

分类号: TP391

密 级: 公开

学校代码: 10363

学 号: 2210110114



安徽工程大学

Anhui Polytechnic University

硕士学位论文

题 目

基于毫米波雷达和视觉的

三维目标检测算法研究

论 文 作 者

江彤

指 导 教 师

时培成

学 科 (专 业)

机械工程

研 究 方 向

无人驾驶环境感知

论文提交日期: 2024 年 5 月 30 日

分类号：TP391

学校代码：10363

密 级：公开

学 号：2210110114

基于毫米波雷达和视觉的三维目标检测算法研究

RESEARCH ON 3D OBJECT DETECTION ALGORITHM
BASED ON MILLIMETER-WAVE RADAR AND VISION

学生姓名：江彤

校内导师：时培成（教授）

专 业：机械工程

研究方向：无人驾驶环境感知

论文答辩日期：2024 年 5 月 26 日

安徽工程大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风。所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全意识到本声明的法律后果由本人承担。

学位论文作者签名：江彤

日期：2024年5月30日

安徽工程大学学位论文版权使用授权书

学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅或借阅。本人授权安徽工程大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 ☐，在 年解密后适用本版权书。

本学位论文属于

不保密 ☐。

学位论文作者签名：江彤

日期：2024年5月30日

指导教师签名：时培成

日期：2024年5月30日

基于毫米波雷达和视觉的三维目标检测算法研究

摘 要

随着自动驾驶技术的迅速发展, 传感范围小、场景信息少且处理方式单一化的单传感器检测无法全方位地适应复杂的真实环境。视觉传感器在二维目标识别上虽然具有优势, 但单独依靠图像难以获得目标的空间位置和速度信息; 毫米波雷达传感器通过测量场景中移动目标的距离、角度和径向速度, 获取目标的深度信息和速度信息, 但无法识别目标的几何形状和语义信息。多项研究证明, 将视觉传感器和毫米波雷达结合比单独的视觉检测具有更高的性能, 因此, 本文基于毫米波雷达和视觉传感器进行三维目标检测算法研究, 从毫米波雷达点云数据处理、特征提取以及点云和图像特征融合三个方面对现有的研究进行改进和优化, 提高了目标检测算法的鲁棒性和结果的准确性。具体工作内容和贡献概述如下:

(1) 整理了毫米波雷达基本工作原理和数据关联方法。本文围绕毫米波雷达和相机融合的三维目标检测理论开展研究, 分析了毫米波雷达基本测速、测距以及测量角度的工作原理, 列举出具备毫米波雷达点云的数据集, 并探讨了将毫米波雷达点云和相机图像进行融合的数据关联技术, 为后续视觉图像和毫米波雷达点云特征融合算法提供理论支撑。

(2) 研究基于毫米波雷达点云的目标检测算法。针对稀疏的毫米波雷达点云数据目标难以检测的问题, 本文提出一种基于 RPCNET 算法的毫米波雷达点云目标检测网络。该网络以原始毫米波雷达点云数据作为输入, 在雷达区域生成网络(RRPN)中检测雷达点云并生成候选目标区域; 然后采用 DGCNN 网络进行点云特征提取并确定锚框位置和尺寸, 最后应用非极大值抑制采样, 以确保最终结果仅包含最具代表性的边界框。本研究在具有挑战性的 NuScenes 和 RadarScenes 数据集上对提出的方法进行了评估, 实验结果表明, 在夜间光线较差的情况下, 该方法仍然在各类别的检测精度上均有较大幅度的提升。

(3) 研究基于毫米波雷达和相机融合的三维目标检测算法。为实现在自动驾驶场景中精准地检测出三维目标, 本文提出一种基于毫米波雷达和相机特征融

合的三维目标精准检测方法 CenRadfusion，该算法针对 CenterFusion 网络进行了改进和进化：为确保融合体系结构的完整性，将毫米波雷达点云投影到图像平面上，并将其作为附加通道添加到 CenterNet 图像检测网络的输入中，初步形成三维检测框；为过滤无关点云滤除白噪声，提高数据质量和目标检测的可靠性，采用毫米波雷达点云密度聚类技术，对毫米波雷达点云进行聚类，获得标签；为强调网络的重要特征，添加了通道注意力模块，自适应地对每个特征通道进行加权，有效融合毫米波雷达点云信息和图像特征。本实验在具有挑战性的 NuScenes 数据集上的实验结果表明，与原始 CenterFusion 算法相比较，汽车，卡车和摩托车的检测 AP 值均有了提升。

关键词：自动驾驶，毫米波雷达点云，3D 目标检测，计算机视觉，特征融合，多模态信息融合

RESEARCH ON 3D OBJECT DETECTION ALGORITHM BASED ON MILLIMETER-WAVE RADAR AND VISION

ABSTRACT

With the rapid development of autonomous driving technology, a single sensor with a small sensing range, less scene information, and a single processing method can no longer adapt to the complex real environment. Although visual sensors have advantages in 2D target recognition, it is difficult to obtain the spatial position and velocity information of the target by relying on images alone, and millimeter-wave radar sensors can obtain the depth and velocity information of the target by measuring the distance, angle and radial velocity of the moving target in the scene, but cannot recognize the geometry and semantics of the target. A number of studies have proved that the combination of vision sensor and millimeter-wave radar improves performance compared with visual detection alone, so this paper uses millimeter-wave radar and vision sensor for 3D target detection. The existing research is improved and optimized from three aspects: millimeter-wave radar point cloud data processing, feature extraction, and point cloud and image feature fusion, which improves the robustness of the target detection algorithm and the accuracy of the results. The specific work content and contributions are summarized below:

(1) The basic working principle and data association methods of millimeter-wave radar are summarized. This paper focuses on the 3D target

detection theory of millimeter-wave radar and camera fusion, analyzes the working principles of basic velocity measurement, ranging and measurement angle of millimeter-wave radar, lists the datasets with millimeter-wave radar point clouds, and discusses the data association technology of fusing millimeter-wave radar point clouds and camera images, so as to provide theoretical support for the subsequent visual images and millimeter-wave radar point cloud feature fusion algorithms.

(2) Research on the object detection algorithm based on millimeter-wave radar point cloud. In order to solve the problem that it is difficult to detect targets in sparse millimeter-wave radar point cloud data, this paper proposes a millimeter-wave radar point cloud target detection network based on RPCNET algorithm. The network uses the original millimeter-wave radar point cloud data as input, detects the radar point cloud in the Radar Region Generation Network (RRPN) and generates candidate target areas, then uses the DGCNN network framework to extract the point cloud features and determine the anchor frame position and size, and finally applies non-maximum suppression sampling to ensure that the final result contains only the most representative bounding box. In this study, the proposed method was evaluated on the challenging NuScenes and RadarScenes datasets, and the experimental results showed that the proposed method still greatly improved the detection accuracy of all types under the condition of poor light at night.

(3) 3D object detection algorithm based on millimeter-wave radar and camera fusion is studied. In order to achieve accurate detection of 3D targets in autonomous driving scenarios, this paper proposes a 3D target accurate detection method based on millimeter-wave radar and camera feature fusion, CenRadfusion, which is improved and evolved for the

CenterFusion network: in order to ensure the integrity of the fusion architecture, the millimeter-wave radar point cloud is projected onto the image plane and added to the input of the CenterNet image detection network as an additional channel to preliminary form a 3D detection frame, and the white noise is filtered out to filter out irrelevant point clouds, so as to improve the data quality and the reliability of object detection. In order to emphasize the important features of the network, a channel attention module is added to adaptively weight each feature channel, which effectively fuses the millimeter-wave radar point cloud information and image features. Experimental results on the challenging NuScenes dataset show that compared to the original CenterFusion algorithm, the AP values of cars, trucks, and motorcycles are improved.

