

DOI: 10.3901/JME.2023.20.281

# 智能汽车环境感知方法综述\*

彭湃 耿可可 王子威 柳智超 殷国栋  
(东南大学机械工程学院 南京 211189)

**摘要:** 智能汽车是全球汽车产业的未来发展方向,是推动我国自主汽车产业高质量发展的应有之义。智能汽车依靠人工智能、泛在传感等先进技术的赋能,实现对驾驶人认知感知、决策规划及控制执行的增强或替代。对道路环境的实时、精准和鲁棒的感知是汽车智能化的基石,近十年间智能汽车感知领域的巨大飞跃多是由深度学习技术推动的。针对近年智能汽车环境感知技术的发展现状,首先梳理智能汽车环境感知系统软硬件架构,对感知、计算设备以及算法部署平台进行总体概述;其次,围绕目标检测与语义分割、多目标跟踪、目标意图识别与轨迹预测、环境建图四方面关键任务,总结近年具有里程碑意义的研究方法与技术看方;最后,分析当前智能汽车环境感知技术所面临的问题和挑战,并对未来发展趋势与关键技术进行展望。  
**关键词:** 智能汽车;环境感知;目标检测与语义分割;跟踪和预测;环境建图  
**中图分类号:** U46

## Review on Environmental Perception Methods of Autonomous Vehicles

PENG Pai GENG Keke WANG Ziwei LIU Zhichao YIN Guodong  
(School of Mechanical Engineering, Southeast University, Nanjing 211189)

**Abstract:** Autonomous vehicles are the future development direction of the global automotive industry and are essential for promoting the high-quality development of China's independent automotive industry. Autonomous vehicles rely on advanced technologies such as artificial intelligence and ubiquitous sensing to enhance or replace the driver's cognitive perception, decision-making planning, and control execution. Real time, accurate, and robust perception of the road environment is the cornerstone of automotive intelligence, and the huge leap in the field of autonomous vehicles perception in the past decade has been mostly driven by deep learning technology. A review of the development of autonomous vehicles environmental awareness technology in recent years is provided. Firstly, it summarizes the software and hardware architecture of autonomous vehicles environmental awareness systems, and provides an overall overview of perception, computing devices, and algorithm deployment platforms; Secondly, milestone research methods and technical solutions in recent years were summarized around four key tasks: object detection and semantic segmentation, multi objects tracking, object intention recognition and trajectory prediction, and environmental mapping; Finally, the problems and challenges faced by current autonomous vehicles environmental perception system were analyzed, and future development trends and key technologies were prospected.  
**Key words:** autonomous vehicles; environmental perception; object detection and semantic segmentation; tracking and prediction; environmental mapping

## 0 前言

当今世界正经历百年未有之大变局,新一轮科技革命和产业变革方兴未艾,智能汽车已成为全球

汽车产业发展的战略方向<sup>[1-3]</sup>。发展智能汽车能够极大地推动我国汽车产业转型升级,壮大经济增长新动能,其相关技术的突破利于提升我国产业基础能力,增强新一轮科技革命和产业变革引领能力,可加快制造强国、科技强国、网络强国、交通强国、智慧社会等国家战略规划的建设,对增强新时代我国的综合实力具有重要的战略意义。为此,相关政

\* 国家自然科学基金资助项目(51975118, 52025121, 52272414)。20230613收到初稿,20230830收到修改稿

府部门相继发布了《交通强国建设纲要》、《智能汽车创新发展战略》等一系列政策文件,指导我国智能汽车产业深化试点示范,进一步完善政策环境和相关基础设施建设,持续推动我国智能汽车实现高质量发展。

智能汽车主要依靠人工智能、视觉计算、雷达和全球定位及网联通信等技术,使汽车具备自主环境感知<sup>[4-5]</sup>、决策规划<sup>[6]</sup>和控制<sup>[7-8]</sup>的能力。当前,智能汽车面临的最基础但又最具挑战性的任务,就是对道路交通环境进行实时、精确且鲁棒的感知。为此,智能汽车通过配备多种模态的传感器,对道路交通各类目标的类别、位置、运动状态、轨迹等进行精准检测,进而为智能汽车添加一双全局的眼睛。本文首先简要介绍了环境感知系统软硬件架构,然后归纳梳理了近年来智能汽车环境感知领域的国内外最新研究成果,如图 1 所示,从目标检测与语义分割、多目标跟踪、意图识别与轨迹预测、环境建图四个方面关键任务及技术方法进行详细阐述,最后对智能汽车环境感知系统研究进行了总结和展望。

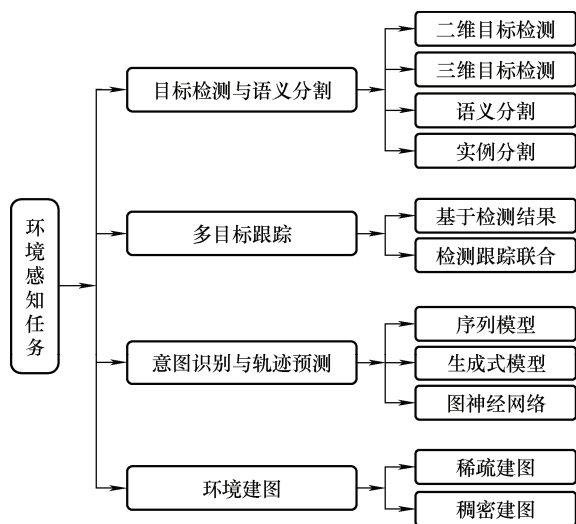


图 1 环境感知关键任务与方法

## 1 智能汽车环境感知系统架构

为了解决智能驾驶与高级辅助驾驶中的关键感知任务,智能汽车环境感知系统通过多种传感器获取周围信息,使用多种计算设备与不同的软件框架搭载算法,完成对感知信息的获取,为后续的决策、规划提供基础。本节从硬件架构与软件架构两个方面对智能汽车环境感知系统进行简介。

### 1.1 硬件架构

#### 1.1.1 传感器

随着仪器科学的发展,车载传感设备不断地升级与完善,这些传感器可以获取不同模态的环境信息,通过多种模态感知数据的融合,为不同的感知任务提供对应解决方案。目前主流传感设备包括可见光相机、激光雷达、毫米波雷达、超声波雷达以及红外热像仪等,其特点如表 1 所示。

表 1 各类传感器特点总结

感知设备	语义信息	空间信息	探测距离	成本	抗噪能力
可见光相机	高	低	较远	低	弱
激光雷达	中	高	中	高	中
毫米波雷达	中	高	远	中	强
超声波雷达	低	中	近	低	弱
红外热像仪	较高	低	较远	高	强

可见光相机应用最为广泛,能够提供道路交通环境的视觉信息,可用于检测各种障碍物、车道线及交通标志等,部署成本低但距离信息精度不高;激光雷达提供高精度的三维点云数据,可用于障碍物检测和车辆定位,但目前依旧受限于成本,市场规模有限;毫米波雷达应用较为成熟,具有探测远、响应快和成本低的特点,但对静止物体检测能力较弱;超声波雷达适用于短距离感知任务,如自动泊车和防撞预警,成本低廉,已规模化应用;红外热像仪在低照度环境下表现出色,能够实现全天候目标检测,是较为新兴的智能汽车车载传感设备。

#### 1.1.2 车载计算设备

智能汽车搭载的车载计算设备通常为工控机或嵌入式计算设备,用于车载传感器信息采集、处理与车辆控制等,相较于一般的计算设备,具有体积小、抗振强、耐高温、低功耗等特点,可在恶劣环境中保持稳定工作。常用的嵌入式计算设备包括英伟达 Jetson 系列、华为 Atlas 系列、赛灵思 Zynq 系列等,均拥有较大算力,可实现深度学习模型的部署。

### 1.2 软件架构

智能汽车环境感知系统的软件架构主要包含三个模块,即传感器数据处理模块、感知算法部署模块以及数据通信模块,其关系如图 2 所示。

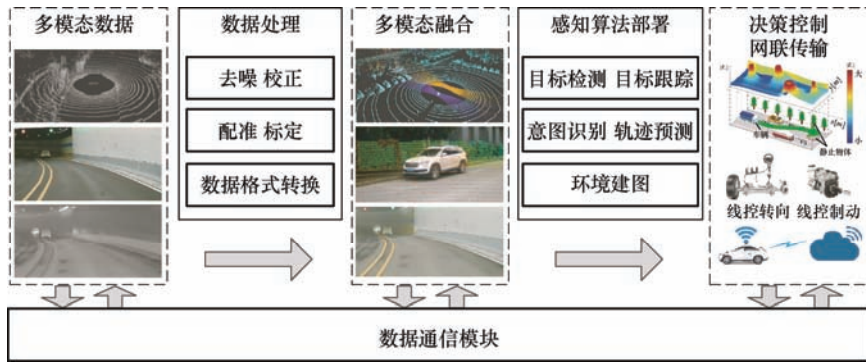


图2 智能汽车环境感知系统软件架构

### 1.2.1 传感器数据处理

智能汽车配备、使用的传感器种类丰富、数量不一，在获取到传感器的原始数据后，为了适配后续的多模态感知算法，往往需要数据处理模块对传感器的原始数据进行预处理。处理的过程包括对单个传感器原始数据的去噪与校正<sup>[9-10]</sup>，对多个传感器间的融合关系进行的配准与标定<sup>[11-12]</sup>，对不同数据格式进行的转换等。处理完成后的数据便于感知算法使用，进而完成各项智能汽车环境感知任务。

### 1.2.2 感知算法部署

目前主流的智能汽车环境感知算法基本都基于GPU架构的深度学习计算，对多种传感器的数据进行特征提取与多种任务的实现。常用的深度学习训练、部署环境包括PyTorch、Tensorflow、Caffe、Keras、MXnet等，通常使用Python、Matlab、C++等编程语言，调用其提供的API，完成模型的搭建与训练，实现包括目标检测、跟踪、状态预测等在内的多种智能感知任务<sup>[13-14]</sup>。

### 1.2.3 数据通信

智能汽车环境感知系统是复杂的信息交互系统，无论是传感器间的信息传递，还是感知信息与其他系统的交互，都对软件平台的信息处理能力有较高的要求。当前相关企业推出了一系列车控软件平台，用于在智能汽车上完成各模块间的交互任务，主流的软件平台包括特斯拉Autopilot、华为MDC、英伟达DRIVE、百度Apollo等。其中百度Apollo是完全开放的软件平台，其基于机器人操作系统(Robot operating system, ROS)，并弥补了ROS系统在大数据传输、网络通讯架构等方面的缺陷。

## 2 环境感知任务及其方法

环境感知是车辆智能化的基础，其关键任务包含目标检测与语义分割、多目标跟踪、意图识别与

轨迹预测、环境建图等，围绕以上四个方面关键任务，学术界及工业界进行了一系列开发与研究，提出了不同的环境感知算法。

### 2.1 目标检测与语义分割

#### 2.1.1 目标检测

目标检测是从背景中分离出所有感兴趣的目標，并确定目标的具体类别和位置。由于各类目标的外观、大小、位姿和尺度差异，外加天气、光照、遮挡等因素影响，实时、精确的目标检测一直是智能驾驶环境感知领域的挑战之一。按照输出的定位边界框维度划分，可分为二维目标检测和三维目标检测，如图3所示。

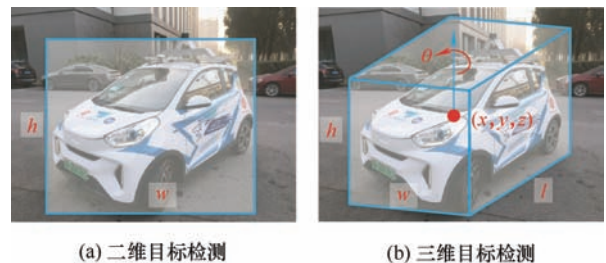


图3 二维和三维目标检测

#### 2.1.1.1 二维目标检测

对于二维目标检测，主要采用视觉传感器作为感知设备，对图像进行特征提取、分类和定位。如图4所示，其发展历史以2014年为分水岭可以分为两个阶段：传统的基于机器学习的目标检测阶段和当前基于深度学习的目标检测阶段。

传统目标检测方法<sup>[15-17]</sup>主要包括候选区域选择、特征提取以及目标分类三个流程。首先通过多尺度滑动窗口<sup>[18]</sup>框选出候选区域，然后利用VJ<sup>[19]</sup>、Haar<sup>[20]</sup>、HOG<sup>[21]</sup>、DPM<sup>[22]</sup>等特征算子提取图像特征，最后采用自适应增强(Adaptive boosting, AdaBoost)<sup>[23]</sup>、支持向量机(Support vector machine, SVM)<sup>[24]</sup>、级联学习<sup>[25]</sup>等分类器对特征进行分类输出。虽然传统目标检测方法仅能提取图像低维特征，

