -30° 至 10° 垂直视场角；惯性导航系统部署于车辆后半段部分。

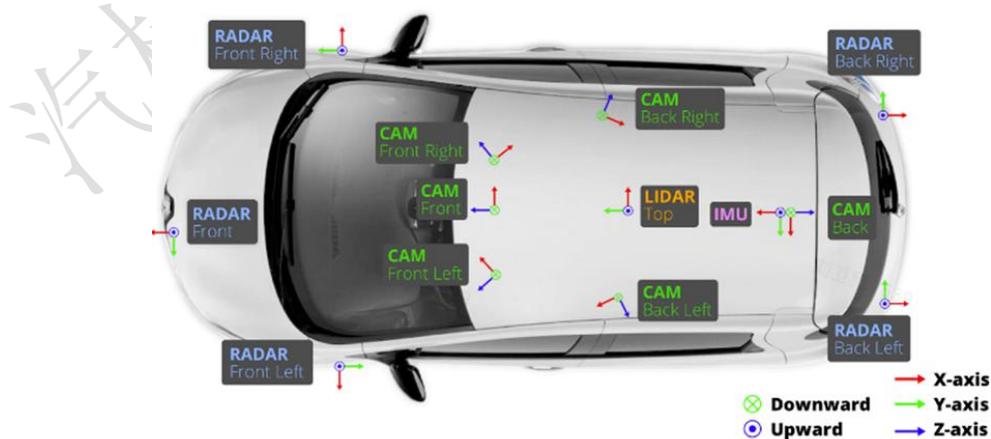


图 2-3 NuScenes 数据集采集车传感器布置

ONCE 数据采集车如图 2-4 所示配置两台 60° 摄像机（CAM1, 9）和 5 台 120° 摄像机（CAM3-8），以及一台激光雷达。60° 摄像机分别置于车辆头尾，另外五台摄像机中，一台朝向正前方，另外四台四周均匀布置。

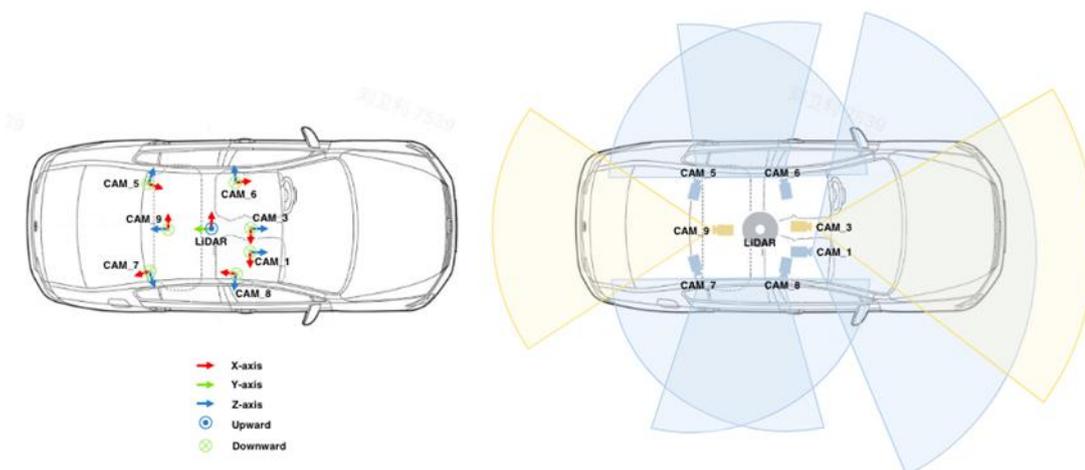


图 2-4 ONCE 数据集采集车传感器布置

H3D 数据采集车顶前端布设三台彩色摄像机，两边摄像机的视场角为 90°，中间的摄像机视场角为 80°；在车顶中央布设 64 线激光雷达，其水平视场角为 360°，垂直视场角为 26.8°；在数据采集车尾部中段布设动态运动分析仪。具体配置方案如图 2-5。

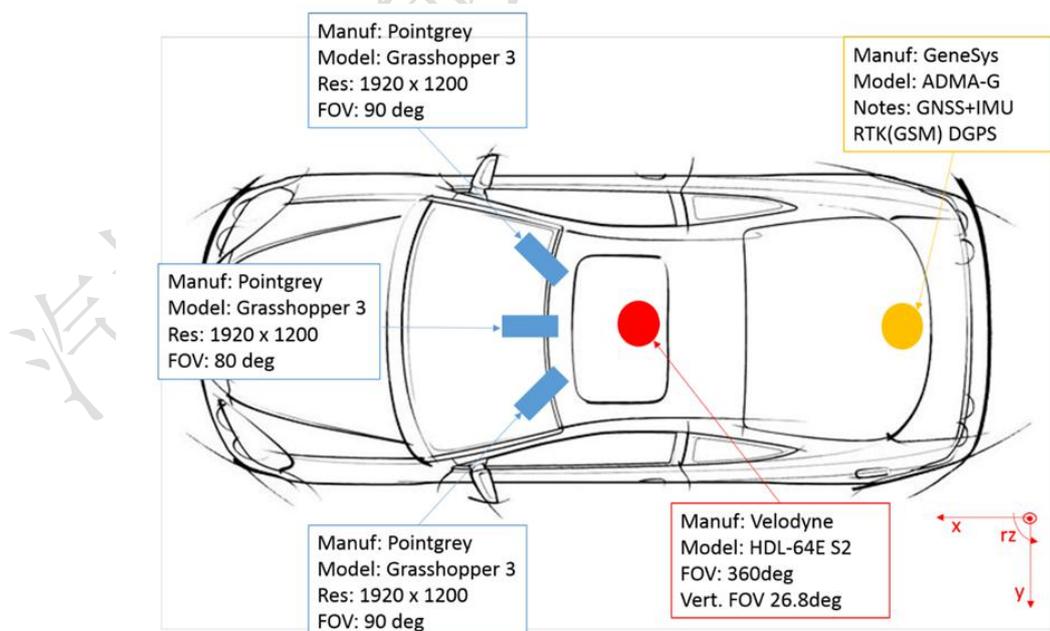


图 2-5 H3D 数据集采集车传感器布置

图 2-6 展示了国内某商用车企业数据采集车的传感器布置方案，其中前向毫米波雷达安装于车辆前保险杠处，与车辆中轴线夹角为 0° ；毫米波雷达安装于车辆四角，与车辆中轴线夹角为 60° ；Mobileye EQ4 智能摄像机安装于车辆前风挡玻璃处，与车辆中轴线夹角 0° ；森云高清摄像机 1 个采集前向视频，安装位置同 Mobileye EQ4 智能摄像机，其余 4 个高清摄像机两个一组分别安装于车辆两侧，用于采集车辆两侧视频；ibeo 激光雷达 4 个安装位置同角毫米波雷达，与车辆中轴线夹角为 70° ，另外 1 个 ibeo 激光雷达安装于车辆正前方保险杠处，与车辆中轴线夹角 0° ；Innovision 300 线激光雷达安装于车辆顶部正前方，与车辆中轴线夹角 0° ；RT3000 高精度定位系统，照度传感器与相关中控机安装于车辆驾驶室内合适位置。



图 2-6 国内某商用车企业智能车数据采集传感器布置

2.1.4 采集系统配置

如图 2-7 所示为常用车载采集系统架构，主要包括外部感知设备、与外部设备相连接的交换机、同步盒、电源模块，数据融合单元，数据采集与存储系统，最终上传到工控机或云端服务器进行存储。

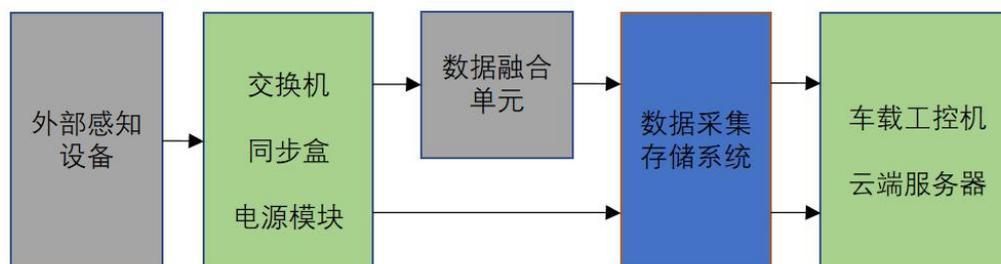


图 2-7 常用车载采集软硬件架构

如图 2-8 所示为某企业的智能车数据采集系统。其中，多路激光雷达以太网连接至同步盒、交换机，通过激光雷达融合单元将数据传输到数据采集存储系统；128 线激光雷达通过以太网直接将数据传输到数据采集系统；10 路摄像机连接至推流盒之后通过以太网连接数据采集存储系统；智能摄像头通过以太网直接将数据传输到数据采集系统。另外，采集软件系统还包括硬件间的时间空间同步，实时定位信息采集等。

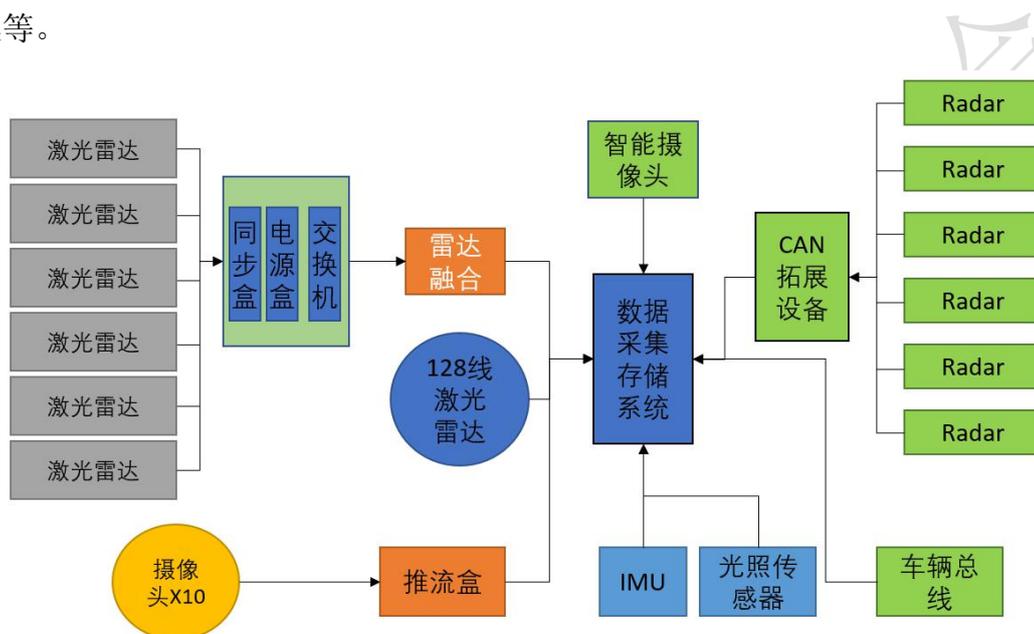


图 2-8 某企业智能车数据采集系统

2.1.5 其他硬件系统

车辆其余硬件系统主要为组合导航。

常用组合导航水平指标如下：

- a. 工作温度：-40°C-75°C；
- b. 防水等级：IP67；
- c. 数据更新率：100Hz；
- d. 陀螺量程： $\geq \pm 300^\circ/\text{s}$ ；
- e. 陀螺零偏稳定性： $< 2^\circ/\text{h}$ ；
- f. 卫星系统：支持 GPS、伽利略、格洛纳斯及北斗卫星系统；
- g. 输出接口：支持 Ethernet，UART，CAN；

- h. 无线通信：支持 WLAN、4G 等；
- i. 姿态精度：0.1°（基线长度 $\geq 2\text{m}$ ）；
- j. 定位精度：单点 L1/L2： $\leq 1.5\text{m}$ ，DGPS： $\leq 0.5\text{m}$ ，RTK： $\leq 1\text{cm}+1\text{ppm}$ ；
- k. 加速度计量程： $\geq \pm 8\text{g}$ ；
- l. 加速度计零偏稳定性： $< 5\mu\text{g}$ ；

2.2 路端感知数据采集

2.2.1 传感器选型

2.2.1.1 激光雷达

激光雷达在车路协同系统中用来精确测量目标位置、运动状态、形状，探测、识别、分辨和跟踪目标。主要的路侧感知激光雷达分为机械式激光雷达和固态激光雷达，机械式激光雷达全方向发射，可以实现 360 度全覆盖感知，但是价格较高，生产周期长，常应用于车端或路侧需广角感知的场景；固态激光雷达包括光相控阵雷达，MEMS 微机电系统雷达，Flash 技术雷达等，结构简单，成本较低，但是只能实现一侧的雷达感知。在路端，两种激光雷达均有应用，常用传感器水平指标如下：

1) 通用水平：

参考 2.1.2 节激光雷达通用水平。

2) 性能水平：

表 2-6 汇总了路端激光雷达性能水平。

表 2-6 路端激光雷达性能水平调研

激光雷达型号	国产某型号	国产某型号	国产某型号
应用数据集	DAIR-V2X	/	/
雷达线束	300 线	125 线	128 线
有效点频	/	75 万点/秒(单回波) 150 万点/秒(双回波)	153.6 万点/秒(单回波) 307.2 万点/秒(双回波)
帧率	6-20Hz	10Hz	10/20Hz
测距能力	280m	0.5m 至 200m	1 至 180m

距离精度	3cm	±5cm	±5cm
垂直视场角	40°	25°	25.4°
垂直角分辨率	0.13°	0.2°	0.2°
水平视场角	100°	120°	120°
水平角分辨率	0.09°至 0.17°(±5°) 0.19°至 0.33°(其他)	0.2°	0.1°(10Hz) 0.2°(20Hz)
工作温度	-40°C至 85°C	-40°C至 85°C	-40°C至 85°C

2.2.1.2 摄像机

摄像机用于在车路协同系统中获取图像和视频数据，主要包括 RGB 相机、红外相机、事件触发相机等。常用传感器水平指标如下：

1) 通用水平：

参考 2.1.2 节中摄像机通用水平。

2) 性能水平：

摄像机性能水平与 2.1.2 节摄像机性能水平类似，主要应用视野更广的摄像机。

- a. 像素：≥200 万像素；
- b. 帧率：≥20Hz；
- c. 工作温度：-20°C~65°C；
- d. 水平视场角：≥80°；
- e. 垂直视场角：≥40°；
- f. 测距能力：≥60m。

2.2.1.3 毫米波雷达

毫米波雷达能够较准确地检测目标的位置、速度等动态信息并且不易受天气状态的干扰，提升在车路协同系统的适用场景和鲁棒性。相关核心指标参考车端“2.1.2 毫米波雷达”相关指标。

2.2.2 传感器配置数量及部署安装方案

路端传感器主要部署场景包括城市路口、城市道路、高速道路等，对于不同的

道路场景，需进行针对性的规划设计。如表 2-7 路端场景可以分成以下几类：

表 2-7 路端场景分类

路端场景	描述
十字路口	城市中十字型路口
T 型路口	城市中 T 型路口
城市道路	城市内长度大于 200m 的道路

具体部署安装方案可以分为横杆部署和路侧竖杆部署两类，横杆部署上可以将摄像机和激光雷达向下倾斜 0-25°，实现对于路面的感知覆盖。横杆部署可以将摄像机和激光雷达水平偏转 0-90°，实现前向、后向的数据感知。

2.2.2.1 十字路口传感器配置数量及部署安装方案

十字路口可以分为有绿化带和无绿化带的情况。图 2-9 为有绿化带时十字路口，道路总宽度较宽，需要在每条道路上布设激光雷达和摄像机。一台激光雷达或毫米波雷达和一台或两台摄像机为一组，四个方向共需要 8 台激光雷达、8 台或 16 台摄像机设备。

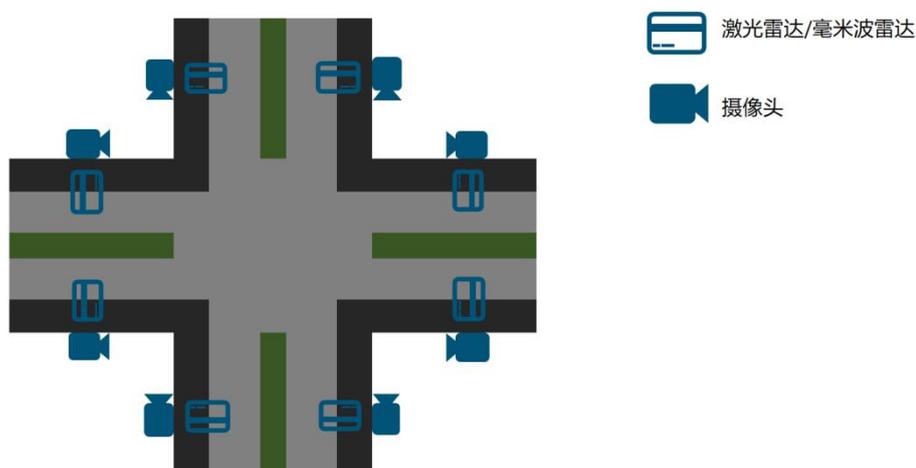


图 2-9 十字路口有绿化带布设方案

如图 2-10，十字路口无绿化带，道路宽度较小时，可以选择在四个方向红绿灯杆上各布置一台激光雷达或毫米波雷达和一台或两台摄像机，共需要 4 台激光雷达，4 台或 8 台摄像机。

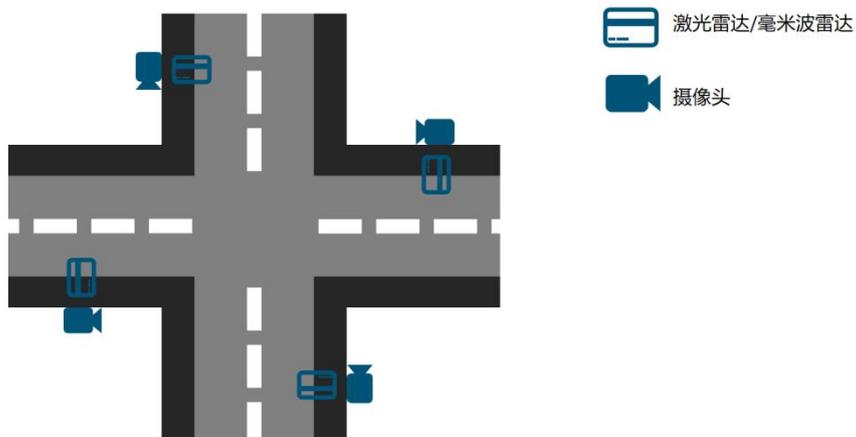


图 2-10 十字路口无绿化带布设方案一

对于道路较小的十字路口，可以选择在道路路口对侧设置激光雷达，如图 2-11，这样只需要安装两组激光雷达或毫米波雷达，节省成本。同时在每个方向上安装摄像机，实现对路口的感知完全覆盖。

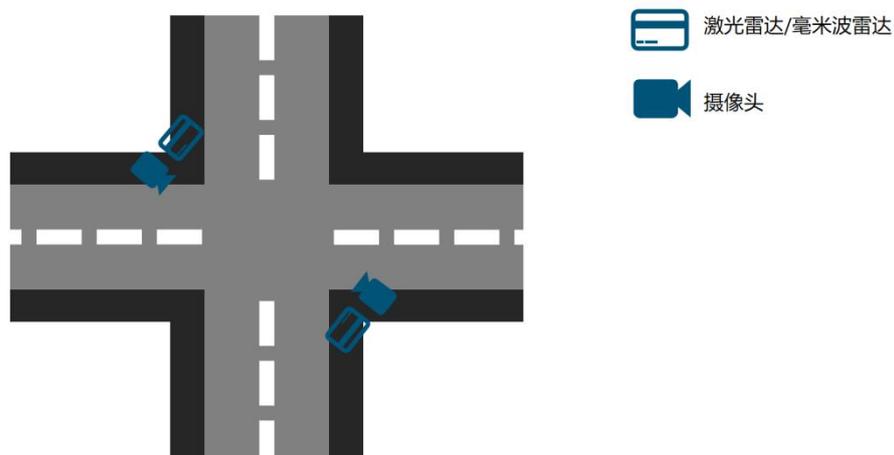


图 2-11 十字路口无绿化带布设方案二

2.2.2.2T 型路口传感器配置数量及部署安装方案

如图 2-12，对于有绿化带的 T 型路口，每个进口方向布设一台激光雷达或毫米波雷达，搭配一台或两台摄像机，每个出口方向布设一台激光雷达，搭配一台或两台摄像机，三个方向共需要布设 6 台激光雷达，6 台或 12 台摄像机设备。

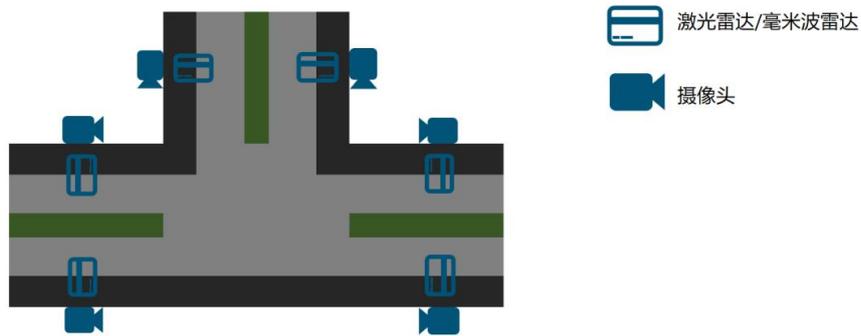


图 2-12 T 型路口有绿化带布设方案

对于无绿化带的 T 型路口如图 2-13 所示，每个进口方向布设一台激光雷达或毫米波雷达，搭配一台或两台摄像机，三个方向共需要布设 3 台激光雷达，3 台或 6 台摄像机设备。

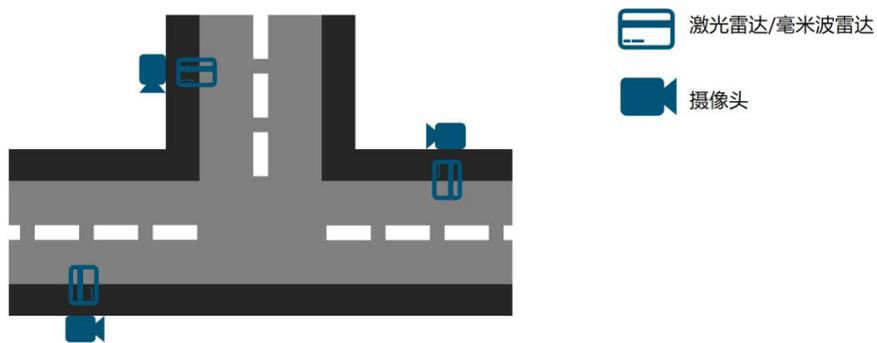


图 2-13 T 型路口无绿化带布设方案一

另一种布设方案如图 2-14 所示，对于无绿化带 T 型路口，可于横向路侧中间位置布置设备布设杆件，一处布设 1 台激光雷达或毫米波雷达，3 台相机。

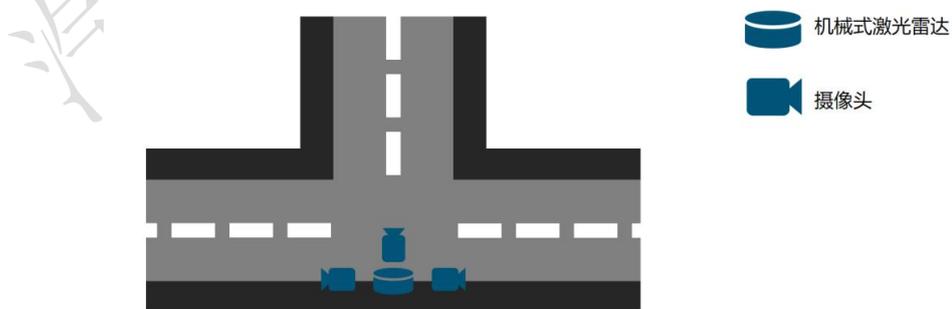


图 2-14 T 型路口无绿化带布设方案二

2.2.2.3 城市道路传感器配置数量及部署安装方案

城市道路上主要的布置方案有两种，分别是架设横杆，直接将传感器布设在道路上方，以及在道路边上布置设备杆。

如图 2-15 所示，在道路中架设横杆，间隔取 120-200m，每跟横杆上布置一台激光雷达或毫米波雷达，两台摄像机。总体布设成本较高，对于布设环境的要求也比较严格。



图 2-15 城市道路布设方案一

在道路侧布设设备杆，间隔 80-150m，每根布设杆上布置一台激光雷达或毫米波雷达，两台摄像机，如下图 2-16 所示，可以实现对于道路的感知覆盖。由于设备布置在路侧，感知结果需要进行进一步的处理。



图 2-16 城市道路布设方案二

2.2.3 采集软件配置

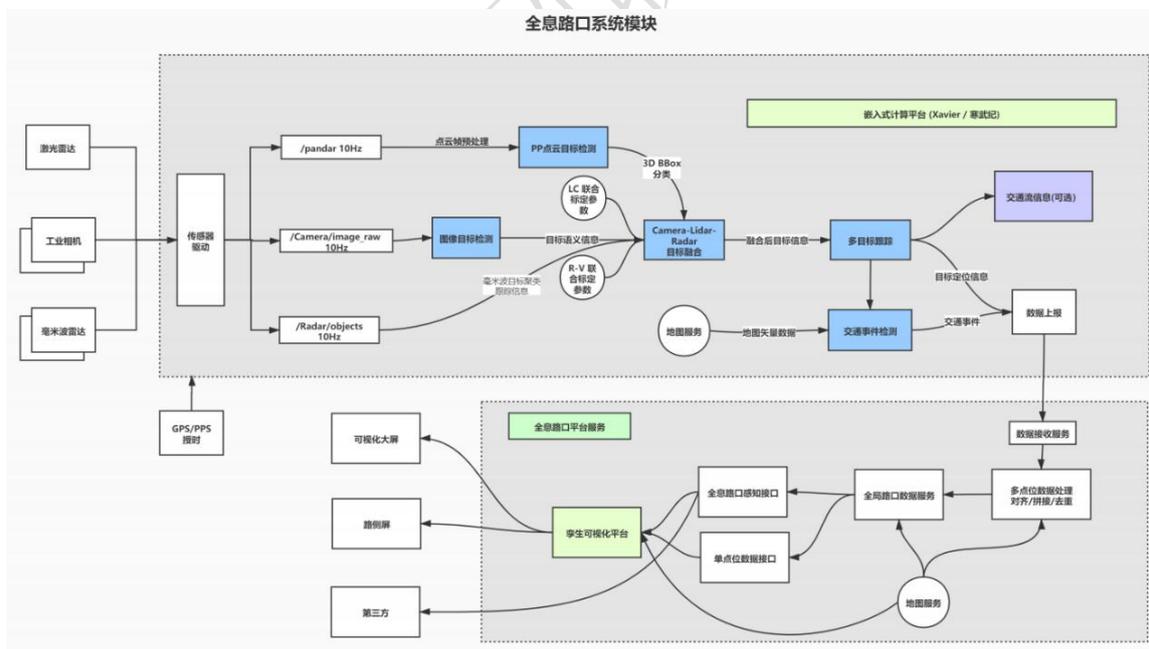


图 2-17 全息路口模块系统架构图

图 2-17 为用于路端感知的全息路口系统架构，各个模块的详细介绍如下：

1) 数据获取模块：通过激光雷达、工业相机，获取路端的点云图像和视觉图

像：

2) 目标融合模块：通过传感器驱动及数据处理，将点云信息、视觉信息以及毫米波聚类跟踪信息进行目标融合；

3) 数据处理模块：融合后的目标信息一方面通过多目标跟踪模块输出交通流信息和目标定位信息；另一方面集合高精度地图模块获取的地图矢量数据，通过交通事件检测模块，输出交通事件信息。其中，目标定位信息和交通事件信息会实时上报给数据接收服务器；

