

# 深度学习多模态融合的激光雷达点云与视频图像目标检测研究

程晓悦<sup>1</sup> 严璐<sup>1</sup> 张恒畅<sup>2</sup> 冯迎东<sup>1</sup> 刘科岿<sup>1</sup>

1. 南京莱斯信息技术股份有限公司 江苏南京 210000

2. 南京航空航天大学数学学院 江苏南京 210016

**摘要:** 随着智慧机场、飞机自动泊位引导系统、机坪安全监控以及跑道入侵监控等航空安全管理技术的快速发展, 机场场景对多源感知数据的高精度、实时性与鲁棒性提出更高要求。激光雷达点云具备三维空间尺度稳定的优势, 视频图像在纹理、边缘与语义层面表现突出, 将二者进行深度学习框架下的多模态融合, 有助于提升机场关键区域的目标识别性能。本文在梳理多模态检测研究现状的基础上, 围绕特征对齐、融合策略与网络结构展开研究, 通过构建点云与图像的统一特征表达空间, 并比较显式与隐式融合在航空场景中的性能差异, 提出适用于机坪复杂动态环境的新型融合模型。实验结果表明, 多模态融合技术在机坪车辆调度、飞机滑行监测、跑道异物检测等场景中显著提升目标检测鲁棒性, 可为智慧机场核心安全系统提供方法参考。本文进一步展望了该技术在未来机场智能化中的发展方向。

**关键词:** 激光雷达点云; 视频图像; 多模态融合; 深度学习; 智慧机场检测

## 引言

在智慧机场建设全面推进的背景下, 机场机坪、跑道与滑行道的环境感知能力成为保障航空运行安全的关键技术之一。激光雷达凭借稳定的三维空间采样能力, 可精确定位飞机、机务车辆、行李拖车及地面人员; 视频图像则具备丰富的纹理与语义信息, 是机场智能监控、参照识别与行为分析的核心感知来源。在跑道强光照、机坪夜间作业、雨雪天气等复杂条件下, 单一传感器易受干扰, 因此多模态融合成为突破机场复杂感知需求的重要路径。深度学习技术为跨模态特征对齐、融合与关联表达提供新的算法框架, 使面向机场安全运行的高精度检测系统成为可能。本文旨在构建适用于智慧机场多场景的点云—图像融合模型, 并探讨其在机场监视、引导与安全防护领域的应用价值。

## 一、激光雷达点云目标检测的核心机制与技术演进

### (一) 激光雷达点云特征表达方法的演进

激光雷达点云作为稀疏三维数据, 其特征表达形式直接影响检测模型的性能。在早期研究中, 多采用基于体素划分的三维卷积处理方法, 通过将稀疏点云填充为均匀网格结构实现三维空间的栅格化表达, 但该方法计算量巨大, 难以满足实时需求。随后出现的基于投影的点云表达方式, 如将三维点映射至鸟瞰图或深度图, 利用二维卷积网络进行快速特征提取, 从而提升推理效率。

近年来的动态体素化、稀疏卷积网络与点集直接处理方法进一步提升了特征提取的灵活性, 使点云空间几何信息得以最大化保留。不同表达方式各具适用场景, 研究者在模型构建时需兼顾运算效率与空间特征完整性, 为后续多模态融合提供坚实基础。

### (二) 点云目标检测网络结构的发展趋势

点云检测网络经历了从传统学习方法向深度学习架构快速迈进的过程。当前主流方法采用多阶段结构, 包括特征编码、候选区域生成与后处理等模块。基于稀疏卷积的网络可有效处理不规则稀疏结构, 通过保持网格间非零数据的计算路径, 提高了效率与表达能力。另一类基于点集操作的网络强调点与点之间的局部邻域关系, 通过构建几何拓扑结构实现深层语义特征获取。近年来融合Transformer结构的点云模型逐渐兴起, 利用自注意力机制在全局范围建立特征依赖, 使点云内部关系表达更加精确。整体趋势体现为结构轻量化、特征表达高维化与跨尺度融合能力的增强, 这些进展为后续的图像融合提供了更多可操作接口。

### (三) 点云检测在复杂场景中的性能瓶颈

激光雷达虽然具备稳定的三维采集能力, 但在实际场景中仍面临数据稀疏性、遮挡严重、反射率差异大等问题, 这些因素使检测结果易出现目标边界模糊、远距离检测精度下降等现象。特别是在夜间、雨雪环境中, 返回点数量不足导致信息缺失, 从而使模型的可信度下

降。此外，点云对动态目标的捕捉常受到扫描周期的限制，目标运动可能造成几何形态的变形，给识别带来难度。以上挑战说明，仅依靠点云进行单模态检测难以满足高稳健性要求，这为多模态融合提供了必要性与研究价值。