泛化性能、检测精度和速度难以满足智能驾驶需求,但是其中的经典思想和技术,如特征金字塔、边界框回归、非极大值抑制、难分负样本挖掘等,对后

来的深度学习目标检测框架产生了深远影响,可以说,深度学习目标检测是站在传统目标检测这个巨人肩膀上的。

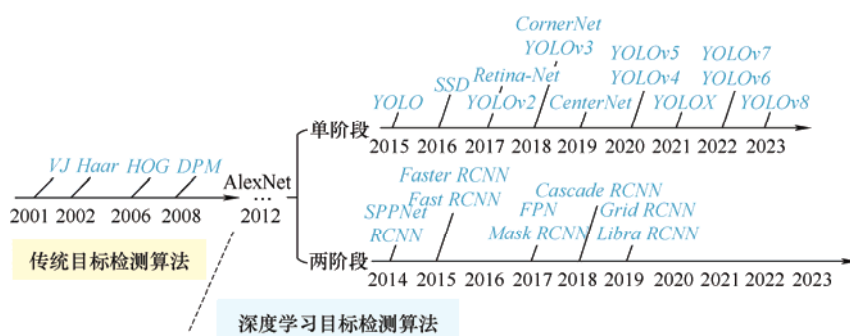


图4 图像目标检测算法发展历程

2014年以来,得益于强大的特征提取、表达和学习能力,深度神经网络在目标检测领域取得了一系列突破性进展,极大地推动了智能汽车环境感知

技术的发展。当前,深度学习目标检测方法主要有两阶段检测算法和单阶段检测算法两条技术路线,如图5所示。

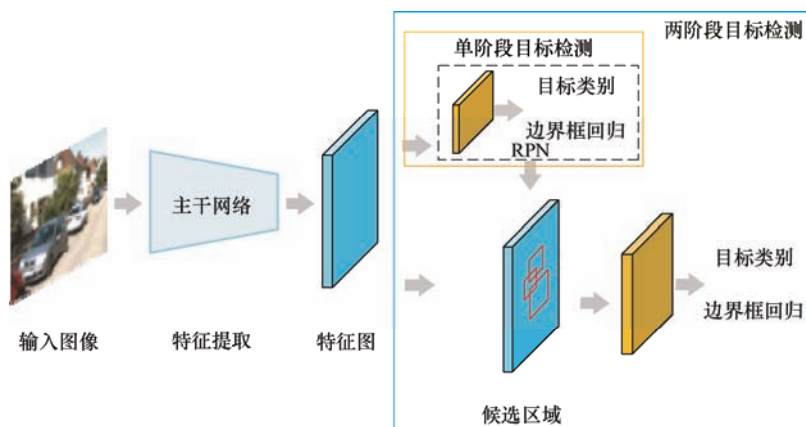


图5 两阶段和单阶段目标检测网络架构

两阶段检测算法继承了传统目标检测算法流程,首先生成一系列候选框来确定可能存在目标的区域,然后对每个候选框进行回归定位与分类。作为深度学习目标检测的开山之作,GIRSHICK等<sup>[26]</sup>提出的两阶段目标检测网络R-CNN首先通过选择性搜索算法<sup>[27]</sup>生成2 000个候选区域,然后将这些候选区域缩放到统一尺寸后利用卷积神经网络提取特征,最后使用分类器对区域内目标进行分类。R-CNN在Pascal VOC 2007数据集上的平均检测精度(mean Average Precision, mAP)相比于当时最先进的传统目标检测方法DPM-v5的33.7%提高了25%,取得了质的飞跃。此后的SPPnet<sup>[28]</sup>、Fast-RCNN<sup>[29]</sup>、Faster-RCNN<sup>[30]</sup>、FPN<sup>[31]</sup>、Cascade RCNN<sup>[32]</sup>等采用更统一和先进的网络架构,更快速的候选框生成算法以及更深层的特征提取网络,提升了目标检测的速度和精度。然而,对于计算资源有限的智能驾驶

车辆来讲,虽然两阶段方法精度很高,但是其高昂的计算成本、较低的检测速度等问题成为制约该类算法部署在智能汽车上的关键因素。

对于单阶段目标检测方法,其将目标检测视为回归或分类问题,采用一个统一的端到端框架直接输出目标类别和位置,相比于两阶段检测方法其精度有所降低,但是显著提高了检测速度。其中,最著名的单阶段目标检测网络莫过于YOLO(You only look once)系列<sup>[33]</sup>。YOLO由REDMON等<sup>[34]</sup>在2015年提出,其将单个神经网络应用于整幅图像,把图像划分为多个区域,并同时预测每个区域内目标的边界框坐标及置信度。YOLO在Pascal VOC 2007数据集上最快能运行到每秒155帧,并且平均检测精度为52.7%,而同期的Faster-RCNN最快目标检测速度仅为每秒17帧,平均检测精度为59.9%。到目前为止,YOLO系列已经更新到YOLOv8,在检



测速度和精度方面优于绝大多数现有目标检测算法, 在 COCO 数据集上实现了 53.9% 的平均检测精度(YOLOv5 为 50.7%), 检测速度达到每秒 280 帧。除了 YOLO 系列之外, SSD<sup>[35]</sup>、CornerNet<sup>[36]</sup>、CenterNet<sup>[37]</sup>、Efficientdet<sup>[38]</sup>等也是很有代表性的单阶段检测算法。

由于其高检测速度和易部署特性, 当前有很多研究基于单阶段检测算法进行智能汽车目标检测。LI 等<sup>[39]</sup>针对训练数据和测试数据之间的域漂移问题, 基于 YOLOv5 的设计了类别一致的正则化模块以及图像级和实例级特征的自适应模块, 可以对齐辅助域和目标域之间的特征分布, 显著提高智能驾驶应用的各种域漂移场景下的目标检测性能。XU 等<sup>[40]</sup>针对智能驾驶场景下 SSD 对小目标检测精度不理想的问题, 以及单阶段检测算法在处理密集检测器训练过程中遇到的极端前景-背景类不平衡的问题, 通过结合空洞卷积和特征融合改进了 SSD 的网络结构, 扩大了接收域, 丰富了浅层的语义信息,

提高了检测精度。WANG 等<sup>[41]</sup>针对道路小目标检测和遮挡问题, 重新设计了 CenterNet 的主干网络和检测头网络, 提出了平均边界模型, 利用边界特征信息定位目标, 实车试验验证了该网络的实时性和鲁棒性。CHEN 等<sup>[42]</sup>分析了几种主流目标检测架构, 包括 Faster R-CNN、R-FCN 和 SSD 等, 以及几种典型的特征提取网络 ResNet、MobileNet、Inception 等, 基于 KITTI 数据集系统地评估了几种目标检测架构和特征提取网络的速度-精度-内存权衡问题。

### 2.1.1.2 三维目标检测

对智能驾驶系统而言, 仅依靠对图像平面的二维目标检测是不够的, 还需要更精确的三维空间定位和尺寸估计, 即三维目标检测。相比于二维目标检测, 其增加了目标三维尺寸、深度、姿态等信息的估计。根据其使用的数据类型, 三维目标检测可以分为基于图像的方法, 基于点云的方法以及图像和点云融合方法, 典型框架如图 6 所示。

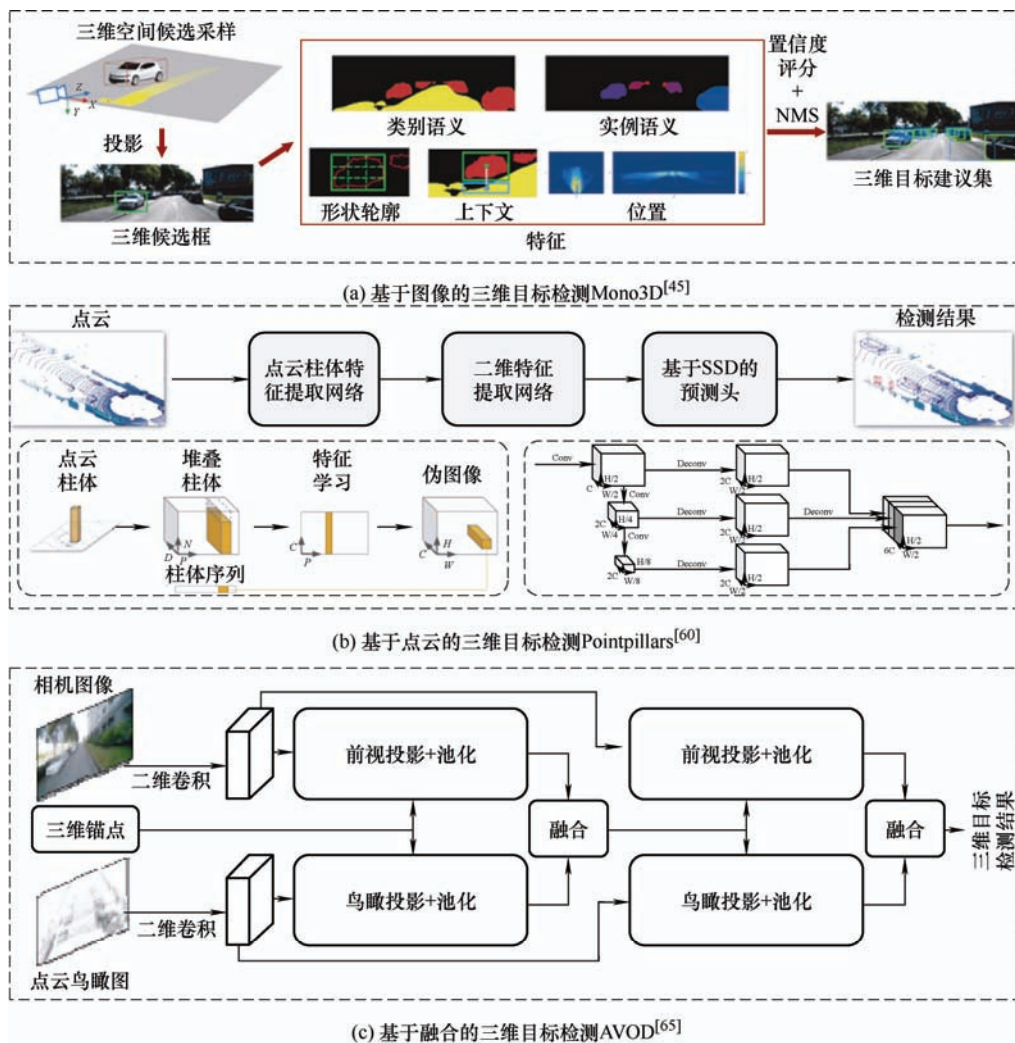


图 6 典型三维目标检测框架

基于图像的方法可分为基于单目视觉和基于双目立体视觉两种方法。基于单目视觉的三维目标检测难点在于二维图像缺乏场景深度信息,因此需要通过目标的几何模型<sup>[43]</sup>或者多视角约束<sup>[44]</sup>来补偿深度信息。CHEN 等<sup>[45]</sup>基于目标与地面接触的几何约束生成候选框,结合语义分割、上下文信息、大小和位置先验以及典型目标形状对候选框进行评分,对高分候选框进行细化得到最终的检测结果,但是该方法在损失计算中存在误差累积问题,因而检测精度不是很高,同时由于候选框密集采样和融合多个先验特征使得网络计算量较大,检测速度较慢。MOUSAVIAN 等<sup>[46]</sup>提出了一种依赖先验集合知识的 Deep3Dbbox 网络,相比于 Mono3D<sup>[45]</sup>,其利用二维目标检测方法简化了网络结构,降低了计算量,提升了推理速度,但由于深度信息缺失,检测精度并没有太多提升。

以上算法都没有考虑目标检测过程中对于遮挡、截断问题的处理。为此, XIANG 等<sup>[47]</sup>提出了一种新型三维体素模式(3D voxel pattern, 3DVP),通过像素强度值、体素形式三维形状以及遮挡掩膜来建立目标的三维模型,所构建的检测框架能够估计图像目标的三维姿态、形状以及目标间的遮挡关系。在此工作的基础上,他们又提出 SubCNN<sup>[48]</sup>网络,通过引入多尺度图像金字塔提高了小目标检测的精度。KUNDU 等<sup>[49]</sup>通过从 CAD 模型集合中学习低维形状空间来利用特定目标的形状先验,提出 3D-RCNN 网络从二维图像中重建三维目标的形状和姿态。但是以上方法需要目标先验的三维模型,而模型的获取较为困难,且这些方法的多目标检测精度也较低。

单目视觉三维目标检测因其在数据收集和传感器成本上的巨大优势,依然是不少领域中主流的选择方案。但是先天缺乏深度信息,其三维目标检测精度始终有所不足,且在目标遮挡和远距离检测场景下存在不少困难。

相比于单目视觉算法,基于双目立体视觉的三维目标检测利用双目或者深度相机获取较为准确的深度信息,其检测精度有明显提升,检测方法可以分为两类:① 基于视觉图像和深度图的双模态深度融合方法;② 基于双目图像三维空间卷积网络的方法。CHEN 等<sup>[50]</sup>提出了一种基于可见光图像和深度图像 HHA<sup>[51]</sup>融合检测的 3DOP 网络,结合智能驾驶领域的三维目标先验特征(语义信息、点云密度、上下文信息等),能够产生高质量的目标三维边界框和位姿信息。但是像 HHA 之类特征图的生成会带来