4) 孪生可视化平台模块：通过对目标定位信息和交通事件信息数据的对齐、拼接、去重处理，发送至孪生可视化平台，连接可视化大屏或者路侧屏，进行直观的图像展示。

2.2.3.2 部署方案

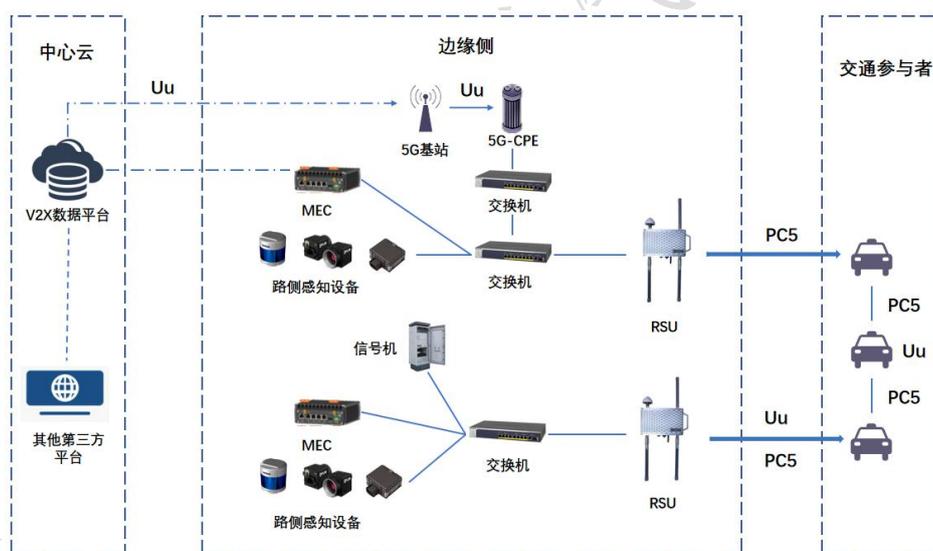


图 2-18 部署架构图

图 2-18 为典型 V2X 系统的部署架构，各个子系统的详细介绍如下：

1) 边缘侧子系统：在 MEC（Mobile Edge Computer，移动边缘计算机）上，搭载激光雷达、工业相机以及毫米波雷达等路侧感知设备，形成完整的边缘计算硬件系统。将感知处理软件系统部署到 MEC 上，系统将生成路侧交通参与者的感知信息。该感知信息，一路通过交换机发送到 RSU（Road Side Unit），RSU 再将信息转发给附近的车辆；另一路通过多层交换机和 5G 系统发送到中心云子系统。

2) 中心云子系统：中心云子系统接收来自边缘侧子系统的传感器原始数据和交通参与者感知信息，可以对路侧进行实时仿真、交通数据分析、自动驾驶仿真以及交通违法判定等任务。这里的传感器原始数据，可以作为数据闭环的原始传感器数据，同时也可以用于感知训练集的构建。

3) 交通参与者子系统：智能网联汽车接收到 RSU 发送过来的路侧感知信息，可以作为自车感知信息的辅助，也可以对汽车的感知盲区进行信息增强，提高智能网联汽车的整体安全性。

2.2.4 其他硬件系统

2.2.4.1 边缘计算单元

边缘计算单元通过以太网口等接口获取激光雷达、摄像头、毫米波雷达的原始数据，运行相关程序算法进行融合计算并输出。常用边缘计算单元水平指标如下：

- a. 支持激光雷达、摄像头、毫米波雷达、RSU 等等多传感器和通信设备接入和数据处理；
- b. 算力：可弹性支持八核心第九代英特尔/至强/酷睿 i9/i7/i5 处理器；
- c. 内存：支持 32GB；
- d. 接口支持：4 个独立以太网口，支持 iAMT12.0，2 个 SATA III，4 个 RS-232/422/485 串口、16 个通道隔离 DIO；
- e. 电源：9V 至 55V 宽范围直流输入；
- f. 通讯：支持 WLAN/4G；
- g. 网络：可扩充支持 5G 网络、10G PoE+；
- h. 存储要求：应支持本地存储，存储大小建议 $\geq 128\text{GB}$ ；
- i. 平均无故障间隔时间：不小于 20000h；
- j. 防尘防水：IP65 级别以上；
- k. 工作温度： $-20^{\circ}\text{C}\sim 60^{\circ}\text{C}$ 。

3 感知训练数据集应用场景和规模

3.1 国内外主流的感知算法模型及神经网络分类

3.1.1 图像感知算法

3.1.1.1 图像目标检测算法

1) 基于锚框 (anchor) 的目标检测算法

基于锚框的检测算法可以分为侧重于精度的二阶段检测算法(如 R-CNN^[1]、Fast R-CNN^[2])和侧重于速度的一阶段检测算法(如 SSD、YOLO)。

a. Fast R-CNN

R-CNN 系列目标检测算法是基于区域的方法。R-CNN 将检测抽象为两个过程，一是基于图片提出若干可能包含物体的区域；二是在提出的这些区域上运行分类网络得到每个区域内物体的类别。Fast R-CNN^[2]在 R-CNN 基础上，提出用 Region Proposal Network 网络取代 Selective Search 算法使得检测任务可以由神经网络端到端地完成，有效避免了 R-CNN 重复计算过多的问题，如图 3-1 所示。

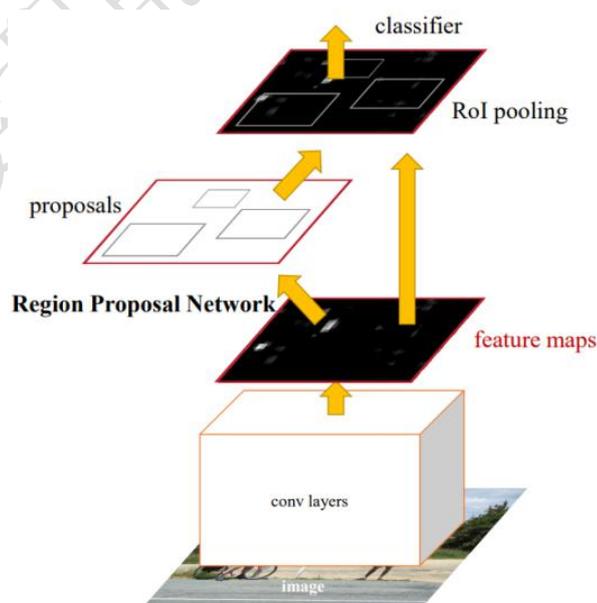


图 3-1 Faster-RCNN 网络结构图

Fast R-CNN 使用数据集：PASCAL VOC07 数据集，使用了数据集 80k 个样本。

优点：只对图片进行一次 CNN 特征提取，共享特征图进行多任务操作，大大减少了计算量，减少了运行时间；用 softmax 取代 SVM 分类器，将原来分步处理合并成一个连续的过程，降低了过程的复杂度；同样使用较大规模的网络，Fast RCNN 和 RCNN 相比，训练时间从 84 小时减少为 9.5 小时，测试时间从 47 秒减少为 0.32 秒。在 PASCAL VOC 2007 上的准确率相差无几，约在 66%-67% 之间。

b. SSD 算法

SSD 利用了固定大小与宽高的锚框（PriorBox）作为区域生成，在 6 个不同的尺度的特征图上都设置锚框，并且在浅层特征图上设置较小的锚框负责检测小物体，在深层特征图上设置较大的锚框来负责检测大物体。SSD 算法使用 VGGNet^[3] 的基础网络，并增加几个卷积层，然后利用 3×3 的卷积核在 6 个特征图上对输入图像进行预测，得到预选框的分类与回归的预测值，最后直接预测出最终的结果，具体流程如图 3-2 所示。

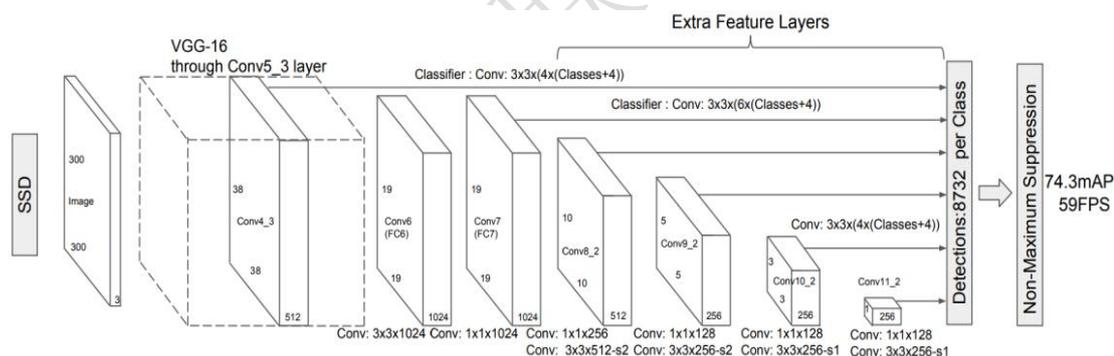


图 3-2 SSD 检测算法结构图

SSD 算法测试数据集：MS-COCO 目标检测数据集。

优点；SSD 算法在检测不同大小的物体时使用的是不同尺度的特征图，这使得 SSD 算法可以兼顾大目标和小目标的检测工作，适用于几乎所有的物体来进行目标检测，所以 SSD 算法的 mAP 高达 70%以上。

c. YOLO 系列检测算法

YOLO 系列检测算法是经典的一阶段检测算法，图 3-3 为其网络架构。YOLO 的作者借鉴了部分 SSD 的思路在 YOLOv2^[4]中使用了全新的特征提取网络 Darknet-19、自适应锚框、多尺度训练等方法，在精确度和检测速度上都有了较大

的提升。此后，作者又在 YOLOv2 的基础上提出了 YOLOv3^[5]，通过引入类似 FPN^[6] 的多尺度网络结构，进一步提升了对小目标的检测精度。

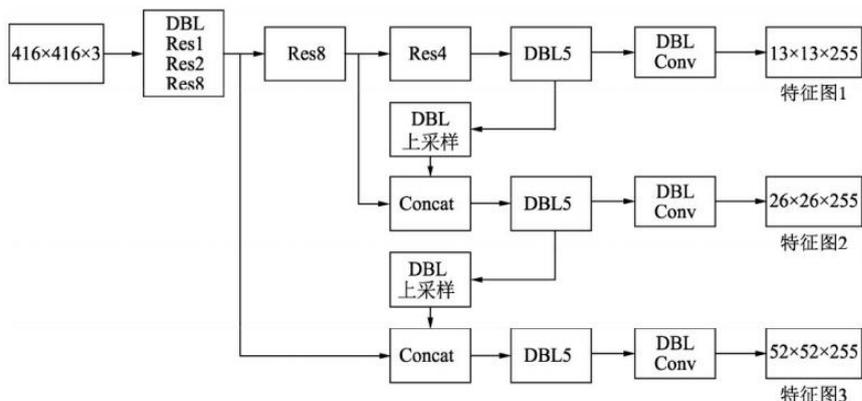


图 3-3 YOLOv3 检测算法结构图

YOLO 测试数据集： PASCAL VOC 检测数据集。

优点：YOLO 将目标检测当做回归问题进行分析，实现了端到端的训练，在实时性上有了较大的提升。抽取整张图片信息，背景错误（把背景错认为物体）比较少，泛化能力也强。

2) 无锚框的目标检测算法

无锚框的检测算法大致可以分为两类，一类是直接预测边框（FCOS^[7]、Foveabox^[8]），另外一类是基于关键点实现目标检测（CornerNet^[9]、ExtremeNet^[10]、CenterNet^[11]）。

a. FCOS 算法

FCOS 通过回归特征图上每个位置距离目标框的上下左右距离来实现检测。一方面，FCOS 通过多尺度和回归幅度限制的方法来缓解一个位置落在了多个目标框内的问题；另一方面，FCOS 采用了 Center-ness 的方法，为每个位置学习一个 Center-ness 分数，之后乘以预测类别分数作为非极大值抑制的输入参数，来解决目标检测框数量过多的问题。图 3-4 为 FCOS 检测算法结构图。

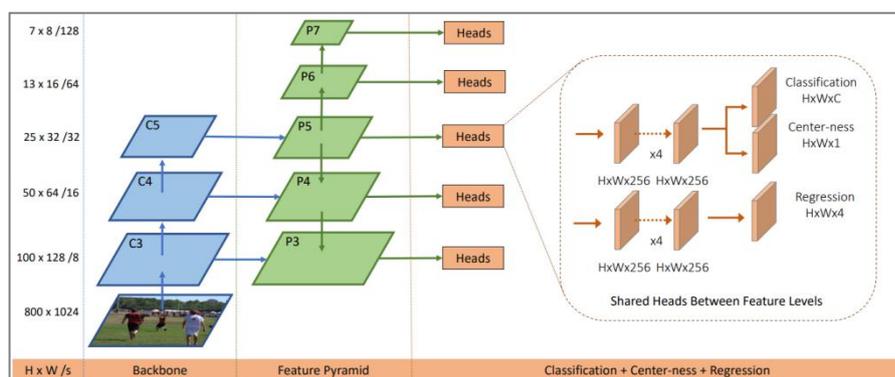


图 3-4 FCOS 检测算法结构图

FCOS 测试数据集：PASCAL VOC 检测数据集，其中使用了 115k 样本作为训练集，5k 张样本作为验证集，最终使用了 20k 张样本作为测试。

优点：FCOS 完全避免了与 anchor 相关的所有计算和超参数，并以每像素预测的方式解决了对象检测，类似于语义分割等其他密集预测任务。在 One-Step 网络中获得一流的性能，并且可以替代 Two-step 网络中的 RPN 模块，效果甚至优于 RPN 结构。

b. CenterNet

CenterNet 将目标检测当作关键点检测来处理，即用目标的中心点来代表整个目标，通过目标的中心点来预测目标的中心点偏移量、宽高。使用高斯核生成热图 (heat map) 来预测目标的类别，如图 3-5 所示。CenterNet 是一个多任务的算法，不仅仅可以实现目标检测，还能实现 3D 检测、人体姿态估计等任务。

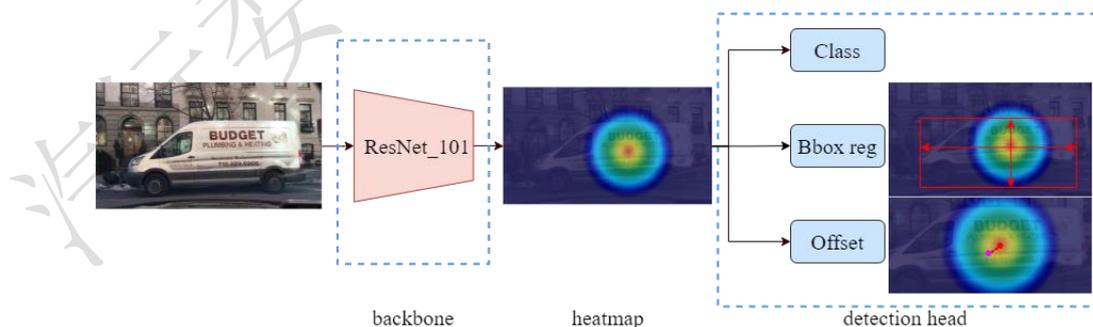


图 3-5 CenterNet 检测算法结构图

CenterNet 使用数据集：MS-COCO 验证数据集,使用了 80k 张样本作为训练集，35k 张样本作为验证集，最终使用 5k 张样本作为测试。

优点：CenterNet 是一个基于 Anchor-free 的目标检测算法。该算法相比单阶段

与双阶段目标检测算法，包括 Faster-RCNN、RetinaNet 和 Yolov3 具有更高的精度。由于该算法去除了耗时的 Anchors 与 NMS 后处理操作，因而该算法具有较快的运行速度。

3.1.1.2 图像目标跟踪算法

目标跟踪任务根据单帧出现的跟踪对象数量，可以分为单目标跟踪算法(Single Object Tracking 简称 SOT)和多目标跟踪算法(Multiple Object Tracking 简称 MOT)。

1) SOT

SOT 主流的研究思路有两种，第一种，将跟踪粗暴地看成是配对问题，即把第一帧当成模板，去匹配其他帧^[12]。第二种思路是通过第一帧给出的目标“理解”目标，在后续帧中，不需要再输入模板，即只有一个输入，网络可以根据自己理解的模板，在新图片中预测出目标。

2) MOT

a. SORT 算法

SORT 算法采用了 FasterRCNN 检测框架，采用基于卡尔曼滤波框架的线性等速运动模型来近似对象的帧间位移，通过计算当前帧的 N 个目标边界框与维护的 M 个跟踪轨迹的估计边界框之间的 IOU 得到损失矩阵，通过匈牙利算法优化求解分配。

SORT 使用数据集：MOT benchmark 数据集。

优点：SORT 可以有效地关联目标，并提升跟踪的实时性。SORT 的核心主要是卡尔曼滤波和匈牙利算法的结合版，可以达到较好的跟踪效果。在当时，追踪速度达到了 260HZ，相比其他方法速度提升了 20 倍。

b. DeepSORT 算法

DeepSORT 在 SORT 的工作基础上进行拓展，通过一个训练好的重识别卷积神经网络 CNN 生成关联矩阵，从而在 SORT 中融合了外观信息。在跟踪的关联问题上，DeepSORT 沿用了 SORT 结合卡尔曼滤波与匈牙利算法的主流范式。对于关联成本矩阵，DeepSORT 通过两个适当的度量组合以融和运动信息与外观信息。

DeepSORT 测试数据集：MOT challenge 数据集。

优点：DeepSort 是基于 Sort 目标跟踪进行的改进，改善有遮挡情况下的目标追

踪效果；同时，也减少了目标 ID 跳变的问题。

3.1.1.3 图像语义分割算法

1) 传统分割方法

传统分割算法以基于阈值和基于边缘检测的方法为主。基于阈值的图像分割方法利用灰度特征和需要分割的目标轮廓的复杂程度计算若干个阈值，并与灰度像素逐一比较进而进行像素分类。基于边缘检测的分割方法同样也是基于灰度特征，但是通过目标轮廓边缘灰度值跳变的特点来判断目标边缘，一个封闭的边缘轮廓即为一个目标。

2) 基于深度学习的语义分割方法

基于卷积神经网络的语义分割方法通过不断迭代更新卷积层中的可学习参数提取图像的高语义环境特征，在分割任务表现出更高的鲁棒性和实用性。

a. HRNet

HRNet 网络通过并行连接多个分辨率的子网在整个网络中保持高分辨率表征，然后将每个子网都与其他并行子网反复融合获得更丰富的高层特征。

HRNet 测试数据集：COCO 关键点检测数据集，训练时使用了 207k 个样本，性能评估时使用了 25k 样本。

优点：通过多尺度融合,HRNet 能够生成高分辨率的特征地图,并将其应用于目标检测、语义分割和姿势估计,从而获得显著的精度

b. HRNet+OCR

HRNet+OCR^[13]以 HRNet 作为骨干网络。其通过并行连接高低分辨率子网，能够保持高分辨率而不是通过从低到高的过程恢复分辨率，减少了下采样再上采样过程中的图像空间信息的丢失，可进行重复的多尺度融合以提高分辨率的表示，获取信息更全面，如图 3-6 所示。

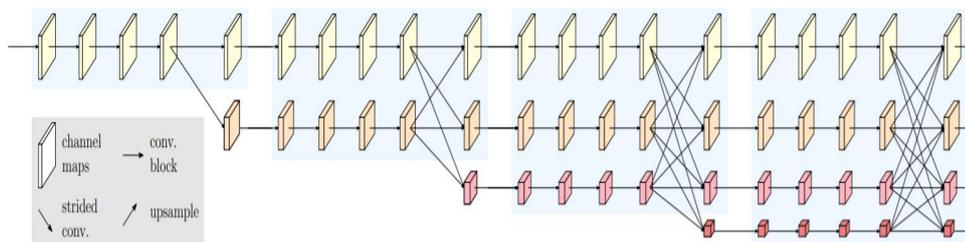


图 3-6 HNet+OCR 多分支并行结构示意图

HRNet+OCR 测试数据集：PASCAL Context and Cityscapes 数据集、LIP 人类解析数据集。在 PASCAL Context 子数据集上，使用了 4998 个样本作训练，5105 个样本作测试；在 PASCAL Cityscapes 子数据集上，使用了 2975 个样本作训练，500 个样本作验证集，1525 个样本为测试集；在 LIP 数据集上，使用了 30462 个样本作训练集，10000 个样本作测试集。

3.1.1.4 图像车道线检测算法

1) 基于分割的方法

a. SCNN

为了区分不同的车道线，SCNN^[4]将不同的车道线作为不同的类别，从而将车道检测转化为多类分割任务。SCNN 设计了一个切片 CNN 结构，以使消息跨行和列传递，如图 3-7 是 SCNN 的网络结构图。

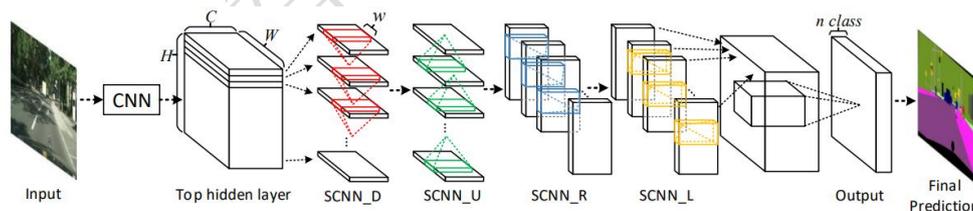


图 3-7 SCNN 网络结构图

SCNN 测试数据集：车道线检测数据集、PASCAL Cityscapes 数据集。在车道线检测数据集上，使用了 88880 个样本作为训练集，9675 个样本作为验证集，34680 个样本作为测试集；在 PASCAL Cityscapes 数据集上，使用了 2975 张用于训练，500 张用于验证，1525 张用于测试。

优点：SCNN 的顺序消息传递方案比传统的密集 MRF / CRF 具有更高的计算效率；消息传递使用的是残差，使得 SCNN 易于训练；SCNN 灵活的结构，可以应用

于任何级别的深度神经网络。

b. RESA

RESA^[15]对 SCNN 结构进行改进，加入切片间不同步幅大小的信息传递，同时解耦相邻层之间的时序上的依赖，增加并行处理能力，如图 3-8 为 RESA 网络结构图。

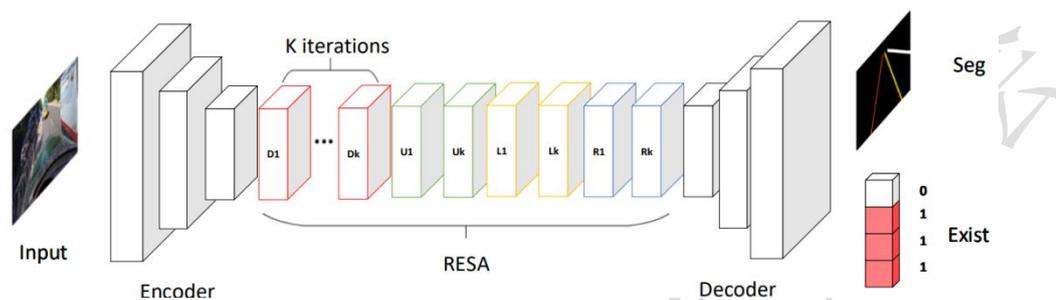


图 3-8 RESA 网络结构图

RESA 测试数据集: CULane 数据集和 Tusimple 数据集。在 CULane 数据集上，使用了 88880 个样本作为训练集，9675 个样本作为验证集，34680 个样本作为测试集；在 Tusimple 数据集上，使用了 3236 张用于训练，358 张用于验证，2782 张用于测试。

优点: RESA 以并行方式传递信息，从而显著减少时间成本；信息将在 RESA 中以不同的步伐传递，因此可以在传播期间收集不同的切片特征图而不丢失信息；RESA 简单灵活，可并入其他网络。

2) 基于检测的方法

基于检测的方法分为基于 Anchor 的方法（如 LineCNN^[16]和 LaneATT^[17]等）和基于实例检测的方法（如 CondLaneNet^[18]、UFAST^[19]等）。

a. LineCNN

LineCNN 类比了 FasterRCNN 中 RegionAnchors 的做法，设置了很多 LineAnchors，即在左-下-右三个方向的边缘像素上设置一系列的 anchor。LineCNN 直接在 anchor 上预测分数、长度以及车道线在固定竖直步长上的横坐标偏移，如图 3-9 所示。

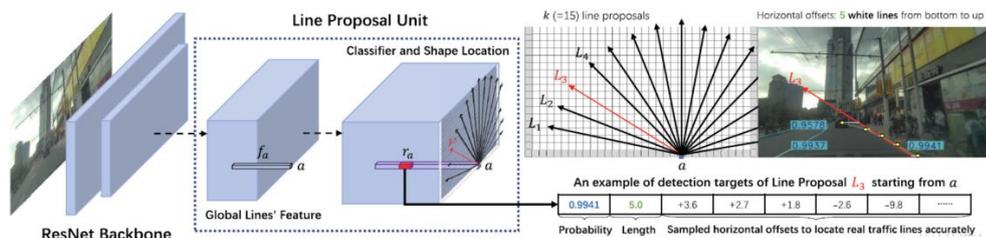


图 3-9 LineCNN 网络结构图

LineCNN 测试数据集： MIKKI2 数据集、TuSimple 数据集。在 MIKKI 数据集上使用了 1M 个样本作为训练集，1k 个样本作为测试集；在 TuSimple 数据集上，使用了 3626 个样本作为训练集，其余作为测试集。

优点: LineCNN 是一个端到端的车道线检测模型，该模型的使用 ResNet 提取特征，使用 RPN 思想实现车道线检测，其核心是提出 LPU 结构, LPU 解决了车道线误差计算的问题，使得该检测模型达到不错效果。

b. LaneATT

LaneATT 是一种基于线性型 Anchor 的池化方法，结合注意力机制来获取更多的全局信息。如图 3-10 所示是该网络的结构示意图，LaneATT 使用 Resnet 作为特征提取，生成一个特征映射，然后汇集起来提取每个锚的特征。这些特性与一组由注意力模块产生的全局特征相结合，通过结合局部和全局特征，这在遮挡或没有可见车道标记的情况下可以更容易地使用来自其他车道的信息。最后，将组合特征传递给全连接层，以预测最终的输出车道。

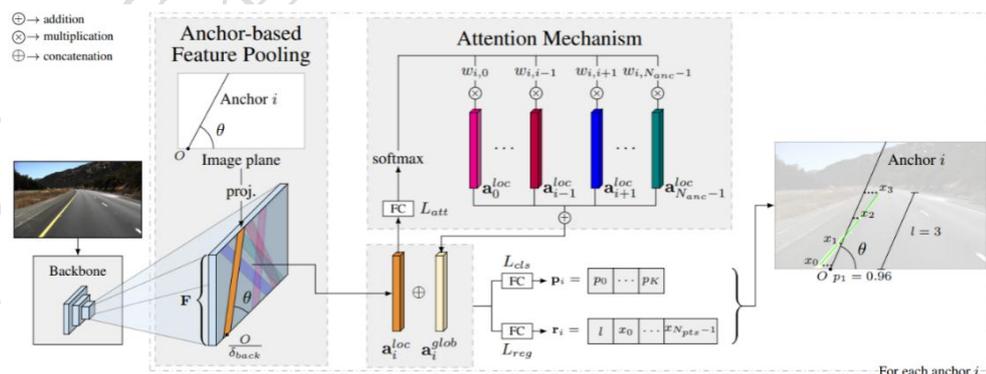


图 3-10 LaneATT 网络结构图

LaneATT 测试数据集： TuSimple 数据集、CULane 数据集和 LLAMAS 数据集。在 TuSimple 数据集上，使用了个 3628 样本作为训练集，使用了个 358 样本作为验证集，使用了 2782 个样本作为测试集；在 CULane 数据集上，使用了 88880 个样本

