·激光雷达感知与自动驾驶多模态3D目标检测技术专题·

自动驾驶中LiDAR与相机融合的目标检测技术研究进展*

庞亚军[1,2]　张继文[1,2]　宁一凡[1,2]

（1.河北工业大学电子信息工程学院，天津 300401；2.河北省先进激光技术与装备重点实验室，天津 300401）

【摘要】针对自动驾驶场景中环境感知能力的安全性和可靠性面临的技术瓶颈，系统梳理自动驾驶中激光雷达（LiDAR）与相机融合的目标检测技术路线及发展动态，通过分析单一传感器在复杂路况下的局限性，论证多模态融合感知的必然趋势。研究发现：在遮挡、远距离和复杂环境条件下，传感器融合技术能够显著提升目标检测效果；在nuScenes、KITTI等公开数据集的验证试验中，传感融合技术有效增强系统在多变场景中的感知鲁棒性与检测性能；当前技术仍面临实时性不足、算法复杂度高及成本制约等挑战，未来仍需优化融合算法的精度和实时性，推动多模态协同感知的标准化。

主题词：自动驾驶　目标检测　激光雷达　相机　数据融合

中图分类号：TP391.41　　**文献标志码：**A　　**DOI:** 10.19620/j.cnki.1000-3703.20250225

Research Progress on Object Detection Technology with LiDAR and Camera Fusion in Autonomous Driving

Pang Yajun[1,2], Zhang Jiwen[1,2], Ning Yifan[1,2]

（1. College of Electronical and Information Engineering, Hebei University of Technology, Tianjin 300401; 2. Hebei Key Laboratory of Advanced Laser Technology and Equipment, Hebei University of Technology, Tianjin 300401）

【Abstract】 Facing the technical bottlenecks in ensuring the safety and reliability of environmental perception in autonomous driving, this paper systematically reviews the technological roadmap and recent developments of LiDAR-camera fusion - based object detection. By analyzing the limitations of single-sensor perception under complex road conditions, the study demonstrates the inevitability of multimodal sensor fusion. The findings indicate that sensor fusion can significantly improve object detection performance in scenarios involving occlusion, long distances, and complex environments. Experiments on public datasets such as nuScenes and KITTI further verify that sensor fusion effectively enhances perception robustness and detection accuracy in diverse scenes. However, current technologies still face challenges such as insufficient real-time performance, high algorithmic complexity, and cost constraints. Future research should focus on improving the precision and efficiency of fusion algorithms and promoting the standardization of multimodal collaborative perception.

Key words: Autonomous Driving, Object Detection, LiDAR, Camera, Data Fusion

【引用格式】 庞亚军, 张继文, 宁一凡. 自动驾驶中LiDAR与相机融合的目标检测技术研究进展[J]. 汽车技术, 2026(2): 18-31.

PANG Y J, ZHANG J W, NING Y F. Research Progress on Object Detection Technology with LiDAR and Camera Fusion in Autonomous Driving[J]. Automobile Technology, 2026(2): 18-31.

1　前言

随着自动驾驶技术的快速发展，目标检测已成为检验自动驾驶技术落地的决定性因素之一，不仅要求算法具备强大的特征提取与模式识别能力，还需兼顾实时性与鲁棒性，以适应不同光照条件、天气状况及复杂交通环境的挑战。

早期的环境感知任务主要依赖单一传感器，如相机、

*基金项目：国家自然科学基金项目（61905063）；河北工业大学创新研究院（石家庄）石家庄市科技合作专项（SJZZXC23011）。

通信作者：庞亚军（1988—），副教授，主要研究方向为激光雷达与激光探测技术，yjpang@hebut.edu.cn。

激光雷达（Light Detection and Ranging，LiDAR）。然而，单一传感器感知能力有限，近年来多探索LiDAR与相机联合标定方法，结合二者优势，显著提升感知系统的精度、鲁棒性和场景适应性。在技术迭代推动下，LiDAR与相机融合感知策略已在众多量产车型及商业系统中得到充分验证。Waymo无人驾驶平台采用高性能激光雷达与多路摄像头的融合方案，通过构建鸟瞰图（Bird's Eye View，BEV）空间特征，实现了高精度的目标检测与路径规划[1]；小鹏汽车在其城市自动导航辅助驾驶（Navigation Guided Pilot，NGP）系统中集成激光雷达、摄像头与惯性测量单元（Inertial Measurement Unit，IMU），构建时空协同感知架构，实现了在城市交通、隧道等复杂场景下对行人、非机动车的精准检测[2]；百度Apollo自动驾驶平台在多城市进行Robotaxi常态化运营，其感知模块通过多模态融合网络整合LiDAR与相机信息，有效提升了遮挡处理能力与全时段感知稳定性[3]；华为乾崑智驾ADS 4（Advanced Driving System 4）深度集成激光雷达、摄像头、毫米波雷达等多种传感器，显著提升了复杂场景下的遮挡处理能力和全天候感知稳定性[4]。

LiDAR与相机的多传感器融合感知不仅具备理论优势，更在工业界中展现出强大的落地潜力与实用价值。本文旨在系统综述激光雷达与相机融合目标检测技术在自动驾驶领域的最新进展，全面剖析其技术原理和应用现状，为相关领域研究者提供参考与启发，推动自动驾驶技术向更高的安全性和可靠性迈进。

2 自动驾驶系统及目标检测技术的发展现状

2.1 自动驾驶系统演进与环境感知需求

自动驾驶系统（Autonomous Driving System，ADS）在无需人类驾驶员干预的情况下自动控制车辆行驶，国际自动机工程师学会（Society of Automotive Engineers，SAE）将自动驾驶能力划分为L0~L5 6个等级。为了提升对分级体系的理解，增加各等级对应典型传感器配置与典型车型进行辅助解读，如表1所示。

表1 SAE自动驾驶分级体系详情

等级	名称	功能	典型传感器配置	典型车型
L0	无自动驾驶	车辆完全由驾驶员操控，无自动驾驶辅助功能	无	1960年大众基础款甲壳虫1200
L1	驾驶辅助	车辆具备单一的驾驶辅助功能，如自适应巡航、车道偏离预警等	摄像头+毫米波雷达	2018款日产轩逸1.6 L CVT智行版
L2	部分自动驾驶	车辆可同时控制横向和纵向运动，如自动紧急制动结合自适应巡航等组合功能	摄像头+毫米波雷达+超声波雷达	特斯拉Model 3、本田雅阁2.0T Honda Sensing
L3	有条件自动驾驶	在特定条件下，车辆可完成大部分驾驶任务，但在系统请求时驾驶员需及时响应并接管	摄像头+毫米波雷达+激光雷达（选配）+全球导航卫星系统+IMU	2018款奥迪A8、2021款本田Legend Hybrid EX
L4	高度自动驾驶	在特定区域和条件下，车辆能完成所有驾驶操作，即便出现系统故障也能安全停车	多线激光雷达+多摄像头+多雷达+高精定位模块	Apollo Robotaxi、Waymo One Robotaxi
L5	完全自动驾驶	车辆在任何可行驶条件下都能实现自动驾驶，无需驾驶员，车内甚至无需驾驶位等相关设备	激光雷达阵列+多路摄像头+各类雷达+高精定位与地图系统	尚无正式量产车型

目标检测技术直接影响自动驾驶系统在复杂、多变交通场景中的安全性与稳定性。传统的目标检测系统受限于计算能力与传感器性能，常依赖单一传感器完成环境感知任务，在实际应用中难以保证复杂环境下感知的可靠性。Hwang等[5]基于KAIST多光谱行人检测数据集，使用RGB（Red-Green-Blue）相机在白天条件下的平均精度（Average Precision，AP）超过82%，夜间则不足51%，说明相机在弱光环境下的检测能力大幅降低；Kim等[6]通过实际道路的雨雾环境测试发现，在降雨强度为30~40 mm/h、能见度50 m的浓雾条件下，LiDAR的点云数量和反射强度显著下降，目标物体的点云数量相比晴天的点云数量降低74%。因此，为满足复杂环境的目标检测需求，研究重心已转向多传感器融合策略[7]。

深度学习技术的兴起为多传感器融合注入新动力，LiDAR与相机深度学习算法凭借卓越的特征提取能力，能够深入挖掘LiDAR点云与相机图像间的数据互补优势，提升感知系统的精度与鲁棒性。基于点云投影的数据融合[8-10]、卷积神经网络[11]（Convolutional Neural Network，CNN）和Transformer[12]特征融合等方法，既丰富了多传感器融合的技术理论，也为该技术的实际应用明确了可行的技术路径[13]。