额外的计算量,因此,很多研究着眼于直接利用双目图像进行三维目标检测。QIN 等<sup>[52]</sup>提出了一种面向双目视觉三维目标检测的三角测量学习网络 TLNet,无需由双目图像生成深度图输入。该网络使用一个三维锚框构建双目图像感兴趣区域间的目标对应关系,并从中学习对锚框附近的目标进行三角测量以及三维目标检测。LI 等<sup>[53]</sup>基于 Faster RCNN 提出了 Stereo R-CNN 三维目标检测网络,通过两个 FPN 网络以及立体区域建议网络,对双目图像进行目标检测并生成目标关联对,然后使用左侧图像感兴趣区域特征来预测三维目标关键点,最后使用三维边界框、二维边界框和关键点之间的投影关系获取三维检测结果。

尽管双目视觉能够获取目标的深度信息,但是相比较于激光雷达的直接测距模式存在固有缺陷。对于智能驾驶而言,目标的三维空间位置和尺寸的估计精度对行驶安全相当重要,而激光雷达能够提供大范围的高精度点云数据,因而很多研究基于激光雷达进行智能驾驶三维目标检测。目前基于点云的三维目标检测方法根据处理点云数据的方式划分,主要有点云投影、点云体素和原始点云三种形式,如图 7 所示。

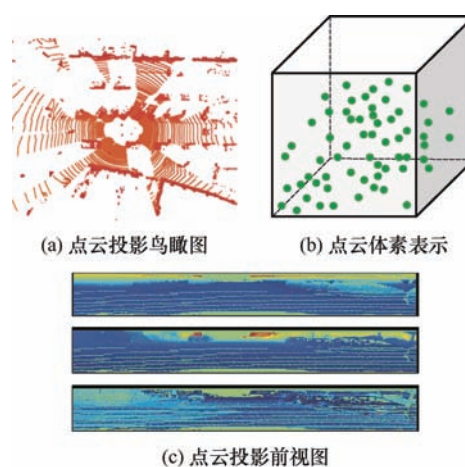


图 7 基于点云的三维目标检测典型数据表示方式

基于点云投影的方法首先将三维点云转换为二维图像,然后利用二维目标检测算法进行处理,最后使用位置和维度回归恢复目标检测三维边界框。SIMONY 等<sup>[54]</sup>提出了 Complex-YOLO 三维目标检测网络,首先将点云投影为彩色鸟瞰图,RGB 通道值分别为点云高度、反射强度和密度,然后基于 YOLO 网络对鸟瞰图进行特征提取,最后使用区域建议网络进行三维边界框回归。YANG 等<sup>[55]</sup>提出的 PIXOR 网络则是按不同高度将点云投影为鸟瞰图,然后使用二维全卷积网络估计目标的位置和姿态。



有研究使用点云的前视投影图<sup>[56]</sup>, 利用二维目标检测框架进行特征提取, 但是由于前视图的遮挡和尺度变化问题, 其检测效果不是很好, 需要结合鸟瞰图实现三维目标检测。

基于点云体素的三维目标检测首先将点云格栅化为离散的网格表示, 得到规则的体素结构, 再使用卷积网络进行处理。ZHOU 等<sup>[57]</sup>提出了 VoxelNet 网络, 利用体素特征编码层处理点云体素, 再使用三维卷积层聚合体素特征, 最后使用 RPN 实现三维目标检测。在此基础上, YAN 等<sup>[58]</sup>提出了 SECOND 网络, 使用三维稀疏卷积核, 并增加了数据增强模块, 提升了三维目标检测的速度与性能。与以上网络结构类似, HE 等<sup>[59]</sup>提出的 SA-SSD 网络加入了提取结构特征的辅助网络, 提高了三维边界框的定位精度。LANG 等<sup>[60]</sup>的 PointPillars 网络将点云创新性的表示为柱体结构, 通过 PointNet<sup>[61]</sup>将柱体内的点云聚集成柱状特征, 生成二维鸟瞰图进行特征提取。

基于原始点云的方法直接将点云作为输入, 试图减少由三维投影或体素结构引起的信息损失, 最著名的要属 PointNet 网络<sup>[61]</sup>, 该网络利用对称函数来解决点云无序性问题, 最大池化层将独立学习的点特征聚合为全局点集特征, 进而进行后续的三维检测任务。但是 PointNet 不能很好地提取局部精细的特征, 后续的 PointNet++网络<sup>[62]</sup>通过添加分层的结构, 类似图像检测中的特征金字塔, 扩大了感受范围, 进而提取不同尺度下的精细特征。然而, 该网络结构较为复杂, 计算成本较高, 限制了其在大规模点云场景的应用。SHI 等<sup>[63]</sup>提出了 PointRCNN 网络, 首先将点云分割为前景点和背景点, 自下而上地直接从点云中生成少量高质量的三维候选区域, 然后学习区域局部空间特征, 并与每个点的全局语义特征相结合, 进行精确的三维边界框细化和置信度回归。基于点云的方法由于点云固有的稀疏特性以及无序性, 使得网络计算量和内存消耗大, 运行效率较低, 并且由于缺乏纹理信息, 易出现目标错误分类。

为了获取更佳的检测性能, 很多研究基于视觉传感器图像的纹理信息和激光雷达点云的几何信息之间的互补特性开发了图像和点云融合的三维目标检测方法。根据融合策略可分为特征级融合和结果级融合。CHEN 等<sup>[64]</sup>提出的 MV3D 网络先从点云投影的鸟瞰图上估计三维候选区域, 然后对候选区域在前向图、鸟瞰图以及图像中的对应特征进行融合, 最后基于融合特征对目标位置、尺寸和姿态进行精

细化回归。在 MV3D 网络的基础上, KU 等<sup>[65]</sup>提出了 AVOD 网络, 首先融合预设三维锚点在图像和鸟瞰图上的特征, 然后基于融合特征估计三维候选目标区域, 该方法为了提高小目标的检测精度, 在图像特征提取网络中还应用了特征金字塔。SINDAGI 等<sup>[66]</sup>提出了 MVX-Net 网络, 将点云和体素投影到图像上, 进而融合图像的特征, 再利用 VoxelNet<sup>[67]</sup>预测三维目标。

对于结果级融合, 通常分别利用各自模态的网络对点云和图像进行特征提取。QI 等<sup>[68]</sup>提出的 F-PointNet 先检测图像上的二维目标, 然后对二维目标边界框对应的三维视锥体中的点云应用 PointNet++网络进行三维边界框回归。这种结构利用图像进行预检测, 降低了待处理点云的规模, 但是串行的工作机制依赖二维目标检测结果, 易受环境干扰。为此, WANG 等<sup>[69]</sup>提出 F-ConvNet 网络, 将单个视锥体划分为连续的多个视锥体并预测多个目标, 一定程度上解决了目标遮挡问题。

理论上图像和点云融合的方法结合了更多有用信息, 可以达到比单模态方法更好的检测性能。但是目前融合算法较为简单且计算量较大, 有待开发更好的融合方法。

### 2.1.2 语义分割

语义分割的目标是对场景中各类目标进行逐像素或逐点的类别分配, 但是不区分同一类别之间的对象。而实例分割是在此基础上, 进一步区别同类间的目标, 实现个体的分割与分类, 如图 8 所示。为了统一语义分割和实例分割, KIRILLOV 等<sup>[70-72]</sup>又提出了全景分割。虽然在激光雷达点云语义分割领域也有很多研究工作<sup>[73-74]</sup>, 但是本文主要关注在图像分割领域的研究。



图 8 语义分割和实例分割

#### 2.1.2.1 语义分割

在语义分割网络的发展过程中, 全卷积网络 (Fully convolutional network, FCN) 和 U-Net 网络是两个具有里程碑意义的设计。LONG 等<sup>[75]</sup>提出的 FCN 将分类网络末端的全连接层替换为卷积层, 使得卷积层神经网络能够处理语义分割问题。RONNEBERGER 等<sup>[76]</sup>继承 FCN 的思想, 提出的

U-Net 网络,其网络结构由编码器、解码器和跳跃连接三个部分组成,奠定了诸多语义分割网络的基础架构,其后基于 U-Net 架构的创新网络不断涌现,例如 V-Net<sup>[77]</sup>、SegNet<sup>[78]</sup>、RefineNet<sup>[79]</sup>和 FastFCN<sup>[80]</sup>等。

早期语义分割架构只关注有限感受野的局部信息,没有引入足够的上下文信息以及不同感受野下的全局信息,易出现误分割。为此,ZHAO 等<sup>[81]</sup>提出了 PSPNet 网络,该网络引入金字塔池化模块来获取图像的局部和全局特征,整合不同区域的上下文信息以获取全局上下文信息,提高了分割的可靠性。此外,CHEN 等<sup>[82]</sup>在 DeepLab v2 网络中使用空洞空间金字塔池化结构,以融合不同级别的语义信息,空洞卷积<sup>[83]</sup>提高了卷积核的感受野,能够更好地利用上下文信息。谷歌的 Transformer<sup>[84]</sup>是自然语言处理领域中最具代表性的模型之一,ZHENG 等<sup>[85]</sup>首次将 Transformer 引入计算机视觉领域,在语义分割任务中替代卷积神经网络进行特征提取,取得了当前最佳的语义分割性能。

近年来语义分割任务在智能驾驶领域受到了越来越多的关注,除行人、车辆分割外,如可行行驶区域检测<sup>[86-88]</sup>也是典型的图像语义分割任务。可行行驶区域是指车辆安全行驶不发生碰撞条件下可以行驶的道路路面部分。PENG 等<sup>[89]</sup>基于 Deeplab v3+进行城市场景可行行驶区域检测,在编码器网络中,消除了全连接层结构,以减少网络参数,并在池化操作前保留最大池索引;在解码器网络的上采样过程中重用最大池索引,在保留图像边界信息的同时减少了内存开销,在 Cityscapes 数据集<sup>[90]</sup>的测试结果证明了该方法的分割性能。TEICHMANN 等<sup>[91]</sup>提出了集分类、目标检测和语义分割一体的 MultiNet 网络,在 KITTI 数据集<sup>[92]</sup>道路分割任务中当时取得了最佳表现。

#### 2.1.2.2 实例分割

实例分割是结合目标检测和语义分割的更高层级任务,其可分为两阶段实例分割和单阶段实例分割。两阶段实例分割网络典型代表为 Mask R-CNN<sup>[93]</sup>,该网络在 Faster R-CNN 的基础上新增了掩膜预测分支。与之类似,Cascade Mask R-CNN<sup>[94]</sup>也是在二维目标检测网络 Cascade R-CNN 基础上添加掩膜分支。但是以上两种方法均存在大目标掩膜边缘预测粗糙的问题。后续的工作,如 MS R-CNN<sup>[95]</sup>尝试添加预测掩膜可靠性得分的支路,重新计算每个掩膜得分并重新排序,以提高分割性能;PointRend<sup>[96]</sup>在上采样操作中对目标边缘进行优化,

得到了更精细的边界掩膜。

近几年很多单阶段目标检测网络在性能上已经超过两阶段检测方法,因而有算法基于单阶段目标检测框架进行实例分割。作为首个实时运行的实例分割网络,BOLYA 等<sup>[97]</sup>提出的 YOLACT 将实例分割分成两个并行子任务,生成一组原型掩膜,同时预测每个实例的掩膜系数,然后将原型与掩膜系数线性组合来生成实例掩膜。其后续版本 YOLACT++<sup>[98]</sup>通过对预测的掩膜进行重新排序,提高了网络性能的同时依旧保持很高的实时性。后来的 BlendMask<sup>[99]</sup>、EmbedMask<sup>[100]</sup>、CondInst<sup>[101]</sup>均是在该网络基础上进行的改进。

## 2.2 多目标跟踪

对智能汽车环境感知系统而言,在目标检测的基础上进一步实现对感兴趣目标的跟踪,能够有效提升感知精度和稳定性。多目标跟踪(Multiple object tracking, MOT)是在复杂场景、目标数量未知情况下,对视频或图像的目标检测结果进行身份识别号(Identity, ID)赋予,使前后帧的相同目标 ID 一致,以便完成后续的目标精准查找、状态估计、轨迹预测等感知任务。MOT 是计算机视觉领域的一项关键技术,也是智能汽车环境感知中的重要任务,其核心在于前后帧间的目标关联,主要难点在于目标间的相互遮挡、运动模糊、目标尺度变化等。

多目标跟踪算法根据实现的形式不同可以有不同的分类方式<sup>[102-103]</sup>,在智能汽车的感知系统中,通常在目标检测的基础上,使用多种软件算法,在线地完成前后帧之间的目标关联,进而实现多目标的跟踪。根据在完成前后帧目标关联时提取特征的阶段不同,可以分为基于检测结果的跟踪方法(Tracking-by-detection, TBD)以及检测与跟踪联合(Joint detection and tracking, JDT)的方法,其过程对比如图 9 所示。