作为训练集，使用了 9675 个样本作为验证集，使用了 34680 个样本作为测试集；在 LLAMAS 数据集上，使用了 58269 个样本作为训练集，使用了 20844 个样本作为验证集，使用了 20929 个样本作为测试集。

优点：LaneATT 相比其他模型具有更快的训练和推理时间（达到 250 帧/秒），实现了实时与更精确的车道检测。

c. CondLaneNet

CondLaneNet 是一种自上而下的车道线检测框架。它首先检测车道实例，然后动态预测每个实例的线形。针对车道实例级识别问题，提出了一种基于条件卷积和逐行公式的条件车道检测策略。如图 3-11 所示是该网络的结构示意图，该网络的骨干网采用标准 ResNet 和 FPN 进行多尺度特征提取。

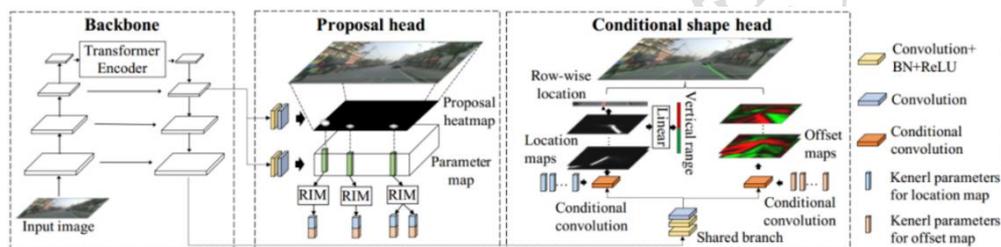


图 3-11 CondLaneNet 网络结构图

CondLaneNet 测试数据集：CurveLanes 数据集、CULane 数据集和 TuSimple 数据集。在 CurveLanes 数据集上，使用了 100k 个样本作为训练集，使用了 20k 个样本作为验证集，使用了 30k 个样本作为测试集；在 CULane 数据集上，使用了 88.9k 个样本作为训练集，使用了 9.7k 个样本作为验证集，使用了 34.7k 个样本作为测试集；在 TuSimple 数据集上，使用了 3.3k 个样本作为训练集，使用了 0.4k 个样本作为验证集，使用了 2.8k 个样本作为测试集。

优点：具有较好的实时性和准确性，在 CULane 上获得 78.14 F1 分数和 220 FPS。

d. UFAST

UFAST 将车道检测过程看作一个基于全局特征的行选择问题，能够有效的降低计算量，提高计算速度，轻量版能够达到 300+fps。

UFAST 测试数据集：CULane 数据集、TuSimple 数据集。在 CULane 数据集上，使用了 88880 个样本作为训练集，使用了 9675 个样本作为验证集，使用了 34680

个样本作为测试集；在 TuSimple 数据集上，使用了 3268 个样本作为训练集，使用了 358 个样本作为验证集，使用了 2782 个样本作为测试集。

优点：UFAST 自顶向下的设计能够更好的利用车道线的先验知识，提高检测实时性，同时在处理严重遮挡等情况下能够获得连续的车道线检测实例。但预设 Anchor 形状会影响检测的灵活性。

3) 基于参数曲线的方法

基于参数曲线的方法使用预先设定的参数曲线，对车道线形状进行检测，相关工作有：PolyLaneNet^[20]、BezierLaneNet^[21]等。

a. PolyLaneNet

PolyLaneNet 是一种用于端到端车道线检测估计的卷积神经网络，如图 3-12 所示。PolyLaneNet 从安装在车辆中的前视摄像机获取输入图像，并输出代表图像中每个车道标记的多项式和每个车道的置信度得分。该方法与现有的最新方法相比具有竞争优势，同时速度更快，不需要后处理即可获得车道估算值。

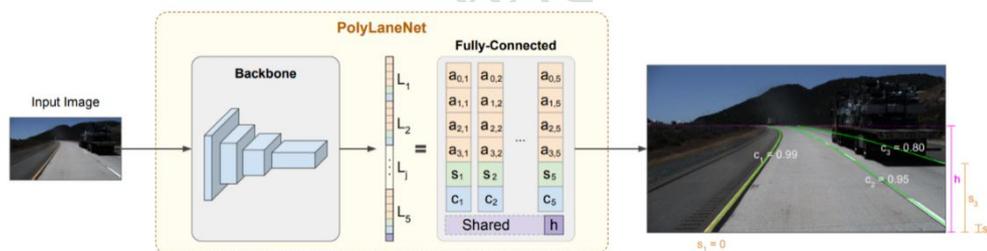


图 3-12 PolyLaneNet 网络结构图

PolyLaneNet 测试数据集：TuSimple、LLAMAS 和 ELAS。在 TuSimple 数据集上，使用了 3268 个样本作为训练集，使用了 358 个样本作为验证集，使用了 2782 个样本作为测试集；在 LLAMAS 数据集上，使用了 58269 个样本作为训练集，使用了 20844 个样本作为验证集，使用了 20929 个样本作为测试集；在 ELAS 数据集上，使用了 11036 个样本作为训练集，使用了 5957 个样本作为测试集。

优点：PolyLaneNet 基于曲线的方法可以自然地学习整体车道表示，具有较高的推理速度，但在准确度上有待提高。

b. BezierLaneNet

BezierLaneNet 方法主要对现有多项式曲线生成方法进行优化，使用 Bezier 曲

线拟合车道线，其网络结构如图 3-13 所示。

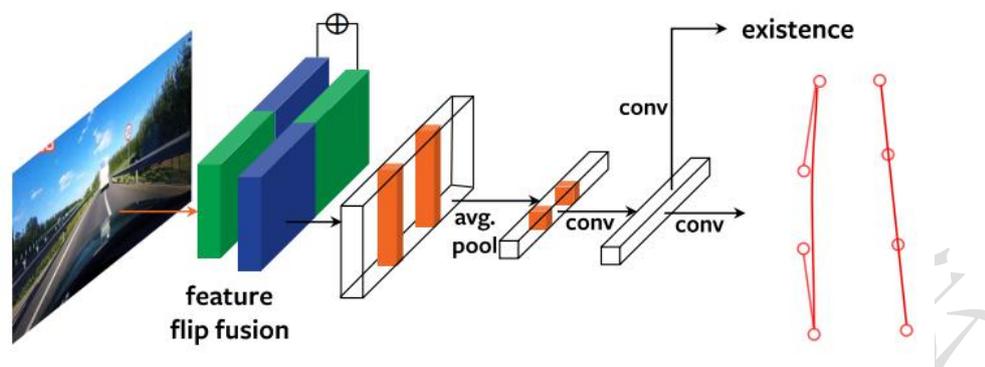


图 3-13 BezierLaneNet 网络结构示意图

BezierLaneNet 测试数据集: CULane、LLAMAS 和 TuSimple 数据集。在 CULane 数据集上，使用了 88880 个样本作为训练集，使用了 9675 个样本作为验证集，使用了 34680 个样本作为测试集；在 LLAMAS 数据集上，使用了 58269 个样本作为训练集，使用了 20844 个样本作为验证集，使用了 20929 个样本作为测试集；在 TuSimple 数据集上，使用了 3268 个样本作为训练集，使用了 358 个样本作为验证集，使用了 2782 个样本作为测试集。

4) 基于关键点的方法

基于关键点的方法直接对车道线的实例进行检测，使用后处理对实例进行划分。相关的工作有：FOLOLane^[22]、GANet^[23]等。

a. FOLOLane

FOLOLane 把车道线检测视作关键点评估+关联任务，在降低计算成本的同时避免了 Segmentation 策略中噪声大、信息冗余的问题。如图 3-14 所示是该网络结构示意图，通过轻量级的分割网络，输出两个分支，一个分支输出 heatmap 图表示像素是否是关键点，另一个分支输出 offsets 用于精准补偿关键点的位置。输出的网络通过关联算法完成局部到全局的曲线关联，形成多条完整的曲线。

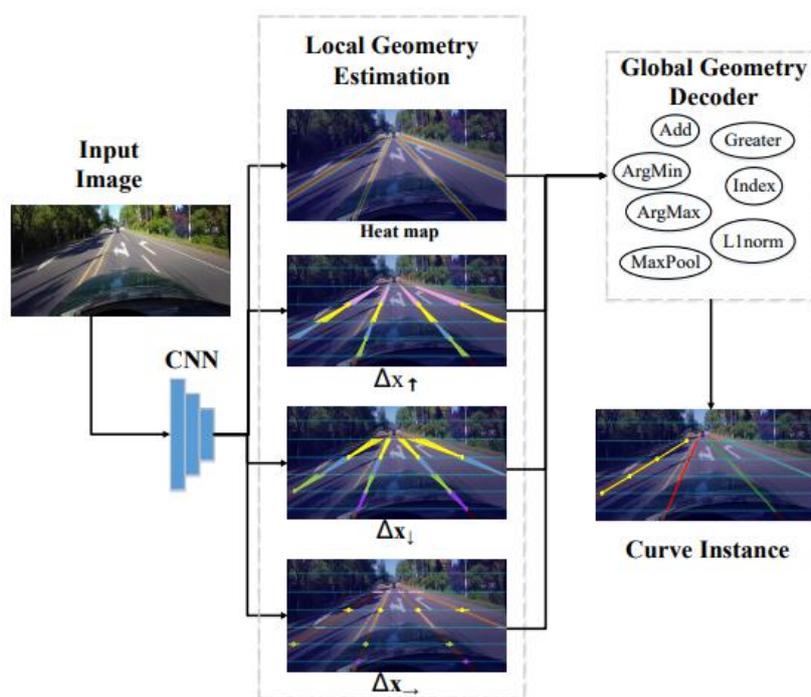


图 3-14 FOLOLane 网络结构示意图

FOLOLane 测试数据集：CULane 数据集、TuSimple 数据集。在 CULane 数据集上，使用了 88880 个样本作为训练集，使用了 9675 个样本作为验证集，使用了 34680 个样本作为测试集；在 TuSimple 数据集上，使用了 3268 个样本作为训练集，使用了 358 个样本作为验证集，使用了 2782 个样本作为测试集。

b. GANet

GANet 网络结构采用如图 3-15。主干网络（backbone）、自注意力模块（SA）、FPN 模块依次提取图片的多尺度特征。网络中的关键点头（keypoint head）和偏移量头（offset head）分别预测关键点的置信度图（confidence map）和关键点到车道线起始点的偏移量图（offset map）。GANet 在推理过程中通过对这二者进行采样和组合，可以将关键点分配到所属的车道线，得到最终的车道线预测结果。

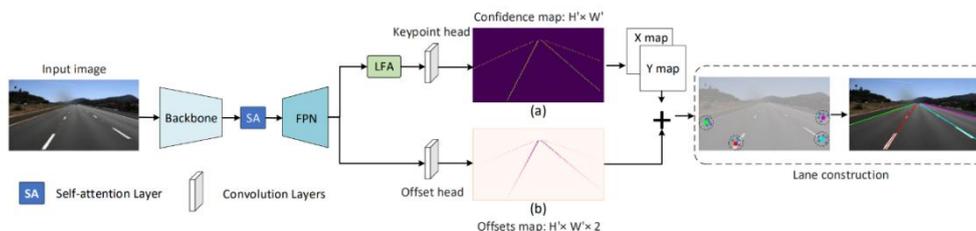


图 3-15 GANet 网络结构示意图

GANet 测试数据集：Scene Flow 数据集和 KITTI 数据集。

优点：GANet 将经典 SGM 算法参数化表示，基于数据学习参数；同时摒弃深度卷积网络权值共享的机制，添加基于图像纹理控制的权重操作层，在立体匹配数据集取得优异性能。

3.1.1.5 图像车位线检测算法

1) 传统车位线检测

传统的车位线检测方法可以分为基于线的检测和基于角的检测。其中基于线的检测主要是通过识别地面上的标记来检测车位线。但该方法必须提供正确的车位线标记，否则不能够很好的完成检测。基于角的检测的方法^{[24]-[32]}通过使用各种技术(如霍夫变换、拉东变换、随机样本一致性 (RANSAC) 或距离变换) 查找线来检测停车位标记。

2) 深度学习车位线检测

a. DeepPS

2018 年 Zhang 等人^[33]提出一种基于深度卷积神经网络的车位线检测算法 DeepPS，该算法利用 CNN 进行车位线检测的网络结构，并且将 DeepPS 算法成功地部署到荣威 E50 电动汽车。该算法利用矩形描述符来提取停车位顶点的矩形邻域内的模式。

b. DMPR-PS

Huang 等人^[34]提出了一个名为定向标记点的概念，首次采用方向描述符，设计了一种网络结构 DMPR-PS。该方法包括定向标记点回归和停车槽推断两个步骤，具体步骤如图 3-16 所示。首先，DMPR-PS 通过在单个正向评估中预测给定环绕视图图像的所有方向标记点的位置、形状和方向，获得标记点及其邻域的信息。然后，DMPR-PS 根据检测到的标记点的几何关系，推断出图像中的停车槽。

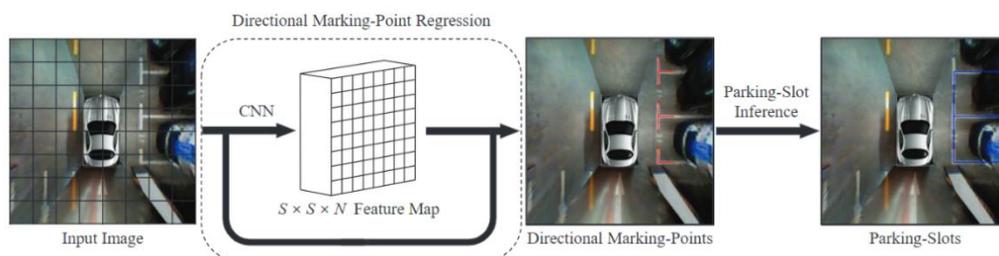


图 3-16 DMPR-PS 网络结构图

DMPR-PS 测试数据集：ps2.0 基准数据集。

优点：DMPR-PS 在测试数据集上精度为 99.42%，召回率为 99.37%，同时在英伟达泰坦 Xp 上实现了每帧 12ms 的检测速度。

3) 基于上下文的两阶段的停车位检测方法

Do 等人^[35]等人在 2020 年提出了一种基于上下文的两阶段的停车位检测方法，是一种端到端的可训练停车位检测方法。总体结构如图 3-17 所示，其结构由两个深度网络模块组成：一个停车位的上下文识别器和停车位检测器。识别器识别停车环境，包含车位类型、角度和车位的可用性；车位检测器通过多个基于类型的具有旋转 anchor box 和旋转非最大抑制的微调检测器来定位车位线的确切位置。

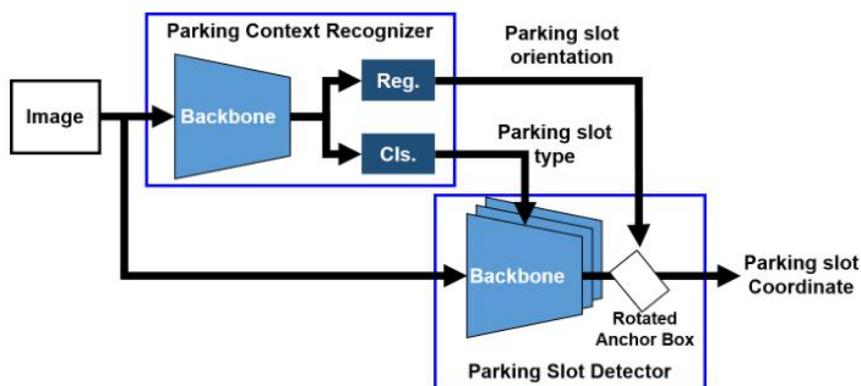


图 3-17 二阶段整体结构框架

该方法测试数据集：PIL_PARK 数据集。

4) 基于注意力图神经网络的停车位检测方法

2021 年 Min 等人^[36]提出了一种基于注意力图神经网络的停车位检测方法。该方法将 GNN 应用于车位线检测的工作，将环视图像中的标记点作为图结构数据，并利用图神经网络聚合标记点之间的相邻信息。图 3-18 为该方法的总体结构。

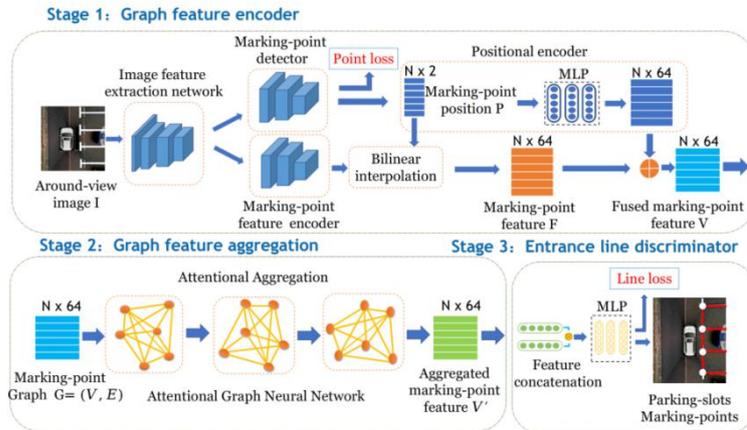


图 3-18 总体结构框架

该方法测试数据集：ps2.0 基准数据集。使用了 7780 个样本作为训练，2290 个样本作为测试。

3.1.1.6 图像深度估计算法

基于深度学习的单目视觉深度估计方法根据有无真实深度信息作为监督可以分为有监督的单目深度估计、无监督的单目深度估计、半监督的单目深度估计（双目图像）等，其他的还有基于条件随机场（CRF）的方法，联合语义分割的方法，引入其他信息辅助的方法等。

1) David^[37]等于 2015 年首次将神经网络用于单目深度估计任务，提出使用两个尺度的神经网络对单张图片的深度进行估计。

该方法测试数据集：NYU Depth、KITTI。其中，在 NYU Depth 数据集上使用了 294 个样本作为训练集，使用了 215 个样本作为测试集；在 KITTI 数据集上使用了 22400 个样本作为训练集，同样数量的不同样本作为测试集。

2) Eigen 等人^[38]提出了一个统一的多尺度网络框架，分别将其用于深度预测，表面法向量估计和语义分割 3 个任务。

该方法测试数据集：NYU Depth、Sift Flow 和 PASCAL VOC。其中，在 VOC 数据集上使用了 8498 个样本作为训练集，使用了 736 个样本作为测试集。

3) Zhou 等^[39]提出了利用一段由单目相机拍摄的图片序列作为训练集，采用无监督的方式训练神经网络进行单目深度估计的方法。如图 3-19 所示。

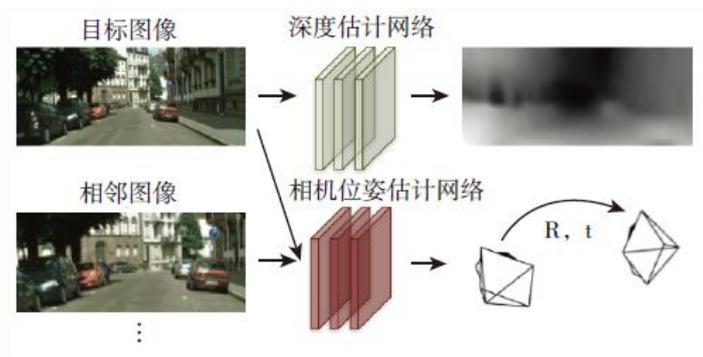


图 3-19 通过单目视频同时估计深度及相邻帧图片间的相机位姿变化

该方法测试数据集：KITTI、PASCAL Cityscapes。

4) Clement 等^[40]利用图像重建损失对网络进行训练，之后由神经网络输出视差图，再利用预测出的视差图计算深度图。

该方法测试数据集：KITTI 数据集、Cityscapes 数据集。其中在 KITTI 数据集上，29000 个样本用于训练，其余用作评估。

3.1.1.7 图像 BEV 感知算法

a. BEVDet

BEVDet 首先通过卷积的方式对多视角图像进行特征提取（Image-View Encoder），再通过基于 LSS 的深度估计方法将多视角的图像特征投影到 BEV 空间下（View Transformer），再用和第一步类似的卷积操作对 BEV 特征进行编码（BEV Encoder），最后通过一个 3D 目标检测头完成最终的目标检测（Head）。BEVDet 算法的网络结构图如图 3-20 所示。

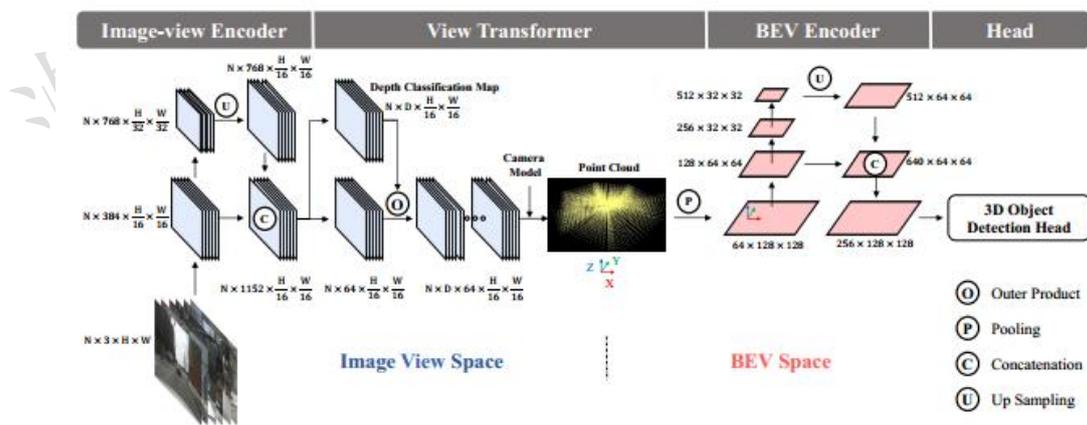


图 3-20 BEVDet 网络结构图

BEVDet 算法测试数据集: nuScenes 数据集, 它由 1000 个序列组成, 每个序列大约 20 秒, 采样频率为 20 帧/秒, 分为 700/150/150 个场景分别用于训练/验证/测试。

优点:BEVDet 将在 BEV 空间下进行 3D 目标检测, 它很好地平衡了运算速度和准确率。BEVDet-Base 在 nuScenes 数据集上取得了 39.3% mAP 和 47.2% NDS 的成绩, 在计算速度相当的情况下其感知精度显著优于其他算法。

b. BEVFormer

BEVFormer 算法对多视角图像进行特征提取后, 从 BEV 空间的 3D query 出发, 得到参考点和采样点, 再通过相机的内外参变换矩阵投影到多视角 2D 图像上和相应的特征进行交互得到 BEV 特征并进行目标检测和语义分割的多任务感知。BEVFormer 还通过对前一帧和当前帧的 BEV 特征进行交互, 获得当前帧缺失的时序特征, 以解决当前帧存在的目标遮挡或不稳定等问题。如图 3-21 所示为 BEVFormer 的网络结构图。

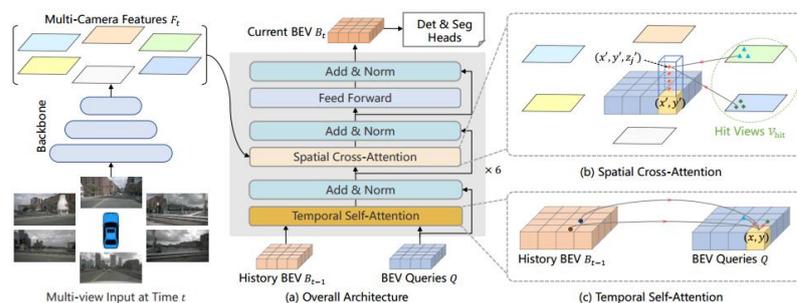


图 3-21 BEVFormer 算法网络结构图

BEVFormer 算法测试数据集: nuScenes 数据集, 它由 1000 个序列组成, 每个序列大约 20 秒, 采样频率为 20 帧/秒, 分为 700/150/150 个场景分别用于训练/验证/测试。

优点:使用了可形变注意力机制, 在指定范围内做自注意力, 增强了网络特征提取的能力, 并且利用时序信息将过去时刻的 BEV 特征与当前时刻对齐, 再利用可变形注意力提取 BEV 特征, 大大提升了对遮挡物体的检测能力。

3.1.1.8 Occupancy 感知算法

a. MonoScene

MonoScene 算法提出了一种从单目图像直接预测三维语义栅格和语义补全的框架。它可以从单目 RGB 图像推理得到场景的稠密几何语义信息。MonoScene 算法首先经过 2D UNet 网络提取不同分辨率下的图像特征，然后在 FLoSP 模块将 2D 特征转换成 3D 特征，并通过 3D UNet 加强深度维度之间的联系，并在其中加入 3D CRP 模块以提升空间语义的响应能力，最后完成 3D 语义的输出。如图 3-22 所示为 MonoScene 算法的网络结构图。

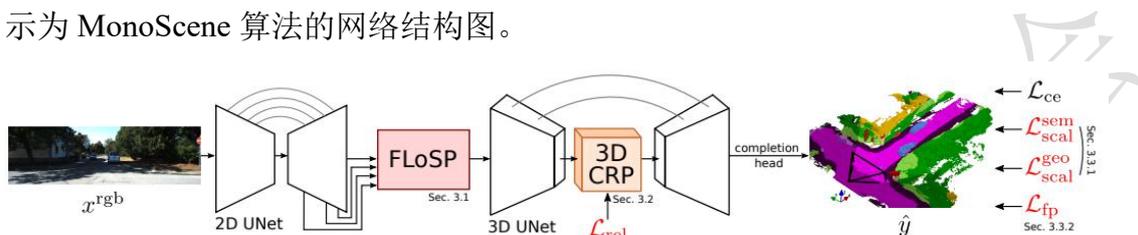


图 3-22 MonoScene 算法网络结构图

MonoScene 算法测试数据集：NYUv2 和 SemanticKITTI 数据集。其中 NYU v2 使用 795 张作为训练集，654 张作为测试集；SemanticKITTI 使用 3834 张作为训练集和 815 张作为验证。

优点：MonoScene 算法不依赖于深度信息或雷达等传感器，而是只输入单帧图像，实现了在从单目图像直接预测三维语义分割的任务。MonoScene 通过将多尺度的 2D 特征投影到给定的 3D 特征图上，能够获得更好地 3D-2D 之间的联系。

b. TPVFormer

TPVFormer 算法将六张环视图像作为输入，通过 Image Backbone 得到不同尺度的特征层 Feature Maps，再通过 TPVFormer 模块可以获得 TPV（Top-View、Pillar-View、Front-View）特征，并使用注意力机制实现 TPV 特征之间的交互，并最终在三个 TPV 特征上应用检测头如图 3-23 所示为 TPVFormer 算法的网络结构图。

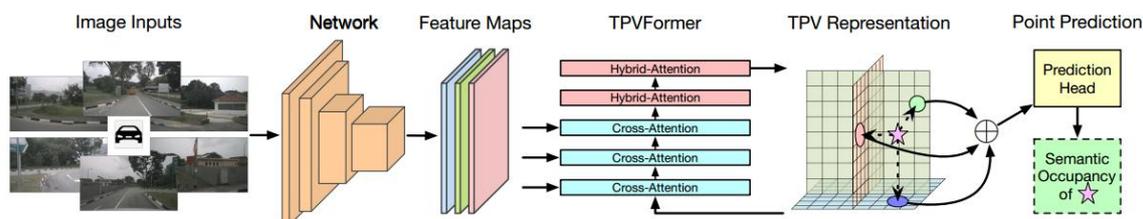


图 3-23 TPVFormer 算法网络结构图

MonoScene 算法测试数据集：Panoptic nuScenes 和 Semantic KITTI 数据集。

优点：TPVFormer 算法提出了一种 TPV 表示，通过在 BEV 和两个额外的垂直平面对每个点的投影特征交互，使得 TPV 特征可以更全面、准确地捕捉环境中的物体信息。且 TPVFormer 算法仅使用图像作为输入，达到了与雷达方法相当的语义分割效果。