2.2 相机检测技术的发展

相机利用透镜汇聚光线在图像传感器成像，并将光信号转为电信号，生成高分辨率的图像数据，包含丰富颜色、纹理及语义信息。在正常光照条件下，相机采集的数据可用于识别并区分车辆类型、行人和交通标志等[14-15]。而且相机的成本较低，是自动驾驶系统的核心传感器之一。

在早期相机的目标检测任务中，检测方法依赖手工设计特征，如哈尔特征、方向梯度直方图特征[16-17]，此类方法常结合支持向量机（Support Vector Machine，SVM）和AdaBoost等分类器，广泛用于行人检测和车辆识别任务。而基于深度学习的目标检测方法因检测精度和鲁棒性显著提升，已逐步成为行业主流选择。

2014年，Girshick等[18]提出基于区域的卷积神经网络（Regions with CNN Features，R-CNN），先选择性搜索生成候选区域，再利用CNN提取各区域的特征，最后通过分类器完成目标检测。该方法在PASCAL VOC 2007数据集上的平均精度均值（mean Average Precision，mAP）达到53.3%，同时AP也有所突破。但由于重复提取候选区域特征，推理速度仅为0.05 帧/s，无法满足自动驾驶的实时性需求。

2015年，Girshick等[19]提出快速区域卷积神经网络（Fast Region-based Convolutional Neural Network，Fast R-CNN），该网络引入感兴趣区域（Region of Interest，ROI）池化层，采用先全局特征、后局部特征提取，大幅减小计算量，提升了计算效率。Fast R-CNN在PASCAL VOC 2007数据集的mAP达到66.9%，推理速度提升至0.5 帧/s，更适用于高效场景处理。

2016年，Redmon等[20]提出YOLO（You Only Look Once）系列，将目标检测转化为回归问题，通过单阶段网络直接预测目标边界框和类别概率，在PASCAL VOC 2007数据集上的mAP达到63.4%，推理速度达45 帧/s。然而，YOLO对小目标检测存在精度不足的问题，以YOLOv3[21]为例，在KITTI数据集的车辆（Car）类别中等难度的验证中，模型的2D mAP达到79.2%，推理速度为25 帧/s，但在行人（Pedestrian）类别中等难度下验证的2D mAP仅为45.3%，说明模型对远距离行人或小型障碍物的检测性能有限。同年，Liu等[22]提出单阶段多框检测器（Single Shot Multibox Detector，SSD），通过在多尺度特征图上进行目标检测，增强了对小目标的检测能力。相较于YOLO，SSD在KITTI数据集的行人类别中等难度下验证，2D mAP达到50.1%，在精度上更具优势，适用于远距离或小型物体的检测场景。

2017年，Ren等[23]基于Fast R-CNN引入区域提议网络（Region Proposal Network，RPN），提出了Faster R-CNN模型。在PASCAL VOC 2007数据集上，该模型的mAP达到73.2%；在KITTI数据集的Car类别中等难度下的2D mAP约为81.5%，精度优于YOLO和SSD，但推理速度仅为5 帧/s。同年，Lin等[24]提出RetinaNet，系统性地解决了单阶段检测器中前景-背景类别严重不平衡的问题。通过引入焦点（Focal）损失，显著降低了大量易分类背景样本对整体损失的主导作用，从而提升了对难分类前景目标的检测能力。在COCO验证集上，模型采用ResNet-50-FPN主干网络的AP达到36.0%，采用ResNet-101-FPN主干网络的AP提升至39.1%。相比Faster R-CNN，RetinaNet的检测精度具有明显优势。

尽管相机在语义理解和成本方面具有优势，其应用仍面临局限性。相机对光照条件和恶劣天气较敏感，在强光、弱光或遮挡场景下的图像质量下降，目标检测任务难度骤增；相较于LiDAR，由于相机获取二维图像，无法直接提供目标物体的距离信息，虽然能够通过立体视觉、结构光、运动视差等技术间接估计深度[25-27]，但在复杂场景的深度估计中表现不佳。此外，现有的深度相机虽然能够直接输出像素级的深度信息，为二维图像提供空间几何约束，但其测距范围有限，仅适用于近距离测量。而自动驾驶对远距离目标的探测能力具有刚性需求，这超出了深度相机的有效工作距离。

2.3 LiDAR检测技术的发展

LiDAR基于光的飞行时间（Time of Flight，ToF）测量技术[28-29]，通过发射、接收激光束测量目标距离。LiDAR能够高精度地获取目标及场景的三维空间信息，在环境感知中表现出色，同时，其对环境光照条件的敏感性较低，即使在黑暗环境中，也能保持稳定的工作状态，为自动驾驶系统提供可靠的环境感知数据[30]。在路径规划、避障等关键任务中，使车辆能够精准识别障碍物，从而规划安全行驶路线[31]。

作为自动驾驶中提供高精度三维信息的传感器，早期LiDAR的目标检测主要依靠点云的几何与统计特征提取方法。例如，通过计算点云的曲率、法向量、平面度等几何属性，采用欧式聚类、区域生长等方法进行目标分割[32]。随着深度学习技术的兴起，基于LiDAR的目标检测逐渐向端到端的神经网络模型转变[33]，显著提升了模型的检测性能。

2017年，斯坦福大学团队[34]提出PointNet深度学习网络，开创了无序点云处理方法。该方法利用多层感知机[35]（Multi-Layer Perceptron，MLP）和最大池化操作提

取全局点云特征，无需点云预处理，适用于目标分类和分割任务，在ShapeNet数据集上的平均交并比（mean Intersection over Union，mIoU）达到83.7%。其优化版本PointNet++[36]通过引入层次特征提取机制，进一步增强了模型对局部特征的学习能力。

2018年，Zhou等[37]提出了首个基于体素的端到端3D目标检测网络VoxelNet，该方法先将稀疏点云进行体素化处理，再通过体素特征编码层（Voxel Feature Encoding，VFE）提取局部几何特征，最后利用三维区域提议网络（3D Region Proposal Network，3D RPN）在体素特征图中生成目标的三维候选框，并完成检测结果预测。VoxelNet在KITTI数据集的Car类别中等难度上验证，3D mAP达到65.11%，在远距离目标检测中表现优异。

2019年，Lang等[38]提出PointPillars，将垂直柱状（Pillars）结构点云投影到平面伪图像，利用2D卷积网络进行特征提取和目标检测。在KITTI数据集的Car类别中等难度验证中，该模型的3D mAP达到79.05%；在NVIDIA Titan Xp图形处理器上的推理速度达到62帧/s，能够实现较好的实时性与精度平衡。相较于VoxelNet，PointPillars在保持精度的同时大幅降低了计算复杂度，更适用于实时自动驾驶。

尽管LiDAR在三维感知和环境适应性方面具有显著优势，但仍面临一些挑战：由于LiDAR生成的点云数据较稀疏，在远距离或小目标检测中，可能无法捕捉目标物体完整的细节特征，从而导致目标检测、识别过程中丢失部分重要信息[39]。同时，LiDAR数据主要提供几何信息，仅能呈现目标的几何轮廓和距离数据，缺乏纹理和语义信息，难以直接判断物体的实际用途或类别属性[40]。此外，高端LiDAR设备的研发部署成本增加，同样成为自动驾驶技术落地的制约因素。

2.4 单一传感器的局限性与融合的必要性

目前，单一传感器难以满足不同场景下的感知需求，相机能够较好地理解语义信息，但其对环境光照的敏感、缺乏目标深度信息，使其对目标的远距离检测、极端环境的目标检测能力受限；LiDAR的点云稀疏性和缺乏语义信息，导致其在远距离目标检测和目标分类中表现不佳，同时高成本也限制其推广应用。

面对复杂交通环境的感知需求，多传感器融合技术已成为突破瓶颈的关键路径。LiDAR与视觉相机在感知维度上形成优势互补，前者能够为视觉图像提供深度基准，有效解决单目视觉的尺度歧义问题；后者通过丰富的纹理特征与语义标签，辅助LiDAR完成目标属性的精确分类。通过融合两者优势，可以显著提升目标检测的精度和鲁棒性。

3 LiDAR和相机融合技术的发展

LiDAR与相机的融合技术逐步发展，形成了多种融合策略。目前，主流的融合技术包含早期、中期和后期3种不同阶段的融合方法，随着融合技术的不断深化，部分研究将多种融合方式中的2种或3种进行交叉融合，全面汇聚不同层次的独特优势，最终形成了兼具综合性的多层次融合技术。现有的融合技术能够在不同阶段整合LiDAR和相机的信息，为自动驾驶的环境感知提供了多样化的解决方案，多传感器融合技术逐步深化，提升了目标检测的精度和鲁棒性。