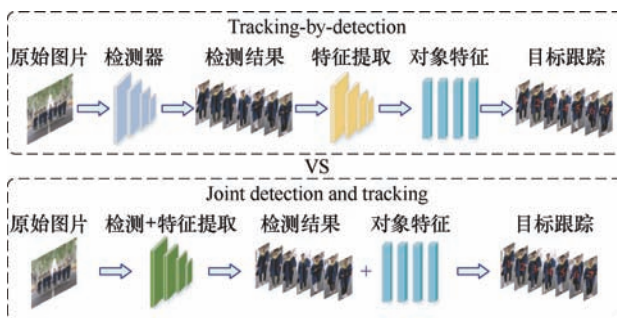


图 9 TBD 算法与 JDT 算法过程对比

### 2.2.1 基于检测结果跟踪的 MOT 算法

基于检测结果对多目标进行跟踪的 MOT 算法首先使用目标检测网络完成每一帧的目标检测,再



根据检测得到的包围框等信息得到图像中的所有目标, 然后进一步对这些目标进行特征提取等处理, 完成对不同目标的 ID 赋予, 由此将多目标跟踪问题转换为相邻帧间的目标关联问题。

SORT(Simple online and real-time tracking)算法<sup>[104]</sup>、DeepSORT(Deep learning-based SORT)算法<sup>[105]</sup>是 TBD 框架下最经典的多目标跟踪算法, 其使用卡尔曼滤波预测下一帧目标的状态, 利用重叠度、外观等特征建立相似度矩阵, 进而基于匈牙利算法等方法完成目标关联, 实现多目标跟踪。VOIGTLAENDER<sup>[106]</sup>在 R-CNN 基础上提出 TrackR-CNN, 对实例分割的网络实现了多目标跟踪, 这是 MOT 首次在实例分割领域的实现。后续的研究多以 TrackR-CNN 为基础进行改进, 实现更精确的多目标跟踪。KIM 等<sup>[107]</sup>将相机、激光雷达的检测结果进行融合, 针对不同距离的物体提出跟踪方法 EagerMOT, 在三维目标的多目标跟踪任务中取得了不错的成绩。WANG<sup>[108]</sup>提出 SMILEtrack, 在目标检测模块后添加 SLM 模块, 用于提取物体的重要外观特征, 加强了对物体运动与外观特征的建模, 实现了较高精度的多目标检测。AHARON 等<sup>[109]</sup>考虑了相机的运动补偿并建立了更精确的状态向量, 实现了一个二级判断的目标关联方法, 并在 MOTchallenge<sup>[110]</sup>等数据集上取得了良好的跟踪精度。DU 等<sup>[111]</sup>以 DeepSORT 为基础, 从特征嵌入和

轨迹关联的角度对跟踪器进行了改进, 解决了缺失关联与缺失检测的问题。CAO 等<sup>[112]</sup>通过添加时间步计算虚拟轨迹, 解决长时间物体非线性运动带来的累积噪声, 使卡尔曼滤波跟踪器的效果达到了先进水平。在此基础上, MAGGIOLINO 等<sup>[113]</sup>引入外观线索, 提出 Deep OC-SORT 进一步提升了多目标跟踪的精度。

基于检测结果的多目标跟踪算法将目标检测任务与多目标跟踪任务分为两个阶段, 在检测完成后针对检测结果完成多目标跟踪, 主要是利用目标的运动特征、外观特征进行相似度计算以及数据关联, 以得到准确的目标轨迹, 然而 TBD 往往与目标检测算法同时执行, 对计算设备的性能要求较高。近年来, 为了平衡检测、跟踪的精度与计算速度, 学者提出了一系列检测与跟踪联合框架下的 MOT 算法。

### 2.2.2 检测与跟踪联合的 MOT 算法

检测与跟踪联合的 MOT 算法将检测与跟踪作为单阶段任务, 通过将检测网络、跟踪网络的特征层共享来实现目标检测与每个对象特征的获取, 用单个网络完成目标检测与多目标跟踪两个任务, 以实现更高效的目标关联。

JDE(Joint detection and embedding)<sup>[114]</sup>算法是首次将目标检测与多目标跟踪在同一个网络框架内完成的 MOT 算法, 如图 10 所示。

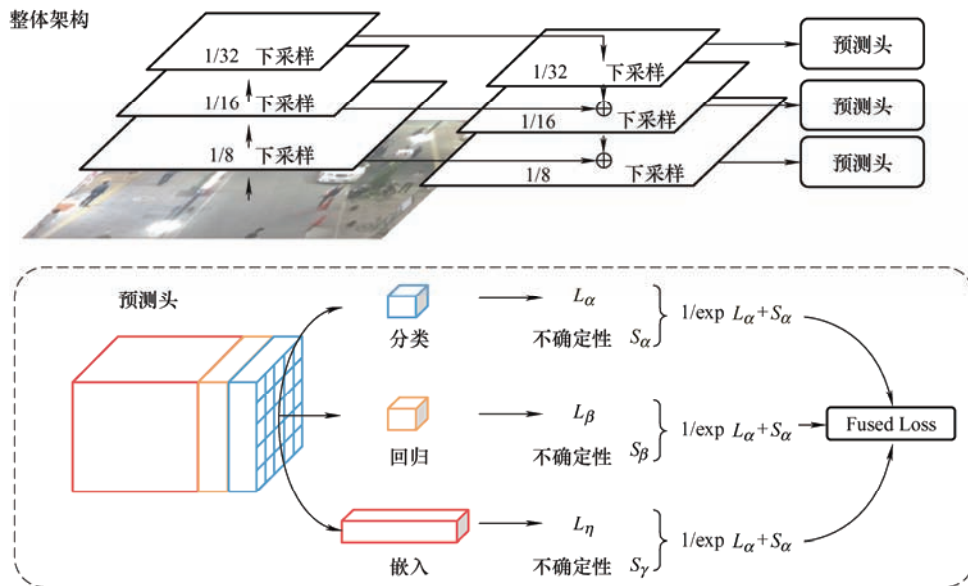


图 10 JDE 算法过程<sup>[114]</sup>

JDE 算法通过使用多任务学习的损失函数, 将目标检测网络预测头的分支用于输出额外的特征, 进而进行跟踪 ID 的赋予, 实现目标关联。在此基础上, CHAABANE 等<sup>[115]</sup>提出 DEFT 算法, 将目标检

测与外观匹配两个任务使用单个网络模型实现, 并将多目标的运动约束用 LSTM 网络建模, 高效地完成了单网络框架下的目标检测与跟踪。ZHOU 等<sup>[116]</sup>以目标检测网络 CenterNet 为基础, 利用历史帧的

线索完成对消失、遮挡目标的恢复,提出 CenterTrack 方法,将目标检测与匹配的信息同时输出,省去特征匹配的步骤,直接完成了多目标跟踪,相较于 JDE 算法更为方便。YIN 等<sup>[117]</sup>使用一个两阶段的三维目标检测与跟踪器,首先用关键点检测器检测对象中心,再使用贪心算法对最近点进行中心点匹配,高效地完成了三维目标的检测与跟踪。ZHANG 等<sup>[118]</sup>同样对 CenterNet 进行改进,使用基于无锚检测架构的目标检测网络,平衡了联合算法中目标检测与跟踪的竞争问题,实现了更高的检测、跟踪精度。PANG 等<sup>[119]</sup>提出 SimpleTrack 算法,提高了在目标遮挡等短暂目标丢失情况下 JDE 算法的鲁棒性与实时性。

检测与跟踪联合的 MOT 算法利用单个网络模型完成目标检测与多目标跟踪的任务,较之基于检测结果的跟踪方法大大提升了计算速度,降低了模型部署的成本,且也拥有较好的跟踪精度。除此之外,近两年有学者通过注意力机制<sup>[120-122]</sup>、图神经网络<sup>[123-124]</sup>等方法,在目标遮挡、背景干扰等复杂环境下,实现了较好的多目标跟踪性能。

总的来说,多目标跟踪任务作为智能感知系统

中继目标检测后的又一重要任务,对其的相关研究非常充分,从最初的基于检测结果的跟踪方法 TBD 到后来单个网络完成检测、跟踪两个过程的联合方法 JDE,再到图神经网络、注意力机制等多种方法的加入,MOT 算法面对遮挡、模糊等问题时的跟踪精度不断提升,已经被广泛的运用在了智能汽车感知系统中。在未来的研究中,需要针对 MOT 算法的核心问题,也就是目标关联模块的有效性进行研究,以实现更好的多目标跟踪。

## 2.3 目标意图识别与轨迹预测

完成对交通目标的检测与跟踪后,对其行驶意图进行识别并预测该目标未来运动轨迹,可以提升智能汽车在动态环境中安全驾驶的能力。目标意图识别与轨迹预测通常基于环境感知系统已经获得的目标当前、历史信息以及周边的交通环境信息,对未来该目标可能的意图、轨迹等进行预测<sup>[125]</sup>。如图 11 所示,根据完成该任务使用方法的不同,可以分为传统的、基于模型驱动的方法,基于经典机器学习的方法,基于数据驱动的深度学习的方法,以及强化学习、融合网络等其他方法。

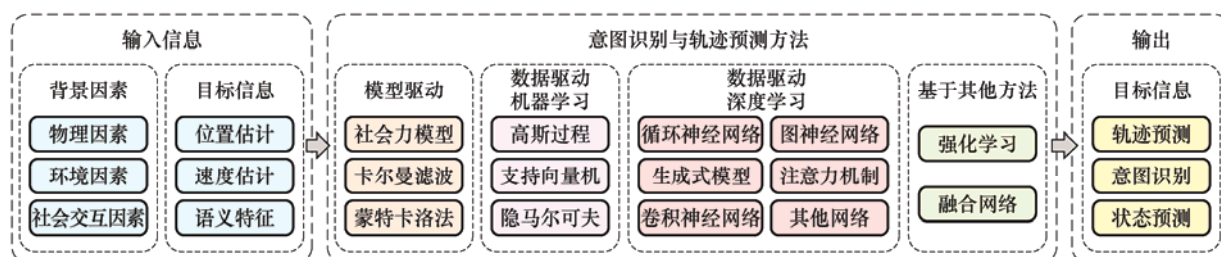


图 11 目标意图识别与轨迹预测方法

基于模型完成对目标的意图识别与轨迹预测时,通常对行人、车辆等交通参与者的动力学、运动学特性进行建模,将目标当前的状态直接应用于模型中,对未来其意图与轨迹完成预测。常用于预测的模型有社会力模型<sup>[126-127]</sup>、元胞自动机<sup>[128]</sup>、蒙特卡洛方法<sup>[129]</sup>、线性回归模型<sup>[130]</sup>等,这些方法虽然能以较低的计算资源完成轨迹的预测,但精度依赖于模型的参数,往往无法应用于复杂场景下,且只能在短期(不超过 1 s)的未来预测取得较准确的结果。近年来学者更倾向于借鉴这些基于模型方法的思想,再使用数据驱动的方法完成对目标未来状态的预测。

经典的基于机器学习的方法用数据训练模型,通过手动设计的特征提取器来挖掘数据特征,进而完成对未来意图的识别与轨迹的预测,随着考量要素的增加,此类方法的准确性也越来越高,常用的

方法包括动态贝叶斯网络<sup>[131-132]</sup>、隐式马尔可夫模型<sup>[133-134]</sup>、高斯过程<sup>[135]</sup>、支持向量机<sup>[136]</sup>等,这些方法较之模型驱动的方法精度有所提升,但需要前期定义的策略较多,仍不适用于复杂的交通场景下,或是长期的预测任务中。

基于深度学习的目标意图识别与轨迹预测方法,不仅能更好地提取目标本身的信息,更能将周围环境、不同交通参与者之间的交互等要素考虑在网络模型中,得到更精确的预测结果。各种序列模型、生成式模型、图神经网络等多种深度学习模型都被应用于该任务中<sup>[137]</sup>。

### 2.3.1 序列模型

序列模型是指在完成对目标意图识别与轨迹预测的过程中,逐步地完成对未来时间的状态预测,即先预测下一时刻的状态,再使用该状态完成后续时刻的状态。由于交通目标的预测任务是典型的时

序问题, 循环神经网络(Recurrent neural network, RNN)最早的被运用于该任务中, 其中长短期记忆网络(Long short term memory, LSTM)及其变种是最常用的预测网络。ALAH 等<sup>[138]</sup>提出 S-LSTM (Social-LSTM), 使用一个社交池(Social pooling)的结构, 将周围行人的轨迹特征用于对目标行人预测中, 实现了更精确的行人轨迹预测, 后续很多研究都受到该工作的启发。KAWASAKI 等<sup>[139]</sup>将卡尔曼滤波与 LSTM 网络结合, 基于车道特征完成了对车辆轨迹的预测。XING 等<sup>[140]</sup>考虑到不同驾驶风格的影响, 使用高斯混合模型识别不同驾驶风格, 再结合 LSTM 完成对前车的轨迹预测。有学者使用多组 LSTM 对不同特征进行编码, 对目标状态预测问题进行研究。DAI 等<sup>[141]</sup>使用一组神经网络完成对车辆未来车道的预测, 再使用额外的一组 LSTM 学习车辆之间的相互作用, 完成了基于 LSTM 的车辆轨迹预测。ZHANG 等<sup>[142]</sup>使用两组 LSTM, 一组完成对车辆转向的预测分类, 一组用于未来轨迹的预测, 对车辆未来的意图与轨迹预测进行了深入的研究。