3.1.1.9 图像感知算法总结

本节调研了图像感知算法，包括图像目标检测算法、图像目标跟踪算法、图像语义分割算法、图像车道线检测算法、图像车位线检测算法以及图像深度估计、BEV 和 Occupancy 感知算法。常用数据集包括 PASCAL VOC 数据集、MOT 数据集、COCO 数据集、CULane 数据集和 Tusimple 数据集、ps2.0 数据集、KITTI 数据集、nuScenes 数据集等，具体如下表 3-1 所示：

表 3-1 图像感知算法总结

算法类型	算法名称	测试数据集
图像目标检测算法	Fast R-CNN	PASCAL VOC07
	SSD 算法	MS-COCO 目标检测数据集
	YOLO 系列检测算法	PASCAL VOC 检测数据集
	FCOS 算法	PASCAL VOC 检测数据集
	CenterNet	MS-COCO 验证数据集
图像目标跟踪算法	SORT 算法	MOT benchmark 数据集
	DeepSORT 算法	MOT challenge 数据集
图像语义分割算法	HRNet 网络	COCO 关键点检测数据集
	HRNet+OCR	PASCAL Context and Cityscapes 数据集、LIP 人类解析数据集
图像车道线检测算法	SCNN	PASCAL Cityscapes 数据集
	RESA	CULane 数据集和 Tusimple 数据集
	LineCNN	MIKKI2 数据集、TuSimple 数据集
	LaneATT	TuSimple 数据集、CULane 数据集 LLAMAS 数据集
	CondLaneNet	CurveLanes 数据集、CULane 数据集 TuSimple 数据集
	UFAST	CULane 数据集、TuSimple 数据集
	PolyLaneNet	TuSimple 数据集、LLAMAS 数据集 ELAS 数据集
	BezierLaneNet	CULane 数据集、LLAMAS 数据集 TuSimple 数据集
	FOLOLane	CULane 数据集、TuSimple 数据集

	GANet	Scene Flow 数据集、KITTI 数据集
图像车位线检测算法	DeepPS	ps2.0 数据集
	DMPR-PS	ps2.0 基准数据集
	上下文两阶段方法	PIL_PARK 数据集
	注意力图神经网络检测方法	ps2.0 基准数据集
图像深度估计算法	David 等	NYU Depth、KITTI
	Eigen 等	NYU Depth、Sift Flow 和 PASCAL VOC
	Zhou 等	KITTI、PASCAL Cityscapes
	Clement 等	KITTI 数据集、Cityscapes 数据集
BEV 感知算法	BEVDet	nuScenes 数据集
	BEVFormer	nuScenes 数据集
Occupancy 感知算法	MonoScene	
	TPVFormer	Panoptic nuScenes 和 Semantic KITTI 数据集

3.1.2 激光点云算法

3.1.2.1 点云目标检测算法

1) LaserNet^[41]

Range View 数据是指将激光雷达水平和垂直方向角度值作为 X-Y 坐标的数据，激光雷达的点云来自于多条激光扫描线。比如说 64 线的激光雷达，那么在垂直方向（Inclination）上就有 64 个离散的角度。激光雷达在 FOV 内扫描一遍，会有多个水平方向（Azimuth）的角度。比如说水平分辨率是 0.1° ，那么扫描 360° 就会产生 3600 个离散的角度。这里也可以粗略把 Inclination 和 Azimuth 理解为地球上的纬度和经度。把水平和垂直方向的角度值作为 X-Y 坐标，就可以得到一个二维图像。图像中的像素值是相应角度下的反射点的特性，比如距离，反射强度等。这些特性可以作为图像的 channel，类似于可见光图像中的 RGB。相比于 Point View 或 BEV 而言 Range image 具有紧凑性，因为这两种表示方法的稀疏性非常明显，而 Range View 则没有稀疏性。LaserNet 是典型的基于 Range View 的检测算法。

LaserNet 基于 Lidar 的 range view 数据处理，使用全卷积网络来预测每个点的 3D box 的多模态分布，然后有效地融合这些分布以生成每个目标的预测，如图 3-24 为 LaserNet 模型结构

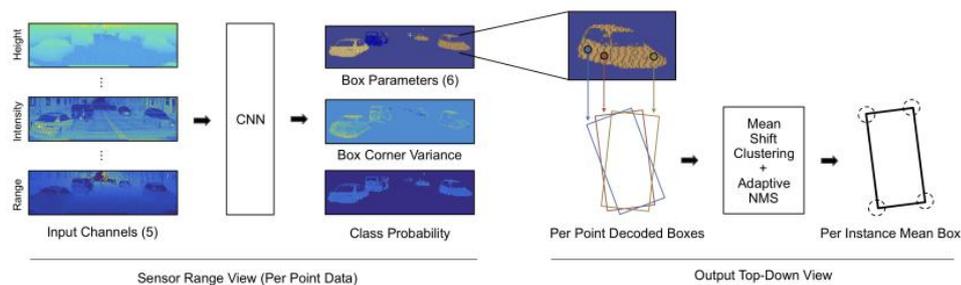


图 3-24 LaserNet 模型结构

LaserNet 用点的 laser id 作为 range view 图像中的纵坐标。离散化 azimuth 作为横坐标。lasernet 的输入是五通道的数据，range, height, azimuth, intensity, flag(表示图像的这个位置是否有有点)，当多个点落在一个位置，保留最近的点。模型结构如上图所示，可以看作是 anchor-free 结构，每个点输出结果然后做 mean shift + NMS (同一个物体上的所有点应该预测一个相近的分布，因为有噪声的存在不可能完全一样，可以通过 mean shift 聚类降低这个噪声，分别在每个类别和混合模型的每个组件上进行 mean shift 聚类)。

LaserNet 测试数据集：KITTI 数据集，使用训练集中的 5985 个样本，并保留其余样本以进行验证。

优点：实验表明，将每个检测建模为一个分布，而不是一个单一的确框，可以获得更好的整体检测性能。基准测试结果表明，与其它方法相比，该方法的运行时间明显较低。

2) PIXOR^[42]

PIXOR 解决了在自动驾驶环境下从点云实时检测三维物体的问题，从鸟瞰图 (BEV) 中表示场景来更有效地利用 3D 数据，是一种无需处理的、单级的检测器，输出从像素级别的神经网络定向解码估计 3D 对象，如图 3-25。模型的输入形式、网络架构和优化器是为了平衡高精度和实时效率而特别设计的。

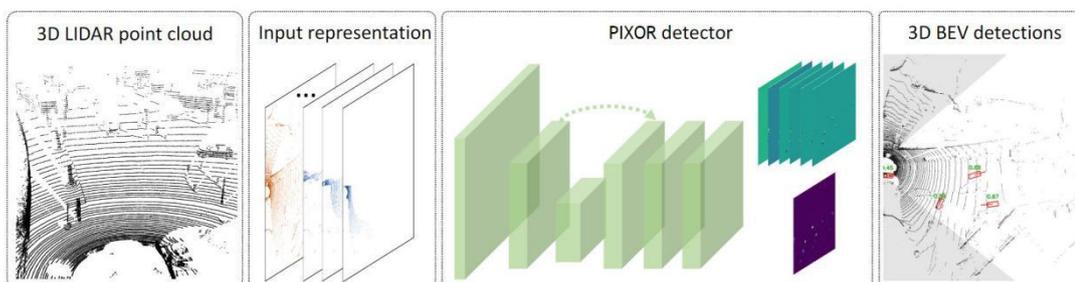


图 3-25 PIXOR 模型结构

PIXOR 使用了一种全卷积神经网络，该网络在单阶段输出像素预测，每个预测对应于 3D 物体估计。由于网络结构中无超参数，这些设计使得 PIXOR 非常简单，具有很好的泛化性。具体地说，不需要预先设计 anchor，也不需要调整从第一阶段传递到第二阶段的采样的 proposal 的数量以及相应的 NMS 的 IoU 阈值。

PIXOR 测试数据集: KITTI-BEV 目标检测数据集和大规模 3D 车辆检测基准数据集。大规模 3D 车辆检测基准数据集为作者收集，名为 TOR4D 的大型 3D 车辆检测数据集，该数据集具有与 KITTI 不同的传感器配置，并且是在北美城市收集的。检测区域为本车前后 100m，左右 40m，总共收集了 6500 个序列，分为 5000/500/1000 作为训练/验证/测试。训练序列以 10 Hz 的频率采样到帧中，而验证和测试序列以 0.5 Hz 的频率采样。结果，训练集中有超过 120 万帧，验证集和测试集中有 5969 和 11969 帧。所有车辆都标有鸟瞰图边界框。

优点：两个数据集表明，所提出的检测器在平均精度（AP）方面明显优于其他最先进的办法，而且速度上仍高于 28fps。

3) Pointpillars^[43]

它利用 PointNets 来学习组织在垂直列中的点云表示。包括 3 个主要阶段：将点云转换为稀疏伪图像的特征编码器网络；2D 卷积主干网络，将伪图像处理为高级表示；检测并会回归 3D box 的检测头。首先，采用 Pillar 编码方式编码 PointCloud：在点云的俯视图的平面进行投影使之变成伪 2D 图，对这种投影进行编码用的是 Pillar 方法，即在投影幕上划分为 $H * W$ 的网格，然后对于每个网格所对应的柱子中的每一个点取原特征 $(x, y, z, r, x_c, y_c, z_c, x_p, y_p)$ 共 9 个，再然后每个柱子中点

多于 N 的进行采样，少于 N 的进行填充 0，形成了 $(9, N, H * W)$ 的特征图。其次，使用 2D Convolution 进行处理：这一部分算得上是网络中的 backbone，backbone 包含两个子网络一个是 top-down 网络，另一个是 second 网络。其中 top-down 网络结构为了捕获不同尺度下的特征信息，主要是由卷积层、归一化、非线性层构成的，second 网络用于将不同尺度特征信息融合，主要由反卷积来实现。最后，使用 SSD 的检测头对目标进行检测。

数据集：所有实验都使用 KITTI 目标检测基准数据集。该数据集由具有激光雷达点云和图像的样本组成。作者只在激光雷达点云上训练，但与使用激光雷达和图像的融合方法进行比较。样本最初分为 7481 个训练样本和 7518 个测试样本。对于实验研究，作者将官方训练分为 3712 个训练样本和 3769 个验证样本，而对于测试提交，作者从验证集中创建了一个包含 784 个样本的最小集，并在剩余的 6733 个样本上进行了训练。

优点：广泛的实验表明，PointPillars 在速度和精度方面都大大优于以前的编码器。尽管仅使用激光雷达，PointPillars 在 3D 和鸟瞰图 KITTI 基准测试方面显著优于现有技术。

4) PointRCNN^[44]

PointRCNN 是基于 PointNet++ 的两阶段目标检测模型，第一阶段采用 PointNet++ 将点云分类为前景点和背景点，并为每个前景点生成一个 3D 包围盒，第二阶段对第一阶段的包围盒进行非极大抑制 (Non-Maximum Suppression, NMS)，并细化调整包围盒参数。

PointRCNN 实验数据集：KITTI 目标检测数据集。按照常用的训练/验证拆分方法，将训练样本分为训练集 (3712 个样本) 和验证 (3769 个样本)。

优点：能够保留点云的细节特征

缺点：由于自动驾驶场景中点云数量可以达到数十万，基于点的编码方式实时性较差，多层感知机的空间编码能力也不如卷积神经网络

5) YOLO3D^[45]

YOLO3D 算法是基于 YOLOv2 进行改进基础上，将 3D 点云在二维透视图像空间进行映射，并在在二维透视图像空间的一次回归元结构成功的基础上，对其进行

扩展，从 LiDAR 点云生成面向三维对象的 bounding box。主要贡献是将 YOLOv2 的损失函数扩展为包含偏航角、笛卡尔坐标系中的 3Dbox 中心和 box 高度的直接回归问题。

YOLO3D 测试数据集：KITTI 基准数据集。

优点：架构可以仅从两个通道的鸟瞰图输入中检测所有对象，并且仅使用一个网络，实现了 40 fps 的实时性能，以及中型汽车在 0.5 IOU 阈值上的 75.3% mAP。

6) VoxelNet^[46]

VoxelNet 是最早提出基于体素进行点云特征编码的模型，首先在三维空间划分体素，对每个体素内的点云，首先进行随机降采样，再采用多层 PointNet 结构作为体素特征编码器(Voxel Feature Encoding, VFE)，进行体素特征编码。之后在通过三维卷积主干提取多尺度空间特征，通过区域提议网络 (Region Proposal Network, RPN) 回归包围盒类别、位置和尺寸。

由于点云的稀疏性，划分的大多数体素为空，对此 VoxelNet 采取的算法策略是将非空体素收集为密集张量，再进行体素特征编码，之后再根据存储的体素坐标将体素特征重构为稀疏张量，如图 3-26 所示。

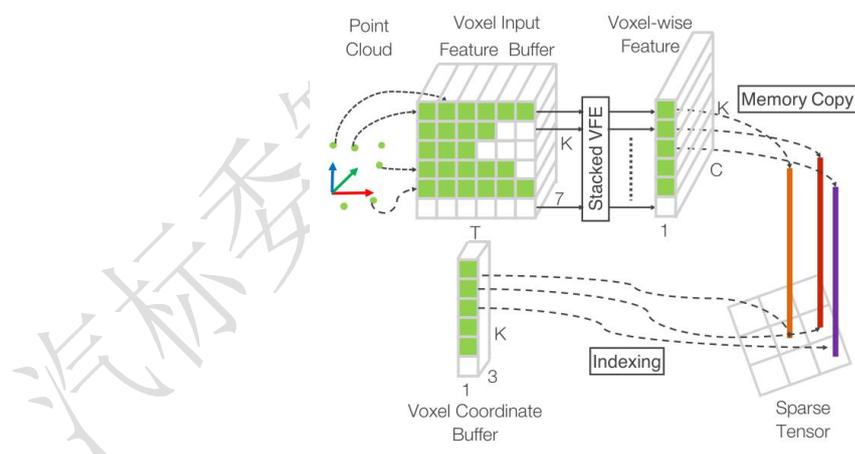


图 3-26 VoxelNet 体素编码结构示意图

VoxelNet 实验数据集：KITTI 3D 目标检测。该数据集包含 7481 帧训练点云和 7518 帧测试点云，涵盖三个类别：汽车、行人和自行车。对于每一类的检测结果，基于简单、中等和困难三个难度级别进行评估，这三个难度级别是根据物体大小、遮挡状态和截断状态确定的。

优点：结构较为简单，扩展性强，推理速度快，在准确性和效率之间取得了很好的平衡，被业界广泛应用。

7) PV-RCNN^[47]

PV-RCNN 结合了基于点和基于体素的特征编码方法的优点，创新地提出了体素到关键点场景编码方案、栅格点的多尺度 RoI 特征提取层，如图 3-27 所示。

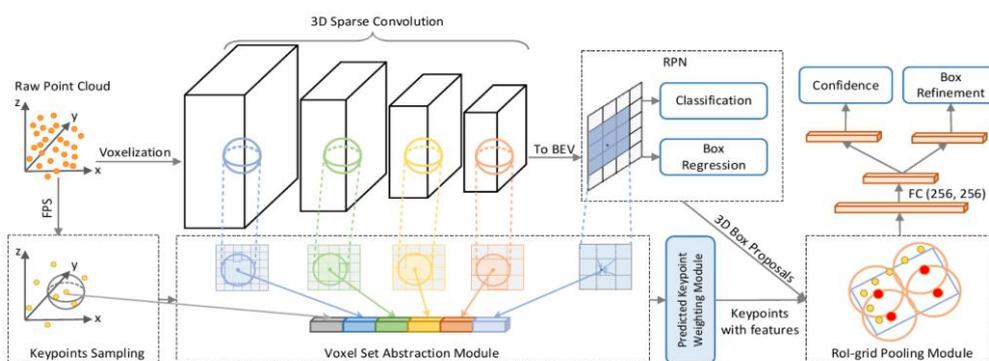


图 3-27 PVRCNN 结构示意图

PV-RCNN 实验数据集：KITTI、Waymo。

优点：基于体素的主干可以高效地编码多尺度特征，并生成高质量的 3D 提议，基于 PointNet 的点集提取模块能以灵活的感受野保留准确的位置信息。

8) CenterPoint^[48]

CenterPoint 提出将 3D 目标作为点来表示和检测，首先使用关键点检测器检测物体的中心，然后回归到其他属性，包括 3D 大小、3D 方向和速度。在第二阶段，它使用对象上的附加点特征来改进这些估计。

CenterPoint 实验数据集：Waymo、nuScenes。Waymo 数据集的点云包含 64 线激光雷达，每 0.1s 生成 180k 个点。nuScenes 包含 1000 个序列，分别有 700、150、150 个序列用于训练、验证和测试。每个序列大约 20 秒长，激光雷达频率为 20 帧/秒。数据集为每帧激光雷达提供车辆姿态信息，但只提供每 10 帧的包围盒注释。nuScenes 使用 32 线激光雷达，每帧产生大约 30k 点。总共有 28k、6k、6k 帧点云分别用于训练、验证和测试。这些标注包括 10 个具有长尾分布的类。

三维目标检测结果通常表示为 3D 包围盒，然而三维世界中的目标不遵循任何特定的方向，基于包围盒的检测器很难枚举所有的方向。

3.1.2.2 点云目标跟踪算法

目标跟踪就是在连续的数据序列中，建立所需要跟踪目标的位置关系，得到目标完整的运动轨迹。通常是给定上一帧的目标位置特征，在下一帧中对该目标的位置以及边界框的尺寸大小进行预测。

目标跟踪的常用算法有很多，主要有两种，分别为生成式模型：建立目标模型或者提取目标特征，利用外观表征在后续帧中进行相似性搜索，最小化跟踪目标和候选目标之间的误差进行定位。判别式模型：通过对比目标模型和背景信息的差异，将目标模型提取出来，从而得到当前帧中的目标位置。传统的跟踪算法大多属于生成式模型，而深度学习的算法大多属于判别式模型，几种经典的算法如下：

1) 基于均值漂移的跟踪：均值漂移是基于概率密度分布，使目标的搜索一直沿着概率梯度上升的方向，迭代收敛到概率分布的峰值上。

2) 基于卡尔曼滤波的跟踪：通过检测或者其他方式得到目标的观测位置，并根据运动方程从上一时刻的位置信息预测当前时刻的位置信息，将观测结果和预测结果进行加权平均得到最终的目标位置信息。

3) 基于粒子滤波的跟踪：对目标基于粒子分布进行建模，定义一种相似度量来确定与目标之间的匹配程度。然后按照高斯分布撒粒子，统计其相似度来确定目标可能的位置信息。

4) 基于核相关滤波的跟踪：通过设计一个滤波模板，利用该模板与目标候选区域做相关运算，最大输出响应的位置即为当前帧的位置信息。

5) 基于深度学习的跟踪：利用卷积神经网络将目标从背景中提取出来，利用其对目标强大的特征抽象能力以及对目标运动过程的拟合能力实现跟踪任务。

以下介绍几个基于深度学习的三维目标跟踪网络：

1) PointTrackNet^[49]

PointTrackNet 是一个端到端的目标检测和多目标追踪的深度学习模型，将前后两帧原始点云作为输入，直接实现三维目标框的回归以及目标关联，使用 PointNet++ 网络的结构作为主干网络，使用逐点关联的方法将三维目标框特征进行匹配，最终完成多目标追踪，其流程如图 3-28 所示。

PointTrackNet 实验数据集：KITTI 三维目标跟踪数据集。该数据集由 21 个训练和验证片段和 29 个需要在线评估的测试片段组成。训练和验证集共有 7987 帧点云、636 段车辆轨迹和 30601 个车辆目标。将训练和验证集分成两部分，Seq-0000 到 Seq-0015 进行训练，Seq-0016 到 Seq-0020 进行验证。

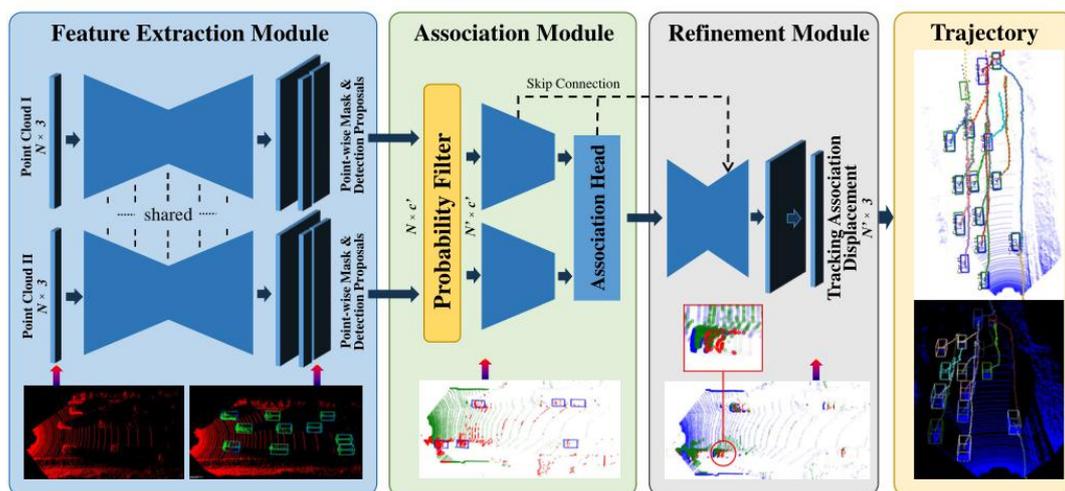


图 3-28 PointTrackNet 结构示意图

优点：在不规则和快速变化的情况下具有比较好的结果。

2) JRMOT^[50]

JRMOT 融合了 2D 和 3D 探测器和描述符特征，利用多模态递归卡尔曼模型，将它们集成到一个联合概率数据关联框架中架构实现在线的、实时三维多物体跟踪。如图 3-29 所示，该三维多物体跟踪系统由检测和跟踪(数据关联和滤波)组件组成；T, D, F 分别表示轨迹，检测，特征，上标表示空间；系统集成来自二维 RGB 图像和三维点云融合成一个具有实时性能的三维多目标递推估计跟踪器。

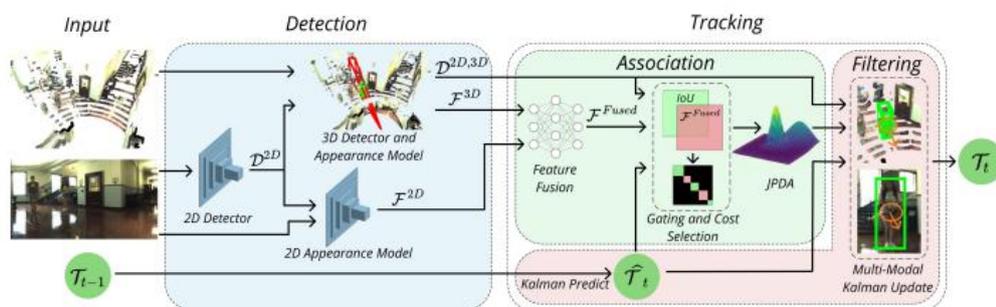


图 3-29 JRMOT 结构示意图

JRMOT 测试验证数据集：KITTI 数据集和自制的 JRDB 数据集。JRDB 数据集是一个新的大规模 2D+3D 数据集和基准标注超过 200 万个检测框和 3500 个时间一致的 2D+3D 轨迹，跨越 54 个室内和室外的场景。

优点：MOTA 较高，计算时间短，性能增益是融合并充分利用 2D RGB 和 3D 点云信息中的互补信息的结果。结合 2D 和 3D 信息的 JRMOT 方法的好处是，对于距离传感器相对较远的轨迹来说更加清晰，其中 3D 点云数据是稀疏的，但 2D RGB 是一个丰富的信息源。

3) AB3DMOT^[51]

AB3DMOT 是 3D 跟踪的经典算法。与 Sort 类似，AB3DMOT 同样简单组合了两个组件——匈牙利算法和卡尔曼滤波算法。将以二维边界框表示的目标拓展成以三维边界框表示，然后以目标三维边界框之间的交并比 3DIoU 为度量指标，使用匈牙利算法匹配相邻两帧之间的目标。

AB3DMOT 测试验证数据集：KITTI 三维 MOT 数据集。该数据集提供 LiDAR 点云和 ground truth 三维边界盒轨迹。由于 KITTI 测试集只支持二维 MOT 评估，而且其 ground truth 未向用户公布，因此作者使用 KITTI 数据集进行三维 MOT 评估。

优点：整个算法的组件都相当简单，使用最简单的组合，达到较高的计算效率，又不会有较大的精度损失。

3.1.2.3 点云语义分割算法

1) 传统语义分割算法

传统的点云语义分割根据其采用的数学方法，可大致分为以下四类：

边缘检测法：在图像分割领域，边缘检测算法已经非常成熟，因此这种方法也被用于三维点云语义分割任务。早在 1999 年，Jiang 等人提出了一种基于扫描线逼近的边缘检测器算法，通过边缘检测对表面进行快速分割；2006 年 Rabbani 等人在此理论之上通过法线和曲率等信息检测边缘信息，从而进行快速分割获得分割结果。然而这种方法在很大程度上受边缘检测器的限制，且从雷达扫描数据到点云数据转换的过程中会不可避免的丢失一部分信息，所以分割精度有一定限制。

曲面生长法：此算法先检测种子平面或曲面，再通过相似性度量对各种子曲面

周围的离散点进行分组，扩大种子所在的面从而分割区域。Vosselman 于 2006 年提出了一种平面（曲面）生长算法，方法分为两步：确定种子曲面和生长种子曲面。同年 Rabbani 等人提出了一个利用平滑约束对点云模型进行分割的算法，该算法利用曲面法线和点的连通性，再结合 KNN 算法或固定距离的相邻点进行快速分割。但该方法的局限性在于种子的好坏难以判断，且不同的种子也会导致分割结果不同。

聚类方法：此方法的原理是通过在特征空间中聚类点的方法进行区域分割，常用的聚类的方法有 K-means 聚类，最大似然法和模糊估计法。Chehata 等人使用 K-means 聚类算法分割点云数据，创新性的提出分层分裂聚类方法来分离地面与建筑物。聚类方法是根据点在特征空间中的特征向量进行计算，因此该方法具有一定的健壮性。

基于图分割的方法：该方法基于同一分割区域中的点比不同分割区域中的点连接更为紧密这一理论基础，通过保证不同分割面之间的点相似性最小来分割区域。200 年，Bucksch 等人将八叉树和图分割理论结合起来对点云数据进行分割。此外，2009 年 Golovinskiy 等人提出了一个基于最小割的点云分割算法，该算法通过建立 K-近邻图，将背景与物体分割出来。

综上所述，传统的点云语义分割算法依赖于数学算法和优化方法，在日益庞大的数据集上的分割效率也越来越低，因此需要更快速和更精确的分割方法，但其中对点云的一些处理思想值得借鉴。

2) 基于深度学习的点云语义分割算法

a. RandLA-Net^[52]

RandLA-Net 是一个轻量级的高效大规模 3D 点云语义分割网络，该网络可以直接推断大规模点云的每个点的语义信息，其关键思想是使用随机点采样而不是其他更复杂的采样方法，然后引入一种新颖的局部特征聚合模块，以逐步增加每个特征的感受野，从而有效地保留更多几何细节。RandLA-Net 能够保证计算和存储效率的同时，克服采样的随机性带来的关键点丢失的问题。