3.1 LiDAR和相机数据的早期融合

早期融合是指在数据层面将LiDAR的原始点云与相机的图像数据直接结合，最大程度保留两者的基础信息。该方法能够充分利用LiDAR和相机的原始数据特性，为后续自动驾驶中的环境感知奠定坚实基础，在自动驾驶中已得到广泛应用。

3.1.1 基于点云投影的融合方法

2018年，Xu等[41]提出PointFusion，通过深度学习框架结合相机图像的语义信息与LiDAR点云的几何信息。该方法的核心在于分别提取两种模态的特征后，利用一个融合网络生成精确的3D边界框。在KITTI数据集中验证，PointFusion的密集型变体在中等难度的3D检测任务中的mAP达到59.46%，其最终版变体的3D mAP达到63%，相比之下，基于LiDAR单模态独立检查的mAP仅为42.31%，充分证明了融合方法的有效性。

2019年，Sindagi等[42]提出MVX-Net架构，该方法基于VoxelNet的体素化点云处理框架，借助早期融合机制实现图像特征向点云体素表示的投影。该网络包含PointFusion和VoxelFusion融合策略，PointFusion通过将3D点投影至图像提取特征，在点级别完成融合；VoxelFusion先将3D体素投影到图像平面，再进行体素级特征融合。在KITTI数据集中困难等级的3D检测任务中，VoxelFusion与PointFusion的3D AP分别达到66.8%、67.4%，均超过基线VoxNet（64.6%），解决了目标遮挡、目标过小等问题。

2020年，Zhang等[43]针对3D车辆检测任务，设计了一种多模态自适应特征融合网络MAFF-Net，通过动态调整图像和LiDAR特征的权重优化融合效果。该方法

在KITTI数据集的Car类别中等难度下，3D mAP达到75.04%，实现了较好的检测性能，进一步推动了早期融合技术的发展。

3.1.2　基于语义增强的融合方法

2020年，Vora等提出了创新性的顺序融合方法PointPainting，其架构如图1所示[44]：第一阶段精准识别并分割待检测目标；第二阶段为PointPainting融合环节，即将第一阶段所生成的语义信息与点云数据进行融合；第三阶段将整合的信息输入经典的3D点云检测网络。在KITTI数据集上针对Pedestrian困难类别进行验证，PointPainting结合PointRCNN检测器在3D检测任务的3D mAP达到54.58%，较原始PointRCNN提升5.8%；对数据集的骑行者（Cyclist）困难类别验证，超越原始PointRCNN（61.92%），实现了66.98%的3D mAP。

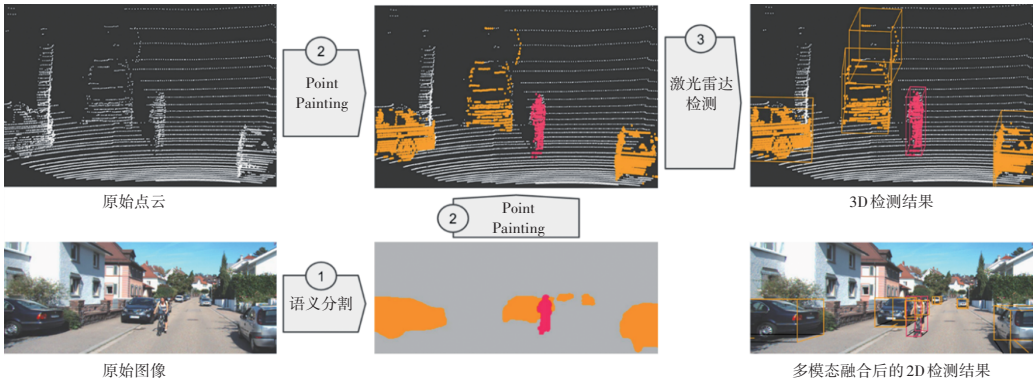


图1　PointPainting架构

2024年，Zhao等[45]提出的LIF-Seg将多模态融合的焦点转向3D LiDAR语义分割任务，通过早期融合使相机的图像特征与点云特征对齐并整合，各点均生成语义标签，提升点云的语义理解能力。在nuScenes验证集上，LIF-Seg的mIoU约为78.2%，相同任务中单一模态的Salsanext方法的mIoU仅为72.2%。

3.2　LiDAR和相机数据的中期融合

中期融合在特征层面进行操作，通过深度学习算法整合LiDAR提取的物体几何特征和相机提取的物体视觉特征，所融合的特征既包含三维空间信息和视觉纹理信息，为物体目标的检测和识别提供更丰富的特征描述。

3.2.1　基于多视图表示的融合方法

2017年，Chen等[46]提出MV3D（Multi-View 3D）网络，利用多视图编码和深度融合的原理，通过BEV生成3D候选框，结合深度融合方案与多视图区域特征，整合LiDAR点云和相机图像的特征信息。该方法将ImageNet预训练的VGG-16网络中采样权重初始化参数，参数量约为VGG-16网络的75%，经GeForce Titan X图形处理器训练，一张图像的推理时间约为0.36 s，在KITTI数据集的中等难度上验证，MV3D实现了89.05%的3D AP。

2018年，Ku等[47]受2D目标检测中特征金字塔网络（Feature Pyramid Network，FPN）的启发，提出AVOD（Aggregate View Object Detection network）架构。通过生成高分辨率特征图，结合多模态融合区域提议网络，在NVIDIA TITAN X图形处理器上的训练速度可达到10帧/s，其中，预处理耗时20 ms，推理耗时80 ms，速度较F-PointNet提升1.7倍。在KITTI数据集的Car类别中等难度验证中，3D AP达到71.88%。同年，Qi等[48]提出Frustum PointNets，创新性地将2D边界框扩展为3D视锥体，在视锥体内进行3D实例分割和无模态3D边界框回归。在KITTI数据集Car类别的中等难度下验证，Frustum PointNets的v1模型和v2模型的3D AP分别达到82.44%、84.02%，均超过MV3D。

2025年，Yu等[49]基于BEV实现了传感器融合与运动规划的一体化设计，提出了BevDrive模型。该模型由双向引导的BEV特征构建模块、双注意力BEV特征融合模块和基于BEV的运动规划模块构成。模型使用带权重衰减的自适应矩估计优化器（Adam with Decoupled Weight Decay，AdamW），在4块NVIDIA Tesla T4图形处理器上进行40个周期的训练。在CARLA Town05 Long基准测试中，BevDrive的驾驶得分、路线完成度等指标均优于其他方法；在Navsim基准测试中取得了感知-驾驶度量分数（Perception-Driving Metric Score，PDMS）的最高分数（83.8分）；在驾驶能力准则（Driving Ability Criterion，DAC）指标上表现突出，达到92.5分。经不同实际驾驶场景下的定量分析显示，在所有时间范围内，BevDrive的平均L2误差最低（0.08 m），优于LateFusion、Transfuser等基线方法。

3.2.2　基于深度优化的融合方法

2022年，Bai等[50]基于Transformer的架构融合

LiDAR和相机特征，提出TransFusion模型，通过软关联机制解决图像条件不佳和传感器校准问题。在训练过程中，对3D主干网络、第1个解码器层和前馈神经网络进行20个周期的预训练后，对融合模块进行6个周期的联合训练。全部试验在8个NVIDIA Tesla V100图形处理器上完成，批量大小为16。在nuScenes数据集上验证，TransFusion实现了68.9%的mAP和71.7%的nuScenes检测分数（nuScenes Detection Score，NDS）。同年，Li等[51]提出DeepFusion，通过融合深度相机和LiDAR特征，引入InverseAug和LearnableAlign技术解决特征对齐问题，并在V100图形处理器上训练模型，如图2所示[51]。在Waymo公开数据集上验证，DeepFusion在L1难度与L2难度上的AP分别达到83.6%、76%。