卷积神经网络(Convolutional neural network, CNN)在计算机视觉等领域取得了极大的成功, 部分学者使用该网络对目标的历史轨迹进行建模, 进而完成对目标未来状态的预测。NIKNIL 等<sup>[143]</sup>建立了一个时间序列的框架, 首次实现了 CNN 网络在轨迹预测领域的使用。BAI 等<sup>[144]</sup>使用膨胀卷积、因果卷积与残差连接的方式, 提出时间卷积网络(Temporal convolutional network, TCN), 较之 LSTM 网络拥有更好的特征提取能力。此外部分基于 CNN 的轨迹预测网络使用鸟瞰视角<sup>[145-146]</sup>, 以热力图以及前后帧图像的特征完成对未来轨迹概率分布的预测。

由于 RNN 可以更好地提取时序数据的特征, 而 CNN 能够更多地考虑周边环境特征对目标的影响, 当前较多的研究将该两种网络结合, 以实现更精确的目标状态预测。DEO 等<sup>[147]</sup>将社交池的结构使用卷积神经网络实现, 稳健地学习车辆运动中的相互依赖关系, 完成了参考周边环境的车辆换道预测。XIE 等<sup>[148]</sup>使用卷积生成的检测结果消除预测轨迹中的异常信息, 以有效地提取不同轨迹间的交互特征, 再使用网格算法优化网络, 实现了 CNN-LSTM 架构下的车辆轨迹预测。为了更好的完成轨迹预测任务, DANIEL 等<sup>[149-151]</sup>使用卷积神经网络对地图信息进行处理, 以排除无法实现的轨迹, 使预测轨迹更贴近真实轨迹, 再使用 RNN 的方式完成了对行人、车辆的轨迹预测。

基于序列模型的意图识别与轨迹预测方法通常将目标的历史轨迹信息作为主要特征提取对象, 再辅以周边的交通参与者与环境信息, 逐帧地完成对目标未来状态的预测, 可以得到一条确定的未来轨迹, 该类方法目前研究较多, 在数据集上可以达到不错的预测效果。

### 2.3.2 生成式模型

行人与车辆的轨迹往往具有相似性与规律性, 且轨迹预测任务往往要求提供多模态的未来轨迹, 因而部分学者使用生成式模型对未来轨迹进行预测, 常用的模型包括生成对抗网络<sup>[152]</sup>(Generative adversarial network, GAN)、各类自编码器(Auto encoder, AE)等, 该类模型往往一次生成未来一段时间的轨迹, 且可以同时生成多条多模态的未来轨迹, 为智能汽车提供更多的信息。

斯坦福大学提出的 Social GAN<sup>[153]</sup>是使用 GAN 网络完成轨迹预测任务的开山之作, 对历史轨迹进行编码作为生成器, 再使用判别器来判断产生的轨迹数据是否真实合理, 较之 S-LSTM 有了较大的提升。HEGDE 等<sup>[154]</sup>考虑到车辆尺寸、驾驶员行为等因素对车辆的社会因素进行了建模, 在鸟瞰数据集上取得了较好的预测精度。WANG 等<sup>[155]</sup>提出 TS-GAN 模型, 使用自创建的卷积网络进行车辆的时空信息的提取, 实现了高精度的多模态车辆轨迹预测。TANG 等<sup>[156]</sup>将动态注意力机制引入条件变分自编码器(Conditional variational auto encoder, CVAE)中, 有效地完成了对轨迹预测概率分布的编码与解码。PECnet<sup>[149]</sup>、SGNet<sup>[151]</sup>、Y-net<sup>[157]</sup>等行人轨迹预测网络将条件变分自编码器与卷积神经网络结合, 在 ETH&UCY 数据集上取得了较为先进的预测结果。此外, CASAS 等<sup>[158]</sup>使用传感器的原始数据作为输入, 使用 CVAE 的方法进行多模态的轨迹预测, 也取得了一定的成绩。

生成式模型在完成轨迹预测任务时, 通常对目标的历史数据进行编码, 再使用解码器来完成未来轨迹的生成, 相较于基于序列模型的方法, 生成式模型可以以此生成多模态的未来轨迹, 也可以缓解模型过拟合的问题, 在该任务中也有较多的研究。

### 2.3.3 图神经网络

使用序列模型完成对目标未来状态的预测时, 往往将每个对象视为一个节点, 因此需要考虑到不同节点之间相互关系, 才能更好地完成交互关系下的状态预测。图神经网络(Graph neural network, GNN)非常适合解决这类和相互交互有关的问题。图卷积神经网络(Graph convolutional network, GCN)



是当前最流行的图神经网络方法,它将卷积神经网络与图神经网络相结合,可以很好地研究不同目标间的相互关系,大部分轨迹预测相关的研究都基于 GCN 展开。LI 等<sup>[159-160]</sup>将每个车辆看成一个节点,提出 GRIP 和 GRIP++ 算法,将不同车辆间的交互关系送入 LSTM 中进行编码与解码,可以实现复杂场景下的交通参与者轨迹预测。MOHAMED 等<sup>[161]</sup>同样使用 GCN,将行人间的交互关系建模为图来代替原来的聚合方法,在行人轨迹预测领域取得了一定的成果。AN 等<sup>[162]</sup>使用半全局图卷积的方法实现对车辆间深度动态交互的构建,提升网络性能的同时减少了运算时间。LI 等<sup>[163]</sup>提出基于注意力机制的时频域图卷积神经网络,更好地学习了车辆多模态特征之间的依赖关系。

图卷积神经网络多与其他神经网络一同使用,用于研究不同交通参与者之间的多模态交互关系,是目前轨迹预测与意图识别领域研究的热门所在。除了图神经网络外,注意力机制<sup>[164]</sup>也常常被作为神经网络的内嵌模块,使网络可以自适应的学习轨迹、环境特征中的关键部分,以更好地完成目标轨迹的预测。

在目标意图识别与轨迹预测领域中,序列模型相关的研究开始较早,他们往往考虑到周边环境、交通参与者之间的相互作用关系,逐步地完成对目标长时间的行为判断或轨迹预测。生成式模型可以一次提供多条完整的未来轨迹,完成对目标的多模态轨迹预测,也有很多的相关研究。图神经网络、注意力机制等现在多被用来研究不同交通参与者间的相互关系,以及对网络关键特征的判断,目前很多学者将它们与其他网络结合,完成了更精确的未来状态预测,具有较好的研究前景。

## 2.4 环境建图

环境建图作为智能车辆导航定位系统的有效补充,也是当前环境感知领域的研究热点。通过构建道路交通环境地图,智能车辆能够准确地感知和理解周围的环境的拓扑结构、几何特征和语义信息<sup>[165-166]</sup>。拓扑结构信息通常包含路面坡度、车流速度和交叉路口延误等,能够帮助车辆选择最佳行驶路线<sup>[167]</sup>。而丰富的几何特征则为环境感知系统提供了动态目标(如车辆、行人等)和静态目标(如建筑物、桥梁等)信息<sup>[168]</sup>。对地图中的感兴趣目标进行语义标注,如车辆、行人、交通标志、建筑物等,能够帮助智能车辆更准确地识别和跟踪不同类型的目标<sup>[169]</sup>,为智能驾驶系统的决策和规划提供更可靠的输入。

同时定位与地图构建(Simultaneous localization and mapping, SLAM)技术被广泛应用在智能驾驶领域,用于实时建立和更新地图,为智能车辆提供精确的定位和环境感知能力。在环境建图过程中,SLAM 算法首先利用车载传感器(如激光雷达、摄像头、IMU 等)获取环境的感知数据,然后通过分析连续的传感器数据,利用运动估计算法(如扩展卡尔曼滤波<sup>[170]</sup>、粒子滤波<sup>[171]</sup>等),估计车辆在不同时域上的运动轨迹。然后从感知数据中提取特征点或地标,如边缘、角点、线段等,将运动估计和特征点进行数据关联,建立地图与车辆运动轨迹之间的联系;最后利用线性优化算法(如图优化<sup>[172]</sup>等),通过最小化误差函数,不断地更新地图和车辆运动轨迹。SLAM 建图方法可以进一步分为稀疏建图和稠密建图。

### 2.4.1 稀疏建图

稀疏建图方法通过选取地图中的关键点或特征点来表示环境,而不是使用完整的点云数据,如图 12 所示。在稀疏建图中,仅选择小部分具有较高信息量或显著特征的点用于地图构建<sup>[173]</sup>。这种方法可以减少数据处理和存储的复杂性,并提高实时性能。稀疏建图方法主要关注于提取和匹配特征点、建立关联关系以及优化地图的拓扑结构,适用于实时应用和资源受限的环境。

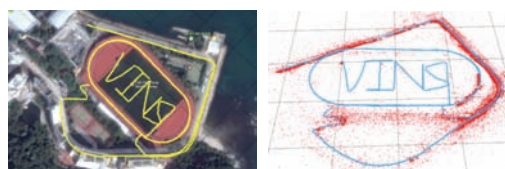


图 12 稀疏建图<sup>[178]</sup>

DAVISON 等<sup>[174]</sup>提出了首个实时单目视觉 SLAM 系统 Mono-SLAM,后端采用扩展卡尔曼滤波算法跟踪从前端获取的稀疏特征点,能够在未知场景中实时跟踪快速运动的单目相机三维轨迹。KLEIN 等<sup>[175]</sup>提出了一种并行跟踪和建图系统(Parallel tracking and mapping, PTAM),首次通过非线性优化方法区分前端和后端,提出关键帧机制,将关键图像串联起来,优化运动轨迹和特征方向。随后的许多视觉 SLAM 系统设计也采用了类似的方法。MURR-ARTAL 等<sup>[176]</sup>提出了 ORB-SLAM,这是一种比较完整的基于关键帧的单目 SLAM 方法,该方法将整个系统分为跟踪、建图、重定位和回环四个部分。后续的优化方法 ORB-SLAM2<sup>[177]</sup>包括地图复用、闭环检测和重定位功能,采用了轻量级的地图构建方法,可以实时地将新的视觉信息融合到地

图中, 增强了地图的一致性和准确性。QIN 等<sup>[178]</sup>采用了一种基于紧耦合、非线性优化的方法, 获得高精度的稀疏地图。CAMPOS 等<sup>[179]</sup>提出的 ORB-SLAM3 可以完全依赖于最大后验估计, 即使在 IMU 初始化期间, 也能在大小型、室内和室外环境中实现实时建图操作。

#### 2.4.2 稠密建图

稠密建图方法旨在生成具有高密度点云的地图, 其中每个空间位置都有对应的点云数据, 并在地图中以高分辨率密集地表示这些点云, 如图 13 所示。稠密建图方法通常需要处理大量的数据, 并在地图构建过程中对点云进行滤波、配准和插值等处理, 以获得平滑而精细的地图表示。



图 13 稠密建图

NEWCOMBE 等<sup>[180]</sup>提出了 DTAM, 该方法是首次利用直接法实现了稠密三维地图的构建, 但是其基于灰度不变的假设容易受到光照影响而失效。ZHOU 等<sup>[181]</sup>在 DTAM 的基础上, 用深度卷积神经网络取代了 TV-L1 优化和相机跟踪, 取得了优于主流方法的结果。BLOESCH 等<sup>[182]</sup>通过在深度图像上训练自动编码器来实现密集场景几何体的更通用的紧凑表示, 解决了密集结构光学方法 (Structure-from-motion, SfM) 问题, 但实时性较差。TATENNO 等<sup>[183]</sup>通过将 LSD-SLAM<sup>[184]</sup>里面的深度估计和图像匹配都换成基于 CNN 的方法, 取得了较为鲁棒的结果。TANG 等<sup>[185]</sup>提出了一种新颖的网络体系结构, 网络根据输入图像生成若干基础深度图, 然后通过光束法平差 (Bundle adjustment, BA) 将最终深度优化为这些基础深度图的线性组合, 以解决 SfM 问题, 从而实现稠密像素重建。KOESTLER 等<sup>[186]</sup>通过使用预先训练的 MVSNet 式神经网络对单目图像进行深度估计, 然后通过 Frame-to-model 的相机跟踪方法来解耦位姿和深度问题, 仅用单目图像就可以重建三维场景。TEED 等<sup>[187]</sup>通过将密集光流估计结构<sup>[188]</sup>与视觉里程计结合, 在多个数据集 (如 Euroc<sup>[189]</sup>和 TartanAir<sup>[190]</sup>数据集) 下获得了较好的结果。此外, SUCAR 等<sup>[191-192]</sup>通过解耦位姿和深度估计, 使用 RGB-D 图像和神经辐射场生成精确的三维重建场景<sup>[193]</sup>。

综上所述, 稀疏建图通常选择关键点或特征点进行地图表示, 数据处理和计算的复杂度较低, 可以实现较快的建图速度, 但地图的细节和精度比较差, 同时对于环境变化和噪声相对较敏感, 鲁棒性不高。而稠密地图包含大量的环境信息, 但其对计算存储资源要求较高, 实时性无法得到保证。在实际应用中, 可以根据需求选择适当的建图策略, 甚至将稠密建图和稀疏建图相结合, 以达到更好的建图效果。

### 3 总结与展望

智能汽车的研究和发展将促进智能交通领域的转型和发展, 为人类的出行方式与社会生活带来了革命性的变化。智能汽车环境感知作为智能驾驶系统的关键技术所在, 仍将是长期热点研究方向。本文从感知关键任务的角度进行切入, 详细介绍了目标检测与分割、多目标跟踪、意图识别与轨迹预测、环境建图的相关感知方法和研究现状。