RandLA-Net 测试验证数据集：Semantic3D 数据集和 SemanticKITTI 数据集。

Semantic3D 数据集由 15 个用于训练的点云和 15 个用于在线测试的点云组成。每个点云都有多达 108 个点，在现实世界的 3D 空间中覆盖了多达 160×240×30 米。

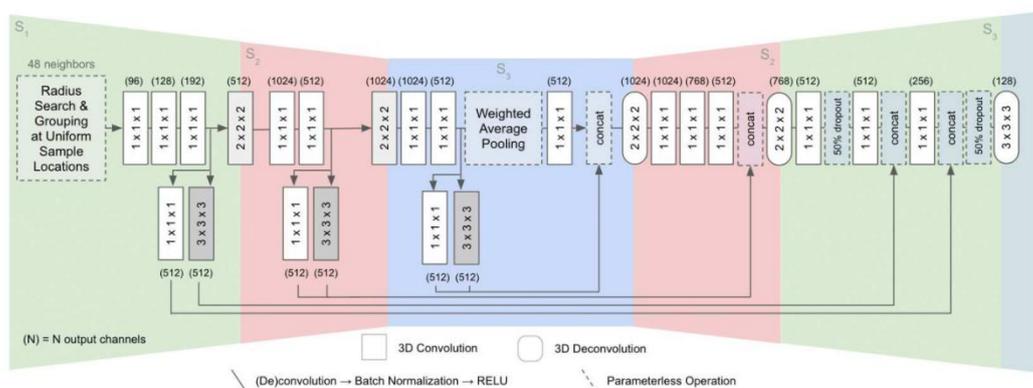
原始 3D 点分为 8 类，包含 3D 坐标、RGB 信息和强度。

SemanticKITTI 由 43552 个密集注释的激光雷达扫描组成，共 21 个序列。每帧点云有多达 105 个点，空间跨度有 160×160×20 米。

优点：用随机抽样来显著减少内存占用和计算成本，在 RandLA-Net 中引入局部特征聚合模块，有效地在大范围内保留有效的特征。实验结果表明，RandLA-Net 能够高效分割点云，证明了使用轻量级的网络架构有效地分割大规模点云的可行性。

b. FCPN^[53]

FCPN 在三维空间进行均匀采样，每个采样的位置收集邻域内固定数量的点用来提取点特征。这里与 VoxelNet 中的方式略有不同：VoxelNet 中的网格也是均匀位置采样，邻域的大小是固定的，但点的数量则是不固定的。所以，VoxelNet 中的网格数据是稀疏的，而 FCPN 中的网格数据则是稠密的。但是，无论哪种方式，最后得到的数据形式都是 4D 张量，这种数据形式可以采用 3D 卷积来处理，并通过一个 U-Net 的结构来提取不同尺度的信息，最终得到原始量化尺度下的分割结果，如图 3-30 所示。每个网格会被分配一个类别标签，网格中所有的点共享此标签。



图

图 3-30 FCPN 结构示意图

FCPN 实验数据集：室外的 Semantic3D，SemanticKITTI 和室内的 S3DIS。Semantic3D 数据集由 15 个用于训练的点云和 15 个用于在线测试的点云组成。每个点云都有多达 108 个点，在现实世界的 3D 空间中覆盖了多达 160×240×30 米。原始 3D 点分为 8 类，包含 3D 坐标、RGB 信息和强度。SemanticKITTI 由 43552 个密集注释的激光雷达扫描组成，共 21 个序列。每帧点云有多达 10⁵ 个点，空间跨度有

160×160×20 米。按照惯例，00~07 和 09~10(19130 帧)序列用于训练，08(4071 帧)序列用于验证，11~21(20351 帧)序列用于在线测试。原始的 3D 点只有 3D 坐标，没有颜色信息。S3DIS 数据集由 6 个大区域中的 271 个房间组成。每个点云都是一个中型的单间(20×15×5 米)，有密集的 3D 点。

优点：数据规整，可以采用标准的卷积操作来处理，多尺度的信息也可以很容易的通过特征金字塔结构来获取。

缺点：对体素大小的设置比较敏感，存在数据缺失问题。

3.1.2.4 点云车道线检测算法

1) 三维转二维

优点：降低了三维数据的复杂度，能利用已有的图像处理技术

缺点：在将三维点云转换为二维强度图像的过程中会造成原有三维信息的丢失，改变原始数据的精度，同时也会给车道线检测技术增加额外的运算时间

2) 直接处理三维

Jiyoung Jun^[54]等提出了一种基于 32 线激光雷达的期望最大化方法检测车道线，并随着配备激光雷达传感器的自动驾驶车辆的前进而实时更新三维线参数。他们首先通过设置垂直坡度阈值区分可行驶区域，再利用高反射强度的特征筛选车道线点，最后通过以固定的车道宽度搜索一组平行线来确定最终的车道线数据。

评估数据集：自采数据集。

通过 32 通道 3D 激光雷达(Velodyne HDR-32E)的车辆采集，Velodyne HDR-32E 有 32 个激光扫描器，水平视场为 360°垂直视场为 40°和角度分辨率为 1.33°，旋转速率为 10hz，每帧点云大约 69000 个点。数据采集车辆以低于 60 公里/小时的速度沿着韩国城南市和仁川市的两条不同路线行驶，路程分别为 7 公里和 2 公里，包括了路口、坡道、隧道、人行横道和地面上的各种道路标志。

优点：利用 3D 点云数据实现同时检测到多个车道，包括车辆当前所在的车道，通过累积检测到的车道来生成车道级数字地图。

3.1.2.5 激光点云算法总结

本节调研了激光点云感知算法，包括点云目标检测算法、点云目标跟踪算法、点云语义分割算法和点云车道线检测算法。常用数据集包括 KITTI 数据集、Waymo 数据集、nuScenes 数据集、Semantic3D 数据集等，具体如表 3-2 所示：

表 3-2 激光点云感知算法总结

算法类型	算法名称	测试数据集
点云目标检测算法	LaserNet	KITTI 数据集
	PIXOR	KITTI-BEV 目标检测数据集
	PointPillars	KITTI 目标检测基准数据集
	PointRCNN	KITTI 目标检测数据集
	YOLO3D	KITTI 基准数据集
	VoxelNet	KITTI 3D 目标检测
	PV-RCNN	KITTI 目标检测数据集 Waymo 目标检测数据集
	CenterPoint	Waymo 目标检测数据集 nuScenes 目标检测数据集
点云目标跟踪算法	PointTrackNet	KITTI 三维目标跟踪数据集
	JRMOT	KITTI 三维目标跟踪数据集 自制的 JRDB 数据集
	AB3DMOT	KITTI 三维 MOT 数据集
点云语义分割算法	RandLA-Net	Semantic3D 数据集 SemanticKITTI 数据集
	FCPN	室外的 Semantic3D 数据集 SemanticKITTI 数据集 室内的 S3DIS 数据集
点云车道线检测算法	Jiyoung Jun 等	主要为自采数据集

3.1.3 融合算法

3.1.3.1 图像点云融合检测算法

1) MV3D

MV3D 算法^[55]来自 CVPR 2017 论文 “Multi-View 3D Object Detection Network for Autonomous Driving”。该算法利用的是点云的俯视图和前视图，这样能够减少计算量，又不至于丧失过多信息，再将两者与图像进行融合最终完成目标检测。算法由两部分组成：3D Proposal Network 和 Region-based Fusion Network，如图 3-31 所

示。

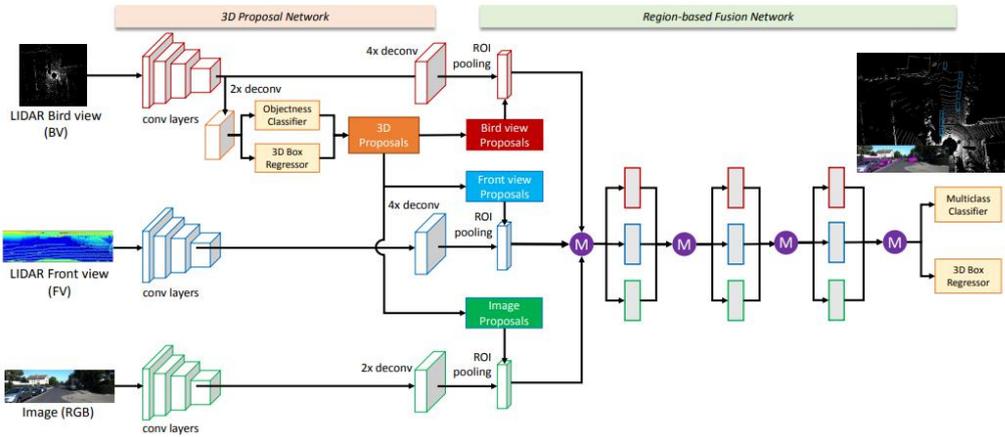


图 3-31 MV3D 流程图

MV3D 测试数据集：KITTI 数据集。KITTI 的三维目标检测数据集包含 7481 张训练图片，7518 张测试图片，以及相应的点云数据。

2) Frustum-PointNet

Frustum-PointNet 算法^[56]由 Charles R. Qi 提出，该算法舍弃了通过投影将三维点云转化为图像或通过量化转化为体积网格，直接对原始点云进行操作。为了有效地运用到大规模检测中，首先利用 2D 目标检测到物体的二维边界框来提取物体的三维边界视锥体，以此来减少在三维点云目标检测中的搜索空间。该算法由三部分组成：Frustum Proposal、3D Instance Segmentation、Amodal 3D Box Estimation，如图 3-32 所示。

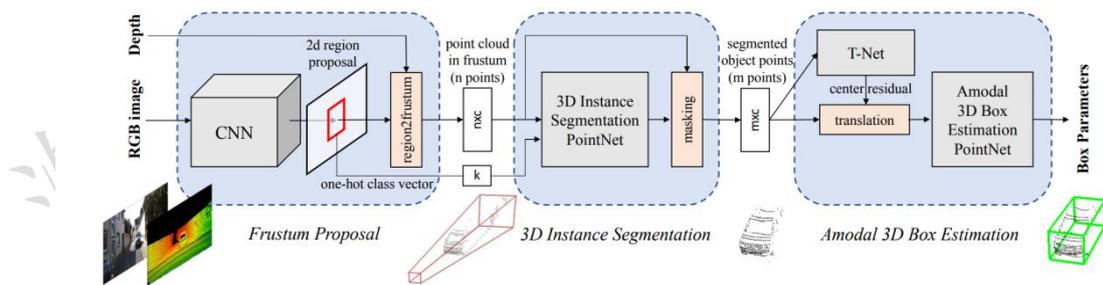


图 3-32 Frustum-PointNet 流程图

该算法在实验中利用了 KITTI 和 SUN-RGBD 数据集进行三维目标检测。KITTI 的三维目标检测数据集包含 7481 张训练图片，7518 张测试图片，以及相应的点云数据，共包含 80256 个标注目标；SUN RGB-D 数据集包含 10335 张不同场景的室

内图片，训练和测试集分别包含 5285 和 5050 个图像，共有 146617 个 2D 多边形标注和 58657 个 3D 边框。

3) PointFusion

PointFusion 算法^[57]来自 CVPR 2018 论文 “PointFusion: Deep Sensor Fusion for 3D Bounding Box Estimation”。该算法由三部分组成：提取点云特征的 PointNet 网络、提取图像特征的 CNN 网络以及一个将两者结合的融合网络，如图 3-33 所示。通过 CNN 和 PointNet 分别对图像与点云数据进行处理提取特征，再通过融合网络将结果进行融合，最终输出一个三维边界框，完成目标检测。

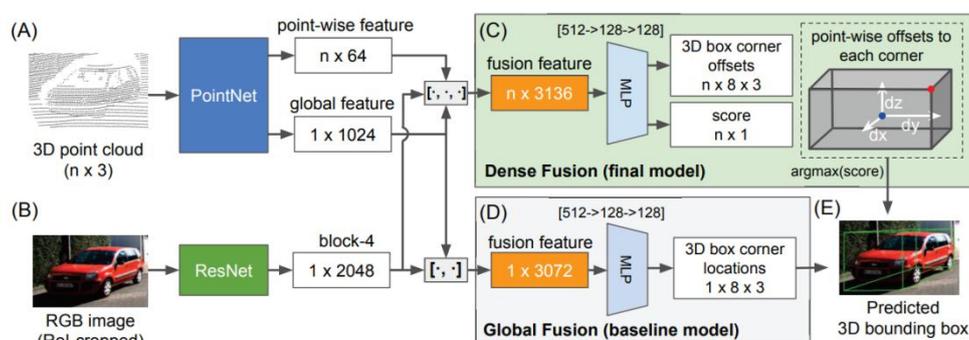


图 3-33 PointFusion 流程图

在算法的实验中同样也是利用了 KITTI 和 SUN-RGBD 数据集。KITTI 的三维目标检测数据集包含 7481 张训练图片，7518 张测试图片，以及相应的点云数据，共包含 80256 个标注目标；SUN RGB-D 数据集包含 10335 张不同场景的室内图片，训练和测试集分别包含 5285 和 5050 个图像，共有 146617 个 2D 多边形标注和 58657 个 3D 边框。

3.1.3.2 图像点云融合跟踪算法

1) DSM

该算法^[58]来自 ICRA2018 论文“End-to-end Learning of Multi-sensor 3D Tracking by Detection”，提出了一种端到端的目标跟踪算法，同时利用图像和点云，通过深度结构化模型（DSM）对连续帧的目标进行匹配，完成目标的追踪，该算法由 Detection Net、Matching Net、Scoring Net 以及 Linear Program 组成，如图 3-34 所示。

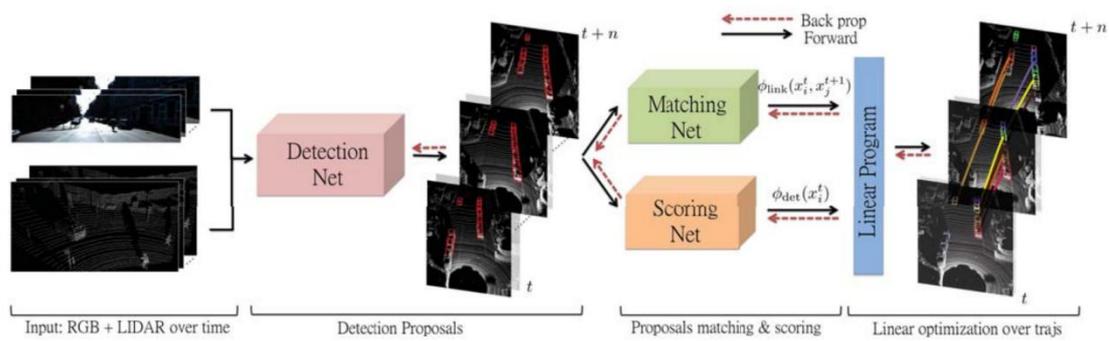


图 3-34 DSM 流程图

该算法的实验中使用了 KITTI 的追踪数据集，由 50 个序列（21 个用于训练，29 个用于测试）组成，训练集有 8026 帧。

2) mmMOT

mmMOT 算法^[59]来自 CVPR2019 论文“Robust Multi-Modality Multi-Object Tracking”，该算法提出了一种具有鲁棒性的多模态 MOT 框架，单独地将点云和图像的特征进行提取，包含了 object detection, feature extraction, adjacency estimation 以及 min-cost flow optimization，如图 3-35 所示。

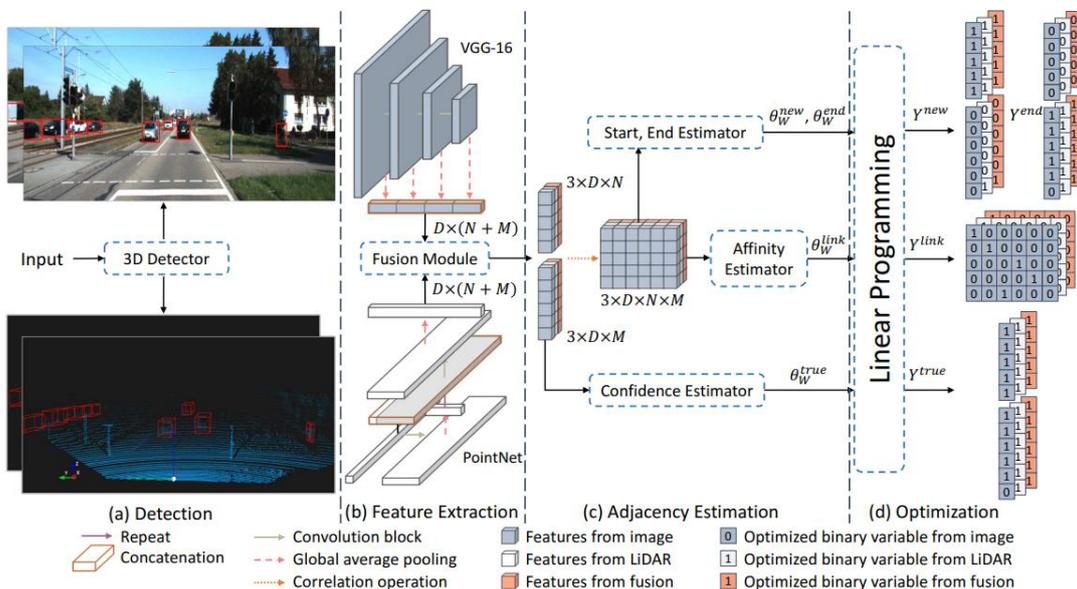


图 3.35 mmMOT 流程图

在该算法的实验中使用了 KITTI 追踪数据集，该数据集包含 21 个训练序列和 29 个测试序列，分别选择 10 个训练序列和 11 个测试序列，共包含 3975 帧训练集和 3945 帧测试集。

3) EagerMOT

EagerMOT 算法^[60]来自 ICRA2021 论文“EagerMOT: 3D Multi-Object Tracking via Sensor Fusion”，该算法首先利用现用的目标检测算法获取点云和图像的边界框，再经过融合模块、匹配模块、追踪周期管理，如图 3-36 所示。

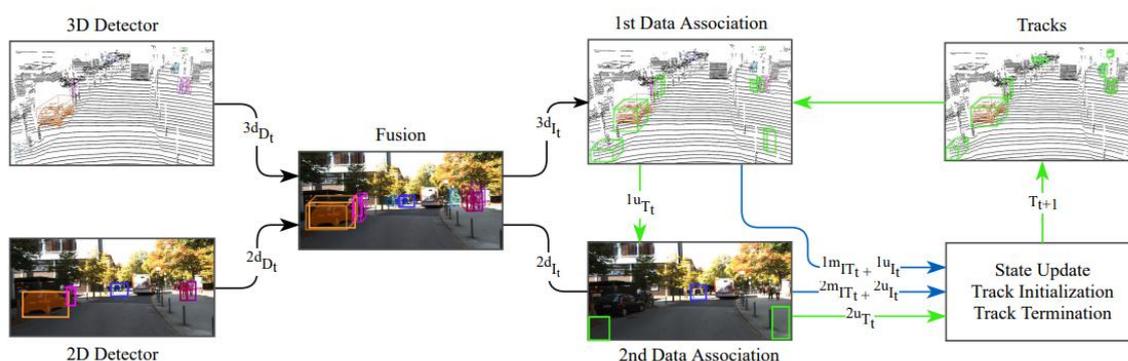


图 3-36 EagerMOT 流程图

该算法的实验中使用了 KITTI 的追踪数据集以及 NuScenes 数据集，NuScenes 数据集有 1000 个场景组成，850 个场景用于训练，150 个场景用于测试，每个场景有 40 帧组成。

3.1.3.3 融合算法总结

本节调研了融合感知算法，包括图像点云融合检测算法和图像点云融合跟踪算法。常用数据集包括 KITTI 数据集、SUN-RGBD 数据集、nuScenes 数据集等，具体如表 3-3 所示：

表 3-3 融合感知算法总结

算法类型	算法名称	测试数据集
图像点云融合检测算法	MV3D	KITTI 数据集
	Frustum-PointNet	KITTI 数据集 SUN-RGBD 数据集
	PointFusion	KITTI 数据集 SUN-RGBD 数据集
图像点云融合跟踪算法	DSM	KITTI 跟踪数据集
	mmMOT	KITTI 跟踪数据集
	EagerMOT	KITTI 跟踪数据集 NuScenes 跟踪数据集

3.2 模型任务所应用的场景维度

3.2.1 道路场景类型

包含高速公路/城市道路/乡村道路/停车场/园区/港头等。可分为公共道路、封闭道路两类。

1) 公共道路

a. 高速公路



图 3-37 高速公路

b. 城市道路



图 3-38 城市道路一



图 3-39 城市道路二

c. 山区蜿蜒道路



图 3-40 山区道路一



图 3-41 山区道路二

d. 乡村道路



图 3-42 乡村道路

2) 封闭道路

a. 封闭园区道路



图 3-43 封闭园区道路

b. 厂房内部道路

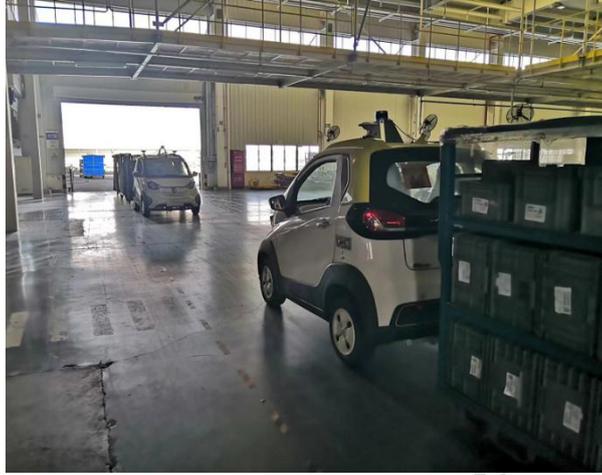


图 3-44 厂房内部道路

c. 停车场道路



图 3-45 地下停车场道路



图 3-46 户外停车场

某企业自建数据集占比情况统计：

高速及城市快速路 33.3%，国省道及城市地面道路 33.3%(包括郊区、城市外围或新区的主干道等)，停车场 33.3%(包括地面、地下单层/跨层、停车楼等)

3.2.2 关键道路特征

包含各类桥（平桥，拱桥，驼峰桥）/隧道/红绿灯十字路口/施工现场/事故现场/人流量高道路/弯道/傍山道路/立交桥道路/陡坡/非正式道路(山路或者自主修建道路)/泥泞路/河堤路。

1) 分叉路



图 3-47 分叉路

2) 铁架桥



图 3-48 铁架桥一



图 3-49 铁架桥二

3) 隧道



图 3-50 隧道

(4) 匝道



图 3-51 匝道

(5) 环形路



图 3-52 环形路

某企业自建数据集占比情况统计：

高速及城市快速路：桥 49%，隧道 25%，施工路段 25%，事故现场 1%；

国省道及城市地面道路：路口 65%(包括掉头、十字路口、人行横道、小路汇入)，桥 5%，施工路段 10%，公交车站/学校路段 10%，路边停车 10%；停车场：故障车/管控/违停/阻碍道路 10%，停车场跨层匝道 20%，停车场车位间通道 60%，停车场出入口匝道 10%。

3.2.3 路面情况

包含积水/水洼/泥土/冰面/雪路/雾/雾霾/强光/弱光/光线交替/泥石流。可分为干燥路面、湿润路面、非直道路面、破损及不规则路面等。

1) 干燥路面

a. 减速带



图 3-53 减速带一



图 3-54 减速带二

b. 水沟路面（水沟在道路上，上面有铁支架或石板等类似东西）



图 3-55 水沟路面一



图 3-56 水沟路面二

2) 湿润路面

包含积水路面、雨后柏油路



图 3-57 积水路面



图 3-58 雨后柏油路

3) 非直道路面

a. 坡道路面



图 3-59 坡道路面一



图 3-60 坡道路面二

b. 侧倾路面



图 3-61 侧倾路面

c. 转弯路面 (U型弯)



图 3-62 转弯路面

d. 连续上下坡起伏道路



图 3-63 连续上下坡起伏道路一



图 3-64 连续上下坡起伏道路二

e. 连续 S 弯路面



图 3-65 连续 S 弯路面一



图 3-66 连续 S 弯路面二

f. “坡道&侧倾&S 弯”组合起来的道路

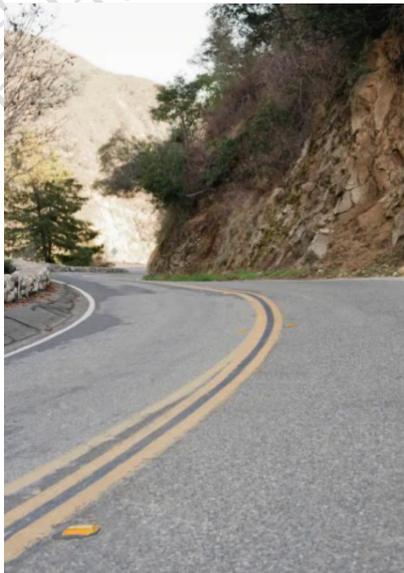


图 3-67 “坡道&侧倾&S 弯”组合起来的道路

4) 破损及不规则路面

包含凹凸不平路面、下雪路面



图 3-68 破损路面



图 3-69 凹凸路面



图 3-70 下雪路面

车道线情况可按如下几个维度分类:

1) 不平行



图 3-71 不平行车道线一



图 3-72 不平行车道线二



图 3-73 不平行车道线三

2) 车道数量变化



图 3-74 车道数量变化

3) 路沿



图 3-75 路沿一



图 3-76 路沿二

4) 汇流口



图 3-77 汇流口

5) 分流口



图 3-78 分流口

6) 双线（分虚实线）



图 3-79 双线实线



图 3-80 双线实线与虚线一



图 3-81 双线实线与虚线二

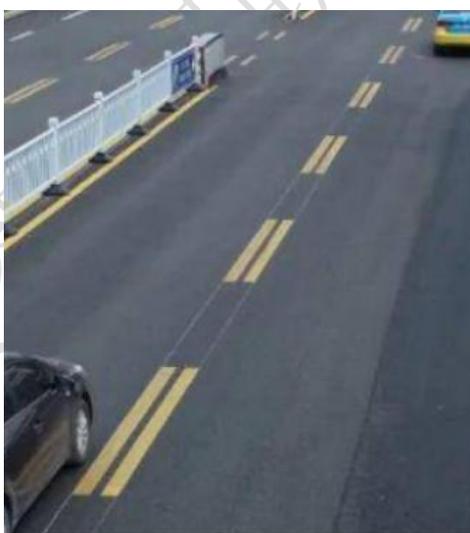


图 3-82 双线虚线

某企业自建数据集占比情况统计：

车道线情况:道路车道线清晰 90%，不清晰 10%；

高度落差情况:无大落差地面 80%，大落差地面(如台阶)20%；

车位线情况:停车场车位线无遮挡 90%，被遮挡 10%。

3.2.4 天气与光照

1) 天气条件

天气可分为：晴天、小雨、中雨、大暴雨、雾霾天气、下雪天、沙尘暴天气
某企业自建数据集占比情况统计：晴天与阴天 70%，雨天 20%，雪天 10%

2) 光照条件：

白天/上午/下午/黄昏/夜晚/强光/白天阴影/逆光/光线交替（弱光转强光/强光转弱光）。

a. 夜晚下细雨的城区道路（路灯偏黄色的场景）；



图 3-83 夜晚细雨城市道路

b. 晴朗天气早上或傍晚的逆阳光行驶场景；



图 3-84 逆光行驶

c. 白天进入或驶出隧道场景



图 3-85 白天进出隧道场景

某企业自建数据集占比情况统计：

日夜时间:白天上午 30%，白天下午 50%，夜晚 20%；光照条件:良好光照(100-10000lux)60%，一般光照(10-100lux)10%，极差光照(<10lux)10%，过强光照(直射相机使之过曝)10%，左右显著不均匀光照(一侧过强光照另一侧一般光照)10%；阴影条件:仅仅少量阴影或无大片阴影 90%，带大片阴影 10%。