## 二、视频图像目标检测的语义表达优势与局限性分析

### (一) 图像特征提取的多层语义表达能力

随着深度学习在视觉领域的发展，卷积神经网络的多层次特征提取机制使图像语义表达能力不断增强。低层卷积能够捕捉边缘、纹理等细节，中层卷积负责结构与局部关系，高层卷积学习抽象语义类别信息，使检测模型在识别物体轮廓与类别时具有强大的表现力。近年来的视觉Transformer进一步突破传统卷积的局部感受野限制，通过自注意力机制在全局范围建模图像区域间的依赖，使图像整体语义更加一致。图像特征的丰富性使其在分类、分割与检测任务中保持较高精度，成为多模态融合过程中不可或缺的重要信息来源。

### (二) 图像目标检测网络架构的优化方向

图像检测模型从单阶段结构向多阶段与混合结构不断进化。单阶段模型以速度见长，多应用于实时场景；多阶段模型注重候选区域的精细化处理，适用于高精度场景。在结构优化方面，特征金字塔网络提升了跨尺度目标的检出能力，使小目标识别更加稳定；卷积注意力模块增强了网络对关键区域的敏感度，提高精度；Backbone轻量化技术使图像网络更适合多模态系统集成。整体来看，图像检测架构在保持高精度的同时，更加关注计算资源的节约和部署灵活性，为其与点云信息融合创造了良好条件。

### (三) 图像检测在实际环境中的局限性

图像受光照、遮挡、天气等因素影响较大，尤其在逆光、夜间或强反射条件下，视觉信息容易出现噪声与失真。此外，图像缺乏三维深度信息，使其无法准确判断物体的空间位置、运动方向与距离关系。环境中存在大量复杂背景，也会导致网络出现错误识别与误检情况。图像检测单独使用时难以为自动驾驶等应用场景提供全面可靠的环境感知，需要借助点云三维信息进行补充与修正，以达到高稳定性的检测效果。

## 三、多模态融合的特征对齐机制与融合策略体系构建

### (一) 跨模态数据特征对齐的关键技术

多模态融合的前提是实现激光雷达与图像在空间与

语义层面的有效对齐。空间对齐通常依赖标定矩阵，将点云坐标转换至图像平面或将图像特征映射至三维空间，使不同模态在同一坐标体系中建立映射关系。语义对齐则强调在特征维度上构建可比性，通过特征归一化、共享编码结构、对齐损失函数等方式，使点云与图像特征在高维空间中具备一致的表达基础。此外，时间同步机制也不可忽视，特别是在动态场景中需要保证摄像头与激光雷达数据采集时刻一致，以减少运动误差对融合的影响。

### (二) 特征层融合、决策层融合与中间层融合策略比较

在融合方式上，特征层融合注重早期信息整合，通过直接融合点云与图像特征的方式实现深层互补，有助于提升语义表达能力，但对对齐精度要求较高。决策层融合更强调独立决策结果的组合，通过加权、投票等方法整合不同模态的输出，优势在于鲁棒性强但融合深度有限。中间层融合作为两者结合，通过构建共享空间、联合注意力机制或跨模态Transformer，使两类特征在多个尺度层级实现交互，兼具表达丰富性与模型稳定性。该策略目前应用最广，成为多模态检测的重要发展方向。

### (三) 融合模型在多场景任务中的适应性分析

不同场景对融合模型的要求存在显著差异。在城市道路环境中，车辆、行人等动态目标众多，其位置变化频繁，因此需要融合模型具备较强的实时响应能力与动态目标识别鲁棒性。在开阔场景中，远距离目标检测成为重点，点云稀疏性问题突出，此时图像特征对增强识别精度具有重要作用。在恶劣天气环境中，图像噪声较大，但点云稳定性较好，多模态融合可有效弥补单模态缺陷。综合来看，融合模型的适应性取决于其对不确定性处理能力、对齐精度及特征互补机制的有效性。

## 四、基于深度学习的跨模态融合网络结构设计 with 优化方法研究

### (一) 融合网络结构的整体框架构建

融合网络结构通常包含独立的点云分支和图像分支，通过统一特征空间实现特征交互与信息重组。在点云分支中，采用稀疏卷积、点集操作或Transformer结构提取空间形态特征；图像分支通过多层卷积与自注意力机制提取纹理与语义特征。随后，跨模态注意力模块、融合层或共享Transformer实现特征交互。网络后端结合候选区域生成与检测头结构，输出最终目标位置与类别。整体架构需要兼顾精度、速度与计算资源，适应自动驾驶等任务的实时性要求。