虽然智能汽车的环境感知技术已然取得了长足的发展, 但随着汽车智能化的提升, 对环境感知精度和实时性的要求也日益增长。例如, 当车辆遇到一些低辨识度场景 (黑夜、大雾等) 或者城市拥堵复杂路段时, 系统能否确保动态目标和周围环境的感知准确性、实时性。毋庸置疑, 这对智能汽车感知来说, 具有很大的挑战性。本文对智能汽车感知未来发展方向的初步展望如下。

(1) 多传感器融合感知。近年来传感器技术的适用性和有效性在智能驾驶系统方面取得了显著的进步。但在复杂恶劣环境下, 单一传感器可能会受大雾或雨滴等因素影响, 导致数据失效<sup>[194]</sup>。实际的解决方案是将多个互补的传感器 (如红外热像仪、激光雷达、毫米波雷达等) 集成在一起, 协同工作以克服各自的缺点。不同的恶劣环境可能需要不同类型的传感器来有效感知, 例如, 雪暴可能需要更多的毫米波雷达数据, 而大雾可能需要更多的红外传感器数据。多传感器融合可以帮助解决单一传感器的局限性和盲区问题, 提高感知的鲁棒性和可靠性。因此, 多传感器融合感知是智能汽车领域中一个重要的研究方向, 它可以提高智能汽车的感知能力。

(2) V2X 协同感知。目前, 许多智能汽车系统都是从单一的角度来感知环境, 而没有从道路上其他车辆、行人和其他交通设施的角度获取额外的道路交通信息<sup>[195]</sup>。当物体被遮挡或距离过远, 仅通过

单车感知系统无法进行准确检测和分类。V2X 协同感知通过无线通信技术将车辆与周围交通参与者包括其他车辆、行人、交通设施等的信息进行交互和共享,实现了感知范围的扩展,从而提高了感知的准确性。V2X 协同感知为智能交通系统和智能驾驶技术的实现提供了重要支持。该项技术将车辆转变为信息节点,能够共享环境信息,从而使道路上的交通更加智能和高效。在未来的交通系统中,V2X 协同感知将扮演着不可或缺的角色,为更安全、更智能的出行创造更多可能性。

(3) 三维视角感知。大多数智能驾驶算法的传统方法在正面或透视图(二维视角)中执行检测、分割、跟踪预测等感知任务。随着传感器配置的日益复杂,集成来自不同传感器的多源信息并以统一的视图表示特征变得至关重要<sup>[196]</sup>。传统的二维视角感知缺乏深度信息、无法捕捉立体结构并且无法处理动态变化场景。与传统的二维视角感知相比,三维视角下的感知能够提供更丰富、更准确的环境信息(包含深度信息、语义信息、纹理信息等),进一步增强对场景的认知能力。近些年,三维鸟瞰视图(Bird's eye view, BEV)<sup>[197-199]</sup>和三维占用<sup>[200-202]</sup>感知引起了工业界和学术界的广泛关注。得益于 BEV 场景直观、容易进行多源传感器特征融合且有利于后续任务,智能汽车的三维 BEV 感知框架将是一种未来趋势。三维视角下的感知也将进一步推动智能交通系统的发展和实现更安全、高效的智能驾驶技术。

## 参 考 文 献

- [1] 李克强,戴一凡,李升波,等. 智能网联汽车(ICV)技术的发展现状及趋势[J]. 汽车安全与节能学报, 2017, 8(1): 1-14.  
LI Keqiang, DAI Yifan, LI Shengbo, et al. State-of-the-art and technical trends of intelligent and connected vehicles[J]. Journal of Automotive Safety and Energy, 2017, 8(1): 1-14.
- [2] 陈虹,郭露露,宫洵,等. 智能时代的汽车控制[J]. 自动化学报, 2020, 46(7): 1313-1332.  
CHEN Hong, GUO Lulu, GONG Xun, et al. Automotive control in intelligent era[J]. Acta Automatica Sinica, 2020, 46(7): 1313-1332.
- [3] 熊璐,杨兴,卓桂荣,等. 无人驾驶车辆的运动控制发展现状综述[J]. 机械工程学报, 2020, 56(10): 127-143.  
XIONG Lu, YANG Xing, ZHUO Guirong, et al. Review on motion control of autonomous vehicles[J]. Journal of Mechanical Engineering, 2020, 56(10): 127-143.
- [4] 彭湃,耿可可,殷国栋,等. 基于传感器融合里程计的相机与激光雷达自动重标定方法[J]. 机械工程学报, 2021, 57(20): 206-214.  
PENG Pai, GENG Keke, YIN Guodong. Automatic recalibration of camera and lidar using sensor fusion odometry[J]. Journal of Mechanical Engineering, 2021, 57(20): 206-214.
- [5] 江浩斌,沈峥楠,马世典,等. 基于信息融合的自动泊车系统车位智能识别[J]. 机械工程学报, 2017, 53(22): 125-133.  
JIANG Haobin, SHEN Zhengnan, MA Shidian, et al. Intelligent identification of automatic parking system based on information fusion[J]. Journal of Mechanical Engineering, 2017, 53(22): 125-133.
- [6] 姜武华,辛鑫,陈无畏,等. 基于信息融合的自动泊车系统多工况车位识别和决策规划[J]. 机械工程学报, 2021, 57(6): 131-141.  
JIANG Wuhua, XIN Xin, CHEN Wuwei, et al. Multi-condition parking space recognition based on information fusion and decision planning of automatic parking system[J]. Journal of Mechanical Engineering, 2021, 57(6): 131-141.
- [7] 王艺,蔡英凤,陈龙,等. 基于模型预测控制的智能网联汽车路径跟踪控制器设计[J]. 机械工程学报, 2019, 55(8): 136-144.  
WANG Yi, CAI Yingfeng, CHEN Long, et al. Design of intelligent and connected vehicle path tracking controller based on model predictive control[J]. Journal of Mechanical Engineering, 2019, 55(8): 136-144.
- [8] 章仁燮,熊璐,余卓平. 智能汽车转向轮转角主动控制[J]. 机械工程学报, 2017, 53(14): 106-113.  
ZHANG Renxie, XIONG Lu, YU Zhuoping. Active steering angle control for intelligent vehicle[J]. Journal of Mechanical Engineering, 2017, 53(14): 106-113.
- [9] LI F, SHI W, TU Y, et al. Automated methods for indoor point cloud preprocessing: Coordinate frame reorientation and building exterior removal[J]. Journal of Building Engineering, 2023, 76: 107270.
- [10] LI X, ZHANG B, SANDER P V, et al. Blind geometric distortion correction on images through deep learning[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 4855-4864.
- [11] YAN G, LIU Z, WANG C, et al. Opencalib: A multi-sensor calibration toolbox for autonomous driving[J]. Software Impacts, 2022, 14: 100393.



- [12] PERŠIĆ J, PETROVIĆ L, MARKOVIĆ I, et al. Online multi-sensor calibration based on moving object tracking[J]. *Advanced Robotics*, 2021, 35(3-4): 130-140.
- [13] ERICKSON B J, KORFIATIS P, AKKUS Z, et al. Toolkits and libraries for deep learning[J]. *Journal of Digital Imaging*, 2017, 30: 400-405.
- [14] KIM S, WIMMER H, KIM J. Analysis of deep learning libraries: Keras, pytorch, and mxnet[C]// 2022 IEEE/ACIS 20th International Conference on Software Engineering Research, Management and Applications (SERA). IEEE, 2022, 54-62.
- [15] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object detection with discriminatively trained part-based models[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 32(9): 1627-1645.
- [16] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features[C]// *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR, 2001*, 1: 511-518.
- [17] PRASAD D K. Survey of the problem of object detection in real images[J]. *International Journal of Image Processing (IJIP)*, 2012, 6(6): 441-466.
- [18] VEDALDI A, GULSHAN V, VARMA M, et al. Multiple kernels for object detection[C]// 2009 IEEE 12th International Conference on Computer Vision. IEEE, 2009: 606-613.
- [19] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features[C]// *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2001*, 1: 511-518.
- [20] LIENHART R, MAYDT J. An extended set of Haar-like features for rapid object detection[C]// *International Conference on Image Processing. IEEE, 2002*, 1: 900-903.
- [21] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]// 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, 1: 886-893.
- [22] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object detection with discriminatively trained part-based models[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 32(9): 1627-1645.
- [23] FREUND Y, SCHAPIRE R E. Experiments with a new boosting algorithm[C]// *Proceedings of the Thirteenth International Conference on International Conference on Machine Learning*. 1996, 96: 148-156.
- [24] HEARST M A, DUMAIS S T, OSUNA E, et al. Support vector machines[J]. *IEEE Intelligent Systems and Their Applications*, 1998, 13(4): 18-28.
- [25] VIOLA P, JONES M J. Robust real-time face detection[J]. *International Journal of Computer Vision*, 2004, 57: 137-154.
- [26] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014: 580-587.
- [27] UIJLINGS J R R, VAN DE SANDE K E A, GEVERS T, et al. Selective search for object recognition[J]. *International Journal of Computer Vision*, 2013, 104: 154-171.
- [28] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904-1916.
- [29] GIRSHICK R. Fast r-cnn[C]// *Proceedings of the IEEE International Conference on Computer Vision*. 2015: 1440-1448.
- [30] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [31] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017: 2117-2125.
- [32] CAI Z, VASCONCELOS N. Cascade r-cnn: Delving into high quality object detection[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018: 6154-6162.
- [33] TERVEN J, CORDOVA-ESPARZA D. A comprehensive review of yolo: From yolov1 to yolov8 and beyond[J/OL]. *ArXiv*, [2023-08-07]. <https://doi.org/10.48550/arXiv.2304.00501>.
- [34] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016: 779-788.
- [35] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]// *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The*

- Netherlands, October 11-14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.
- [36] LAW H, TENG Y, RUSSAKOVSKY O, et al. Cornernet-lite: Efficient keypoint based object detection[J/OL]. ArXiv, [2020-09-16]. <https://doi.org/10.48550/arXiv.1904.08900>.
- [37] DUAN K, BAI S, XIE L, et al. Centernet: Keypoint triplets for object detection[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 6569-6578.
- [38] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 10781-10790.
- [39] LI G, JI Z, QU X, et al. Cross-domain object detection for autonomous driving: A stepwise domain adaptative YOLO approach[J]. IEEE Transactions on Intelligent Vehicles, 2022, 7(3): 603-615.
- [40] XU X, ZHAO J, LI Y, et al. Banet: A balanced atrous net improved from ssd for autonomous driving in smart transportation[J]. IEEE Sensors Journal, 2020, 21(22): 25018-25026.
- [41] WANG H, XU Y, WANG Z, et al. Centernet-auto: A multi-object visual detection algorithm for autonomous driving scenes based on improved centernet[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2023, 7(3): 742-752.
- [42] CHEN L, LIN S, LU X, et al. Deep neural network based vehicle and pedestrian detection for autonomous driving: A survey[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(6): 3234-3246.
- [43] LI B, OUYANG W, SHENG L, et al. Gs3d: An efficient 3d object detection framework for autonomous driving[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 1019-1028.
- [44] CHABOT F, CHAOUCH M, RABARISOA J, et al. Deep manta: A coarse-to-fine many-task network for joint 2D and 3D vehicle analysis from monocular image[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2040-2049.
- [45] CHEN X, KUNDU K, ZHANG Z, et al. Monocular 3D object detection for autonomous driving[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 2147-2156.
- [46] MOUSAVIAN A, ANGUELOV D, FLYNN J, et al. 3D bounding box estimation using deep learning and geometry[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 7074-7082.
- [47] XIANG Y, CHOI W, LIN Y, et al. Data-driven 3D voxel patterns for object category recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1903-1911.
- [48] XIANG Y, CHOI W, LIN Y, et al. Subcategory-aware convolutional neural networks for object proposals and detection[C]// 2017 IEEE Winter Conference on Applications of Computer Vision (WACV). 2017: 924-933.
- [49] KUNDU A, LI Y, REHG J M. 3D-RCNN: Instance-level 3D object reconstruction via render-and-compare[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 3559-3568.
- [50] CHEN X, KUNDU K, ZHU Y, et al. 3D object proposals using stereo imagery for accurate object class detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(5): 1259-1272.
- [51] GUPTA S, GIRSHICK R, ARBELÁEZ P, et al. Learning rich features from rgb-d images for object detection and segmentation[C]// Proceedings of European Conference on Computer Vision, 2014: 345-360.
- [52] QIN Z, WANG J, LU Y. Triangulation learning network: From monocular to stereo 3D object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 7615-7623.
- [53] LI P, CHEN X, SHEN S. Stereo R-CNN based 3D object detection for autonomous driving[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 7644-7652.
- [54] SIMONY M, MILZY S, AMENDEY K, et al. Complex-yolo: An euler-region-proposal for real-time 3D object detection on point clouds[C]// Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, September 8-14, 2018, Proceedings, Part I. Springer International Publishing, 2019: 197-209.
- [55] YANG B, LUO W, URTASUN R. Pixor: Real-time 3D object detection from point clouds[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7652-7660.
- [56] SHANG J, CHEN Y, NIE J. Lasernet: A method of laser stripe center extraction under non-ideal conditions[J].