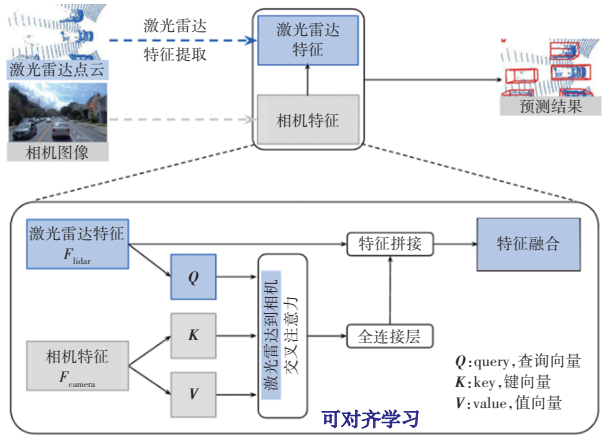


图2　DeepFusion融合流程

InverseAug将数据增强阶段后的关键点投影至2D相机坐标系，是DeepFusion实现多模态特征精确对齐的关键，其流程如图3所示。该方法通过逆向应用所有数据增强操作，将3D关键点从增强后的坐标系转换为原始坐标系，利用激光雷达和相机参数将3D关键点投影到相机特征上。

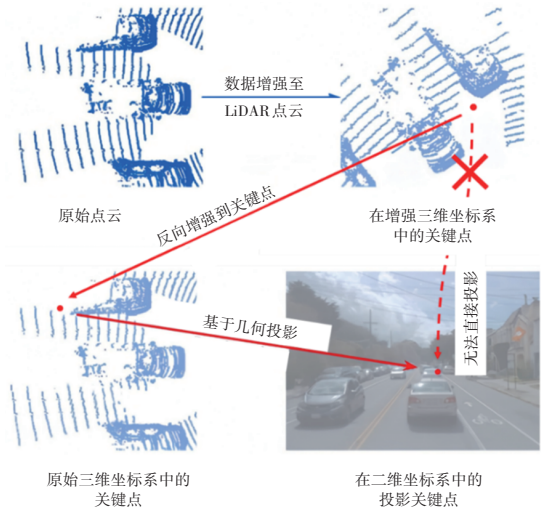


图3　InverseAug流程

2024年，Zhao等[52]基于BEV融合框架，通过改进相机和激光雷达编码器提升融合效果，提出了SimpleBEV模型（见图4）并在Titan V100图形处理器上进行训练。在nuScenes数据集上的3D mAP达到75.7%，超越了TransFusion（68.9%）。

在图4中，SimpleBEV包含相机与LiDAR 2个分支，其中，相机分支分为主分支和辅助分支，主分支提取图像特征并将其投影到BEV空间，辅助分支仅在训练阶段使用，用于提升深度估计的稳定性。此外，该模型通过级联深度估计网络提高相机深度预测精度，并利用LiDAR深度信息进行校正。SimpleBEV的深度估计模块采用两阶段级联结构，如图5所示[52]，第一阶段根据输入图像生成初始深度图，将与其对应的特征图拼接后，输入第二阶段网络以进一步细化深度估计结果；LiDAR点云被转换至相机坐标系并投影到图像平面，生成稀疏深度图。在融合生成最终深度图时，网络对LiDAR投影区域赋予更高权重，从而有效提升深度估计的整体精度与鲁棒性。

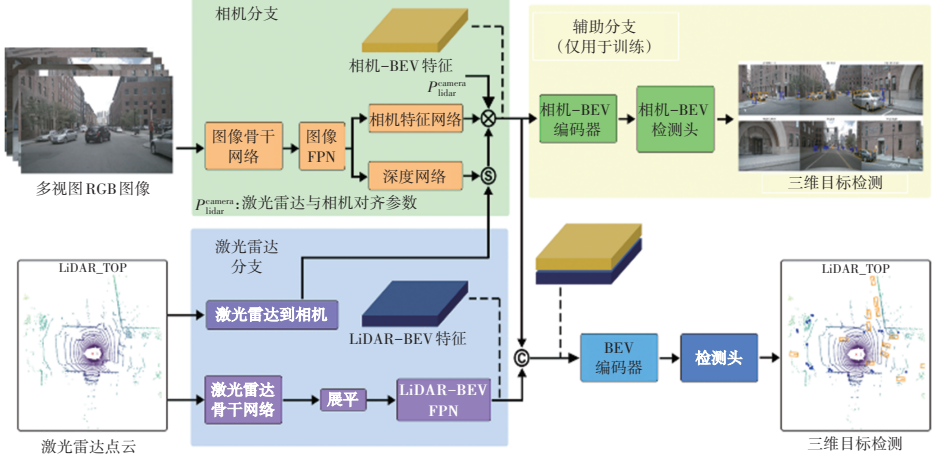


图4　SimpleBEV融合架构

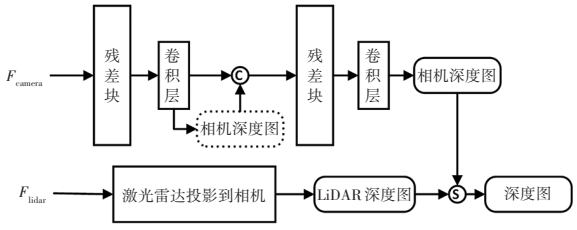


图5　深度估计流程

2025年，Li等[53]提出BEVFormer模型，用于LiDAR和相机数据学习BEV表示的多模态时空变换器编码器。该框架包含网格状BEV查询、多模态空间交叉注意力和时间自注意力层，有效聚合来自LiDAR和相机以及历史BEV特征的时空特征。在nuScenes测试集上验证，该模型实现了74.1%的NDS，相比仅使用LiDAR的基线模型的NDS提升了3.9%。

3.3　LiDAR和相机数据的后期融合

后期融合在决策层面进行，LiDAR和相机独立处理检测数据，生成各自的目标检测结果后，再通过融合算法整合结果，具有较高的灵活性和可扩展性，在自动驾驶中可同样展现出较强的适应性。

3.3.1　基于注意力机制的融合方法

2021年，Prakash等[54]提出TransFuser模型，基于注意力机制融合单视图图像和LiDAR数据，并利用Transformer自注意力机制捕捉3D场景的全局上下文。在CARLA模拟器试验中，相比几何融合方法碰撞率降低76%，成功解决了现有传感器融合方法在复杂场景下性能不佳的问题。

2024年，Dam等[55]提出AYDIV模型（见图6），由全局上下文融合对齐Transformer（Global Contextual Fusion Alignment Transformer，GCFAT）、稀疏融合特征注意力（Sparse Fused Feature Attention，SFFA）和体素网格注意力（Volumetric Grid Attention，VGA）组件构成，GCFAT将摄像头视为局部特征，将来自图像和激光雷达点云的深度信息视为全局查询；SFFA以协同特征对齐方式，将GCFAT提取的图像特征与体素化的激光雷达特征进行对齐；VGA在网格空间中结合了稀疏体素化的激光雷达特征和SFFA的图像特征，分别用于融合深度估计与RGB图像特征、体素化LiDAR和图像特征、伪点云和图像特征，该模型在Waymo数据集上验证实现了78.77%的航向平均精度均值（mean Average Precision with Heading，mAPH），在Argoverse2数据集上验证实现了37.70%的AP，显著提升了远距离检测性能。

3.3.2　基于空间一致性的融合方法

2022年，Drews等[56]提出DeepFusion多模态融合架构（见图7），该架构采用独立的FPN提取特征，分别对LiDAR、相机和雷达采集的数据进行针对性处理，将特征转换为通用BEV视图，在特征对齐和聚合后，检测头提供分类和回归输出。该模型解决了多模态3D目标检测问题，在nuScenes数据集的Car类别验证中实现了81.6%的AP。

2023年，Qin等[57]提出SupFusion训练策略，通过极坐标采样生成高质量特征，引入辅助特征级监督，并结合深度融合模块，从而显著提升融合效果。该方法在KITTI基准测试的3D mAP提升了2%。

2024年，Alfeqy等[58]提出CLF-BEVSORT模型，基于简单在线实时跟踪（Simple Online and Realtime Tracking，SORT）框架在BEV空间融合LiDAR和相机数据，引入新的关联策略，将结构相似性纳入成本函数。在KITTI数据集上的评估，CLF-BEVSORT在Car类别上取得了77.26%的高阶跟踪准确性（Higher Order Tracking Accuracy，HOTA）分数，超越了StrongFusionMOT、DeepFusionMOT等方法。同年，Vats等[59]提出VaLID方法，通过构建一个轻量级的3层全连接前馈神经网络，