KEY WORDS: autonomous driving, millimeter-wave radar point cloud, 3D object detection, computer vision, feature fusion, multi-modal information fusion

目录

摘 要.....	I
ABSTRACT.....	III
目录.....	VI
第 1 章 绪论.....	1
1.1 研究背景及意义.....	1
1.2 国内外研究现状.....	3
1.2.1 基于毫米波雷达的三维目标检测.....	4
1.2.2 基于相机-毫米波雷达融合的三维目标检测	9
1.3 本文研究内容及章节安排.....	13
第 2 章 毫米波雷达检测.....	15
2.1 引言.....	15
2.2 FMCW 雷达信号和系统模型	16
2.3 毫米波雷达测距、测速以及角度测量原理.....	18
2.3.1 毫米波雷达距离测量.....	18
2.3.2 毫米波雷达速度测量.....	20
2.3.3 毫米波雷达角度估算.....	23
2.4 毫米波雷达数据集.....	25
2.5 数据关联.....	26
2.5.1 点云坐标转换.....	26
2.5.2 点云的二维投影.....	26
2.6 本章小结.....	27
第 3 章 基于毫米波雷达点云目标检测方法 RPCNET	28
3.1 引言.....	28
3.2 RPCNET 网络总体架构	29
3.2.1 候选目标区域生成.....	29
3.2.2 点云特征提取池化.....	30
3.2.3 检测框置信度预测.....	32
3.2.4 确定检测框.....	33
3.2.5 损失函数.....	33

3.3 实验环境及评价指标.....	34
3.3.1 实验环境.....	34
3.3.2 NuScenes 数据集评价指标.....	34
3.3.3 RadarScenes 数据集评价指标	35
3.4 实验结果与分析.....	36
3.4.1 在 NuScenes 数据集上的实验结果与分析	36
3.4.2 在 RadarScenes 数据集上的实验结果与分析	37
3.5 本章小结.....	40
第 4 章 视觉和毫米波雷达融合的 3D 目标检测算法 CenRadfusion	41
4.1 CenterFusion 目标检测网络概述	41
4.1.1 CenterNet 网络	42
4.1.2 基于视锥的关联.....	44
4.2 CenRadfusion.....	45
4.2.1 空间扩展投影.....	46
4.2.2 雷达数据预处理.....	47
4.2.3 增加注意力机制.....	48
4.3 实验与结果.....	49
4.3.1 数据准备.....	49
4.3.2 消融实验.....	49
4.3.3 与现有方法的对比.....	51
4.3.4 检测结果显示.....	51
4.4 本章小结.....	52
第 5 章 总结与展望.....	54
5.1 本文总结.....	54
5.2 展望.....	55
参考文献.....	56
攻读硕士学位期间发表学术论文及参与项目	65
致谢.....	66

第 1 章 绪论

1.1 研究背景及意义

随着人工智能和新一代数字技术的飞速发展，汽车企业正经历着一场前所未有的变革，在这场变革中，自动化与电气化成为了推动全球汽车产业格局变革的重要力量^[1]。汽车企业不断提升车辆的自动化智能化水平，同时市场对自动驾驶技术的接受度也在稳步上升，但安全始终是自动驾驶汽车技术发展中不可忽略的底线。自动驾驶汽车的安全行驶依赖于对周围环境的精准感知，其中，三维目标检测成为了一项核心技术，并被认为是自动驾驶技术的重要发展方向^[2]。如图 1-1 所示的是车辆三维目标检测效果图，检测所获得的信息使得车辆在复杂的道路环境中做出正确的驾驶决策。

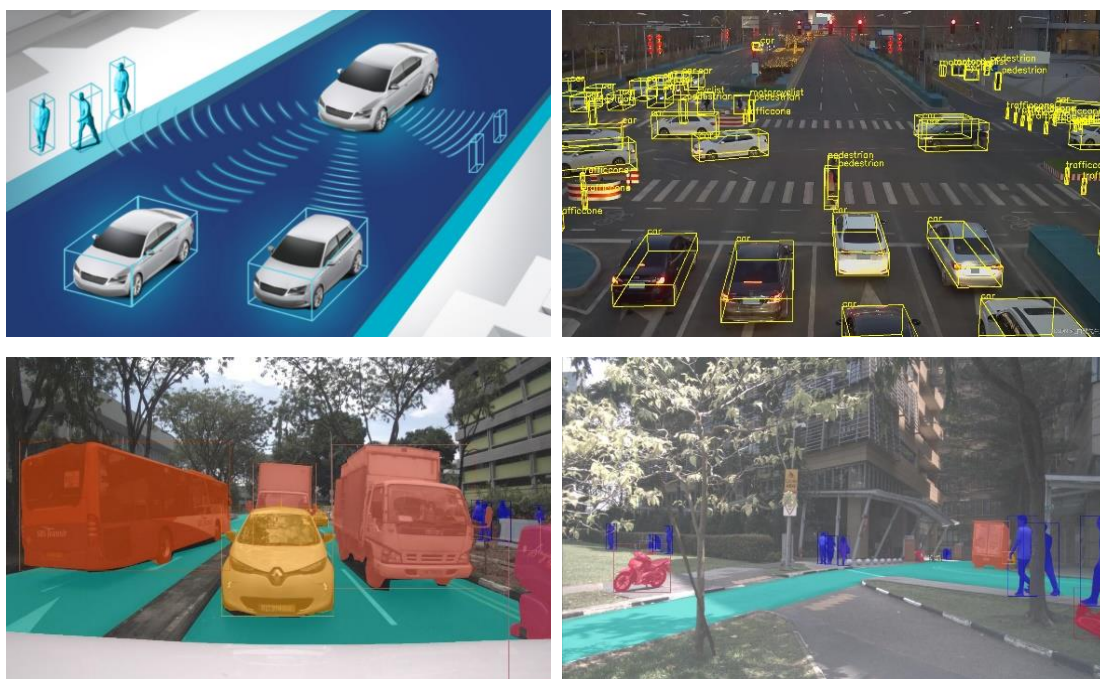


图 1-1 车辆三维目标检测

如图 1-2 自动驾驶三维目标检测的流程所示，自动驾驶系统依赖于先进的环境感知、自主决策和控制执行三大关键技术，通过多种车载传感器感知周围环境，并利用三维目标检测算法和网络模型实时分析和决策，使得车辆能够准确地感知和描述周围环境中物体的位置、形状、大小和类别^[3]。这为自动驾驶汽

车识别道路标志，检测和跟踪其他目标提供了关键支持，使其能够安全地适应复杂的交通环境。

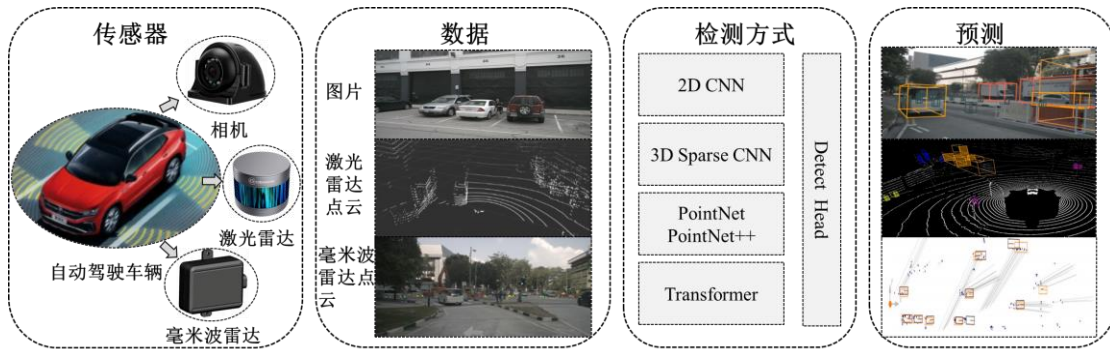


图 1-2 自动驾驶三维目标检测的流程

传感器感知技术对自动驾驶系统的实现具有直接影响。作为常见的车载传感器，相机、激光雷达和毫米波雷达在目标检测方面各有其独特的优势。首先，相机以像素的形式提供目标的丰富外观信息，包括颜色、形状和纹理，这些特征可以清楚地识别障碍物。其次，激光雷达通过激光束进行实时动态测量，以三维点云的形式提供周围环境的空间信息和目标轮廓，通过较高的角度分辨率和探测精度预测目标位置和速度，并且在恶劣天气和极端光照条件下仍然实现可靠检测。最后，毫米波雷达作为自动驾驶汽车上常见且必要的传感器，能够通过多普勒效应测量目标的距离，速度和角度等信息，并且对各种光照和天气条件具有良好的适应性，尤其在穿透雾、烟等遮蔽物方面效果更佳。相较于激光雷达和相机特点，毫米波雷达具有更远的探测范围和更低的价格成本。