- Applied Optics, 2023, 62(13): 3387-3397.
- [57] ZHOU Y, TUZEL O. Voxelnet: End-to-end learning for point cloud based 3D object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 4490-4499.
- [58] YAN Y, MAO Y, LI B. Second: Sparsely embedded convolutional detection[J]. Sensors, 2018, 18(10): 3337.
- [59] HE C, ZENG H, HUANG J, et al. Structure aware single-stage 3D object detection from point cloud[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020: 11873-11882.
- [60] LANG A H, VORA S, CAESAR H, et al. Pointpillars: Fast encoders for object detection from point clouds[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 12697-12705.
- [61] QI C R, SU H, MO K, et al. Pointnet: Deep learning on point sets for 3D classification and segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 652-660.
- [62] QI C R, YI L, SU H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017: 5105-5114.
- [63] SHI S, WANG X, LI H. Pointcnn: 3D object proposal generation and detection from point cloud[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 770-779.
- [64] CHEN X, MA H, WAN J, et al. Multi-view 3D object detection network for autonomous driving[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1907-1915.
- [65] KU J, MOZIFIAN M, LEE J, et al. Joint 3D proposal generation and object detection from view aggregation[C]// 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2018: 1-8.
- [66] SINDAGI V A, ZHOU Y, TUZEL O. Mvx-net: Multimodal voxelnet for 3D object detection[C]// 2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019: 7276-7282.
- [67] ZHOU Y, TUZEL O. Voxelnet: End-to-end learning for point cloud based 3D object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 4490-4499.
- [68] QI C R, LIU W, WU C, et al. Frustum pointnets for 3D object detection from RGB-D data[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 918-927.
- [69] WANG Z, JIA K. Frustum convnet: Sliding frustums to aggregate local point-wise features for amodal 3D object detection[C]// 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2019: 1742-1749.
- [70] KIRILLOV A, HE K, GIRSHICK R, et al. Panoptic segmentation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 9404-9413.
- [71] KIRILLOV A, GIRSHICK R, HE K, et al. Panoptic feature pyramid networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 6399-6408.
- [72] PORZI L, BULO S R, COLOVIC A, et al. Seamless scene segmentation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 8277-8286.
- [73] WU B, WAN A, YUE X, et al. Squeezeseg: Convolutional neural nets with recurrent crf for real-time road-object segmentation from 3D lidar point cloud[C]// 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018: 1887-1893.
- [74] MILIOTO A, VIZZO I, BEHLEY J, et al. Rangenet++: Fast and accurate lidar semantic segmentation[C]// 2019 IEEE/RSJ International Conference on Intelligent Robots And Systems (IROS). IEEE, 2019: 4213-4220.
- [75] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3431-3440.
- [76] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]// Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015: 234-241.
- [77] MILLETARI F, NAVAB N, AHMADI S A. V-net: Fully convolutional neural networks for volumetric medical image segmentation[C]// 2016 Fourth International Conference on 3D Vision. 2016: 565-571.
- [78] BADRINARAYANAN V, KENDALL A, CIPOLLA R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions



- on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.
- [79] LIN G, MILAN A, SHEN C, et al. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017 : 1925-1934.
- [80] WU H, ZHANG J, HUANG K, et al. Fastfcn: Rethinking dilated convolution in the backbone for semantic segmentation[J/OL]. ArXiv, [2019-03-28]. <https://doi.org/10.48550/arXiv.1903.11816>.
- [81] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017 : 2881-2890.
- [82] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Deeplab : Semantic image segmentation with deep convolutional nets , atrous convolution , and fully connected crfs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4): 834-848.
- [83] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[J/OL]. ArXiv , [2016-04-30]. <https://doi.org/10.48550/arXiv.1511.07122>.
- [84] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017: 6000-6010.
- [85] ZHENG S, LU J, ZHAO H, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 6881-6890.
- [86] YASRAB R. Ecrv: An encoder-decoder based convolution neural network (CNN) for road-scene understanding[J]. Journal of Imaging, 2018, 4(10): 116-125.
- [87] CHEN P R, HANG H M, CHAN S W, et al. Dsnet: An efficient cnn for road scene segmentation[J]. APSIPA Transactions on Signal and Information Processing, 2020, 9(27): 127-137.
- [88] ZHANG X, CHEN Z, WU Q M, et al. Fast semantic segmentation for scene perception[J]. IEEE Transactions on Industrial Informatics, 2019, 15(2): 1183-1192.
- [89] PENG Y, HAN W, OU Y. Semantic segmentation model for road scene based on encoder-decoder structure[C]// 2019 IEEE International Conference on Robotics and Biomimetics (ROBIO). IEEE, 2019: 1927-1932.
- [90] CORDTS M, OMRAN M, RAMOS S, et al. The cityscapes dataset for semantic urban scene understanding[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 3213-3223.
- [91] TEICHMANN M, WEBER M, ZOELLNER M, et al. Multinet : Real-time joint semantic reasoning for autonomous driving[C]// 2018 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2018: 1013-1020.
- [92] GEIGER A, LENZ P, STILLER C, et al. Vision meets robotics: The kitti dataset[J]. The International Journal of Robotics Research, 2013, 32(11): 1231-1237.
- [93] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]// Proceedings of the IEEE International Conference On Computer Vision. 2017: 2961-2969.
- [94] CAI Z, VASCONCELOS N. Cascade R-CNN: High quality object detection and instance segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 43(5): 1483-1498.
- [95] HUANG Z, HUANG L, GONG Y, et al. Mask scoring R-CNN[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019 : 6409-6418.
- [96] KIRILLOV A, WU Y, HE K, et al. Pointrend: Image segmentation as rendering[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 9799-9808.
- [97] BOLYA D, ZHOU C, XIAO F, et al. Yolact: Real-time instance segmentation[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 9157-9166.
- [98] BOLYA D, ZHOU C, XIAO F, et al. Yolact++ better real-time instance segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(2): 1108-1121.
- [99] CHEN H, SUN K, TIAN Z, et al. Blendmask: Top-down meets bottom-up for instance segmentation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 8573-8581.
- [100] YING H, HUANG Z, LIU S, et al. Embedmask: Embedding coupling for one-stage instance segmentation[J/OL]. ArXiv , [2019-12-05]. <https://doi.org/10.48550/arXiv.1912.01954>.
- [101] TIAN Z, ZHANG B, CHEN H, et al. Instance and panoptic segmentation using conditional convolutions[J]. IEEE Transactions on Pattern Analysis and Machine

- Intelligence, 2022, 45(1): 669-680.
- [102] RAKAI L, SONG H, SUN S J, et al. Data association in multiple object tracking : A survey of recent techniques[J]. Expert Systems with Applications, 2022, 192: 116300.
- [103] BASHAR M, ISLAM S, HUSSAIN K K, et al. Multiple object tracking in recent times: A literature review[J/OL]. ArXiv , [2022-09-11]. <https://doi.org/10.48550/arXiv.2209.04796>.
- [104] BEWLEY A, GE Z, OTT L, et al. Simple online and realtime tracking[C]// 2016 IEEE International Conference on Image Processing (ICIP). IEEE, 2016: 3464-3468.
- [105] WOJKE N, BEWLEY A, PAULUS D. Simple online and realtime tracking with a deep association metric[C]// 2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017: 3645-3649.
- [106] VOIGTLAENDER P, KRAUSE M, OSEP A, et al. Mots : Multi-object tracking and segmentation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 7942-7951.
- [107] KIM A, OŠEP A, LEAL-TAIXÉ L. Eagermot: 3D multi-object tracking via sensor fusion[C]// 2021 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2021: 11315-11321.
- [108] WANG Y H. SMILEtrack: SiMilarity LEarning for multiple object tracking[J/OL]. ArXiv, [2023-08-20]. <https://doi.org/10.48550/arXiv.2211.08824>.
- [109] AHARON N, ORFAIG R, BOBROVSKY B Z. BoT-SORT : Robust associations multi-pedestrian tracking[J/OL]. ArXiv , [2022-07-07]. <https://doi.org/10.48550/arXiv.2206.14651>.
- [110] DENDORFER P, OSEP A, MILAN A, et al. Motchallenge: A benchmark for single-camera multiple target tracking[J]. International Journal of Computer Vision, 2021, 129: 845-881.
- [111] DU Y, ZHAO Z, SONG Y, et al. StrongSORT: Make deepsort great again[J]. IEEE Transactions on Multimedia, 2023, 1-14.
- [112] CAO J, PANG J, WENG X, et al. Observation-centric sort: Rethinking sort for robust multi-object tracking[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 9686-9696.
- [113] MAGGIOLINO G, AHMAD A, CAO J, et al. Deep oc-sort : Multi-pedestrian tracking by adaptive re-identification[J/OL]. ArXiv , [2023-02-23]. <https://doi.org/10.48550/arXiv.2302.11813>.
- [114] WANG Z, ZHENG L, LIU Y, et al. Towards real-time multi-object tracking[C]// Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16. Springer International Publishing, 2020: 107-122.
- [115] CHAABANE M, ZHANG P, BEVERIDGE J R, et al. Deft: Detection embeddings for tracking[J/OL]. ArXiv, [2021-06-06]. <https://doi.org/10.48550/arXiv.2102.02267>.
- [116] ZHOU X, WANG D, KRÄHENBÜHL P. Objects as points[J/OL]. ArXiv , [2019-04-25]. <https://doi.org/10.48550/arXiv.1904.07850>.
- [117] YIN T, ZHOU X, KRAHENBUHL P. Center-based 3D object detection and tracking[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 11784-11793.
- [118] ZHANG Y, WANG C, WANG X, et al. Fairmot: On the fairness of detection and re-identification in multiple object tracking[J]. International Journal of Computer Vision, 2021, 129: 3069-3087.
- [119] PANG Z, LI Z, WANG N. Simpletrack: Understanding and rethinking 3D multi-object tracking[C] //Computer Vision—ECCV 2022 Workshops: Tel Aviv, Israel, October 23-27, 2022, Proceedings, Part I. Cham: Springer Nature Switzerland, 2023: 680-696.
- [120] ZHOU X, YIN T, KOLTUN V, et al. Global tracking transformers[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 8771-8780.
- [121] XU Y, BAN Y, DELORME G, et al. TransCenter: Transformers with dense representations for multiple-object tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(6): 7820-7835.
- [122] CHU P, WANG J, YOU Q, et al. Transmot : Spatial-temporal graph transformer for multiple object tracking[C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2023: 4870-4880.
- [123] DAI P, WENG R, CHOI W, et al. Learning a proposal classifier for multiple object tracking[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 2443-2452.
- [124] ZAECH J N, LINIGER A, DAI D, et al. Learnable online graph representations for 3D multi-object tracking[J]. IEEE Robotics and Automation Letters,

- 2022, 7(2): 5103-5110.
- [125] HUANG Y, DU J, YANG Z, et al. A survey on trajectory-prediction methods for autonomous driving[J]. IEEE Transactions on Intelligent Vehicles, 2022, 7(3): 652-674.
- [126] WU W, CHEN M, LI J, et al. An extended social force model via pedestrian heterogeneity affecting the self-driven force[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(7): 7974-7986.
- [127] HAN Y, CHAO Q, JIN X. A simplified force model for mixed traffic simulation[J]. Computer Animation and Virtual Worlds, 2021, 32(1): e1974.
- [128] WANG J, LÜ W, JIANG Y, et al. A cellular automata approach for modelling pedestrian-vehicle mixed traffic flow in urban city[J]. Applied Mathematical Modelling, 2023, 115: 1-33.
- [129] TAMANG N, SUN Y. Application of the dynamic Monte Carlo method to pedestrian evacuation dynamics[J]. Applied Mathematics and Computation, 2023, 445: 127876.
- [130] SCHÖLLER C, ARAVANTINOS V, LAY F, et al. What the constant velocity model can teach us about pedestrian motion prediction[J]. IEEE Robotics and Automation Letters, 2020, 5(2) : 1696-1703.
- [131] HE G, LI X, LV Y, et al. Probabilistic intention prediction and trajectory generation based on dynamic bayesian networks[C]// 2019 Chinese Automation Congress (CAC). IEEE, 2019: 2646-2651.
- [132] LI Y, LU X Y, WANG J, et al. Pedestrian trajectory prediction combining probabilistic reasoning and sequence learning[J]. IEEE Transactions on Intelligent Vehicles, 2020, 5(3): 461-474.
- [133] QUINTERO R, PARRA I, LORENZO J, et al. Pedestrian intention recognition by means of a hidden markov model and body language[C]// 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2017: 1-7.
- [134] DENG Q, SÖFFKER D. Improved driving behaviors prediction based on fuzzy logic-hidden markov model (fl-hmm)[C]// 2018 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2018: 2003-2008.
- [135] MÍNGUEZ R Q, ALONSO I P, FERNÁNDEZ-LLORCA D, et al. Pedestrian path, pose, and intention prediction through gaussian process dynamical models and pedestrian activity recognition[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 20(5): 1803-1814.
- [136] KUMAR P, PERROLLAZ M, LEFEVRE S, et al. Learning-based approach for online lane change intention prediction[C]// 2013 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2013: 797-802.
- [137] HUANG R, XUE H, PAGNUCCO M, et al. Multimodal trajectory prediction : A survey[J/OL]. ArXiv, [2023-02-21]. <https://doi.org/10.48550/arXiv.2302.10463>.
- [138] ALAHI A, GOEL K, RAMANATHAN V, et al. Social lstm : Human trajectory prediction in crowded spaces[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016 : 961-971.
- [139] KAWASAKI A, SEKI A. Multimodal trajectory predictions for urban environments using geometric relationships between a vehicle and lanes[C]// 2020 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2020: 9203-9209.
- [140] XING Y, LV C, CAO D. Personalized vehicle trajectory prediction based on joint time-series modeling for connected vehicles[J]. IEEE Transactions on Vehicular Technology, 2019, 69(2): 1341-1352.
- [141] DAI S, LI L, LI Z. Modeling vehicle interactions via modified LSTM models for trajectory prediction[J]. IEEE Access, 2019, 7: 38287-38296.
- [142] ZHANG T, SONG W, FU M, et al. Vehicle motion prediction at intersections based on the turning intention and prior trajectories model[J]. IEEE/CAA Journal of Automatica Sinica, 2021, 8(10): 1657-1666.
- [143] NIKHIL N, TRAN MORRIS B. Convolutional neural network for trajectory prediction[C]// Proceedings of the European Conference on Computer Vision (ECCV) Workshops. 2018: 186-196.
- [144] BAI S, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[J/OL]. ArXiv, [2018-04-19]. <https://doi.org/10.48550/arXiv.1803.01271>.
- [145] CHOU F C, LIN T H, CUI H, et al. Predicting motion of vulnerable road users using high-definition maps and efficient convnets[C]// 2020 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2020: 1655-1662.
- [146] CUI H, NGUYEN T, CHOU F C, et al. Deep kinematic models for kinematically feasible vehicle trajectory predictions[C]// 2020 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2020 : 10563-10569.