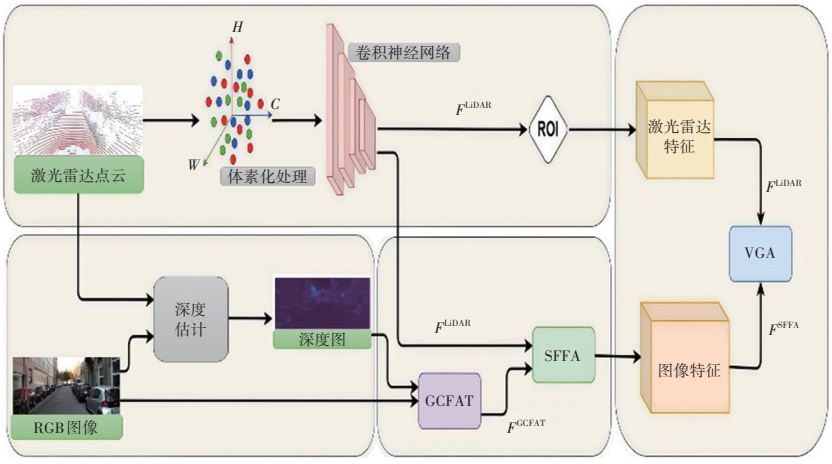


图6　AYDIV模型多模态融合架构

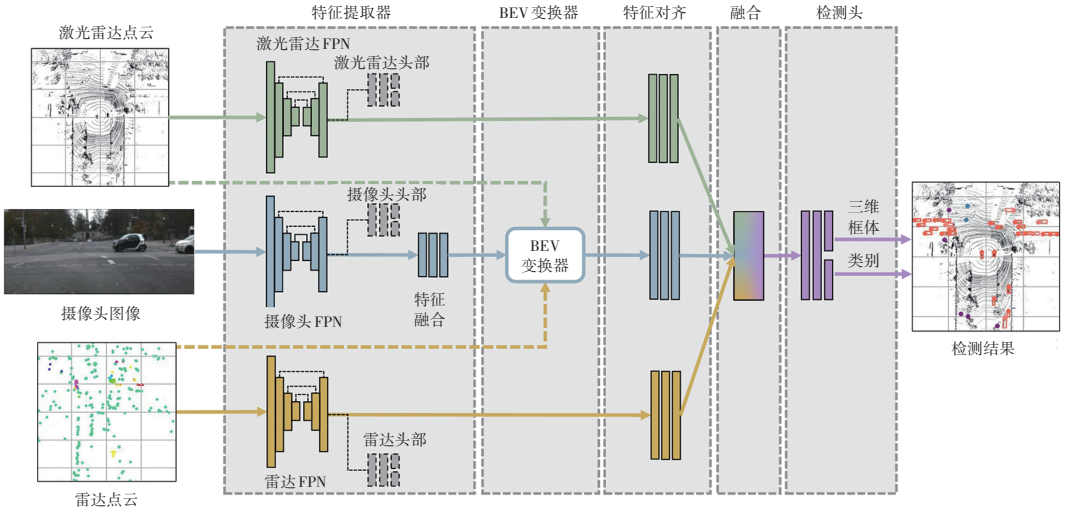


图7　DeepFusion融合架构

用于判别候选LiDAR目标的可信度。根据相机检测结果对LiDAR检测进行可靠性验证，从而降低传感器噪声带来的误检风险。在KITTI数据集上验证，VaLID的假阳性率平均降低了63.9%，提升了模型整体检测性能。

2025年，针对恶劣天气条件的3D目标检测研究，Zhang等[60]提出融合框架FogFusion。该方法通过结合深度补全网络与雾卷积（Depth Completion Network with Fog Convolution，DCFC）生成虚拟点云，以增强雾天条件下稀疏的LiDAR数据密度；采用灵活柱状体体素（Flexible Cylindrical Voxel，FCV）编码和柱状融合模块（Cylindrical Fusion Module，CFM）实现多模态特征的有效整合。在检测结果阶段，自适应权重调整融合相机和LiDAR的输出，有效应对雾天对图像质量的影响。在KITTI和KITTI-C数据集上的训练结果表明，FogFusion在雾天条件下的检测能比基线模型提升超过3.32%。

3.4　LiDAR和相机数据的多层次融合

多层次融合结合了早期、中期和后期融合的优点，通过在不同层级逐步融合LiDAR和相机数据，实现更全面、更精准的环境感知效果，该技术的高效性和鲁棒性为自动驾驶中的复杂场景提供了支持。

2020年，Philion等[61]提出主要用于BEV语义分割任务的LSS（Lift，Splat，Shoot）方法，通过深度估计网络提取多视角图像的隐含深度信息，将相机特征图转换至3D自车坐标系，再融合LiDAR点云特征，最终生成BEV空间的语义分割。

2022年，Liang等[62]提出BEVFusion（见图8），该方法通过解耦相机网络和LiDAR特征处理流程，将相机和LiDAR的原始输入分别编码为统一的BEV空间特征，通过动态融合模块对2种特征进行整合，最后输入检测头完成目标预测。在nuScenes数据集的训练结果表明，相较于主流3D检测与融合基线（包括PointPillars、CenterPoint和TransFusion等），BEVFusion在添加高斯噪声的鲁棒性测试中将mAP提升15.7%~28.9%，超越了当前最先进的TransFusion方法。

2024年，Dong等[63]提出SuperFusion网络（见图9），该网络在数据层、特征层和BEV层3个层面融合LiDAR

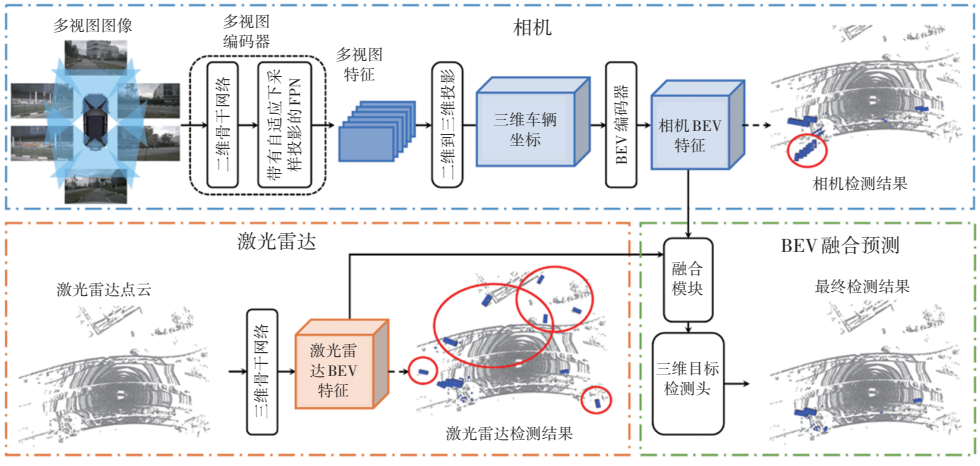


图8　BEVFusion的多层次融合模块架构

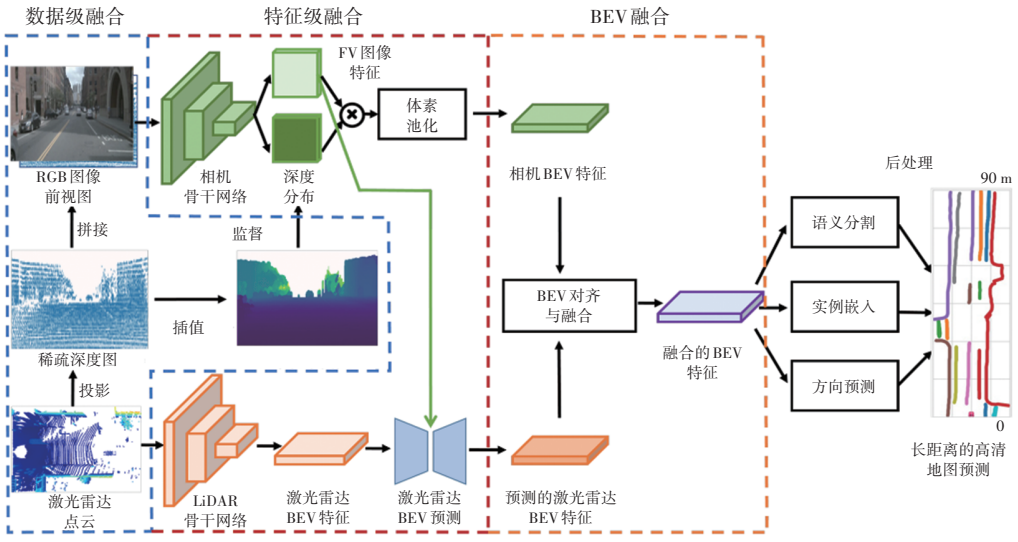


图9　SuperFusion的数据融合流程

与相机数据，利用LiDAR点云的深度信息提升图像深度估计的准确性，同时借助图像特征引导长距离LiDAR特征预测，解决远距离点云稀疏的问题。在nuScenes数据集上验证，SuperFusion在高清地图生成任务上优势显著，尤其在远距离道路元素的生成精度上表现优异，在60~90 m距离的类别验证中，Driver的交并比（Intersection over Union，IoU）可达29.2%。但由于SuperFusion使用了Transformer融合模块，模型训练相对较重，实时性不足，在计算资源和显存占用方面高于一些传统单模态3D检测方法。