Chao 等人^[4]详细探讨了毫米波雷达、激光雷达和相机在不同应用场景中的优缺点，如图 1-3 相机，毫米波雷达和激光雷达的特性比较所示，传统的单一传感器在目标检测方面各有其局限性，难以达到理想的实时性和准确性效果。因此，传感器融合技术成为了提高目标检测效果的最佳选择之一，就系统成本而言，毫米波雷达和相机都相对经济有效，被大量应用于汽车上。从传感器特性互补上，具体而言，毫米波雷达能够在各种天气条件下提供距离和速度信息，但目标分类的能力较差^[5]，而相机具有良好的色彩感知和分类能力，当光线不足、恶劣天气或镜头沾上灰尘或者水滴导致的成像结果不清楚，目标检测性能会受到显著影响。综上所述，基于毫米波雷达和相机各自的优缺点，互补的优势可以有效地提高车辆的感知精度和目标检测鲁棒性，从而提升自动驾驶的安

全性和可靠性。

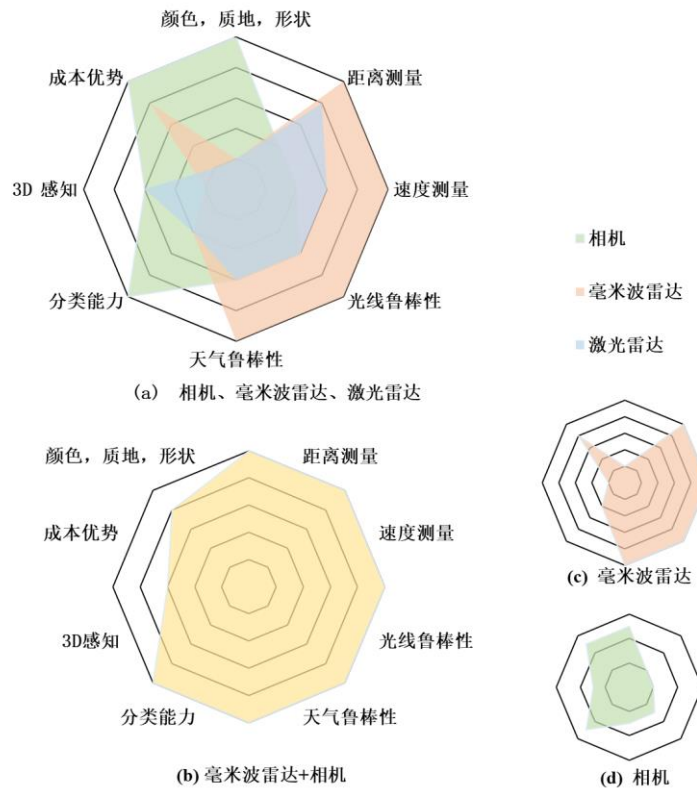


图 1-3 传感器特性对比。(a)相机，毫米波雷达，激光雷达特性的对比(b)雷达和相机融合特性；(c)毫米波雷达特性；(d)相机特性

鉴于上述研究背景，本文致力于对毫米波雷达和相机传感器融合的三维目标检测方法进行研究。通过探索新的深度神经网络框架和设计新的多模态信息融合方法，旨在提高三维目标检测的准确性，从而提升复杂交通场景下环境感知的安全性。

1.2 国内外研究现状

自动驾驶技术的进步推动了三维目标检测研究的浪潮，促进了多种创新方法的发展。当前国内外学者根据不同的传感器或传感器组合的数据特点，提出了多种有效的三维目标检测方法。从本文研究内容出发，首先介绍基于点云检测的不同方式，本节以三个方面展开对基于毫米波雷达的三维目标检测算法的介绍，分别为基于点的方法，基于网格的方法和基于图形的方法；其次按照融合位置的不同分为前视图和鸟瞰图，列举了基于相机-毫米波雷达的三维目标检测算法。

1.2.1 基于毫米波雷达的三维目标检测

现阶段基于激光雷达点云的三维目标检测研究已经取得了显著的研究进展，然而，毫米波雷达点云只能提供目标的粗略位置，无法准确地表示该目标的轮廓信息。基于毫米波雷达点云的目标检测和语义分割算法，大部分是从基于激光雷达点云的检测算法中获得灵感。

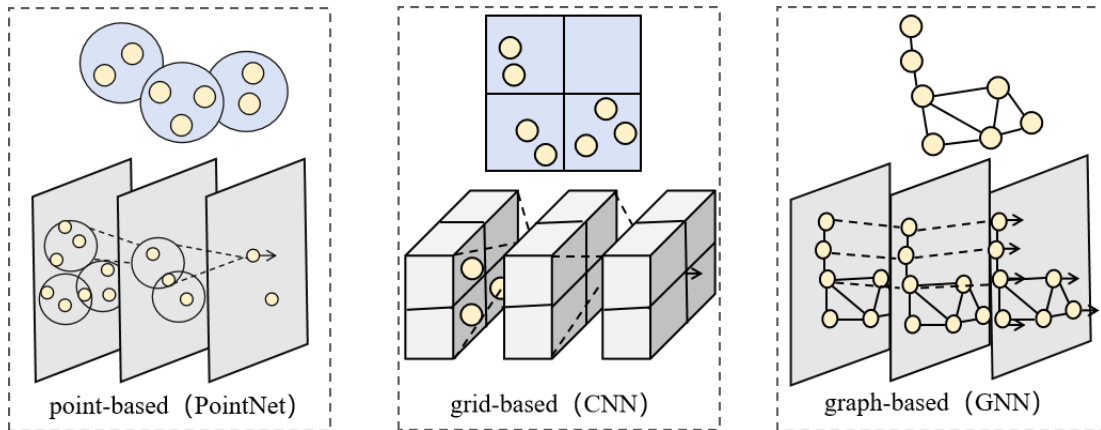


图 1-4 深度神经网络检测点云内目标的三种代表算法

由图 1-4 可见，利用深度神经网络检测点云内目标的方法可分为三大类：

（一）基于点的方法：直接处理点云数据，通过学习和提取每个点的特征进行检测对应目标；

（二）基于网格的方法：将点云数据转换为规则网格结构进行处理，以便于应用传统的卷积神经网络进行目标检测；

（三）基于图形的方法：利用图形结构进行建模和处理^[6]，通过学习点云中点与点之间的分布关系来完成目标检测的任务。

（一）基于点的方法

基于点的方法如表 1-1 所示，通常使用原始点云作为输入，并利用如 PointNet^[7]，PointNet++^[8]和 Frustum PointNets^[9]作为主干网络，以尽可能多地保留原始点云的几何形状。Qi 等人提出了 PointNet，是基于点集合学习方面的先驱之一，能够对每个点的特征进行编码，然后对整个点云进行聚合，以产生全局特征表示，在不需要显式拓扑结构的情况下能够对点云数据进行端到端的处理，是点云处理领域的重要里程碑。PointNet++是对 PointNet 的改进版，旨在提高对局部特征的提取能力，其引入了分层结构，允许网络在不同的层级上聚合局部和全局信息，使得密度自适应层，即多尺度分组（Multi-Scale Grouping,

MSG) 和多分辨率分组 (Multi-Resolution Grouping, MRG), 能够自适应地捕获点云的局部结构和细粒度模式。Frustum PointNets 首先将 3D 场景中的目标投影到 2D 图像平面上, 然后使用 PointNet 或类似的网络对每个目标的点云数据进行处理, 最后进行目标检测和定位, 其结合了点云数据和视锥体投影的信息, 以实现在 3D 场景中的目标检测和定位。

表 1-1 基于点检测方法

方法	内容	局限性
2D 检测 ^[10]	仅根据稀疏雷达数据使用 PointNet 检测 2D 物体的方法。	需要先验知识来调整候选目标的特征, 可能导致对不同场景的适应性有限。
High-Resolution ^[11]	对高分辨率汽车雷达进行点云分割, 通过开发基于点的神经网络, 解决了雷达反射点数据格式与传统卷积神经网络不匹配的问题。	分组半径的确定选择对分割准确性有显著影响。
Radar Track ^[12]	利用多帧的雷达点云数据来预测道路用户的轨迹信息, 并通过两个分支输出置信度和跟踪信息, 通过多普勒速度信息提高了跟踪性能, 并使用卡尔曼滤波器进行轨迹关联以保持一致的轨迹。	当前模型在长期预测方面存在挑战, 数据集类别表示不全, 会影响不同场景下的泛化能力。
Scheiner ^[13]	使用雷达数据检测非视距场景中的行人和骑车者, 并利用目标的反射来估计隐藏目标的几何形状和速度进行目标跟踪。	对高度反射的目标定位有限制。
Palfy ^[14]	神经网络将每个毫米波雷达目标和目标级特征周围的雷达立方体裁成块并引入聚类步骤创建候选目标。	目标整合时间和跟踪方面可进一步提升。