3.2.5 交通参与者与交通标志

1) 交通参与者类型

高速及城市快速路:车辆 90%，两轮车(摩托与电瓶)5%，地面养护人员 5%；国道及城市地面道路:四轮机动车 20%，三轮机动车 5%，两轮机动车 5%，两轮非机动车 30%，三轮非机动车 5%，行人 20%，其他参与者 15%。

2) 交通参与者行为：

包括变道、横穿、逆行、转弯、直行、连续变道、急刹车、急加速、急停、车辆熄火、非法倒车、倒车。

某企业自建数据集占比情况统计：

高速及城市快速路的车辆:正常行驶 59%，变道 40%(正常变道 85%，加塞 10%，极端情况加塞 5%)，逆行 1%；

高速及城市快速路的两轮车：在最右车道行驶 40%，在应急车道行驶 30%，在中间或者最左侧车道行驶 30%；

高速及城市快速路的路面养护人员：静止(静止作业 80%)沿车道方向行走(应急车道或者施工道路内部沿着道路行走 15%)横穿(5%);国省道及城市地面道路的车辆：正常行驶 45%，变道 30%，横穿 5%，掉头 10%，停在路边 10%;

国省道及城市地面道路的两轮车：正常行驶 60%，横穿 20%，逆行 10%，变道 10%;

国省道及城市地面道路的行人：在人行道行走 40%，沿车道线附近前行/逆行各 20%，横穿 20%。

3) 道路标志物

a. 锥桶



图 3-86 锥桶

b. 低矮平板拖车（封闭厂区内）

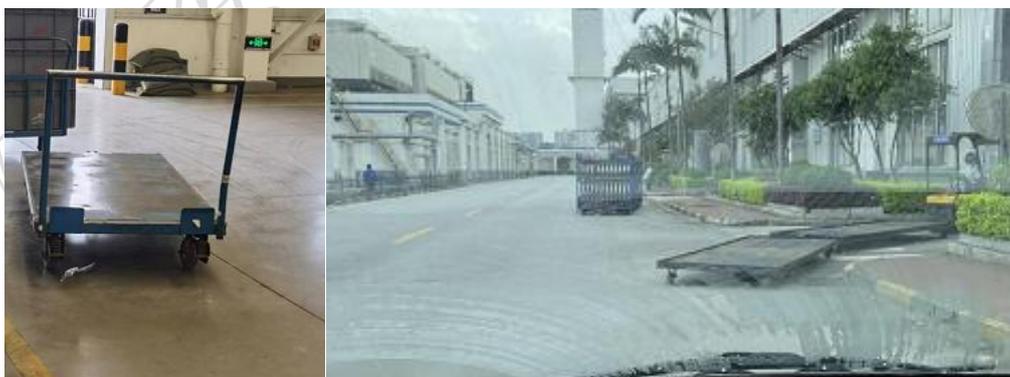


图 3-87 低矮平板拖车

c. 国标路牌



图 3-88 国标路牌

d. 交通灯



图 3-89 交通灯

e. LED 灯限速牌



图 3-90 LED 灯限速牌

f. 货车上的限速牌（误识别测试）



图 3-91 货车上的限速牌

g. 高速公路上的交通标识牌



图 3-92 高速公路交通标识牌

h. 地面标识线（停止线、车道线、斑马线、箭头、限速、自行车道标识等）



图 3-93 地面标识



图 3-94 斑马线

3.2.6 其他

1) 交通流状况

包含交通流正常/拥堵/交通瘫痪/交通事故/非机动车窜挤。

某企业自建数据集占比情况统计：交通流正常 80%，拥堵 20%

2) 城市分布：东南西北

某企业自建数据集占比情况统计：地理方位:东部 20%，西部 30%，北方 20%，南方 30%；省份城市:主要在人口多的省份，一二线城市；地理类型:平原 97%，带纵向坡度 2%，带横向坡度 1%。

3.3 感知训练数据集子集所需量级建议

3.3.1 视觉图像算法所需数据量级

1) 图像目标检测算法数据集

针对目标检测需求，建议为达到 95% 以上的识别准确率和召回率，需要不少于 100 万张用于训练和验证的图像，每张图像平均具备 8 个以上的细分类别目标。具体包括全车、车头车尾、行人、交通标识、异形障碍物等。对当前帧图像很难识别的目标（如较远目标和遮挡严重目标）进行标识，在测试时可进行忽略或单独评测。不同目标差异较大，如行人检测难度相较于其他类型难度偏高，需要更大数据量。尽量覆盖各种场景，如白天、晚上、雨天、雾天等。

2) 图像红绿灯检测算法数据集

红绿灯是视觉感知算法识别的最主要目标之一，针对其需要进行重点分析。针对红绿灯检测需求，建议为达到 99% 以上的识别准确率和召回率，需要不少于 50w 张用于训练与验证的图像，尽可能覆盖多种类型的红绿灯，包括圆形、箭头等，尽量覆盖不同天气情况，对当前帧图像很难识别的目标（如较远目标和遮挡严重目标）进行标识，在测试时可进行忽略或单独评测。

3) 图像目标追踪算法数据集

针对每一类目标，应不少于 20 万张用于训练和验证的图像（要求单目标时间

连续 5s 以上)，保证每帧中均含有有效目标。

4) 图像语义分割算法数据集（车道线、可行驶区域、路面标记）

针对图像语义分割算法需求，应不少于 50 万张用于训练和验证的图像，每张图像平均具备 2 个以上的细分类别目标，主要针对道路相关内容进行采集。

5) 图像车道线检测算法数据集

针对车道线检测，建议为达到 99% 以上的识别率与准确率，需要不少于 50 万张用于训练和验证的图像，尽可能覆盖多样性道路场景，如城市道路（包括复杂路口）。尽量覆盖各种车道线类型、道路情况、天气情况等。对当前帧图像很难识别的车道线（磨损严重）进行标识，在测试时可进行忽略或单独评测。

6) 图像车位线检测算法数据集

针对车位线检测，目前应用较多的数据集包括 ps2.0, PIL-park, PSV 等等，其中 ps2.0 使用四路鱼眼相机结果拼接，共包含了 12165 张车位线图片；PSV 数据集使用鱼眼相机拼接，共包含 4249 张车位线图片；PIL-park 数据集只使用单个相机，共包含 18299 张图片；另有数据集 wuzzh/Parking-slot-dataset 共包含约 15000 张图片（目前还没有完整制作完整），并包含了一些草地车位、砖地车位、立体车位等特殊车位图片。

为了实现精确的停车位检测，建议数据集数据量达到 10 万张以上，如果使用单个摄像机所得图片，建议数据量达到 15 万张以上。

7) 图像深度估计算法数据集

建议至少满足不同场景类型及关键道路特征场景下的数据量 10 万张，精度要高、场景覆盖广。

3.3.2 激光点云算法所需数据量级

1) 点云 3D 目标检测算法数据集（交通参与者）

针对 3D 目标检测，建议为达到 90% 以上的识别准确率和召回率，需要不少于 100w 帧用于训练和验证的点云数据，覆盖 V2X 和车端，每帧点云平均具备 2 个以上的细分类别目标。尽可能覆盖各种天气场景。

2) 点云目标追踪算法数据集

针对每一类目标，应不少于 20w 帧用于训练和验证的点云数据（要求单目标时间连续 5s 以上），每帧点云平均具备 2 个以上的细分类别目标。

3) 点云语义分割算法数据集

至少满足不同场景类型及关键道路特征场景下的数据量 10w 帧。

4) 点云车道线检测算法数据集

至少满足不同场景类型及关键道路特征场景下的数据量 2w 帧。

3.3.3 融合算法所需数据量级

1) 图像点云融合检测算法数据集

根据现有典型算法的论文实验中使用的数据集，包含 KITTI 的目标检测数据集以及 SUN-RGBD 数据集。在图像点云融合检测算法数据集应有图像点云的总数 100w 数量以上，目标数量大概的分布在 5-50 之间，保证平均每帧的目标大概在 10 个左右，覆盖城区高速快速路以及多个城市。

2) 图像点云融合追踪算法数据集

根据现有典型算法的论文实验中使用的数据集，使用了 KITTI 的目标追踪数据集以及 NuScenes 数据集。建议数据集中总帧数应超过 100w 帧，目标追踪的种类至少应包含车、行人类别。

4 感知训练数据集数据属性和要求

4.1 视觉图像算法数据属性和要求

4.1.1 图像 2D 目标检测算法数据集

4.1.1.1 标注范围

图像数据集中的目标以非旋转矩形框的形式标注出来，同时包括目标的类别，子类别，属性，标注类别见表 4-1。

表 4-1 图像数据集类别（详情见“4.1.1.2 标注类别”章节）

类别	描述
Vehicle (车辆)	此类别包含场景中出现的所有车辆类别，包括汽车，自行车，摩托车等子类
Human (人)	此类别包含场景中检测到的所有人，包括行人，骑手，移动单元，人脸等子类别。正在站立/走路的为行人 (pedestrian)，摩托车上的人或自行车上的人为骑手 (Rider)。如果一个人正在推着，牵着，骑着某个物体且并不属于专门的车辆种类 (例如婴儿车，购物手推车，赛格威 (两轮平衡车)，滑板车/小型摩托车，轮椅等等)，则物体应该被归类为移动单元 (mobile item)，人的面部为人脸子类
Traffic_sign (交通标志)	此类别包含所有面向采集车方向的交通标志 (包含交通灯)，包括我们不需要标注标志的背面
Non_vehicle_light (非交通灯)	此类别包含所有与交通无关的灯，如路灯，以及道路或车道灯或反射器
Obstacle_Object (障碍物体)	此类别包含所有可能堵塞可行驶路面的不可行驶物体，以及特定场景下的物体，如建筑物，紧急锥形物，和消防栓等等

标注框：围绕图像中特定对象的封闭区域绘制的非旋转矩形框，以矩形框的左上角 UV 坐标和右下角 UV 坐标表示，填写在对应的属性 Target_Location (目标坐标) 中，以 List 形式表示，如图 4-1：

```
"Target_Location": [  
  [  
    852.6383173296754,  
    488.3539094650207  
  ],  
  [  
    997.3113854595337,  
    690.6858710562415  
  ],  
]
```

图 4-1 目标标注框坐标示意

4.1.1.2 标注类别

本章节对图像 2D 目标数据集中不同的目标类别进行详细定义，实际应用中可在此基础上扩展，定义请见表 4-2。

注：由于不同目标类别属性不同，因此在本章节中对属性作补充描述。

表 4-2 图像 2D 标注目标类别详细定义

类别	子类别	属性 (详情见表 4.3)	描述/示例
Vehicle (车辆)	Car (轿车)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节) Color (颜色) Yaw_angle (航行线)	小型汽车 (掀背车, 轿车/轿跑车), SUV, MPV 或面包车
	Small_sized_truck (小型卡车)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节) Color (颜色) Yaw_angle (航行线)	小型卡车 (泛指长度在 8 米以内的卡车), 包括皮卡和房车
	Larger_sized_truck (大型卡车)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节) Color (颜色) Yaw_angle (航行线)	大型卡车 (泛指长度大于 8 米的卡车)
	Bus (公共汽车)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节) Color (颜色) Yaw_angle (航行线)	所有公共汽车, 包括公交车、旅行巴士, 等等
	Train (火车)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节) Color (颜色) Yaw_angle (航行线)	所有轨道车, 包括火车、电车、巨轮
	Tricycle (三轮车)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节) Yaw_angle (航行线)	任何三轮电动车或汽油动力三轮车辆, 例如电动三轮车、快递三轮车、货运大型汽油三轮车
	Tricycle (三轮车)	Occluded (遮挡)	任何三轮电动车或汽油动力

	车)	Truncated (截断) Additional_details (补充细节) Yaw_angle (航行线)	三轮车辆, 例如电动三轮车、 快递三轮车、货运大型汽油 三轮车
	Motorcycle (摩 托车)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节) Yaw_angle (航行线)	任何两轮电动车或汽油动力 车辆, 例如摩托车&骑手 注: 当摩托车或骑手在拥挤 的场景中完全或大部分被遮 挡时, 可以忽略大部分被遮 挡的对象。
	Bicycle (自行 车)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节) Yaw_angle (航行线)	在道路、人行道或自行车道 上低速行驶的任何两轮人力 或电动车辆。仅标注自行车, 自行车上的人需要单独标注 为“Rider (骑手)” 注: 不要标注放置在车上的 自行车(通过车顶或行李箱 架或拖车挂接装置等连接)。 这些应该被认为是运载它们 的车辆的一部分。
	Special_vehicle (特殊车辆)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节) Yaw_angle (航行线)	缓慢移动或静止的清洁或施 工设备/车辆。只有当它们在 路面上或干扰了交通流量时 才需要标注
	Emergency_veh icle (紧急车辆)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节) Yaw_angle (航行线)	任何紧急车辆, 如警车, 消 防车, 救护车 注: 不要将普通的或未标记 的警车标注为紧急情况, 只 需将它们标注为正常车辆类 别即可。
	Special_vehicle (特殊车辆)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节) Yaw_angle (航行线)	缓慢移动或静止的清洁或施 工设备/车辆, 只有当它们在 路面上或干扰了交通流量时 才需要标注
	Emergency_veh icle (紧急车辆)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节) Yaw_angle (航行线)	任何紧急车辆, 如警车, 消 防车, 救护车 注: 不要将普通的或未标记 的警车标注为紧急车辆, 只 需将它们标注为正常车辆类 别即可。
	Trailer (拖车)	Occluded (遮挡) Truncated (截断)	任何箱式仓库、平板车、飞 机或休闲车式拖车

		Additional_details (补充细节) Color (颜色) Yaw_angle (航行线)	注：当卡车和拖车并没有分成两个明显独立的部分时，标注为拖车 trailer，并配以适当的属性即可。
	Headlight (前灯)	Additional_details (补充细节)	在夜间行驶的来自全部车辆的所有独立的前灯，可见并处于照明状态时 注：最小标注像素为 5*5。
	taillight (尾灯)	Additional_details (补充细节)	在夜间行驶的来自全部车辆的所有独立的尾灯，可见并处于照明状态时 注：最小标注像素为 5*5。
	Turnsignal (转弯信号)	Additional_details (补充细节)	所有车辆上正在照明的转弯信号/指示器。转弯方向用于指示它是否为目标车辆的左转或右转信号 注：最小标注像素为 5*5。
Human (人)	Pedestrian (行人)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节) Posture (姿势)	一帧画面里行走的成年人或儿童。携带的儿童不需要被单独标注，他应该属于明显的行人的一部分
	Rider (骑手)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节)	任何乘坐两轮车或小型交通工具的人(自行车，摩托车，婴儿车，轮椅)
	Mobile_item(移动物体)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节)	任何可以骑、推、拖，并且和人有关联的物体，如果一个人推车或拖着一个移动物体，他所推着或者推着的物体标注为移动物体，这个人单独标注为行人 注： 1) 如果一个物体和人没有关联，或没有被骑手骑，那么它不是一个移动物体，而且应该被标注为物体，如：购物手推车，婴儿车，手推车，带轮子的垃圾车，等等。 2) 如果一个人骑着一个物体(物体并不是自行车或摩托车)，例如是个轮椅，那么这个人应该被标注为骑手

			(rider), 这个轮椅需要被标注为移动物体, 如: 赛格威, 轮椅, 滑板车, 滑板, 等等。
	face (人脸)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节)	注: 最小标注像素为 20*20。
Traffic_ sign (交通标志)	Traffic_light (交通灯)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节) Light_state (灯的颜色)	正确: 独立的灯 错误: 无群组
	Traffic_light (交通灯)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节) Light_state (灯的颜色)	灯的颜色: 红色、黄色、绿色 灯的方向: 垂直或水平 注: 只单独标注交通灯, 无需标注灯群。
	Guidelight (引导灯)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节) Lane_state (通道状态)	可逆车道灯, 中间带方向指示灯, 引导灯的属性会指示你该车道是开放, 关闭, 或标志是关闭/不可操作
	pedestrian light (行人灯)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节)	十字路口处的行走/禁止行走
	Speed_limit (限速)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节) Electronic (电子的) Value (速度值)	任何与速度有关的交通标志 (速度值属性标签应与标牌上打印的速度限制数值相匹配)

		High_limit (限高速的)	
	Stop (停止)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节) Electronic (电子的)	停止标志是八边形, 红色背景, 白色文字
Traffic_ sign (交通标志)	Yield (让行)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节) Electronic (电子的)	让行标志是三角形, 红色和白色背景, 黑色或红色文字
	Warning (警告)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节) Electronic (电子的)	警告标志突出显示在某些场景下, 如弯道, 学校, 死胡同, 或交通信号。也可以警告可能存在的危险, 如颠簸, 自行车, 低空飞行的飞机, 急救车辆, 等等
	Prohibition (禁止)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节) Electronic (电子的)	禁止通行标志, 就是禁止通行的一种交通标示, 提示车辆和行人不能再向前行驶及通过, 以免发生交通事故。红圈内空白是禁止一切机运车辆通行, 红底白横杠是禁止机动车辆驶入, 也有红底内加一个白叉
	pole (立杆)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节)	安装在人行道上或靠近道路边缘的杆子, 指示地铁站的入口。请务必在电杆与地面的连接处 (不仅仅是标志部分) 标注电杆
	rectangle (矩形标牌)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节) Electronic (电子的)	导向板、路标、出口方向标志等。
Non_vehicle_light (非机动车灯)	Street_light (路灯)	Additional_details (补充细节)	路侧所有照明路灯 注: 最小像素尺寸为 5*5。
	Reflector (反射灯)	Additional_details (补充细节)	道路和柱子反射器 注: 白天或光线充足的情况下, 不需要标注反射灯。
Obstacle_Object (障碍物)	Traffic_cone(交通锥桶)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节)	任何类型的交通圆锥 (必须是圆锥形的) 注: 道路分隔物, 交通柱, 交通桶等一般都不是锥形

			的，当它们被放置在可行驶路面时，需要被标记为物体 object。
	Other (其他障碍物)	Occluded (遮挡) Truncated (截断) Additional_details (补充细节)	任何放置在路边或道路的静止物体

4.1.1.3 标注属性

对于各个目标类别及子类别属性，详细定义见表 4-3。

表 4-3 图像 2D 标注目标类别属性定义

属性	值	描述
Occluded (遮挡)	0: 0% (完全可见) 1: 1-50% (部分遮挡) 2: 51-100% (大部分遮挡) 3: 未知	此属性默认值为 0，只有出现部分遮挡或全部遮挡时，将其更改为相应数值
Truncated (截断)	0: 0% (完全可见) 1: 1-50% (部分可见) 2: 51-100% (大部分被截断)	截断指的是物体离开图片边缘的情况。这个属性适用于当车辆，人，交通标志，交通灯的部分或全部被截断的情况。默认值为 0
Additional_details (补充细节)	不限格式和类型	此属性为开放属性，具体内容可以根据需要，在不同的版本中自定义
Color (颜色)	"black","blue","brown","darkgray","green","purple","red","silver","white","yellow", "other"	车辆颜色
Yaw_angle (航向线)	List [起始点 UV 坐标, 结束点 UV 坐标]	用俩个点画一条线来表示车辆当前的航向朝向
Posture (姿势)	0: 站立 1: 行走或奔跑 2: 蹲下, 躺下, 坐下或其它	表示当前人的姿态，默认值是 0
Electronic (电子的)	1: 是 0: 否	电子属性，用于标记电子交通标志，这些标志可以改变以显示不同的文本或图形。该属性应仅用于指示标志是“电子的”
Value (速度值)	10 ... 100 (etc)	速度值属性用于标记限速符号值
High_limit (限高速的)	是否限高速: 1: 是 0: 否	当有些路段限制低速时为 0，默认值为 1

Light_state (灯的颜色)	Red: 红灯 Yellow: 黄灯 Green: 绿灯 Straigh: 直行箭头 Left: 左转箭头 Right: 右转箭头 Uturn: U型弯箭头 Off: 关闭(或未工作)	浅色用于标记它所代表的交通灯堆栈的灯光配置。如果标注多灯配置, 请使用所有必需的属性来指示哪些灯在照明状态。如果灯不工作, 请标注关闭
Lane_state (通道状态)	Open: 车道开通 Closed: 车道关闭(或相反的交通流) Off: 关闭(或未工作)	指示灯用于标记可逆或可关闭车道的灯光配置(例如当车道方向针对高峰时间进行调整时, 或者收费亭车道是开放还是关闭时)

4.1.2 图像目标追踪算法数据集

4.1.2.1 标注范围

同“4.1.1 图像 2D 目标检测算法数据集-4.1.1.2 标注范围、4.1.1.2 标注类别”章节。

4.1.2.2 标注属性

多目标跟踪须将每个目标和其他目标进行区分开来, 具体方法是给每个目标分配一个 ID, 并记录他们的轨迹, 因此图像目标追踪标注需在“4.1.1 图像 2D 目标检测算法数据集-4.1.1.3 标注属性”的基础上, 增加连续的“跟踪 ID”属性来进行描述。

4.1.3 图像语义分割算法数据集

4.1.3.1 车道线

1) 标注范围

车道线的类别定义见表 4-4。

表 4-4 车道线类别

类别	子类别
Lane_line (车道线)	Lane_line (车道线)
	Diversion_line (导流线)
	Zebra_line (斑马线)
	Stop_line (停止线)
	Stop_yieldline (停止让行线)

	Speed_bump (减速让行线)
	Toturn_line (待转线)
	Shunt_line (分流线)
	Confluence_line (合流线)
	Mesh_line (网状线)

标注时，应识别所有符合当前车辆行驶的车道线，ID 和车道线的对应关系见表 4-5:

表 4-5 车道线的 ID 与车道线的对应关系

车道线的 ID	车道线位置
left_i	车辆左侧车道的车道线，i 为从 0 开始的数字。离当前车辆越远的车道线，数字越大。
right_i	车辆右侧车道的车道线，i 为从 0 开始的数字。离当前车辆越远的车道线，数字越大。

以图 4-3 为例:



图 4-3 车道线 ID 标注图示意

2) 标注属性

每个车道线需标注 5 个属性，见表 4-6。

表 4-6 车道线标注属性

属性	值
边界线 (Boundary_line)	1-是 0-否
线型(linear)	Solid:实线 Dotted: 虚线 Zebra: 斑马线 Other: 其他

线数 (Line_number)	Single: 单线 Double: 双线 Other: 其他
颜色 (Color)	White: 白色 Yellow: 黄色
遮挡 (Occluded)	0: 0% (完全可见) 1: 1-50% (部分遮挡) 2: 51-100% (大部分遮挡) 3: 未知

4.1.3.2 可行驶区域

1) 标注范围

对于图像数据集中可行驶区域的分割结果以数个多边形框的形式标注出来，同时应包括目标的类别，子类别，描述，可行驶区域的类别定义见表 4-7。

表 4-7 行驶区域类别

类别	子类别	描述
Drivable_areas (可行驶区域)	Directly_drivable_areas (直接可行驶区域)	符合《中华人民共和国道路交通安全法》的合法区域，区域里所有的点到车的连线都不会被障碍物阻断，且是当前正在行驶的区域（优先于其他汽车或“通行权”的区域）
	Alternative_drivable_areas (替代可行驶区域)	符合《中华人民共和国道路交通安全法》的合法区域，且是当前未行驶的车道，但可以通过更改车道来实现（其他汽车的“通行权”优先于本车的行驶区域）
Undrivable_areas (不可行驶区域)	/	图像中除去当前可行驶区域的其他区域，包括中间有隔离带的对向车道，没有隔离带的反向车道，有隔离带的辅路、自行车道，商铺门口柏油马路，高速入口等 注：不可行驶区域不进行标注。

围绕图像中特定对象的封闭区域绘制的多边形框，保存多边形框的每个顶点坐标，填写在对应属性 coordinates 中，以 List 形式表示，见图 4-4:

```

▼ coordinates:[
  ▼[
    ▼[
      1091.5912208504799,
      332.49657064471876
    ],
    ▼[
      1156.0631001371742,
      206.2962962962963
    ],
    ▼[
      1319.3004115226338,
      246.07681755829907
    ],
    ▼[
      1342.620027434842,
      399.71193415637856
    ],
    ▼[
      1272.6611796982165,
      477.90123456790127
    ],
    ▼[
      1091.5912208504799,
      332.49657064471876
    ]
  ]
]

```

图 4-4 可行驶区域标注 json 示意

2) 标注属性

可行驶区域的属性主要涵盖 3 类，见表 4-8：

表 4-8 可行驶区域标注属性

属性	值
imageDimension（对应图片的维度）	按图片实际像素补充
flags（标签）	可行驶区域类型标签
describe（描述）	扩展预留属性

4.1.3.3 路面标记

1) 标注范围

对于图像数据集中路面标记的分割结果以数个多边形框的形式标注出来，同时应包括目标的类别，子类别，描述，标注类别见表 4-9。

表 4-9 路面标记类别

类别	子类别	描述
Road_marking (路面标记)	Lane_line (车道线)	所有用来分割行驶区域的地面标识线，包括：普通车道、边缘线、停止线、具备警示意义的车道线形态
	Speed_bump (减速带)	横向、纵向减速带都要标出。纵向减速带一些小菱形组成。
	Zebra_line (斑马线)	交叉路口行人保护区域的标识，具有多种形态

	Stop_line (汽车停止线)	一般在路口处
	Diamond_sign (地面菱形标志)	减速标志，用于提示车辆减速
	Arrow_sign (箭头标识)	注意细分每种箭头类别： Traight: 直行 Straight_left: 直行或左转 Left: 左转 Right: 右转 Straight_right: 直行或右转 Straight_turnback: 直行或掉头 Left_turnback: 左转或掉头 Left_right: 仅可左右转 Turnback: 掉头 Other: 其他
	Characters_sign (文字标识)	地面上的数字、文字等信息
	Cycle_sign (圆形标识)	地面上的圆形标识，包括速度等信息
	Other (其他)	其他地面标识

2) 标注属性

路面标记的属性主要涵盖 3 类，见表 4-10：

表 4-10 路面标记标注属性

属性	值
imageDimension (对应图片的维度)	按图片实际像素补充
flags (标签)	路面标记类型标签
describe (描述)	扩展预留属性

4.1.4 图像车道线检测算法数据集

4.1.4.1 标注范围

同“4.1.3 图像语义分割算法数据集-4.1.3.1 车道线”章节。

4.1.4.2 标注属性

同“4.1.3 图像语义分割算法数据集-4.1.3.1 车道线”章节。

4.1.5 图像车位线检测算法数据集

4.1.5.1 标注范围

根据国家规定，停车位的尺寸大小一般为宽 2.5 米，长 5.0 米~5.5 米。主要是为了符合停车需求。

车位的类型主要有三种：垂直型、侧方型、倾斜型。

4.1.5.2 标注属性

车位线的标记属性主要涵盖 8 类，见表 4-11：

表 4-11 车位线标注属性

属性	说明
Shape（形状）	矩形车位，斜向车位等
Color（颜色）	白色实线、黄色实线等
Closed（封闭）	封闭车位、开口车位
Truncated（截断）	完整车位包括四个角点，完整车位线），部分完整车位包括一条或两条完整边线，两个或三个完整角点，两条不完整边线），其他少于两个完整交掉角点）
Occluded（遮挡）	完全遮挡、部分遮挡等
Light（光线）	室内日光灯、室外白天、室外黑夜、室外街灯等
Weather（天气）	正常、雨天、雾天等
Special（特殊）	草地车位、石砖地车位等

4.2 激光点云算法数据属性和要求

4.2.1 点云 3D 目标检测算法数据集

4.2.1.1 标注范围

交通参与者类别应涵盖以下 7 类：car、bus、truck、tricycle、bicycle、pedestrian、other_vehicle，详细说明见表 4-12：