- [147] DEO N, TRIVEDI M M. Convolutional social pooling for vehicle trajectory prediction[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2018: 1468-1476.
- [148] XIE G, SHANGGUAN A, FEI R, et al. Motion trajectory prediction based on a CNN-LSTM sequential model[J]. Science China Information Sciences, 2020, 63: 1-21.
- [149] DANIEL N, LAREY A, AKNIN E, et al. PECNet: A deep multi-label segmentation network for eosinophilic esophagitis biopsy diagnostics[J/OL]. ArXiv, [2021-03-02]. <https://doi.org/10.48550/arXiv.2103.02015>.
- [150] ZHAO H, GAO J, LAN T, et al. Tnt: Target-driven trajectory prediction[C]// Conference on Robot Learning. PMLR, 2021: 895-904.
- [151] WANG C, WANG Y, XU M, et al. Stepwise goal-driven networks for trajectory prediction[J]. IEEE Robotics and Automation Letters, 2022, 7(2): 2716-2723.
- [152] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
- [153] GUPTA A, JOHNSON J, FEI-FEI L, et al. Social gan: Socially acceptable trajectories with generative adversarial networks[C]// Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition. 2018: 2255-2264.
- [154] HEGDE C, DASH S, AGARWAL P. Vehicle trajectory prediction using gan[C]// 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC). IEEE, 2020: 502-507.
- [155] WANG Y, ZHAO S, ZHANG R, et al. Multi-vehicle collaborative learning for trajectory prediction with spatio-temporal tensor fusion[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 23(1): 236-248.
- [156] TANG C, SALAKHUTDINOV R. Multiple futures prediction[C]// Proceedings of the 33rd International Conference on Neural Information Processing Systems. 2019: 15424-15434.
- [157] WANG Y, ZHAO S, ZHANG R, et al. Multi-vehicle collaborative learning for trajectory prediction with spatio-temporal tensor fusion[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 23(1): 236-248.
- [158] CASAS S, GULINO C, SUO S, et al. Implicit latent variable model for scene-consistent motion forecasting[C]// Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XXIII 16. Springer International Publishing, 2020: 624-641.
- [159] LI X, YING X, CHUAH M C. Grip: Graph-based interaction-aware trajectory prediction[C]// 2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, 2019: 3960-3966.
- [160] LI X, YING X, CHUAH M C. Grip++: Enhanced graph-based interaction-aware trajectory prediction for autonomous driving[J/OL]. ArXiv, [2020-05-19]. <https://doi.org/10.48550/arXiv.1907.07792>.
- [161] MOHAMED A, QIAN K, ELHOSEINY M, et al. Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 14424-14432.
- [162] AN J, LIU W, LIU Q, et al. DGInet: Dynamic graph and interaction-aware convolutional network for vehicle trajectory prediction[J]. Neural Networks, 2022, 151: 336-348.
- [163] LI R, QIN Y, WANG J, et al. AMGB: Trajectory prediction using attention-based mechanism GCN-BiLSTM in IOV[J]. Pattern Recognition Letters, 2023, 169: 17-27.
- [164] HADDAD S, WU M, WEI H, et al. Situation-aware pedestrian trajectory prediction with spatio-temporal attention model[J/OL]. ArXiv, [2019-02-13]. <https://doi.org/10.48550/arXiv.1902.05437>.
- [165] MESSAOUD K, DEO N, TRIVEDI M M, et al. Trajectory prediction for autonomous driving based on multi-head attention with joint agent-map representation[C]// 2021 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2021: 165-170.
- [166] WONG K, GU Y, KAMIJO S. Mapping for autonomous driving: Opportunities and challenges[J]. IEEE Intelligent Transportation Systems Magazine, 2020, 13(1): 91-106.
- [167] GUANETTI J, KIM Y, BORRELLI F. Control of connected and automated vehicles: State of the art and future challenges[J]. Annual Reviews in Control, 2018, 45: 18-40.
- [168] BLOCHLIGER F, FEHR M, DYMCHYK M, et al. Topomap: Topological mapping and navigation based on visual SLAM maps[C]// 2018 IEEE International



- Conference on Robotics and Automation (ICRA). IEEE, 2018: 3818-3825.
- [169] SIAM M, ELKERDAWY S, JAGERSAND M, et al. Deep semantic segmentation for automated driving: Taxonomy, roadmap and challenges[C]// 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2017: 1-8.
- [170] ULLAH I, SU X, ZHANG X, et al. Simultaneous localization and mapping based on Kalman filter and extended Kalman filter[J]. Wireless Communications and Mobile Computing, 2020, 2020: 1-12.
- [171] ZHANG F, LI S, YUAN S, et al. Algorithms analysis of mobile robot SLAM based on Kalman and particle filter[C]// 2017 9th International Conference on Modelling, Identification and Control (ICMIC). IEEE, 2017: 1050-1055.
- [172] CHEN S, ZHOU B, JIANG C, et al. A lidar/visual slam backend with loop closure detection and graph optimization[J]. Remote Sensing, 2021, 13(14): 2720.
- [173] 阴贺生, 裴硕, 徐磊, 等. 多机器人视觉同时定位与建图技术研究综述[J]. 机械工程学报, 2022, 58(11): 11-36.
- YIN Hesheng, PEI Shuo, XU Lei, et al. Review of research on multi-robot visual simultaneous localization and mapping[J]. Journal of Mechanical Engineering, 2022, 58(11): 11-36.
- [174] DAVISON A J, REID I D, MOLTON N D, et al. MonoSLAM: Real-time single camera SLAM[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1052-1067.
- [175] KLEIN G, MURRAY D. Parallel tracking and mapping for small AR workspaces[C]// 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality. IEEE, 2007: 225-234.
- [176] MUR-ARTAL R, MONTIEL J M M, TARDOS J D. ORB-SLAM: A versatile and accurate monocular SLAM system[J]. IEEE Transactions on Robotics, 2015, 31(5): 1147-1163.
- [177] MUR-ARTAL R, TARDÓS J D. ORB-SLAM2: An open-source slam system for monocular, stereo, and RGB-D cameras[J]. IEEE Transactions on Robotics, 2017, 33(5): 1255-1262.
- [178] QIN T, LI P, SHEN S. Vins-mono: A robust and versatile monocular visual-inertial state estimator[J]. IEEE Transactions on Robotics, 2018, 34(4): 1004-1020.
- [179] CAMPOS C, ELVIRA R, RODRÍGUEZ J J G, et al. ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multimap SLAM[J]. IEEE Transactions on Robotics, 2021, 37(6): 1874-1890.
- [180] NEWCOMBE R A, LOVEGROVE S J, DAVISON A J. DTAM: Dense tracking and mapping in real-time[C]// 2011 International Conference on Computer Vision. IEEE, 2011: 2320-2327.
- [181] ZHOU H, UMMENHOFER B, BROX T. Deeptam: Deep tracking and mapping[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018: 822-838.
- [182] BLOESCH M, CZARNOWSKI J, CLARK R, et al. CodeSLAM—learning a compact, optimisable representation for dense visual SLAM[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 2560-2568.
- [183] TATENO K, TOMBARI F, LAINA I, et al. CNN-SLAM: Real-time dense monocular slam with learned depth prediction[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 6243-6252.
- [184] ENGEL J, SCHÖPS T, CREMERS D. LSD-SLAM: Large-scale direct monocular SLAM[C]// Computer Vision-ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part II 13. Springer International Publishing, 2014: 834-849.
- [185] TANG C, TAN P. Ba-net: dense bundle adjustment network[J/OL]. ArXiv, [2019-08-25]. <https://doi.org/10.48550/arXiv.1806.04807>.
- [186] KOESTLER L, YANG N, ZELLER Ns, et al. Tandem: Tracking and dense mapping in real-time using deep multi-view stereo[C]// Conference on Robot Learning. PMLR, 2022: 34-45.
- [187] TEED Z, DENG J. Droid-slam: Deep visual slam for monocular, stereo, and rgb-d cameras[C]// 35th Conference on Neural Information Processing Systems, NeurIPS 2021. Neural information processing systems foundation, 2021: 16558-16569.
- [188] TEED Z, DENG J. Raft: Recurrent all-pairs field transforms for optical flow[C]// Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part II 16. Springer International Publishing, 2020: 402-419.
- [189] BURRI M, NIKOLIC J, GOHL P, et al. The EuRoC micro aerial vehicle datasets[J]. The International Journal of Robotics Research, 2016, 35(10): 1157-1163.

- [190] WANG W, ZHU D, WANG X, et al. Tartanair: A dataset to push the limits of visual slam[C]// 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2020: 4909-4916.
- [191] SUCAR E, LIU S, ORTIZ J, et al. iMAP: Implicit mapping and positioning in real-time[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 6229-6238.
- [192] ZHU Z, PENG S, LARSSON V, et al. Nice-SLAM: Neural implicit scalable encoding for SLAM[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 12786-12796.
- [193] MILDENHALL B, SRINIVASAN P P, TANCIK M, et al. Nerf: Representing scenes as neural radiance fields for view synthesis[J]. Communications of the ACM, 2021, 65(1): 99-106.
- [194] FAYYAD J, JARADAT M A, GRUYER D, et al. Deep learning sensor fusion for autonomous vehicle perception and localization: A review[J]. Sensors, 2020, 20(15): 4220.
- [195] YEE R, CHAN E, CHENG B, et al. Collaborative perception for automated vehicles leveraging vehicle-to-vehicle communications[C]// 2018 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2018: 1099-1106.
- [196] LI H, SIMA C, DAI J, et al. Delving into the devils of bird's-eye-view perception: A review, evaluation and recipe[J/OL]. ArXiv, [2022-09-28]. <https://doi.org/10.48550/arXiv.2209.05324>.
- [197] LI Z, WANG W, LI H, et al. Bevformer: Learning bird's-eye-view representation from multi-camera images via spatiotemporal transformers[C]// Computer Vision—ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23-27, 2022, Proceedings, Part IX. Cham: Springer Nature Switzerland, 2022: 1-18.
- [198] YANG C, CHEN Y, TIAN H, et al. BEVFormer v2: Adapting modern image backbones to bird's-eye-view recognition via perspective supervision[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 17830-17839.
- [199] XU R, TU Z, XIANG H, et al. CoBEVT: Cooperative bird's eye view semantic segmentation with sparse transformers[J/OL]. ArXiv, [2022-09-25]. <https://doi.org/10.48550/arXiv.2207.02202>.
- [200] TONG W, SIMA C, WANG T, et al. Scene as occupancy[J/OL]. ArXiv, [2023-06-26]. <https://doi.org/10.48550/arXiv.2306.02851>.
- [201] WEI Y, ZHAO L, ZHENG W, et al. SurroundOcc: Multi-camera 3D occupancy prediction for autonomous driving[J/OL]. ArXiv, [2023-08-27]. <https://doi.org/10.48550/arXiv.2303.09551>.
- [202] CAO A Q, DE CHARETTE R. Monoscene: Monocular 3d semantic scene completion[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 3991-4001.

作者简介: 澎湃, 男, 1993 年出生, 博士研究生。主要研究方向为智能汽车多模态融合感知。

E-mail: pengpai@seu.edu.cn

殷国栋(通信作者), 男, 1976 年出生, 博士, 教授, 博士研究生导师。主要研究方向为先进电动汽车、车辆动力学与控制、智能汽车和车辆主动安全控制。

E-mail: ygd@seu.edu.cn