Danzer 等人^[10]采用了一种基于两阶段方法的 2D 汽车检测和分割技术, 利用了 PointNet 和 Frustum PointNets。该方法将每个点视为一个候选目标, 并通过先验知识对候选目标的特征进行调整。随后, 利用 PointNet 和 Frustum PointNets 对每个候选目标中的点进行分类。最后, 仅对被确定为目标的目标执行边界框预测。Feng 等人^[11]展示了神经网络在毫米波雷达点云分割中的潜力, 他们采用了 PointNet 和 PointNet++ 架构, 该研究重点关注了道路和护栏等目标的分割任

务。

Radar TrackNet^[12]利用收集到的多帧毫米波雷达点云来检测和跟踪道路用户，该方法结合了多普勒速度信息和双分支方法用于置信度和跟踪的输出，实验结果表明该网络优于其他跟踪方法。Scheiner 等人^[13]证明了雷达点云在探测和跟踪非视线（Non-Line of Sight, NLOS）环境中的隐藏目标能力。通过将 BEV 雷达点云转换为伪图像，并应用金字塔网络和放大网络，能够在看到物体碰撞之前估计碰撞物体的几何形状和速度。

Palffy 等人^[14]的研究采用卷积神经网络（Convolutional Neural Network, CNN）从雷达数据立方体中提取局部信息模式的方法，将这些模式作为雷达检测点的协变量。通过这种方式，显著提升了毫米波雷达监测点分类器的效能。

基于点的方法通常会尽可能多的保留原始点云的几何形状，然而与网格方法相比，硬件高效地支持是影响三维空间中点检索的关键。

（二）基于网格的方法

雷达点云数据的处理对于传统方法来说具有挑战性，因其空间非均匀性的分布，为了更有效地处理这类数据，研究人员普遍采用将多帧点云将其转换为鸟瞰图（Bird's Eye View, BEV）伪图像的方法，也称之为网格图^[15]。这种网格图的结构类似于二维图像，使得能够直接应用基于深度学习的算法来完成检测和分割任务。如表 1-2 所示，展示了基于网格的主流方法。

网格图通常分为两种类型：一种是基于占据的网格图，用于表示雷达数据中的障碍物和自由空间；另一种是基于振幅的网格图，显示每个单元反射的雷达截面积（Radar Cross Section, RCS）值。Werber 等人^[16]提出了基于占据和基于振幅的网格图，这两种类型均针对定位任务设计。实验结果表明，基于占据和基于振幅的算法都有效地提供了周围环境信息。在此基础上，研究人员采用了名为 LeNet-5^[17]的卷积神经网络来对车辆和非车辆^[18]进行分类。

随后，Lombacher 等人^[19]结合了基于占据网格地图^{[20][21]}的方法，并应用基于像素的算法对每个网格进行分类，从而揭示了通过雷达网格地图获取语义信息的潜力。与此同时，Dreher 等人^[22]将毫米波雷达点云转换为至多一个点的网格地图，并利用 YOLOv3^[23]进行目标检测。

为了处理静态障碍物，Sless 等人^[24]提出了一种改进的激光雷达地面真实图

方法，并通过监督学习实现了静态障碍物的占据网格映射。该方法利用了具有跳跃连接的编码器-解码器架构，从数据中推导出潜在的联合概率函数，并超越了传统的方法。另外，Schumann 等人^[25]引入了一个双分支网络，将静态和动态目标分离成不同的分支。其中一个分支在网格地图上执行语义分割以识别静态目标，而另一个分支在点云上执行实例分割以检测移动目标。然后将这两个分支的输出合并，生成语义点云。然而，这种双分支结构会导致高计算工作量并增加内存消耗。

表 1-2 基于“网格”检测方法

方法	内容	局限性
基于 振幅 的 网 格 图	Radar Grids ^[15] 使用卷积神经网络在雷达网格中进行目标分类的方法。	网络性能受到输入窗口大小的影响，更大的输入窗口能获得更好的结果；网络无法准确分类未完全出现在雷达网格中的目标。
	Dreher ^[22] 基于 YOLOv3 网络架构的全卷积网络，用于雷达数据中的二维目标检测。	处理数据时累计数据可能会导致移动目标形状的失真。
	双分支 ^[25] 分成静态目标和动态目标两个独立的分支，两个分类器结果合并成一个点云，新的雷达网格地图进行循环实例分割	两个分支的综合融合，会需要更多的运算时间，点云捕捉信息不完全。
基于 占 据 网 格 图	Sless ^[24] 从稀疏雷达数据中学习占据网格映射的新方法。	网络性能收到聚合雷达帧数量的影响，在处理目标类别不平衡时需要进行权重调整。
	Werber ^[16] 提出两种基于雷达数据的网格映射算法。	基于振幅网格图更好展示不同目标的反射特性，占据网格图更清晰结构化车辆周围密集的目标。
	Lombacher ^[19] 利用雷达数据和卷积神经网络进行停放车辆检测的方法。	对于其他障碍物高度遮挡的车辆，形状建模不够完整，导致识别困难。

尽管上述方法取得了一定的进展，但仍然存在一些挑战。由于网格地图是由一系列雷达点云构成的，因此不适合实时映射。此外，点云的稀疏性仍然影

响着对网格地图进行检测和分割的准确性。

（三）基于图形的方法

图（graph）是一种非结构化的抽象数据结构，由节点（node）和连接节点的边（edge）组成的。图神经网络（Graph Neural Network, GNN）是指一类专门基于图结构数据进行特征学习和提取的神经网络模型。在图神经网络中，点云中的点被映射为图的节点，以保留点云的结构信息，并通过图的边来建模点之间的关系^[26]。

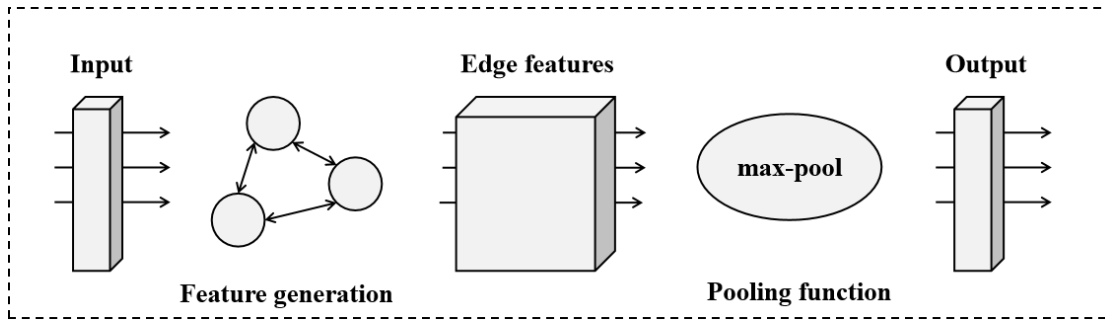


图 1-5 Radar-PointGNN 网络中利用 GNN 卷积图

如表 1-3 所示，在点云数据处理中，基于图的方法被广泛应用。通过构建图结构来实现操作，Shi 和 Rajkumar 等人成功地将这种方法应用于激光雷达点云的目标检测任务中^[27]。另外，Radar-PointGNN^[28]模型如图 1-5 所示，是一种基于毫米波雷达点云的检测新型方法。除了空间坐标外，速度、雷达截面积、测量速度、点密度等属性也被视为每个数据点的特征。Radar-PointGNN 的研究表明，尽管理论上低质量毫米波雷达点云数据的目标检测是可能的，但实验中其不能产生足够准确的结果。类似地，Gong 等人^[29]使用 Point-GNN 方法作为基线，提出了用于人类活动识别的 MMPoint-GNN 方法，MMPoint-GNN 处理由毫米波雷达收集的稀疏点云，证明了 GNN 在毫米波雷达相关任务中的优越性能。利用毫米波雷达数据，采用时空图卷积网络进行步态识别。此外，在文献^[30]中通过建立 DGCNN 模型的一个变体来解决手势识别问题，通过构建局部领域表示每个点的特征，同时将边缘卷积模块首尾相连以获得多层次和更丰富的语义表征。

采用 GNN 对雷达点云进行三维特征提取，结果表明通过将雷达点云映射到上下文表示，图表示比其他点云编码器产生更有效的目标建议。GNN 在检测雷达张量方面也显示出其优势。图张量雷达网络（GTR-Net）^[31]的体系结构利用图卷积操作来聚合点云节点之间的信息。该过程包括根据连接节点各自的边权