表 4-12 交通参与者类别说明

类别名称	类别说明
car	包含轿车、SUV、MPV、皮卡车、五菱宏光类小面，四轮快递车等

bus	包含公交汽车、校车、小巴、大金杯、C型房车等
truck	大型集卡、大型货车、大型半挂车、大型厢式货车、中型厢式货车、油罐车、垃圾车、机动农用三轮车等
tricycle	快递三轮车、电动三轮车等
bicycle	自行车、摩托车、电动自行车
pedestrian	包含成人、儿童等，详见属性
other_vehicle	婴儿车、小推车、坐人的轮椅

4.2.1.2 标注属性

1) 3D 真值框属性

3D 真值框属性应涵盖以下 7 类：x、y、z、l、w、h、heading，详细说明见表 4-13：

表 4-13 3D 真值框属性

属性	含义	备注
x	中心点 x 坐标	
y	中心点 y 坐标	
z	中心点 z 坐标	
l	长	单位 m
w	宽	单位 m
h	高	单位 m
heading	航向角	取值范围：[-pi,pi]

2) 目标属性

目标属性应涵盖以下 3 类：truncated、occluded、abnormal，详细说明见表 4-14。

表 4-14 目标属性说明

属性	含义	值	值描述
truncated	是否截断：目标物是否在边缘且未完全在画面中	0	目标物完整
		1	目标物不完整（出现在固态雷达点云边缘）
occluded	是否被遮挡：目标物是否被其他物体遮挡	0	完全未被遮挡
		1	小部分被遮挡（不超过一半）
		2	大部分被遮挡

		3	未知（不知道是否被遮挡）
abnormal	数据是否异常	0	默认 0，无异常
		1	残影（拼接不完善）

3) 行人姿态属性

行人姿态属性在目标属性上应增加 pose，详细说明见表 4-15。

表 4-15 行人姿态属性说明

属性	含义	值	值描述
pose	姿态	0	站立
		1	坐或蹲
		2	其他（躺或者弯腰）

4) 二轮车属性

二轮车属性在目标属性上应增加 with_rider，详细说明见表 4-16。

表 4-16 二轮车属性说明

属性	含义	值	值描述
with_rider	是否有骑车人	0	无
		1	有

4.2.2 点云目标追踪算法数据集

4.2.2.1 标注范围

同“4.2.1 点云 3D 目标检测算法数据集-4.1.1.1 标注范围”章节。

4.2.2.2 标注属性

1) 轨迹属性

目标轨迹属性应涵盖以下 2 类：轨迹长度、轨迹频率，详细说明见表 4-17。

表 4-17 轨迹属性说明

属性	要求
轨迹长度	大于等于 20 秒
轨迹频率	≥ 5 Hz

2) 目标属性

目标属性应涵盖以下 2 类：frame_id、ID，详细说明见表 4-18。

表 4-18 目标属性说明

属性	说明
frame_id	当前目标所属的帧
timestamp	当前目标对应的时间戳
ID	目标在当前帧的跟踪 ID,同一轨迹中所有目标须保持 ID 一致,且尺寸不变

4.2.3 点云语义分割算法数据集

4.2.3.1 标注范围

点云语义分割所包含的类型见表 4-19。

表 4-19 点云语义分割目标类别

标注大类	一级类别	类型
交通参与者	机动车	乘用车
		客车
		货车
		半挂牵引车
		特殊车辆
	两轮车	自行车
		摩托车
		三轮车
	骑手	自行车骑手
		摩托车骑手
		三轮车骑手
	行人	/
	忽略	/
路障	锥形桶	/
	海马	/
	石墩	/
	隔离带	/
	栅栏	/
	台阶	/
路面	人行道	/
	路沿	/
	可行驶区域	/
绿植	草	/
	灌木	/

绿植	树干	/
	树冠	/
楼宇	楼宇	/
路灯-杆	路灯-杆	/
路灯-灯源	路灯-灯源	/
交通信号灯-灯	交通信号灯-灯	/
交通信号灯-支架	交通信号灯-支架	/
交通标识-牌	交通标识-牌	/
交通标识-杆	交通标识-杆	/
忽略	忽略	/

4.2.3.2 标注属性

同“4.2.1 点云 3D 目标检测算法数据集-4.2.1.1 标注属性”章节。

4.2.4 点云车道线检测算法数据集

4.2.4.1 标注范围

点云车道线检测算法所包含的类型见表 4-20。

表 4-20 点云车道线检测目标类别

类别	子类别
Lane_line (车道线)	Lane_line (车道线)
	Diversion_line (导流线)
	Zebra_line (斑马线)
	Stop_line (停止线)
	Stop_yieldline (停止让行线)
	Speed_bump (减速让行线)
	Toturn_line (待转线)
	Shunt_line (分流线)
	Confluence_line (合流线)
	Mesh_line (网状线)

4.2.4.2 标注属性

每个车道线需标注 5 个属性，见表 4-21。

表 4-21 车道线标注属性

属性	值
----	---

边界线 (Boundary_line)	1-是 0-否
线型(linear)	Solid:实线 Dotted: 虚线 Zebra: 斑马线 Other: 其他
线数 (Line_number)	Single: 单线 Double: 双线 Other: 其他
遮挡 (Occluded)	0: 0% (完全可见) 1: 1-50% (部分遮挡) 2: 51-100% (大部分遮挡) 3: 未知

4.3 融合算法数据属性和要求

4.3.1 图像点云融合检测算法数据集

4.3.1.1 标注范围

图像点云融合检测算法所包含的类型同“4.1.1 图像 2D 目标检测算法数据集-4.1.1.2 标注类别”章节。

4.3.1.2 标注属性

由于图像点云融合的标注框涵盖 2D 与 3D 框，所以图像点云融合检测算法的标注属性需兼容“4.1.1 图像 2D 目标检测算法数据集-4.1.1.3 标注属性”与“4.2.1 图像 2D 目标检测算法数据集-4.2.1.2 标注属性”章节中的属性描述。

4.3.2 图像点云融合追踪算法数据集

4.3.1.1 标注范围

图像点云融合检测算法所包含的类型同“4.1.1 图像 2D 目标检测算法数据集-4.1.1.2 标注类别”章节。

4.3.1.2 标注属性

由于图像点云融合的标注框涵盖 2D 与 3D 框，所以图像点云融合检测算法的标注属性需兼容“4.1.1 图像 2D 目标检测算法数据集-4.1.1.3 标注属性”与“4.2.1 图像 2D 目标检测算法数据集-4.2.1.2 标注属性”章节中的属性描述。

在此基础上，可参考前后帧，对相同物体增加 ID 描述（赋予同一 ID）。

注：在同一连续帧序列中，一个 ID 只对应一个物体，不同物体 ID 不同。

4.4 训练数据标注精确度行业普遍情况

4.4.1 抽检比例

训练数据标注的抽检比例指的是待验收的已标注样本占总的已标注样本的比例，一般用百分数来表示。评估训练数据标注精度时，应对被评估标注数据进行抽样质量评估。抽样样本数量以待质检总帧数（单帧任务）或待质检总序列数（追踪任务）为基准，对该数据进行抽样质检，按照行业普遍情况，质检数据比例应不少于 10%。

4.4.2 准确率

训练数据标注的准确率指的是合格的已标注样本数占总的已标注数的比例。准确率要求见表 4-22。按照行业普遍情况，框准确率、线准确率以及点准确率要求不低于 95%，帧准确率不低于 90%。

表 4-22 训练数据标注准确率

准确率命名	评定方法	准确率要求
框准确率	框准确率=合格框数/总框数	不低于 95%
线准确率	线准确率=合格线条数/总线条数	
点准确率	点准确率=合格点数/总点数	
帧准确率	帧准确率=帧准确率/总帧数	不低于 90%

4.4.3 贴合度 (cm/pixel)

训练数据标注中，贴合度指的是作业人员标注的点、线、框与原始数据中的点、

线、框的真值之间的重合程度。按照行业普遍情况，点云框贴合度应不大于 10cm，图像框贴合度应不大于 5pixel。

4.4.4 航向角

训练数据标注中，航向角指的是，在传感器局部坐标系下，标注框的方向与局部坐标系的航向参考轴之间的夹角。按照行业普遍情况，标注框的航向角与标注目标的航向角偏差范围不大于 3 度（1~2 度）。

4.4.5 连续帧的最短帧间隔时长

最短帧间隔时长指的是传感器连续帧之间的时间间隔，一般用毫秒来计量。按照行业普遍情况，连续帧的最短帧间隔时长应为 200ms。

4.4.6 速度的标注以及最大误差

速度标注最大误差指的是，标注过程中允许的物体速度值与其真值之间差值的最大绝对值。连续帧的采集时间间隔为 200ms，标注框的速度计算方法为当前帧目标框在前后帧 GPS 定位的距离除以时间间隔（200ms），按照行业普遍情况，速度最大误差不大于 3km/h（1~2km/h）。

4.4.7 图像标注的像素误差

训练数据标注中，像素误差指的是 2d 标注框与图像中物体的真值框之间的像素差值。按照行业普遍情况，图像标注的像素误差不大于 5pixel。

4.5 感知训练数据集数据标注格式

4.5.1 图像标注数据导出格式

图像标注数据导出格式见表 4-23。

表 4-23 图像标注数据导出格式

标注项	导出格式
2D 框	宜采用 txt/json

语义分割	宜采用 json/Mask
目标检测	宜采用 xml/txt/json
目标追踪	宜采用 xml/txt
车道线检测	宜采用 json

4.5.2 点云标注数据导出格式

点云标注数据导出格式见表 4-24。

表 4-24 点云标注数据导出格式

标注项	导出格式
3D 框	宜采用 json
语义分割	宜采用 pcd/coco/voc
目标检测	宜采用 pcd/bin/npz
目标追踪	宜采用 xml/json
车道线检测	宜采用 json

汽标委智能网联汽车分会

5 感知训练数据集测评方法

5.1 前言

随着科技的进步，智能网联汽车智能化传感器的普及，通过对感知数据集进行训练以提高传感器应对一般场景的同时，还可以通过无监督学习方式提升智能网联车辆应对一般场景和特殊场景的感知能力，如车辆被遮挡、恶劣天气、车辆形态差异等。

由于目前各种公开感知训练数据集的侧重点不同，不同数据集合经过训练后在同一感知算法进行测评时结果的差异性太大，导致无法实现在统一标准下对不同感知训练数据集进行相关测试及评价。基于目前状况，对训练数据集本身进行评价和数据集测评进行相关标准化研究显得尤为重要。

5.2 感知数据质量评价指标

5.2.1 图像数据质量评价指标

5.2.1.1 图像质量评价

图像质量常见的评价指标有：

- 1) 像素均值：图像像素的平均值；
- 2) 图像标准差：图像像素灰度值相对于均值的离散程度；
- 3) 图像平均梯度：图像的清晰度。

1) 像素均值

像素均值反映了图像的平均亮度，平均亮度越大，图像质量越好，均值 u 按公式(1)计算：

$$u = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N F(i, j) \quad (1)$$

式中：

$F(i, j)$: 评价图像在像素点 (i, j) 的像素值, 单位为 px ;

$M \times N$: 待评价图像的大小, 单位为 dpi 。

2) 像素标准差

像素标准差越大, 表明图像中灰度级分布越分散, 图像质量也就越好。标准差 std 按公式(2)计算:

$$std = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (F(i, j) - u)^2} \quad (2)$$

式中:

$F(i, j)$: 评价图像在像素点 (i, j) 的像素值, 单位为 px ;

$M \times N$: 待评价图像的大小, 单位为 dpi ;

u : 像素均值。

3) 图像平均梯度

图像平均梯度反映图像中细节反差和纹理变换。图像平均梯度越大, 图像越清晰。平均梯度按公式(3)计算:

$$\nabla G = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \sqrt{\Delta x F(i, j)^2 + \Delta y F(i, j)^2} \quad (3)$$

式中:

$F(i, j)$: 评价图像在像素点 (i, j) 的像素值, 单位为 px ;

$M \times N$: 待评价图像的大小, 单位为 dpi ;

Δx : 像素点 (i, j) 在 x 方向上的一阶差分;

Δy : 像素点 (i, j) 在 y 方向上的一阶差分。

5.2.2 点云数据质量评价指标

激光雷达点云数据质量主要从点云的密度和精度进行评价, 常见的评价指标有:

1) 点云密度: 用于描述单位面积上激光雷达点的平均数量。

2) 高程精度: 用于评价激光雷达点云数据的高程与其真实的地面高程之间误差分布离散程度的指标, 采用最大高程误差、高程中误差, 相对高程中误差进行评价。

3) 平面精度: 用于评价激光雷达点云数据的平面位置与其真实的地面位置之

间误差分布的离散程度，采用最大平面位置误差、平面位置中误差，相对平面位置中误差进行评价。

1) 点云密度

激光雷达点云密度利用全测区(水域除外)激光雷达点个数与全测区(水域除外)面积进行计算。激光雷达点云密度，按式(1)计算：

$$\rho = \frac{n - \sum_{i=1}^m n_i}{A - \sum_{i=1}^m A_i} \quad (1)$$

ρ : 激光雷达点云密度，单位为个每平方米个/(m²)；

n : 测区内激光雷达点云总点数，单位为个；

n_i : 第*i*个水域内激光雷达点云点数，单位为个；

m : 测区内水域个数，单位为个；

A : 全测区激光雷达点云覆盖的面积，单位为平方米(m²)；

A_i : 第*i*个水域激光雷达点云覆盖的面积，单位为平方米(m²)。

2) 高程精度

最大高程误差用于评价激光雷达点云数据的高程与其真实的地面高程之间的最大差别，利用测区野外检查点数据进行评定。检查点测量方法与精度应符合 CH/T 1021: 2010 中 6.5 相关要求，且分布均匀，位置明显，避免选择航带重叠区及高程急剧变化处。检查点应位于裸露的平坦地面上；检查点数量视激光雷达点云密度、覆盖范围、覆盖类型等具体情况确定，宜不少于 15 个；检查点高程基准应与激光点云高程基准一致。利用以检查点为中心的邻近激光点内插出检查点位置的高程，进行误差计算。

最大高程误差 d_{max} 按式(2)计算：

$$d_{max} = \max(|Z_i - \hat{Z}_i|) \quad (2)$$

式中：

d_{max} : 最大高程误差，单位为米(m)；

Z_i : 第*i*个检查点对应的激光雷达点内插高程值，单位为米(m)， $i = 1, 2, \dots, n$ ；

n: 检查点个数, 单位为个;

\hat{Z}_i : 第*i*个检查点的实测高程值, 单位为米(m), $i = 1, 2, \dots, n$ 。

高程中误差用于评价激光雷达点云数据的高程与其真实的地面高程之间的差别, 利用测区野外检查点数据进行评定。利用以检查点为中心的邻近激光雷达点内插出检查点位置的高程, 进行误差计算。

高程中误差 Z_{RMSE} 按式(3)计算:

$$Z_{RMSE} = \pm \sqrt{\frac{\sum_{i=1}^n (Z_i - \hat{Z}_i)^2}{n}} \quad (3)$$

式中:

Z_{RMSE} : 高程中误差, 单位为米(m);

Z_i : 第*i*个检查点对应的激光点内插高程值, 单位为米(m), $i = 1, 2, \dots, n$;

n: 检查点个数, 单位为个;

\hat{Z}_i : 第*i*个检查点的实测高程值, 单位为米(m), $i = 1, 2, \dots, n$ 。

相对高程中误差用于评价激光雷达点云数据的高程与平均高程之间的差别, 利用裸露的平面作为检测平面(平面上激光点数量视激光雷达点云密度等具体情况确定, 宜不少于 15 个激光点)进行评定。计算检测平面内激光点的高程平均值, 使用激光雷达点云高程值与均值的差值来评定激光雷达点云相对高程中误差。对于高程值与高程平均值之差大于 2 倍相对高程中误差的激光点应进行校核, 判断该点是否属于检测平面。将不属于检测平面的点去除后, 再次计算相对高程中误差, 避免非检测平面点造成的误差。

相对高程中误差 Z_σ 按式(4)计算:

$$Z_\sigma = \pm \sqrt{\frac{\sum_{i=1}^n (Z_i - \bar{Z})^2}{n-1}} \quad (4)$$

其中,

$$\bar{Z} = \frac{\sum_{i=1}^n Z_i}{n} \quad (5)$$

式中:

Z_σ : 相对高程中误差, 单位为米(m);

Z_i : 第*i*个激光点的高程值, 单位为米(m), $i = 1, 2, \dots, n$;

n: 检测平面中激光点个数, 单位为个;

Z: 激光雷达点云高程平均值, 单位为米(m), $i = 1, 2, \dots, n$ 。

3) 平面误差

最大平面位置误差用于评价样本数据中激光雷达点云数据的平面位置与其真实的地面位置之间的最大差别, 利用测区野外检查点数据进行评定。检查点测量与精度应符合 CH/T 1022—2010 中 6.3 相关要求, 检查点应分布均匀, 位置明显, 位于高程突变的目标地物等明显标识处。检查点数量视激光雷达点云密度, 覆盖范围, 覆盖类型等具体情况确定, 宜不少于 15 个; 检查点坐标系应与激光雷达点云坐标系一致。激光雷达点云在 X 方向和 Y 方向的最大平面位置误差也可用于反映激光雷达点云数据的平面位置与其真实的地面位置之间的最大差别。

最大平面位置误差按式(6)计算, X 方向上的最大平面位置误差按式(7)计算, Y 方向上的最大平面位置误差按式(8)计算:

$$d_{XYmax} = \max \left[\sqrt{(X_i - \hat{X}_i)^2 - (Y_i - \hat{Y}_i)^2} \right] \quad (6)$$

$$d_{Xmax} = \max (|X_i - \hat{X}_i|) \quad (7)$$

$$d_{Ymax} = \max (|Y_i - \hat{Y}_i|) \quad (8)$$

式中:

d_{XYmax} : 最大平面位置误差, 单位为米(m);

X_i : 激光点在 X 方向上的平面位置, 单位为米(m), $i = 1, 2, \dots, n$;

n: 检查点个数, 单位为个;

\hat{X}_i : 检查点在 X 方向上的实际平面位置, 单位为米(m), $i = 1, 2, \dots, n$;

Y_i : 激光点在 Y 方向上的平面位置, 单位为米(m), $i = 1, 2, \dots, n$;

\hat{Y} : 检查点在 Y 方向上的实际平面位置, 单位为米(m), $i = 1, 2, \dots, n$;

d_{Xmax} : 激光雷达点云在 X 方向上的最大平面位置误差, 单位为米(m);

d_{Ymax} : 激光雷达点云在 Y 方向上的最大平面位置误差, 单位为米(m)。

平面位置中误差用于评价激光雷达点云数据的平面位置与其真实的地面位置

之间的差别，利用测区野外检查点数据进行评定。平面位置中误差 XY_{RMSE} 按式(9)计算：

$$XY_{RMSE} = \pm \sqrt{X_{RMSE}^2 + Y_{RMSE}^2} \quad (9)$$

其中，

$$X_{RMSE} = \pm \sqrt{\frac{\sum_{i=1}^n (X_i - \hat{X}_i)^2}{n}} \quad (10)$$

$$Y_{RMSE} = \pm \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \quad (11)$$

式中：

XY_{RMSE} ：激光雷达点云平面位置中误差，单位为米(m)；

X_{RMSE} ：激光雷达点云在 X 方向上的平面位置中误差，单位为米(m)；

Y_{RMSE} ：激光雷达点云在 Y 方向上的平面位置中误差，单位为米(m)；

n：检查点个数，单位为个；

X_i ：激光点在 X 方向上的平面位置，单位为米(m)；

\hat{X}_i ：检查点在 X 方向上的实际平面位置，单位为米(m)；

Y_i ：激光点在 Y 方向上的平面位置，单位为米(m)；

\hat{Y}_i ：检查点在 Y 方向上的实际平面位置，单位为米(m)。

相对平面位置中误差用于评价激光雷达点云数据平面位置的内部变形程度，利用测区野外检查点数据进行评定。利用检查点对两点之间的长度变形来评定激光雷达点云相对平面位置中误差，相对平面位置中误差 XY_σ 。按式(12)计算：

$$XY_\sigma = \pm \sqrt{\frac{\sum_{k=1}^n (d_{i,j} - \hat{d}_{i,j})^2}{n}} \quad (12)$$

其中：

$$d_{i,j} = \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2} \quad (13)$$

$$\hat{d}_{(i,j)} = \sqrt{(\hat{X}_i - \hat{X}_j)^2 + (\hat{Y}_i - \hat{Y}_j)^2} \quad (14)$$

式中：

XY_σ ：相对平面位置中误差，单位为米(m)；

K：激光雷达点云数据范围内的检查点对个数，单位为个；

n : 激光雷达点云数据范围内所有检查点对的总数, 单位为个;

$d_{i,j}$: 激光雷达点云中检查点 i 和检查点 j 之间的平面距离, 单位为米(m), $i =$

$1, 2, \dots, n, j = 1, 2, \dots, m$;

m : 检查点个数, 单位为个;

\hat{d} : 检查点 i 和检查点 j 之间的实际地面平面距离, 单位为米(m);

X_i : 检查点 i 在 X 方向上的平面位置, 单位为米(m);

X_j : 检查点 j 在 X 方向上的平面位置, 单位为米(m);

Y_i : 检查点 i 在 Y 方向上的平面位置, 单位为米(m);

Y_j : 检查点 j 在 Y 方向上的平面位置, 单位为米(m)。

5.2.3 融合感知数据评价指标

提高感知系统性能, 需发挥各种传感器相对优势。自动驾驶系统需对不同传感器数据进行融合处理, 包括毫米波雷达与视觉相机融合, 以应对更加复杂的场景。

5.2.3.1 融合感知数据时空一致性评价

1) 空间标定

一般情况下, 传感器需要进行空间的标定, 主要从位置和方向上, 考虑如下 6 个维度:

- a. X 轴: 在三维空间中, X 轴通常代表水平方向的左右移动;
- b. Y 轴: 在三维空间中, Y 轴通常代表水平方向的前后移动;
- c. Z 轴: 在三维空间中, Z 轴通常代表垂直方向的上下移动;
- d. Roll: Roll 是围绕 X 轴的旋转, 也就是前后倾斜;
- e. Pitch: Pitch 是围绕 Y 轴的旋转, 也就是左右倾斜;
- f. Yaw: Yaw 是围绕 Z 轴的旋转, 也就是左右转向。

由于多源数据融合处理需要对点云、图像等感知数据的坐标系进行校准处理, 因此为了避免数据融合过程中, 坐标系之间的转换导致误差积累, 而影响融合后的数据质量, 需要确保传感器空间标定过程涉及的上述 6 个维度上相对于真实世界的

误差满足误差容忍度要求。

2) 时间同步

为了确保数据融合过程中的时间同步，应保证不同传感器的触发时间相差较小，即确保不同进程时间戳代表的真实的时刻一致。

在时间戳对齐后，仍存在多源数据采样频率不完全匹配的情况，在这种情况下，可以使用插值或外推技术将它们的时间分辨率对齐。

5.2.3.2 拼接图像质量评价

由于环视图像由不同角度图像经过拼接而成，在拼接过程中存在错位、重影等影响。相对于前向、后向、侧向单一方向传统的质量评价其相关评价需进行重新定义。通过对全景图像的拼接损失占比、拼接重影占比、拼接错位长度、凭借缝隙宽度以及拼接后图片峰值信噪比为全景图像提供评价标准。

1) 拼接损失占比

拼接损失占比用于评价图像拼接后损失的多少，其表示方法为拼接损失的图像占整个标定黑背棋盘格面积的比值，损失占比越小则说明其图像损失越小。

$$l = \frac{S_l}{S_c} \quad (1)$$

l : 单个棋盘格的损失占比；

S_l : 单个棋盘格损失的面积；

S_c : 棋盘格本身面积。

2) 拼接重影占比

拼接重影占比用于评价图像拼接后重影部分的多少，其表示方法为单个棋盘格重影面积与棋盘格本身面积的比值，拼接重影占比越小则说明其图形重影占比越小。

$$g = \frac{S_g}{S_c} \quad (2)$$

g : 单个棋盘格的损失占比；

S_g : 单个棋盘格重影面积；

S_c : 棋盘格本身面积。

3) 拼接错位长度

拼接错位长度用于评价图像拼接后错位部分的程度，其表示方法为棋盘格的错位距离与棋盘格横向错位距离与棋盘格纵向错位距离平方和开根。

$$d_d = \sqrt{d_x^2 + d_y^2} \quad (3)$$

d_d : 棋盘格的错位距离;

d_x : 棋盘横向错位距离;

d_y : 棋盘纵向错位距离。

4) 拼接缝隙宽度

拼接缝隙宽度用于评价图像拼接后图片边界的重合程度，其表示方法为图像距离与实际距离缝隙宽度差。

5) 拼接后峰值信噪比

拼接后峰值信噪比用于评价图片拼接后细节的保留程度。对于峰值信噪比，值越大，说明拼接后图片和原始图片差别越小，细节保留程度越高。

$$\text{MSE}(I, K) = \frac{1}{mn} \sum_{i=0}^m \sum_{j=0}^n ||I(i, j) - K(i, j)||^2 \quad (4)$$

m、n: 图像宽高;

$I(i, j)$: 对应位置像素。

5.2.3.3 融合图像质量评价

融合图像质量评价通常是指对多个图像进行融合后，评估融合图像的质量，评价融合图像质量的方法可分为主观评价方法和客观评价方法:

主观评价方法通常由依靠人眼对融合图像效果进行主观判断，评估标准主要包括清晰度、对比度、颜色保真度、视觉失真等。

常用的评价融合图像质量的指标可分为图像融合质量的客观评价方法根据对比对象的不同分为三类^[3]:

基于融合图像统计特征的客观评价指标: 该类指标是根据融合图像本身的统计特征值来衡量融合图像质量。常用的评价指标有: 信息熵、空间频率、平均梯度、均值、标准差等。

基于融合图像与理想参考图像之间的客观评价标: 该类指标是通过对比融合图

像与理想参考图像之间的差异，从而对融合图像质量进行客观评价。常用的评价指标有：均方误差、均方根误差、信噪比、峰值信噪比、相关系数等。

基于源图像与融合图像之间的客观评价指标：该类指标主要是针对融合图像与源图像之间联系程度的大小来对图像质量做出客观评价。常用的评价依据有：图像熵值、互信息、边缘信息保持度、通用图像质量、结构相似性、自然场景分析等。

5.3 感知训练数据集评价方法

5.3.1 感知数据集数据类型及分布评价

训练数据集需要对数据集中出现目标类型及出现比例进行统一，如机动车、非机动车、行人、特种车辆等；天气信息至少包含：晴天、雾、阴天、雪、雨等天气；时间段至少包含：白天、凌晨、黑夜、傍晚；交通设施至少包含：交通信号灯、车道线、道路标识、桥梁、不同类型路口等；目标物密度至少包含：无其他目标物、单个目标物、密集目标物等；道路设施至少包括：桥梁、收费站、加油站等；数据集中数据尺寸大小需统一；数据集中命名规则至少包含：目标框边界坐标描述规则、目标 ID 的命名规则等；数据集中目标框与实际目标的重合度(IOU)等；不同运动状态目标物的标注形式等。

评测数据集应做到场景分布多样性，包含标注类别多样，数据标注精度准确等维度进行评测。以下为不同分布维度下，道路场景数据集细类分布。

1) 按天气，气候，时间分布，视角等分布

表 5-1 道路场景数据集环境因素维度分布百分比

分布维度	细类	举例分布±5%
天气维度分布	正常	70%
	雨	10%
	雪	10%
	雾霾	10%
季节分布	春	25%
	夏	25%
	秋	25%
	冬	25%