同年，Song等[64]提出双向互补融合框架BiCo-Fusion，通过预融合阶段的体素增强模块（Voxel Enhancement Module，VEM）和图像增强模块（Image Enhancement Module，IEM）分别增强激光雷达和相机特征的语义和空间感知能力，在统一融合阶段自适应融合增强后的特征，构建统一表示。在nuScenes测试集上，模型的mAP和NDS分别达到74.5%、72.4%，超越了BEVFusion的表现。

4　各类融合方法的分析及技术挑战

通过分析早期、中期、后期以及多层次融合方法的主流方法模型与关键特性，发现早期融合注重在数据层面整合原始信息，中期融合聚焦于特征层面的深度融合，后期融合则在决策层面融合结果，而多层次融合更是结合了不同层级的优势。融合技术的应用，为自动驾驶环境感知系统构建了精度更高、鲁棒性更强、环境适应性更优的解决方案。

4.1　系统设计复杂度

早期融合在系统设计方面相对简单，主要集中于数据层面的整合，其核心任务是实现点云与图像的对齐，该方法对融合模块的设计要求较低，如坐标转换、几何对齐与数据拼接层，但对数据预处理和校准模块的精度要求较高，系统设计重点解决校准偏差和数据同步问题；中期融合的系统设计复杂度略有提升，需要分别设计LiDAR和相机的特征提取模块，并通过深度学习算法整合特征，且融合模块需平衡特征表示能力和计算效率；后期融合的系统设计更注重模块化，各模态的处理独立进行，融合模块仅需整合检测结果，设计复杂度较低。同时，为使模块化设计便于系统扩展，可通过改进融合算法提升模态间信息的整合深度；相较于常规融合技术，多层次融合的系统设计最复杂，且需在多个层级设计融合模块，协调数据层、特征层和决策层的处理流程，对系统集成能力要求较高，可通过开发模块化框架降低设计复杂性。

4.2　任务适配性

早期融合适合对空间定位精度要求较高的任务，其直接利用原始数据的能力能够提升系统的定位能力，但在需要语义理解的任务中表现有限；中期融合的任务适配性更为均衡，融合后的特征包含空间和语义信息，能够提升目标识别能力，适合需要同时完成定位和分类的任务，但其在某些边缘场景中表现不稳定；后期融合更适合对系统稳定性要求较高的任务，其独立处理的方式确保了单一模态失效时的基本功能，通过增强系统鲁棒性提升安全性，适合强调可靠性的任务，但对需要深度模态交互的任务支持不足；多层次融合的任务适配性最全面，能够根据任务需求动态调整融合策略，同时提升定位、识别和鲁棒性，适用于多任务并行的复杂场景，但其高复杂性可能导致在单一任务上的性能略低于单一

优化的方法。

4.3 部署成本与资源需求

部署成本是实际应用的重要考量因素。早期融合对计算资源的需求较高，需处理大规模原始数据，同时对高精度校准设备的依赖提高了硬件成本；中期融合通过特征提取能够减少数据量，计算资源需求降低，但深度学习模型的训练和推理仍需较高算力支持；后期融合的部署成本相对较低，各模态的独立处理允许使用较轻量化的模型，融合阶段的计算开销较小，适合资源受限的嵌入式系统；多层次融合的部署成本最高，其多阶段融合策略需要更强大的计算硬件支持，同时对模型训练和调优的要求也更高，适合计算资源充足的高端系统。

4.4 融合技术的挑战与应对策略

尽管当前多模态融合感知技术已在精度与鲁棒性方面取得显著进展，但在实际部署中仍面临严峻的实时性与成本挑战。通过对典型融合方法比较分析，结果如表2所示，为相关研究者提供价值参考。

表2 不同层次的LiDAR与相机数据融合方法比较分析

融合层次	典型方法	试验结果	优势	缺点
早期融合	PointFusion	KITTI数据集中等难度上验证，最终变体的mAP达到63%	性能优于单模态，小目标检测出色	计算复杂度高，对数据对齐要求严格
早期融合	PointPainting	KITTI数据集Pedestrian困难难度上验证，mAP达到54.58%	语义信息丰富，遮挡和远距离检测出色	实际部署中，图像分割网络在复杂环境，鲁棒性受限
中期融合	MV3D	KITTI数据集中等难度上验证，AP达到89.05%	3D定位和检测任务表现优异	对多视图数据依赖强，且图片的推理速度较慢
中期融合	TransFusion	nuScenes数据集上验证，mAP达到68.9%、NDS达到71.7%	全局上下文捕捉能力强	训练难度大，对超参数敏感
后期融合	TransFuser	在CARLA模拟器试验中，相比几何融合方法，碰撞率降低76%	端到端自动驾驶性能优异	对传感器校准要求高，对噪声敏感
后期融合	VaLID	KITTI数据集上验证，假阳性率平均降低63.9%	降低假阳性率	依赖于单个检测器的准确性，泛化能力有限
多层次融合	BEVFusion	nuScenes数据集上验证，相较于Transfusion的mAP提升15.7%	结构简单、鲁棒性强，适用于多种场景	计算资源需求高，对实时性有挑战
多层次融合	SuperFusion	nuScenes数据集上验证，60~90 m的IoU达到29.2%	高清地图生成效果显著	实现复杂度高，对数据质量要求严格

在自动驾驶场景中，感知系统面临严苛的实时性要求，系统需将感知推理时延控制在毫秒级，方可满足L3及以上自动驾驶功对低延迟的基本需求。然而，当前主流的多模态感知算法多采用复杂的特征提取与深度融合机制，尤其是基于Transformer架构或多阶段信息处理流程的模型，其端到端推理延迟可达到100~300 ms，远超可接受的实时运行上限。兼顾高精度与低延迟仍是现阶段融合感知模型工程部署面临的核心挑战。

除实时性瓶颈外，多模态感知系统在硬件配置上的经济可行性亦广受关注。高精度三维感知通常依赖于LiDAR获取空间结构信息，而中、高线束激光雷达在制造工艺、芯片封装与光电系统方面仍存在一定技术壁垒，导致其整体成本长期处于较高水平，制约了其在成本敏感型车型中的广泛部署，成为制衡多模态感知系统大规模落地的重要因素之一。

为了应对当前多模态融合感知系统在实时性与部署成本方面面临的挑战，本文提出4种关键解决思路，以在保障系统感知精度的前提下，提升计算效率、降低资源消耗：

a. 算法优化与模型压缩。通过模型结构重构、剪枝、参数量化、知识蒸馏等深度神经网络压缩技术，在不显著牺牲感知精度的前提下，降低模型复杂度与计算负载，有助于提升模型在边缘计算平台上的推理速度。

b. 软硬件协同设计机制。面向不同应用场景，基于异构计算平台的软硬件协同架构可实现感知模块的动态调度与算力自适应分配，以在性能与能耗之间取得平衡。同时，针对融合感知模型的运行特性进行硬件结构优化，如设计专用张量加速单元、构建低延迟神经网络执行框架等，提升整体算力密度与能耗比。

c. 构建轻量级多模态感知框架。在保持多源数据协同建模能力的同时，设计低计算负担、低通信成本的融合框架。为了减少冗余计算与特征对齐开销，可通过稀疏编码机制、注意力引导式融合策略以及跨模态特征压缩机制实现端到端的特征处理链条优化。

d. 建设系统级模块集成与标准化接口。通过建立标准化数据接口、统一时间同步机制与共享感知语义空

间，可显著降低系统通信与冗余计算开销，提升整体运行稳定性与可维护性。此外，标准化接口的建立亦是推动产业规模化部署的重要基础，可促进异构设备间的协同互操作，增强感知系统的跨平台迁移能力与工程落地可行性。

5 结束语

本文系统分析了LiDAR与相机在自动驾驶目标检测中的应用与发展，深入探讨了从单一传感器到多模态融合技术的演进过程及融合技术在复杂场景下的感知优势。研究表明：LiDAR与相机的融合技术通过整合传感器的互补优势，显著提升了目标检测的精度与鲁棒性，尤其在遮挡、远距离目标识别及恶劣环境下的表现更为突出，为自动驾驶系统的环境适应能力提供了重要支持。

未来研究需在实时性、精度与成本间寻求平衡，通过优化融合算法、探索低成本硬件解决方案及推动多传感器协同感知的标准化，进一步提升技术的普适性。通过系统剖析不同融合方法的潜力与局限性，为自动驾驶领域的研究提供了理论参考，推动了多模态感知技术的发展，实现提升交通安全与效率的目标。

参 考 文 献

[1] Waymo. The Waymo Driver Handbook: Teaching an Autonomous Vehicle How to Perceive and Understand the World Around It[EB/OL]. (2021-10-07) [2025-05-19]. https://waymo.com/blog/2021/10/the-waymo-driver-handbook-perception.