值对其特征进行加权。因此，通过聚合相关信息来改进稀疏点的缺陷，从而获得更好的性能。

表 1-3 基于图神经的 3D 目标检测方法

方法	内容	局限性
Radar-PointGNN ^[28]	不仅利用点的特征，还利用了点与点之间的特征，并将其嵌入到图边缘中。	输入特征的限制，以及在边缘特征生成过程中需要使用特定的点对特征进行补偿。
GTR-Net ^[31]	直接使用原始雷达张量数据，不依赖于人为生成的点云数据。	超过 50 米距离范围的目标检测能力尚未充分评估。

利用图的检测方式不仅直观地表示出点之间的关联性，而且对于更复杂的空间局部结构可同时捕获“顶点一边”信息，与使用 PointNet 等为骨干的网络相比，更容易感知目标的 3D 形状，其代价是花费更长的前馈时间。

1.2.2 基于相机-毫米波雷达融合的三维目标检测

在自动驾驶领域的感知环节中，融合多模态传感器数据已成为一项重要的发展趋势。尽管单一传感器各有其优势，但单独使用时受到性能限制。相比于单一传感器，雷达与相机的融合系统能够提供更丰富的深度信息和纹理信息，同时避免了使用激光雷达时高昂的计算和价格成本^[32]，已有大量研究验证了相机-雷达融合技术对目标检测和语义分割准确性和鲁棒性的显著提升。

通常情况下，相机-毫米波雷达融合网络采用双输入分支结构，分别接收毫米波雷达和相机数据，并根据融合阶段的不同分为数据级融合、特征级融合和决策级融合，以获得最终的融合结果。根据数据融合的位置的不同又分为前视图融合和鸟瞰图融合两种类型。

（一）基于相机-雷达前视图融合

前视图融合（Front View, FV）是将毫米波雷达数据投影到图像平面上的方法，其中雷达数据可以是三维点云、部分点云信息或雷达张量。通过在图像平面上生成潜在目标的候选区域，该方法排除了大量的非目标区域，减少了计算负担，提高了目标识别速度。在嘈杂的环境下，Yadav 等人^[33]发现毫米波雷达数据表现了较好的鲁棒性，与相机数据融合后系统性能与毫米波雷达检测方

法相比进一步提高。基于相机-雷达前视图融合的目标检测方法总结见表 1-4。

表 1-4 基于相机-雷达前视图融合的目标检测方法

方法	内容	局限性
RPR ^[34]	调整雷达提议的大小和位置，估计每个雷达提议的分数。	需要额外的计算成本用于调整雷达提议的属性。
RRPN ^[35]	基于雷达检测的实时区域提议，可以用于任何两阶段目标检测网络。	依赖雷达检测，可能受雷达检测范围和精度的限制，无法涵盖视觉数据提供的信息。
Chadwick ^[38]	使用不同焦距相机自动生成的训练数据的方法，提高远处小目标检测性能。	地面真实值集中非常小的目标比例较高。类似图像的雷达点表示存在局限性。
YODar ^[33]	将 YOLOv3 目标检测网络和定制的雷达分割相结合。	由于天气原因，图像的 3D 定位面临挑战。
RVNet ^[41]	雷达和视觉特征分别提取，输出分支后，雷达和图像特征图融合。	雷达特征稀疏，无法划定障碍物。
Dong ^[43]	使用嘈杂标签和基于规则的关联方法训练。	难以准确关联雷达点和边界框；检测结果中信息有限，阻碍了基于学习的融合。

为了解决毫米波雷达点云与图像像素之间的关联问题，Qi 等人^[34]提出雷达建议细化（RPR）网络技术，该网络能够将雷达建议与相机图像特征相结合，调整雷达提议的大小和位置。并估计每个雷达提议的客观性分数以对应背景噪声等问题。雷达区域建议网络（RRPN）^[35]是一种实时雷达目标检测注意机制，通过在图像上投影雷达点数据，然后为每个雷达点生成锚定边界框，借用成熟的图像检测技术作为基础，将雷达点数据转换为类似图像格式，提供更精确的目标检测建议。

雷达数据投影到图像平面上还可用于创建特征图，以补充基于图像的特征。这些方法通过利用额外的输入，包括范围、速度和雷达截面积（RCS）值来提高检测精度。Nabati 和 Qi^[36]通过使用 CenterNet^[37]在图像中的中心点来定位目标，利用雷达特征（如深度、旋转角度、速度）来补充图像特征，然后，利用基于视锥的数据关联策略，将雷达探测结果与图像中的目标进行精确匹配，生成基于雷达的特征图来增强图像特征，实验结果表明，相较于 CenterNet 算法，

整体场景检测评分（NDS）提高了 12%以上。

Chadwick 等人^[38]将雷达点云投影到相机平面上，并将雷达和相机分支的特征由 ResNet^[39]块生成，然后通过连接和添加操作进行融合，这种方法超越了单独使用相机的探测器性能，因为雷达传感器能够持续为远距离目标提供有效的运动指示。在相机-雷达融合中，YOdar^[40]是一种基于不确定性的方法，其中包括降低评分阈值，并将雷达点云分配给图像切片，聚合输出组合与梯度增强分类器，实验结果表明，YOdar 显著提高了黑夜场景的目标检测性能。

此外，一些研究人员将雷达点云投影到图像平面上，形成雷达伪图像。例如，在 RVNet^[41]和 SO-Net^[42]中，这些方法通过相机传感器的固有矩阵将稀疏雷达图像数据转化为图像坐标系中的伪图像，其中包含有关深度、横向速度和纵向速度的信息。RVNet 是一个基于 YOLOv3 的单级目标检测网络，包含雷达和相机两个输入分支，以及用于小障碍物和大障碍物检测的两个输出分支。SO-Net 则重点研究了单一网络内的多任务学习，对 RVNet 中的两个输出分支进行了改进，用于车辆检测和自由空间分割。

Dong 等人^[43]提出了另一种关联网络，将雷达点云和二维边界框投影到图像平面上，从而形成新的伪图像，用于学习来自两个传感器的语义表示信息。该网络通过计算和最小化雷达点云和图像边界框之间的欧氏距离，提高了关联的准确性。

将雷达数据投影到图像平面上可以为目标检测提供建议和特征。然而，由于雷达提供的方位角分辨率较低以及相机的校准误差，投影的雷达点云可能会偏离目标^[44]。虽然扩大感兴趣区域可能有助于解决这个问题，但这可能导致同一区域内检测到多个重叠的目标，使目标匹配变得混乱。此外，由于目标的遮挡，雷达数据在图像透视图上的投影可能会受到限制。

（二）基于相机-雷达鸟瞰图融合

如表 1-5 所示，另一种融合方法是将毫米波雷达数据或相机图像转换为鸟瞰图（BEV）坐标。例如，Meyer 和 Kusch^[45]是首次提出了 4D 毫米波雷达点云数据集，该数据集受到数据大小的限制，仅包含 500 帧，每帧点云生成 6 个高度图和 1 个密度图的 BEV 图像。Cui 等人^[46]将雷达点云投影到前视图和鸟瞰图，并提出了一个 3D 区域建议网络，用于从相机图像和雷达 BEV 图像中生成建议，

这是目前唯一的 4D 毫米波雷达和相机校准方法。随着 4D 毫米波雷达技术的快速发展，4D 毫米波雷达校准是一个潜在的方向，未来将会有更多的突破。

表 1-5 基于相机-毫米波雷达鸟瞰图融合的目标检测方法

方法	内容	局限性
Radsegnet ^[47]	在全天候条件下，尤其在恶劣天气条件下保持更好的性能	聚变系统受相机输入影响。
Simple-BEV ^[48]	批量处理大小和输入分辨率会极大地影响性能。	毫米波雷达数据的稀疏性。
Lim ^[49]	毫米波雷达信号和相机数据的互补性增强目标检测。	毫米波雷达依靠有限的滤波获得信息。

BEV 图像将感知空间离散成网格，这可能会导致有价值的细化边界框信息的丢失。为了解决这个问题，Bansal 等人^[47]在 BEV 图中添加了其他基于点的特征（如速度、RCS 值），从相机语义图、雷达点云和雷达 BEV 网格图中创建了一个语义点网格（Super Point Graph, SPG）。SPG 编码可以独立地从相机和毫米波雷达中提取信息，并将相机图像中的语义信息编码到毫米波雷达点云中。因此，在相机输入变得不可靠的情况下，SPG 编码方法仅使用毫米波雷达数据仍然保持可靠的操作。