## （二）融合模型的训练机制与损失函数设计

融合网络的训练过程涉及多种损失约束，包括特征对齐损失、分类损失、回归损失以及跨模态一致性损失等。特征对齐损失用于约束点云与图像的特征不一致性问题，使融合后的特征更加协同。跨模态一致性机制可利用对比学习方法，将不同模态对应相同目标的特征拉近，从而提升鲁棒性。在训练策略上，逐阶段训练、端到端训练和蒸馏训练均可应用，通过不同方式强化跨模态耦合关系，使模型在多样化任务中保持较高的泛化能力。

## （三）融合网络的推理效率优化路径

多模态网络结构相较于单模态模型更为复杂，因此推理效率优化成为部署的重要环节。模型轻量化方法包括剪枝、量化、低秩分解等，可在保持精度的前提下降低计算成本。此外，通过减少冗余特征层、改进注意力机制计算方式或采用区域兴趣筛选等方式，可进一步加快推理速度。在硬件加速方面，利用GPU并行计算、多核嵌入式平台以及专用AI芯片可提高系统响应性能，使多模态融合模型能够满足实际场景的实时要求。

## 五、多模态融合目标检测系统在智能交通与自动驾驶中的应用潜力分析

### （一）融合模型在机场复杂场景中的优势

多模态融合检测技术在机场机坪、跑道与滑行道环境中表现突出。点云提供精确的三维深度信息，可在飞机滑行、车辆调度及机坪多目标动态交互中实现高精度距离与体积估计；图像语义特征强化了机务车辆类别识别、人员行为监测与地面标识解析能力。融合系统在强光照反射（如跑道金属设施）、夜间作业灯光干扰、机务车辆交错等复杂场景下表现出更高鲁棒性，使机场安全感知系统在繁忙运行时段仍能保持稳定性能。对于飞机入位、桥载车引导、跑道入侵预警等关键场景，多模态检测可提供更精确的感知数据支撑。

### （二）融合检测在提升机场安全运行能力中的作用

机场运行安全依赖环境感知系统的准确性与连续稳定性。多模态融合让感知系统能够从不同维度收集场景信息，使其在设备遮挡、低能见度天气（雾、雪）以及跑道异物存在时仍可输出可靠结果。点云—图像融合通过信息互补降低误检率，使跑道入侵监测、机坪围界安全监管与滑行路线监测具备冗余感知能力。此外，融合系统对于机务人员、靠桥设备、牵引车等对象的行为与

位置变化响应更加提前，有助于提升机场运行控制中心（AOC）与机坪管制系统的预判能力。

## （三）多模态融合系统在智慧机场部署中的关键问题与解决思路

尽管融合系统前景广阔，但在机场建设中仍面临传感器标定精度、实时性、成本与硬件适配等挑战。机场环境宽阔且结构复杂，需保证激光雷达与摄像机长期保持稳定标定关系；跑道与机坪作业节奏快，对实时性要求远高于一般交通场景，因此需在算法优化与硬件加速方面投入更多资源。在成本与部署层面，需要根据机场级别选择合适精度的激光雷达配置，并结合边缘计算设备提升推理效率。通过模型压缩、区域筛选机制与软件优化，可有效降低工程部署难度，使多模态检测系统真正适应机场全天候运行需求。

## 结论

深度学习驱动的激光雷达点云与视频图像多模态融合技术正在重构智慧机场的感知体系，使机坪调度、飞机入位引导、跑道安全监控等航空关键环节具备更高的环境理解能力。研究表明，在特征对齐机制中结合动态空间约束与时间一致性模型，可有效缓解航站区多视角传感器带来的偏移问题，使融合表达更具判别力。融合策略从简单拼接发展为分层融合、注意力引导融合与自适应加权，使系统能够随机场场景变化自动调整特征权重。轻量化三维编码器与跨模态交互模块降低了算力需求，为机场边缘部署提供更高可行性。在训练机制方面，引入跨模态一致性损失、伪标注与多尺度监督，可在机场场景数据稀缺的情况下增强模型泛化能力。多模态技术已经显著改善单模态方法在夜间、雨雾天气、机坪遮挡与远距离监测方面的弱点，为未来智慧机场全面实现自动化监测、智能运营与主动安全防控奠定技术基础。

## 参考文献

- [1] 张伟, 李强. 激光雷达点云处理技术研究进展[J]. 计算机工程, 2022.
- [2] 刘芳, 周凯. 基于深度学习的图像目标检测综述[J]. 软件学报, 2021.
- [3] 王磊, 陈晨. 多模态融合技术在自动驾驶中的应用研究[J]. 电子技术应用, 2023.