[2] Xiaopeng Motors. XPeng Announces Next-Gen Autonomous Driving Architecture at Auto Guangzhou 2020[EB/OL]. (2020-11-20) [2025-05-19]. https://ir.xiaopeng.com/news-releases/news-release-details/xpeng-announces-next-gen-autonomous-driving-architecture-auto.

[3] 萝卜快跑.《高级别自动驾驶应用白皮书》发布，萝卜快跑无人化商业应用成亮点[EB/OL]. (2023-09-26) [2025-05-19]. https://www.apollogo.com/ch/news/autonomous-driving/9918.
Apollo Go. The "White Paper on High-Level Autonomous Driving Applications" has been Released, Apollo Go's Unmanned Commercial Applications Become a Highlight [EB/OL]. (2023-09-26) [2025-05-19]. https://www.apollogo.com/ch/news/autonomous-driving/9918.

[4] 华为. 华为乾崑智驾ADS 4正式发布[EB/OL]. (2025-04-22) [2025-05-19]. https://auto. huawei. com/cn/news/2025/huawei-qiankun-intelligent-technology-launch/.
Huawei. Huawei Qiankun Intelligent Driving ADS 4 Officially Launched[EB/OL]. (2025-04-22) [2025-05-19]. https://auto. huawei. com/cn/news/2025/huawei-qiankun-intelligent-technology-launch/.

[5] HWANG S, PARK J, KIM N, et al. Multispectral Pedestrian Detection: Benchmark Dataset and Baseline[C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015: 1037-1045.

[6] KIM J, PARK B J, KIM J. Empirical Analysis of Autonomous Vehicle's Lidar Detection Performance Degradation for Actual Road Driving in Rain and Fog[J]. Sensors, 2023, 23(6).

[7] ZHAGN H X, LI S X, ZHU X Z, et al. 3D LiDAR and Monocular Camera Calibration: A Review[J]. IEEE Sensors Journal, 2025, 25(7): 10530-10555.

[8] KHAN D, BEAK M, KIM M Y, et al. Multimodal Object Detection and Ranging Based on Camera and Lidar Sensor Fusion for Autonomous Driving[C]// Proceedings of the 27th Asia-Pacific Conference on Communications (APCC). Jeju Island, Korea: IEEE, 2022: 342-343.

[9] ZHAO X M, SUN P P, XU Z G, et al. Fusion of 3D LIDAR and Camera Data for Object Detection in Autonomous Vehicle Applications[J]. IEEE Sensors Journal, 2020, 20(9): 4901-4913.

[10] GU J Y, BELLONE M, PIVOŇKA T, et al. CLFT: Camera-LiDAR Fusion Transformer for Semantic Segmentation in Autonomous Driving[J]. IEEE Transactions on Intelligent Vehicles, 2024, 9(5): 7115-7126.

[11] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-Based Learning Applied to Document Recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[12] VASWANI A, SHAZEER N M, PARMAR N, et al. Attention is All You Need[C]// Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS). Long Beach, CA, USA: NIPS, 2017: 5998-6008.

[13] ABDULMAKSOUD A, AHMED R. Transformer-Based Sensor Fusion for Autonomous Vehicles: A Comprehensive Review[J]. IEEE Access, 2025, 13: 41822-41838.

[14] SECCI F, CECCARELLI A. On Failures of RGB Cameras and Their Effects in Autonomous Driving Applications[C]// Proceedings of the 2020 IEEE 31st International Symposium on Software Reliability Engineering (ISSRE). Coimbra, Portugal: IEEE, 2020: 13-24.

[15] CECCARELLI A, SECCI F. RGB Cameras Failures and Their Effects in Autonomous Driving Applications[J]. IEEE Transactions on Dependable and Secure Computing, 2023, 20(4): 2731-2745.

[16] 宋晓琳, 邬紫阳, 张伟伟. 基于阴影和类Haar特征的动态车辆检测[J]. 电子测量与仪器学报, 2015, 29(9): 1340-1347.

SONG X L, WU Z Y, ZHANG W W. Dynamic Vehicle Detection Based on Shadow and Haar-Like Feature[J]. Journal of Electronic Measurement and Instrumentation, 2015, 29(9): 1340-1347.

[17] WEI Y, TIAN Q, GUO J H, et al. Multi-Vehicle Detection Algorithm through Combining Harr and HOG Features[J]. Mathematics and Computers in Simulation, 2019, 155: 130-145.

[18] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]// Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, OH, USA: IEEE, 2014: 580-587.

[19] GIRSHICK R. Fast R-CNN[C]// 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015: 1440-1448.

[20] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 779-788.

[21] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement[EB/OL]. (2018-04-08) [2025-05-19]. https://arxiv.org/abs/1804.02767.

[22] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector[C]// Proceedings of the European Conference on Computer Vision (ECCV). Amsterdam, The Netherlands: Springer, 2016: 21-37.

[23] 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.

[24] LIN T, GOYAL P, GIRSHICK R, et al. Focal loss for Dense Object Detection[C]// Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017.

[25] LAGA H, JOSPIN L V, BOUSSAID F, et al. A Survey on Deep Learning Techniques for Stereo-Based Depth Estimation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(4): 1738-1764.

[26] CHEN X Z, KUNDU K, ZHU Y K, et al. 3D Object Proposals Using Stereo Imagery for Accurate Object Class Detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(5): 1259-1272.

[27] PHAM C C, JEON J W. Robust Object Proposals Re-Ranking for Object Detection in Autonomous Driving Using Convolutional Neural Networks[J]. Signal Processing: Image Communication, 2017, 53: 110-122.

[28] ADAMO G, BUSACCA A. Time of Flight Measurements via Two LiDAR Systems with SiPM and APD[C]// Proceedings of the 2016 AEIT International Annual Conference (AEIT). Capri, Italy: IEEE, 2016: 1-5.

[29] LV L J, LI D, WU T F, et al. Signal Flux and Time-Of-Flight Estimation Based on Waveform Optimization for Single-Photon LiDAR[J]. Measurement, 2024, 224.

[30] LI Y, IBANEZ-GUZMAN J. LiDAR for Autonomous Driving: The Principles, Challenges, and Trends for Automotive LiDAR and Perception Systems[J]. IEEE Signal Processing Magazine, 2020, 37(4): 50-61.

[31] KHAN M A, MENOUAR H, ABDALLAH M, et al. LiDAR in Connected and Autonomous Vehicles-Perception, Threat Model, and Defense[J]. IEEE Transactions on Intelligent Vehicles, 2024, 9(1): 2389-2407.

[32] DOUILLARD B, UNDERWOOD J, KUNTZ N, et al. On the Segmentation of 3D Lidar Point Clouds[C]// Proceedings of the 2011 IEEE International Conference on Robotics and Automation (ICRA). Shanghai, China: IEEE, 2011: 2798-2805.

[33] ARNOLD E, AL-JARRAH O Y, DIANATI M, et al. A Survey on 3D Object Detection Methods for Autonomous Driving Applications[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(10): 3782-3795.

[34] CHARLES R Q, SU H, KAICHUN M, et al. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 77-85.

[35] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning Representations by Back-Propagating Errors[J]. Nature, 1986, 323(6088): 533-536.

[36] QI C, YI L, SU H, et al. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space[C]// Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS). Long Beach, CA, USA: NIPS, 2017: 5099-5108.

[37] ZHOU Y, TUZEL O. VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, UT, USA: IEEE, 2018: 4490-4499.

[38] LANG A H, VORA S, CAESAR H, et al. PointPillars: Fast Encoders for Object Detection From Point Clouds[C]// Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019: 12689-12697.