Harley 等人^[48]专注于来自多个相机和毫米波雷达的 BEV 地图。他们的方法通过在自车相机图像周围投影三维坐标和投影位置的双早期采样特征来生成具有特征的三维体积。随后，通过将三维特征与光栅格化毫米波雷达图像连接起来，生成 BEV 特征图，是将多个毫米波雷达传感器的所有雷达点云转换为 BEV 坐标，得到高维 BEV 特征图。

除了将毫米波雷达数据投影到 BEV 外，还可以利用同源矩阵将相机图像从前视图转换为鸟瞰图。例如，Lim 等人^[49]利用逆透视变换（Inverse Perspective Mapping, IPM）^[50]将相机图像和二维毫米波雷达张量转换为笛卡尔坐标，使用同源变换将相机图像投影到毫米波雷达平面上。在通过独立的特征提取器分支后，两种模态的特征通过额外的融合层，形成统一的特征图。此外，毫米波雷达点云和相机图像都被投影到鸟瞰图上，并通过独立的特征提取学习共享的特

征^[51]。因此，在 BEV 上的投影数据比前视图在遮挡的情况下存在优势。尽管如此，由于 IPM 是基于平坦路面的假设，但应用于现实世界的场景时，容易产生动态目标的扭曲。

综上所述，基于毫米波雷达点云的目标检测方法在恶劣天气条件下具有较好的稳定性，但存在数据稀疏和分辨率低的问题。基于相机-毫米波雷达融合的 3D 检测方法能够克服单一传感器的局限性，将两种模态数据进行互补融合，提高了目标检测的准确性和鲁棒性。受到该类方法的启发，本文针对毫米波雷达点云特征提取和雷达点云的稀疏性问题，提出了基于相机和毫米波雷达传感器的 3D 目标检测算法的研究。

1.3 本文研究内容及章节安排

本文采用多模态融合技术，将相机图像与毫米波雷达点云数据相结合，实现了对目标的三维检测，这一研究顺应了多传感器信息融合的发展趋势，具有重要的科研价值和应用前景。

第 1 章 绪论，在阐述选题的背景和意义的基础上，对基于毫米波雷达点云的研究算法及基于相机-毫米波雷达融合的三维目标检测技术的研究现状进行了分析，为本文后续算法的研究奠定基础。最后简述本文的研究内容和章节安排。

第 2 章主要介绍了毫米波雷达的基本原理和工作机制。首先介绍了毫米波雷达的分类，其次介绍了毫米波雷达的信号和系统模型，在此基础上，分别对毫米波雷达的测距，测速和角度测量展开介绍，然后引入了现阶段包含毫米波雷达数据的数据集，为后续实验的进展进行铺垫，最后是列举出两种常用的毫米波雷达点云和视觉图像进行数据关联的方法，为后续章节中融合过程具有指导作用。

第 3 章提出了一种基于毫米波雷达点云进行的三维目标检测的方法，RPCNET 网络它不仅使用点本身的信息，还使用点之间的关系。该模型被设计为同时考虑点特征和边缘特征，嵌入在图的边缘。该模型通过初步锚定空间近端点所处的候选目标检测框，通过 DGCNN 网络对图的边缘进行卷积运算来提取局部模式，最后将所提出的 RPCNET 模型在 NuScenes 数据集和 RadarScenes 数据集上优于以往的所有方法。

第 4 章 旨在利用毫米波雷达和相机数据的特征融合，以实现三维目标精准

检测，本方法对经典的 CenterFusion 网络的改进和进化。首先，为了保证融合体系结构的完整性，在系统检测初期，不再仅仅依赖于相机所采集到的图像信息，本章将毫米波雷达点云投影到图像平面上，并将其作为附加通道添加到 CenterNet 图像检测网络的输入中，以生成相关属性初步形成三维检测框。接着，采用了毫米波雷达点云密度聚类技术，是对毫米波雷达点云进行聚类产生标签，过滤无关点云，有效地滤除白噪声，从而提高了数据质量和目标检测的可靠性。最后，添加一个通道注意模块 Squeeze-and-Excitation Networks，对于每个特征通道进行加权，以增强网络的重要特征。

第 5 章 总结与展望。对全文工作及创新点进行总结，并讨论未来的研究方向。

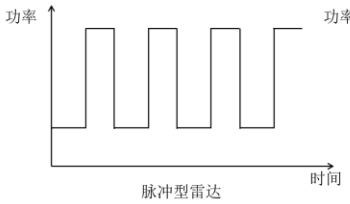
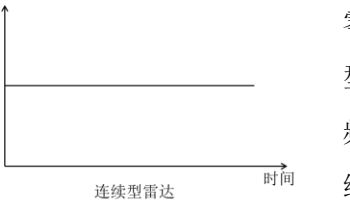
第 2 章 毫米波雷达检测

2.1 引言

雷达技术最早于 19 世纪左右被引入，主要用于无线电探测和测距，是军事和监视安全应用的重要技术^[52]。近年来，毫米波雷达技术的应用领域不断扩大，尤其是在商业、汽车和工业领域得到广泛应用。毫米波雷达系统通过发射电磁波信号，随后接收由目标反射的波来探测周围目标，并估计其距离、径向速度和方位角等相关参数，其中主要通过延时和多普勒频率的精度估计测量距离和径向速度。

毫米波雷达的信号通常是连续的波形或者是一系列的短脉冲序列^[53]。如表 2-1 所示，从雷达信号的波形可以看出，毫米波雷达可以分为两种：一种是脉冲式雷达，一种是调频连续波（Frequency Modulated Continuous wave，FMCW）雷达。脉冲雷达采用间歇性地连续发射一系列的短脉冲信号来获得运动目标的距离和径向速度，同时在发射间隙接受回波，从而实现了一种收发间隔机制。根据发射脉冲与截获脉冲之间的时间延迟计算出毫米波雷达与目标的距离^[54]。

表 2-1 毫米波雷达工作机制

工作方式	脉冲类型	连续波类型		
		CW (恒频连续波)	FSK（频移键控 连续波）	FMCW (调频连续波)
特点	原理与激光雷达类似，探测技术复杂，成本较高	只用于测速但不 可用于测距	可探测单个目标的距离和速度	可对多个目标进行距离、速度的探测；分辨率高，技术成熟
电磁波辐射能量图				零差 线性 步进 调频 连续波

另一方面，连续波雷达是一种同时发射和接收回波信号的雷达系统，能够

发射连续的正弦波信号，这些信号可以是单频连续波（CW）或调频连续波（FMCW）。其频率调节方式包括三角波，锯齿波或编码调制等^[55]。其中，线性调频连续波（LFMCW）是一种常见调频模式，其采用三角波或者锯齿波等形式进行调频。线性调频连续波是一种特殊的结构形式，它通过发射波形信号进行频率调制，从而估计目标的高分辨率距离和径向速度^[56]。现代汽车毫米波雷达中，基于 FMCW 调制方案得到了广泛研究^[57]。这些毫米波雷达在自动驾驶中自适应巡航控制（ACC）、盲区监控系统（BSD）等技术的使用中得到了广泛应用，并逐渐成为受欢迎的传感组件之一。

本章内容首先回顾了传统毫米波雷达信号和系统模型，其次阐述了毫米波雷达的距离测量，速度测量和角度测量的基本原理，接着列出了目前阶段包含毫米波雷达点云的数据集，最后介绍了常用的毫米波雷达点云和视觉图像数据关联的方法。