分布维度	细类	举例分布±5%
时间分布	黎明	10%
	早晨	39%
	下午	31%
	黄昏	4%
	傍晚	16%
	夜晚	40%
视角分布	前视	25%
	侧视	50%
	后视	25%

2) 按交通场景分布

表 5-2 道路场景数据集交通场景维度分布百分比

分布维度	一级分类	举例分布±5%
场景分布	高速	21%
	城市	50%
	郊区	24%
	隧道	4%
	地下停车场	0.10%
	地上停车场	0.10%

3) 按交通参与者统计

表 5-3 道路场景数据集交通参与者维度目标物数量

分布维度	一级分类	目标物数量
交通参与者	两厢车	100w+
	三厢车	100w+
	卡车	100w+
	行人	100w+
	行人-坐	10w+
	摩托车/自行车	100w+
	公交车	100w+

5.3.2 数据集划分

数据集包含训练集、验证集和测试集。训练集用来训练模型，验证集为了确保模型没有过拟合，验证模型在新数据上的表现，同时通过调整超参数，让模型处于最好的状态（验证集非必需的，如无需调整超参数，则直接用测试集来评估效果）。测试集是用来评估模型效果。

5.3.3 数据集划分方式

留出法：按照固定比例将数据集静态的划分为训练集、验证集、测试集；

留一法：每次的测试集都只有一个样本，要进行 m 次训练和预测。这个方法用于训练的数据只比整体数据集少了一个样本，因此最接近原始样本的分布；

k 折交叉验证：一种动态验证的方式，这种方式可以降低数据划分带来的影响，划分具体步骤如下：

1) 将数据集分为训练集和测试集，将测试集放在一边；

2) 将训练集分为 k 份，每次使用 k 份中的 1 份作为验证集，其他全部作为训练集，通过 k 次训练后，得到了 k 个不同的模型，评估 k 个模型的效果，从中挑选效果最好的超参数，使用最优的超参数，然后将 k 份数据全部作为训练集重新训练模型，得到最终模型。

5.3.4 分类问题评估指标

分类问题的评估指标都是围绕二元分类混淆矩阵的四种情况来计算的

混淆矩阵：每一行代表了预测类别，每一行的总数表示预测为该类别的数据的数目；每一列代表了数据的真实归属类别，每一列的数据总数表示该类别的数据实例的数目，如表 5-4 所示：

表 5-4 混淆矩阵

	实际 1	实际 0
预测 1	TP	FP
预测 0	FN	TN

TP(True Positive): 真阳性，实际为 1，且判断为 1（正确）；

FP(False Positive): 假阳性，实际为 0，但判断为 1（错误）；

FN(False Negative): 假阴性，实际为 1，但判断为 0（错误）；

TN(True Negative): 真阴性，实际为 0，且判断为 0（正确）。

根据混淆矩阵的定义，分类问题评估指标和计算方法如下：

(1) 准确率：预测正确的结果占总样本的百分比，

$$\text{准确率} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (1)$$

(2) 精确率（差准率）：所有被预测为正的样本中实际为正的样本的概率，

$$\text{精准率} = \frac{TP}{TP+FP} \times 100\% \quad (2)$$

(3) 召回率（查全率）：实际为正的样本中被预测为正样本的概率，

$$\text{召回率} = \frac{TP}{TP+FN} \times 100\% \quad (3)$$

(4) F1 分数：综合精确率和召回率表现的平衡点，

$$F1 = \frac{2 \times \text{精确率} \times \text{召回率}}{\text{精确率} + \text{召回率}} \quad (4)$$

(5) ROC 曲线：又称接受者操作特征曲线，最早应用于雷达信号检测领域，用于区分信号与噪声。后来人们将其用于评价模型的预测能力，ROC 曲线是基于混淆矩阵得出的，其主要的两个指标是真正率(TPR)和假正率(FPR)，二者计算公式如下：

$$TRP = \text{灵敏度} = \text{召回率} = \frac{TP}{TP+FN} \quad (5)$$

$$FPR = 1 - \text{特异度} = \frac{FP}{FP+TN} \quad (6)$$

(6) AUC：又称曲线下面积，连接对角线，它的面积正好是 0.5，ROC 曲线越陡越好，所以理想值就是 1，一个正方形，而最差的随机判断都有 0.5，所以一般 AUC 的值是介于 0.5 到 1 之间的。AUC 的一般判断标准为：

0.5–0.7：效果较低，例如，可用于预测股票；

0.7–0.85：效果一般；

0.85–0.95：效果很好；

0.95–1：效果非常好，但一般较难实现。

5.3.5 回归问题评估指标

(1) MAE（平均绝对误差）：是最简单的回归度量。它将每个实际值和预测值的差值相加，最后除以观察次数，是目标值和预测值之间的绝对差；

(2) MSE（均方误差）：MSE 取每个实际值和预测值之间的差值，然后将差值平方并将它们相加，最后除以观测数量，它是实际值和预测值之间的平方差；

(3) RMSE（均方根误差）：RMSE 取每个实际值和预测值之间的差值，然后将差值平方并将它们相加，最后除以观测数量。然后取结果的平方根，即 RMSE 是

MSE 的平方根；

(4) MSLE (均值平方对数误差): 计算平方对数误差或损失的期望；

(5) MAPE (平均绝对百分比误差): MAPE 是计算对相对误差损失的预期。

所谓相对误差, 就是绝对误差和真值的百分比;

(6) MSE 会惩罚大错误, 而 MAE 不会。随着 MSE 和 MAE 的值都降低, 模型趋向于一条更好的拟合线。

汽标委智能网联汽车分会

6 标准化建议

目前因感知数据集领域的相关标准较少，尤其是开发车辆感知功能所用的数据集根据目标功能、技术路线等的不同，在行业内普遍存在一定的差异。同时，数据集构建可以主要分为采集和标注两部分，结合现阶段对于测绘数据采集的国家政策法规要求，数据采集部分通常由具备相应资质的企业统一完成，标注工作则由专业标注公司或数据需求方自行完成。因此建议面向数据采集和数据标注分别开展进一步的标准化工作。

数据采集相关工作目前已经开展了“自动驾驶系统测试场景数据采集和分析标准需求研究”项目，建议参考该项目的研究进程，进一步明确相关的标准化需求情况。数据标注现有即将正式实施的推荐性国家标准《GB/T 42755-2023 人工智能 面向机器学习的数据标注规程》，规定了人工智能领域面向机器学习的数据标注框架流程。国内针对车辆感知数据集数据标注的标准仍处于空白状态，建议区别于 GB/T 42755-2023 开展对车辆感知数据的数据标注内容、标注质量等具体指标的行业标准研制工作。考虑到当前行业内面向车辆及自动驾驶的数据集发展尚不完善，行业相关企业的数据集差异较大，建议配合研究成果及后续的数据标注行业标准研制工作，建立“中国智能驾驶标准训练集”供行业广泛参考和应用。

附录

(资料性)

感知数据集数据结构

A 感知数据集数据结构图

感知数据集数据结构应遵循如下架构，见图 A-1:

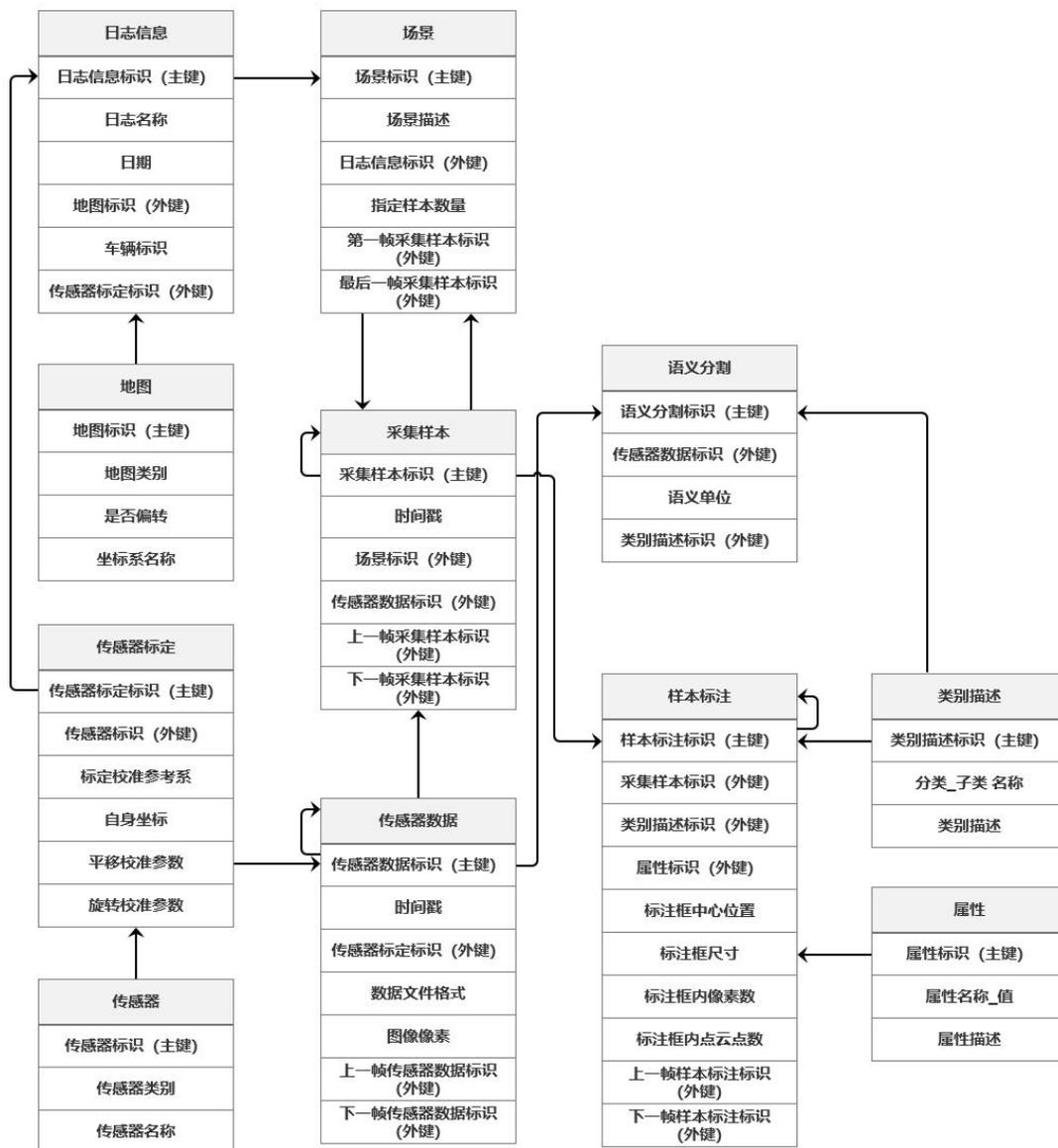


图 A-1 感知数据集数据结构

B 结构元数据说明

- 传感器：具体的传感器类型，数据格式说明见表 B-1。

表 B-1 传感器-数据格式

字段名称	说明
传感器标识（主键）	传感器唯一记录标识符
传感器类别	传感器归属类别（Camera、LiDAR 等）
传感器名称	传感器的名称：品牌_型号

- 传感器标定：以特定参考系为基准校准传感器，数据格式说明见表 B-2。

注：以“传感器标识”与“传感器”建立关联，定向具体传感器标定信息。

表 B-2 传感器标定-数据格式

字段名称	说明
传感器标定标识（主键）	传感器标定唯一记录标识符
传感器标识（外键）	外部键，指向传感器
标定校准参考系	传感器标定所依赖的基准坐标系：车身参考系、WGS84 坐标系等
自身坐标	传感器相对于参考系的自身坐标
平移校准参数	传感器相对于参考系的标定平移参数
旋转校准参数	传感器相对于参考系的标定旋转参数

- 地图：地图基础信息，数据格式说明见表 B-3。

表 B-3 地图-数据格式

字段名称	说明
地图标识（主键）	地图唯一记录标识符，指向具体区域
地图类别	地图表征类别：路口、非机动车道、人行道、高速路段等
是否偏转	默认为 WGS84 坐标系的情况下，是否存在偏转
坐标系名称	坐标系的名称：WGS84 坐标系、GCJ-02 坐标系等

- 日志信息：数据提取内容的日志信息，数据格式说明见表 B-4。

注：以“地图标识”、“传感器标定标识”分别与“地图”、“传感器标定”建立关联，提取信息作为基础日志中的构成部分。

表 B-4 日志信息-数据格式

字段名称	说明
日志信息标识（主键）	日志信息唯一记录标识符
日志名称	日志文件名称
日期	日期（格式：YYYY-MM-DD）
地图标识（外键）	外部键，指向地图
车辆标识	车辆的唯一标识：日志来自于车，则本项必填
传感器标定标识（外键）	外部键，指向传感器标定

- 场景：从日志中提取的场景信息及说明，数据格式说明见表 B-5。

注：以“日志信息标识”与“日志信息”建立关联，由日志中提取用于场景描述的关键信息；同时以“第一帧采集样本标识”、“最后一帧采集样本标识”与采集样

本建立相互关联，定向查询首尾帧的样本信息。

表 B-5 场景-数据格式

字段名称	说明
场景标识（主键）	场景唯一记录标识符
场景描述	对场景的具体描述
日志信息标识（外键）	外部键，指向日志信息
指定样本数量	场景中的样本数量
第一帧采集样本标识（外键）	外部键，指向场景中的第一个采集样本
最后一帧采集样本标识（外键）	外部键，指向场景中的最后一个采集样本

- 传感器数据：例如图像或点云，数据格式说明见表 B-6。

注：以“传感器标定标识”与“传感器标定”建立关联，定向数据来源（传感器）的具体信息；同时以“上一帧传感器数据标识”、“下一帧传感器数据标识”建立自我关联，追溯指定（帧）传感器数据的上下帧信息。

表 B-6 传感器数据-数据格式

字段名称	说明
传感器数据标识（主键）	传感器数据（例如：图像、点云或毫米波）的唯一记录标识符
时间戳	传感器数据的时间戳
传感器标定标识（外键）	外部键，指向传感器标定
数据文件格式	数据帧文件的格式：jpg、pcd 等
图像像素	图像数据帧的像素，宽×高
上一帧传感器数据标识（外键）	外部键。上一帧来自同传感器的样本数据。如果是场景开始则为空
下一帧传感器数据标识（外键）	外部键。下一帧来自同传感器的样本数据。如果是场景结束则为空

- 采集样本：标注的样本帧，数据格式说明见表 B-7。

注：以“场景标识”、“传感器数据标识”分别与“场景”、“传感器数据”关联，为指定帧样本定向场景信息与来源相关信息；同时以“上一帧采集样本标识（外键）”、“下一帧采集样本标识（外键）”建立自我关联，追溯指定样本的上下帧信息。

表 B-7 采集样本-数据格式

字段名称	说明
采集样本标识（主键）	采集样本唯一记录标识符
时间戳	采集样本记录的时间戳
场景标识（外键）	外部键，指向场景信息
传感器数据标识（外键）	外部键，指向传感器数据
上一帧采集样本标识（外键）	外部键，上一帧的采集样本。场景开始位置为空

下一帧采集样本标识（外键）	外部键，下一帧的采集样本. 场景结尾位置为空
---------------	------------------------

● 样本标注：定义样本中所见目标的标注边界框，所有位置数据都是相对于全局坐标系给出的，数据格式说明见表 B-8。

注：以“采集样本标识”、“类别描述标识”、“属性标识”分别与“采集样本”、“类别描述”、“属性”建立关联，为指定采集样本的标注定向标注类别与相关属性信息；同时以“上一帧样本标注标识”、“下一帧样本标注标识”建立自我关联，追溯指定采集样本的上下帧标注信息。

表 B-8 样本标注-数据格式

字段名称	说明
样本标注标识（主键）	样本标注的唯一记录标识符
采集样本标识（外键）	外部键，指向采集样本
类别描述标识（外键）	外部键，指向类别描述
属性标识（外键）	外部键，指向属性信息
标注框中心位置	标注框的中心位置
标注框尺寸	标注框的中心位置
标注框内像素数	标注框内的像素数
标注框内点云点数	标注框内的雷达扫描点数
上一帧样本标注标识（外键）	外部键，上一时刻来自同一个样本的标注标识，如果这是此对象的第一个标注，则为空
下一帧样本标注标识（外键）	外部键，下一时刻来自同一个样本的标注标识，如果这是此对象的最后一个标注，则为空

● 语义分割：定义图像或点云中单位像素或点的语义，数据格式说明见表 B-9。

注：以“传感器数据标识”与“传感器数据”建立关联，定向语义分割的对象数据信息；同时以“类别描述标识”与“类别描述”建立关联，引用类别描述信息。

表 B-9 语义分割-数据格式

字段名称	说明
语义分割标识（主键）	语义分割的唯一记录标识
传感器数据标识（外键）	外部键，指向传感器数据
语义单位	建立语义分割的单位大小
类别描述标识（外键）	外部键，指向类别描述

● 类别描述：物体类别的分类（例如车辆、交通标志等），子类别由“_”划分，数据格式说明见表 B-10

表 B-10 类别描述-数据格式

字段名称	说明
类别描述标识（主键）	类别描述的唯一记录标识

分类_子类_名称	分类名称：车辆_轿车、交通标志_让行
类别描述	类别具体的描述

● 属性：标注的属性说明，在类别保持不变的情况下可以更改，数据格式说明见表 B-11

表 B-11 属性-数据格式

字段名称	说明
属性标识（主键）	属性的唯一记录标识
属性名称_值	属性名称与值：遮挡、截断、颜色等
属性描述	属性具体的描述

汽标委智能网联汽车分委会

参考文献

- [1] R. Girshick, J. Donahue, T. Darrell, et al. " Rich feature hierarchies for accurate object detection and semantic segmentation, " in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580-587.
- [2] R. Girshick. " Fast r-cnn, " in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1440-1448.
- [3] K. Simonyan and A. Zisserman, " Very deep convolutional networks for large-scale image recognition, " arXiv preprint arXiv:1409.1556, 2014.
- [4] J. Redmon and A. Farhadi, " YOLO9000: better, faster, stronger, " in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 7263-7271.
- [5] J. Redmon and A. Farhadi, " Yolov3: An incremental improvement, " arXiv preprint arXiv:1804.02767, 2018.
- [6] T.-Y. Lin, P. Dollár, R. Girshick, et al. " Feature pyramid networks for object detection, " in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2117-2125.
- [7] Z. Tian, C. Shen, H. Chen, et al. " Fcos: Fully convolutional one-stage object detection, " in Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp.9627-9636.
- [8] T. Kong, F. Sun, H. Liu, et al. " Foveabox: Beyond anchor-based object detection, " IEEE Transactions on Image Processing, vol. 29, pp. 7389-7398, 2020.
- [9] H. Law and J. Deng, " Cornernet: Detecting objects as paired keypoints, " in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 734-750.
- [10] X. Zhou, J. Zhuo, and P. Krahenbuhl, " Bottom-up object detection by grouping extreme and center points, " in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 850-859.
- [11] K. Duan, S. Bai, L. Xie, et al. " Centernet: Keypoint triplets for object detection, " in Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 6569-6578.
- [12] Voigtlaender, Paul and Luiten, et al., Siam R-CNN: Visual Tracking by Re-0Detection, Proceedings of the IEEE/CVF Conference on CVPR, 2020
- [13] Y. Yuan, X. Chen, and J. Wang, " Object-contextual representations for semantic segmentation, " in Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16, 2020, pp. 173-190: Springer.
- [14] Parashar A, Rhu M, Mukkara A, et al. SCNN: An accelerator for compressed-sparse convolutional neural networks[J]. ACM SIGARCH computer architecture news, 2017, 45(2): 27-40.
- [15] Zheng, Tu, Hao Fang, Yi Zhang, et al. "RESA: Recurrent Feature-Shift Aggregator for Lane Detection." ArXiv:2008.13719 [Cs], March 25, 2021.
- [16] Li, Xiang, Jun Li, et al. "Line-CNN: End-to-End Traffic Line Detection With Line Proposal Unit." IEEE Transactions on Intelligent Transportation Systems 21, no. 1

(January 2020): 248–58.

[17] Tabelini, Lucas, Rodrigo Berriel, et al. “Keep Your Eyes on the Lane: Real-Time Attention-Guided Lane Detection.” ArXiv:2010.12035 [Cs], November 17, 2020.

[18] Liu, Lizhe, Xiaohao Chen, et al. “CondLaneNet: A Top-to-down Lane Detection Framework Based on Conditional Convolution.” ArXiv:2105.05003 [Cs], June 10, 2021.

[19] Qin, Zequn, Huanyu Wang, et al. “Ultra Fast Structure-Aware Deep Lane Detection.” ArXiv:2004.11757 [Cs], August 4, 2020.

[20] Tabelini, Lucas, Rodrigo Berriel, et al. “PolyLaneNet: Lane Estimation via Deep Polynomial Regression.” ArXiv:2004.10924 [Cs], July 14, 2020.

[21] Feng, Zhengyang, Shaohua Guo, et al. “Rethinking Efficient Lane Detection via Curve Modeling.” ArXiv:2203.02431 [Cs], March 4, 2022.

[22] Qu, Zhan, Huan Jin, et al. “Focus on Local: Detecting Lane Marker from Bottom Up via Key Point.” arXiv, May 28, 2021.

[23] Wang, Jinsheng, Yinchao Ma, et al. “A Keypoint-Based Global Association Network for Lane Detection.” ArXiv:2204.07335 [Cs], April 15, 2022.

[24] H. G. Jung, D. S. Kim, P. J. Yoon, et al. “Parking slot markings recognition for automatic parking assist system,” in Proc. IEEE Intell. Veh. Symp., 2006, pp. 106-113.

[25] Y. Tanaka, M. Saiki, M. Katoh, et al. “Development of image recognition for a parking assist system,” in Proc. World Congr. Intell. Transp. Syst. Serv., 2006, pp. 1-7.

[26] S. Houben, M. Komar, A. Hohm, et al. “On-Vehicle Video-Based Parking Lot Recognition with Fisheye Optics,” in Proc. Int. IEEE Conf. Intell. Transp. Syst., 2013, pp. 7-12.

[27] C. Wang, H. Zhang, M. Yang, et al. “Automatic parking based on a bird’s eye view vision system,” Adv. Mech. Eng., vol. 2014, pp. 1-13.

[28] X. Du and K. K. Tan, “Autonomous reverse parking system based on robust path generation and improved sliding mode control,” IEEE Trans. Intell. Transp. Syst., 2015, vol. 16, no. 3, pp. 1225-1237.

[29] J. K. Suhr, and H. G. Jung, “Automatic Parking Space Detection and Tracking for Underground and Indoor Environments,” IEEE Trans. Ind. Electron., 2016, vol. 63, no. 9, pp. 5687-5698.

[30] S. Lee and S. W. Seo, “Available parking slot recognition based on slot context analysis,” IET Intell. Transp. Syst., 2016, vol. 10, no. 9, pp. 594-604.

[31] S. Lee, D. Hyeon, G. Park, et al. “Directional-DBSCAN: Parking-slot Detection using a Clustering Method in Around-View Monitoring System,” in Proc. IEEE Intell. Veh. Symp., 2016, pp. 349-354.

[32] J. K. Suhr and H. G. Jung, “A Universal Vacant Parking Slot Recognition System Using Sensors Mounted on Off-the-Shelf Vehicles,” Sensors, 2018, vol. 18, no. 4, pp. 1213-1-1213-21.

[33] L. Li, L. Zhang, X. Li, et al. “Vision-based parking-slot detection: A benchmark and a learning-based approach,” in IEEE International Conference on Multimedia and Expo (ICME), 2017, pp. 649-654

- [34] J. Huang, L. Zhang, Y. Shen, et al. Dmpr-ps: A novel approach for parking-slot detection using directional marking-point regression, in 2019 IEEE International Conference on Multimedia and Expo (ICME), 2019, pp. 212–217.
- [35] H. Do and J. Y. Choi, "Context-Based Parking Slot Detection With a Realistic Dataset," in IEEE Access, 2020, vol. 8, pp. 171551-171559.
- [36] C. Min, J. Xu, L. Xiao, et al. "Attentional Graph Neural Network for Parking-Slot Detection," in IEEE Robotics and Automation Letters, 2021, vol. 6, no. 2, pp. 3445-3450.
- [37] David E, Christian P, Rob F. Depth map prediction from a single image using a multi-scale deep network[C]. The 28th Conference on Neural Information Processing Systems, Montréal, Canada, December 8-13, 2014.
- [38] David E, Rob F. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture[C]. The 15th International Conference on Computer Vision, Santiago, Chile, December 13-16, 2015.
- [39] Zhou T H, Matthew B, Noah S, et al. Unsupervised learning of depth and ego-motion from video[C]. The 30th IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, USA, July 22-25, 2017.
- [40] Clement G, Oisin M A, Gabriel J B. Unsupervised monocular depth estimation with left-right consistency[C]. The 29th IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, June 19-21, 2016.
- [41] G. P. Meyer, A. Laddha, E. Kee, et al. "LaserNet: An Efficient Probabilistic 3D Object Detector for Autonomous Driving," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 12669-12678, doi: 10.1109/CVPR.2019.01296.
- [42] Yang B, Luo W, Urtasun R. PIXOR: Real-time 3D Object Detection from Point Clouds[J]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.
- [43] Lang A H, Vora S, Caesar H, et al. PointPillars: Fast Encoders for Object Detection From Point Clouds[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019.
- [44] Zhou Q, Yu C. Point RCNN: An Angle-Free Framework for Rotated Object Detection[J]. Remote Sensing, 2022, 14.
- [45] Ali W, Abdelkarim S, Zahran M, et al. YOLO3D: End-to-end real-time 3D Oriented Object Bounding Box Detection from LiDAR Point Cloud[J]. ECCV 2018: "3D Reconstruction meets Semantics" workshop, 2018.
- [46] Zhou Y, Tuzel O. VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018.
- [47] Shi S, Guo C, Jiang L, et al. PV-RCNN: Point-Voxel Feature Set Abstraction for 3D Object Detection[J]. 2019.
- [48] Yin T, Zhou X, P Krhenbühl. Center-based 3D Object Detection and Tracking[J]. 2020.
- [49] Wang S, Sun Y, Liu C, et al. PointTrackNet: An End-to-End Network For 3-D Object Detection and Tracking From Point Clouds[J]. IEEE Robotics and

Automation Letters, 2020, 5(2):3206-3212.

[50] Sheno A , Patel M , Gwak J Y , et al. JRMOT: A Real-Time 3D Multi-Object Tracker and a New Large-Scale Dataset[J]. 2020.

[51] Weng X , Wang J , Held D , et al. AB3DMOT: A Baseline for 3D Multi-Object Tracking and New Evaluation Metrics[J]. 2020.

[52] Hu Q , Yang B , Xie L , et al. RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds[J]. 2019.

[53] Rethage D , Wald J , Sturm J , et al. Fully-Convolutional Point Networks for Large-Scale Point Clouds: 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part IV[C], 2018.

[54] Jung J , Bae S H . Real-time road lane detection in Urban areas using LiDAR data[J]. Electronics, 2018, 7(11).

[55] Chen X , Ma H , Wan J , et al. Multi-View 3D Object Detection Network for Autonomous Driving[J]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[56] Qi C R , Wei L , Wu C , et al. Frustum PointNets for 3D Object Detection from RGB-D Data[J]. 2017.

[57] Xu D , Anguelov D , Jain A . PointFusion: Deep Sensor Fusion for 3D Bounding Box Estimation[J]. IEEE, 2018.

[58] Frossard D, Urtasun R. End-to-end Learning of Multi-sensor 3D Tracking by Detection[J]. 2018:635-642.

[59] Zhang W , Zhou H , Sun S , et al. Robust Multi-Modality Multi-Object Tracking[C]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2019.

[60] Kim A , Oep A , L Leal-Taixé. EagerMOT: 3D Multi-Object Tracking via Sensor Fusion[J]. 2021.