[39] XU Z Y, LI Y H, ZHU S Q, et al. Expanding Sparse LiDAR Depth and Guiding Stereo Matching for Robust Dense Depth Estimation[J]. IEEE Robotics and Automation

Letters, 2023, 8(3): 1479-1486.

[40] PATEL M, JUNG R. Simulation-Based Performance Evaluation of 3D Object Detection Methods with Deep Learning for a LiDAR Point Cloud Dataset in a SOTIF-Related Use Case[EB/OL]. (2025-03-05) [2025-05-19]. https://arxiv.org/abs/2503.03548.

[41] XU D F, ANGUELOV D, JAIN A. PointFusion: Deep Sensor Fusion for 3D Bounding Box Estimation[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, UT, USA: IEEE, 2018: 244-253.

[42] SINDAGI V A, ZHOU Y, TUZEL O. MVX-Net: Multimodal VoxelNet for 3D Object Detection[C]// Proceedings of the 2019 International Conference on Robotics and Automation (ICRA). Montreal, QC, Canada: IEEE, 2019: 7276-7282.

[43] ZHANG Z H, ZHANG M, LIANG Z D, et al. MAFF-Net: Filter False Positive for 3D Vehicle Detection with Multi-Modal Adaptive Feature Fusion[EB/OL]. (2020-09-23) [2025-05-19]. https://arxiv.org/abs/2009.10945.

[44] VORA S, LANG A H, HELOU B, et al. PointPainting: Sequential Fusion for 3D Object Detection[C]// Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 4604-4612.

[45] ZHAO L, ZHOU H, ZHU X G, et al. LIF-Seg: LiDAR and Camera Image Fusion for 3D LiDAR Semantic Segmentation[J]. IEEE Transactions on Multimedia, 2024, 26: 1158-1168.

[46] CHEN X Z, MA H M, WAN J, et al. Multi-View 3D Object Detection Network for Autonomous Driving[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 6526-6534.

[47] KU J, MOZIFIAN M, LEE J, et al. Joint 3D Proposal Generation and Object Detection from View Aggregation[C]// Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid, Spain: IEEE, 2018: 1-8.

[48] QI C R, LIU W, WU C X, et al. Frustum PointNets for 3D Object Detection from RGB-D Data[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, UT, USA: IEEE, 2018: 918-927.

[49] YU Z, LI J, WEI Y Z, et al. Combining Camera-LiDAR Fusion and Motion Planning Using Bird's-Eye View Representation for End-to-End Autonomous Driving[J]. Drones, 2025, 9(4).

[50] BAI X Y, HU Z Y, ZHU X G, et al. TransFusion: Robust LiDAR-Camera Fusion for 3D Object Detection with Transformers[C]// Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 1080-1089.

[51] LI Y W, YU A W, MENG T J, et al. DeepFusion: Lidar-Camera Deep Fusion for Multi-Modal 3D Object Detection[C]// Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 17161-17170.

[52] ZHAO Y, GONG Z, ZHENG P R, et al. SimpleBEV: Improved LiDAR-Camera Fusion Architecture for 3D Object Detection[EB/OL]. (2024-11-08) [2025-05-19]. https://arxiv.org/abs/2411.05292.

[53] LI Z Q, WANG W H, LI H Y, et al. BEVFormer: Learning Bird's-Eye-View Representation from LiDAR-Camera via Spatiotemporal Transformers[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(3): 2020-2036.

[54] PRAKASH A, CHITTA K, GEIGER A. Multi-Modal Fusion Transformer for End-to-End Autonomous Driving[C]// Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE, 2021: 7073-7083.

[55] DAM T, DHARAVATH S B, ALAM S, et al. AYDIV: Adaptable Yielding 3D Object Detection via Integrated Contextual Vision Transformer[C]// Proceedings of the 2024 IEEE International Conference on Robotics and Automation (ICRA). Yokohama, Japan: IEEE, 2024: 10657-10664.

[56] DREWS F, FENG D, FAION F, et al. DeepFusion: A Robust and Modular 3D Object Detector for Lidars, Cameras and Radars[C]// Proceedings of the 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Kyoto, Japan: IEEE, 2022: 560-567.

[57] QIN Y R, WANG C Q, KANG Z J, et al. SupFusion: Supervised LiDAR-Camera Fusion for 3D Object Detection[C]// Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE, 2023: 21957-21967.

[58] ALFEQY L, ABDELMUNIM H E H, MAGED S A, et al. Kalman Filter-Based Fusion of LiDAR and Camera Data in Bird's Eye View for Multi-Object Tracking in Autonomous Vehicles[J]. Sensors, 2024, 24(23).

[59] VATS V, NIZAM M B, DAVIS J. VaLID: Verification as Late Integration of Detections for LiDAR-Camera Fusion [EB/OL]. (2024-09-23) [2025-05-19]. https://arxiv. org/abs/2409.15529.

[60] ZHANG B, WANG Y, ZHANG C, et al. FogFusion: Robust

3D Object Detection Based on camera-LiDAR Fusion for Autonomous Driving in Foggy Weather Conditions[J]. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 2025.

[61] PHILION J, FIDLER S. Lift, Splat, Shoot: Encoding Images from Arbitrary Camera Rigs by Implicitly Unprojecting to 3D[C]// Proceedings of the European Conference on Computer Vision (ECCV). Glasgow, UK: Springer, 2020: 194-210.

[62] LIANG T T, XIE H W, YU K C, et al. BEVFusion: A Simple and Robust LiDAR-Camera Fusion Framework[C]// Proceedings of the 36th Conference on Neural Information Processing Systems (NIPS). New Orleans, LA, USA: NIPS, 2022.

[63] DONG H, GU W H, ZHANG X J, et al. SuperFusion: Multilevel LiDAR-Camera Fusion for Long-Range HD Map Generation[C]// Proceedings of the 2024 IEEE International Conference on Robotics and Automation (ICRA). Yokohama, Japan: IEEE, 2024: 9056-9062.

[64] SONG Y, WANG L. BiCo-Fusion: Bidirectional Complementary LiDAR-Camera Fusion for Semantic-and Spatial-Aware 3D Object Detection[J]. IEEE Robotics and Automation Letters, 2025, 10(2): 1457-1464.

（责任编辑　瑞　秋）

修改稿收到日期为2025年5月19日。

《汽车技术》征稿启事

《汽车技术》杂志是中国第一汽车集团有限公司主办的国内外公开发行的汽车前瞻与应用技术类月刊，为我国高质量科技期刊分级目录入选期刊、中国科学引文数据库（CSCD）来源期刊、中文核心期刊、中国科技核心期刊、RCCSE中国核心学术期刊（A）、Scopus数据库收录期刊、俄罗斯《文摘杂志》（AJ）收录期刊、日本科学技术振兴机构数据库入选期刊、EBSCO学术数据库收录期刊、欧洲学术出版中心（EuroPub）数据库收录期刊。

《汽车技术》杂志以报道汽车整车及其零部件设计、研究、试验等方面的前瞻与应用技术为主，并兼有理论研究内容，是中国汽车行业核心学术和知识传播与共享的平台。

《汽车技术》将在国家提出的“创新、协调、绿色、开放、共享”发展理念的指引下，把握《节能与新能源汽车技术路线图》和“低碳化、信息化、智能化”的汽车技术主流发展趋势，努力在传统内燃机汽车高效动力系统、轻量化、低阻力领域，新能源汽车和互联智能汽车技术领域，大力吸收优质稿源，为广大科研和工程技术人员服务，为我国汽车工程技术创新能力提升贡献力量。

《汽车技术》欢迎高等院校师生、研发工程技术人员、技术管理人员及相关人员不吝赐稿，反映国家重点扶持项目、自然科学基金项目和其他重点项目等研究成果的稿件将被优先选择刊登。

投稿要求：

1. 文章字数最好控制在6 000~8 000字范围之内；

2. 请按科技论文要求撰写文章摘要，摘要中文字数控制在180字左右；

3. 文章必须附有公开发表的、体现本领域最新研究成果的参考文献，且在文中应标注文献引用处；

4. 文章主要作者应提供其简介，包括出生年、性别、职称、学历、研究方向及技术成果等；

5. 来稿的保密审查工作由作者单位负责，确保署名无争议，文责自负；

6. 请勿一稿多投；

7. 本刊使用网站投稿，请先登陆网站注册成功后投稿，详细投稿要求见本刊网站中“下载中心”栏的“作者指南”，网址：http://qcjs.cbpt.cnki.net。

《汽车技术》编辑部