2.2 FMCW 雷达信号和系统模型

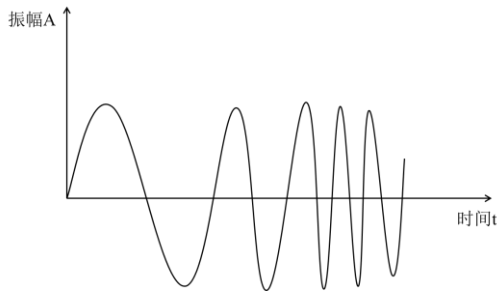


图 2-1 线性调频脉冲，振幅作为时间的函数

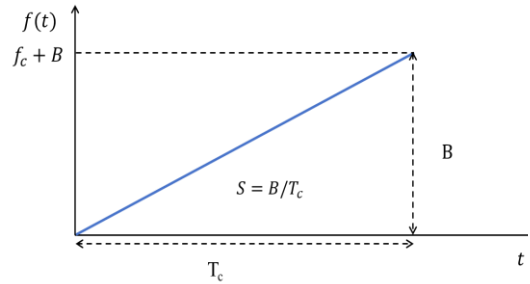


图 2-2 具有起始频率 f_c 、带宽 B 、斜率 S 、时间持续时间 T_c 的 FMCW 线性调频脉冲

在 FMCW 雷达系统中，当信号的频率随时间变化呈线性增加，这种信号被称为线性调频连续波（LFMCW）^[58]。如图 2-1 所示，LFMCW 信号一般具有起始频率、带宽和持续时间的特征。超过持续时间的 FMCW 信号被称为啁啾（chirp）。如图 2-2 所示，在 LFMCW 的信号中，随着时间的推移，频率以速率 S 线性增加，瞬时频率可被表示为：

$$f(t) = \frac{1}{2\pi} \frac{d}{dt} (2\pi f_c t + \pi S t^2) = f_c + S t \quad (2-1)$$

如图 2-3 所示，通过合成器，毫米波雷达系统首先产生成一个线性调频脉

冲，其通过发射天线（ T_X ）发射，到达目标后，信号 $X_T(t)$ 部分能量（即雷达回波）向雷达接收天线（ R_X ）反射，作为信号 $X_R(t)$ 在此被接收。在此基础上，通过对接收到的信号 $X_R(t)$ 进行处理，以提取目标的距离、速度、角度以及等特征，以实现目标的检测、定位、跟踪和识别^[59]，最后，通过混频器将信号 $X_T(t)$ 和 $X_R(t)$ 信号合并，产生中频（IF）信号 $x_{IF}(t)$ 。

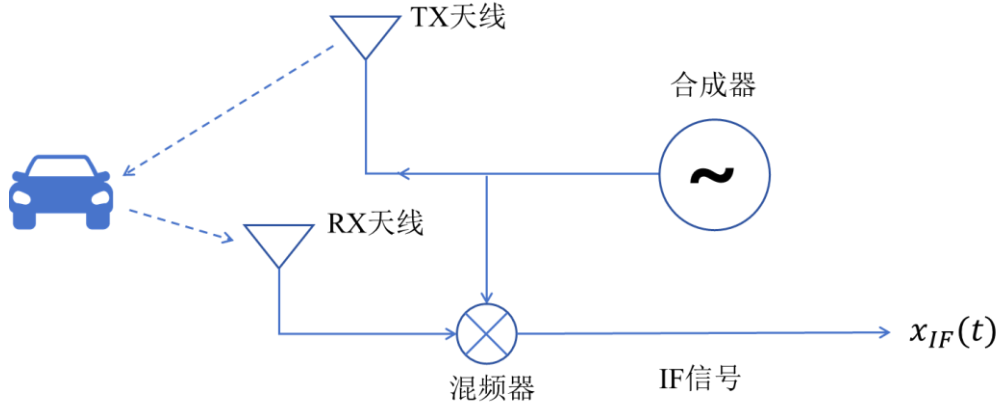


图 2-3 雷达系统的方框图

该混频器是一种将两个信号合并以产生一个新频率信号的电子设备，其两个正弦输入信号 $X_T(t)$ 和 $X_R(t)$ 如下：

$$X_T(t) = \sin(\omega_T t + \phi_T) \quad (2-2)$$

$$X_R(t) = \sin(\omega_R t + \phi_R) \quad (2-3)$$

当 FMCW 信号 $X_T(t)$ 的频率随时间呈线性增加时， $X_T(t)$ 和 $X_R(t)$ 的瞬时频率 $f_T(t)$ 和 $f_R(t)$ 在混频过程中是不同的。输出信号 $X_{out}(t)$ 的瞬时频率即为两个正弦函数的瞬时频率之差。而输出信号的 $X_{out}(t)$ 的相位与两个输入信号相位之差相等：

$$X_{out} = \sin[(\omega_T - \omega_R)t + (\phi_T - \phi_R)] \quad (2-4)$$

混频器的运行机制可以通过图形化方式理解，即在观测期间，线性调频脉冲信号 $X_T(t)$ 和 $X_R(t)$ 的频率表示。毫米波雷达系统中，雷达的发射天线发射线性调频脉冲，如图 2-4 可见，随着时间 τ 的推移，接收天线收到目标反射回来的信号，其频率差为 $S\tau$ ，从而形成中频信号（IF 信号）。

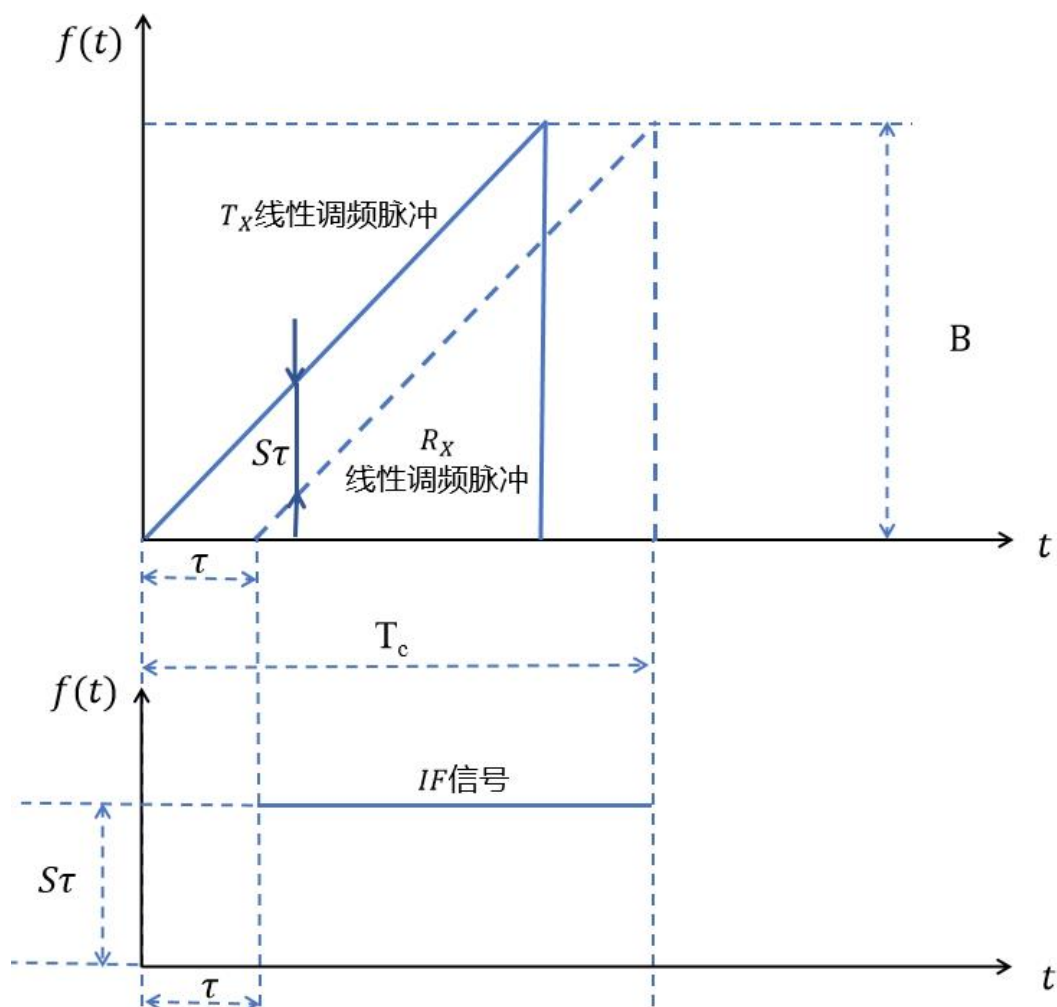


图 2-4 从发射和接收的 chirp 信号中推导出恒频中频信号

混频器生成的 IF 信号在频率和相位上都是由两个输入信号直接相减得到的，因此，可以通过频率-时间图表示，如图 2-4 所示，其呈现为一条频率保持恒定的直线。

2.3 毫米波雷达测距、测速以及角度测量原理

2.3.1 毫米波雷达距离测量

(1) 单个目标测距原理

对于距离雷达的范围为 d 的目标，所发射的信号被该目标反射后延迟后到达接收天线，表达式为：

$$\tau = \frac{2d}{c} \quad (2-5)$$

其中 c 是光速。如图 2-4 所示，当发射和接收的 chirp 信号在时间上重叠时，在有效的接收周期内，中频信号频率（即拍频）保持恒定。即：

$$f_{IF} = f_T(t) - f_R(t) = S\tau \quad (2-6)$$

式 (2-6) 中 S 为图 2-2 中定义的 chirp 信号线性增加频率的斜率。

$$f_{IF} = S\tau = 2Sd/c \quad (2-7)$$

对式 (2-7) 稍作变换，得到目标到雷达的距离为：

$$d = f_{IF}c/2S \quad (2-8)$$

IF信号与频率和相位有关：IF信号的初始相位 ϕ_0 对应图 2-4 左侧垂直虚线表示的时间点，为 TX 线性调频脉冲相位和 RX 线性调频脉冲相位之差：

$$\phi_0 = 2\pi f_c \tau \quad (2-9)$$

由于对于典型的毫米波雷达系统，chirp 信号的起始频率 f_c 远远大于 $S\tau/2$ ，则有：

$$\phi_{IF} = \phi_0 + \pi S\tau^2 \approx 2\pi f_c \tau = 4\pi d/\lambda \quad (2-10)$$

其中，参数 $\lambda = c/f_c$ 是线性调频脉冲在起始频率下的波长。

(2) 多目标测距原理

多目标测距信号图如图 2-5 所示，三种不同的 RX 线性调频脉冲分别接收不同的目标。每个线性调频脉冲都有不同的延迟时间，其延时和与被测目标的距离成正比。不同的RX线性调频脉冲被转换成各固定频率的IF 单音信号。

距离分辨率是对两个或多个目标的一种辨别方法，它是指在一定的空间内，通过快速傅里叶变换（FFT）将其转换为频域，实现被测目标的各个峰值的分离。傅里叶变换原理表明^[60]，在一定程度上依赖于 FMCW 系统的带宽 B ，通过延长IF信号，就能获得更高的分辨率。

傅里叶变换理论还表明，观测窗口 (T) 能分辨间隔超过 $1/THz$ 的频率分量。这意味着频率差值满足方程式 (2-11) 给出的关系，就可以分辨两个 IF 单音信号：

$$\Delta f = \frac{2S\Delta d}{c} > \frac{1}{T_c} \quad (2-11)$$

其中 T_c 是观测时间长度。由公式 (2-11) 可以表达为 $\Delta d > \frac{c}{2ST_c} = \frac{c}{2B}$ （由于 $B = ST_c$ ），距离分辨率 d_{res} 仅取决于线性调频脉冲扫频的带宽：

$$d_{res} = \frac{c}{2B} \quad (2-12)$$

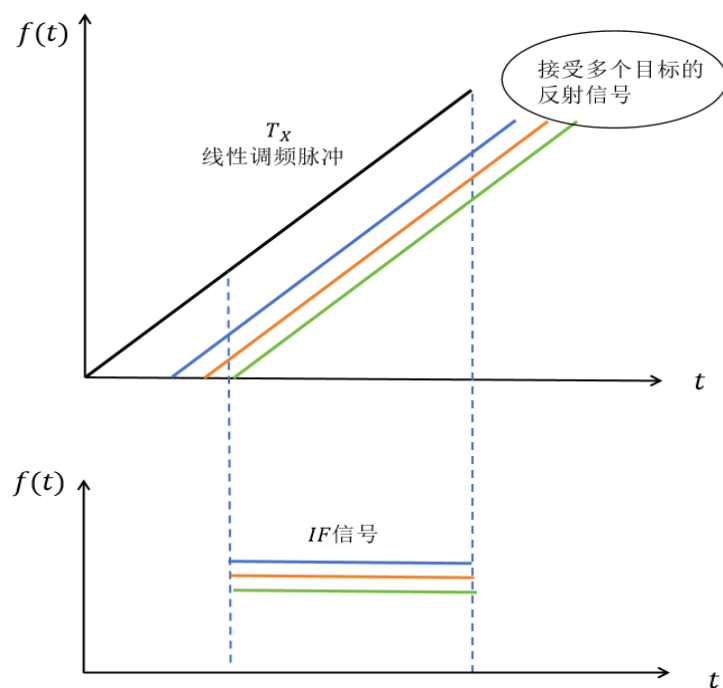


图 2-5 针对多个目标检测反射的多个IF信号

2.3.2 毫米波雷达速度测量

毫米波雷达速度测量技术基于多普勒效应，利用运动目标在运动过程中所产生的相位和频率变化，称为多普勒频移。多普勒频移反映了波在运动中的属性变化规律。当移动台以恒定速率沿一个固定方向移动时，因为传播距离之差，相位和频率会发生变化，目标距离的微小变化 Δd 会导致接收信号的相位变化。因此，根据接收到的信号在一定时期内的相位变化来估计目标的速度（即 $\Delta d/t$ ）。本小节主要从两线性调频脉冲的速度测量，同距不同速的多目标测速以及速度分辨率三个方面进行介绍。

（1）两线性调频脉冲的速度测量

如图 2-6 所示，FMCW 雷达通过间隔发射双线性调频脉冲信号来测量目标，其间隔为 T_c [61]。每个反射的线性调频脉冲经过傅里叶变换处理，与每个线性调频脉冲相对应的距离-傅里叶变换，在同一位置上出现不同的峰值。

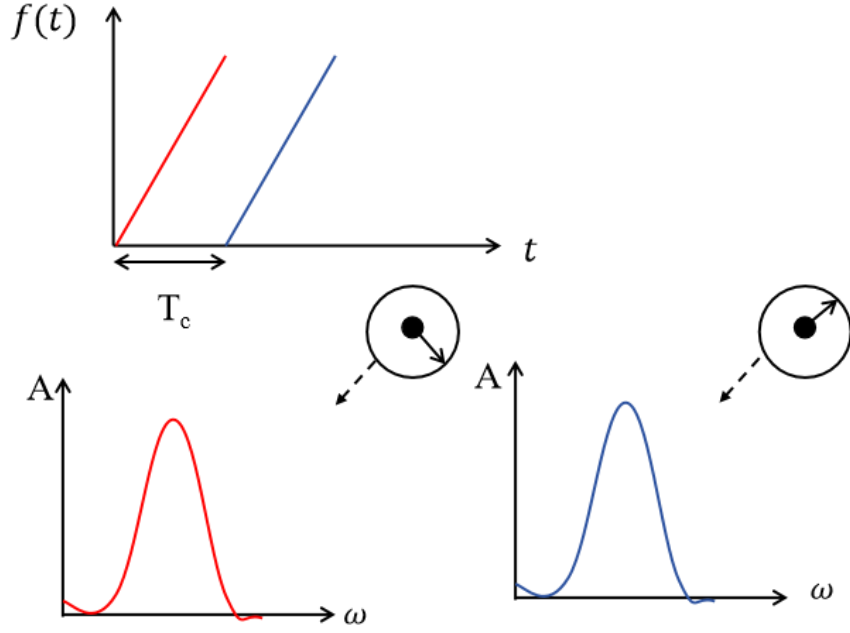


图 2-6 两啁啾之间的速度测量

若目标速度为 v ，在持续时间 T_c 内，目标移动的距离为 $\Delta d = vT_c$ 。根据中频信号模型中的公式(2-11)和公式（2-9），通过双线性调频脉冲得到的中频信号其频率和相位之差为：

$$\Delta f = \frac{2S\Delta d}{c} = \frac{2SvT_c}{c} \quad (2-13)$$

$$\Delta \phi = \frac{4\pi\Delta d}{\lambda} = \frac{4\pi vT_c}{\lambda} \quad (2-14)$$

速度可以通过公式(2-15)：

$$v = \frac{\lambda \Delta \phi}{4\pi T_c} \quad (2-15)$$

由于测速是以相位差为基础，只有当 $|\Delta \phi| < \pi$ ，通过数学方法推导出来 $v < \frac{\lambda}{4T_c}$ 。公式（2-16）提供了双线性调频脉冲以间隔为 T_c 测得的最大相对速度（ v_{max} ）。 v_{max} 越高，两个信号之间的传输时间就越短

$$v_{max} = \frac{\lambda}{4T_c} \quad (2-16)$$

（2）同距不同速的多目标测速

当多个速度各异的目标与雷达之间的距离相同时，双线性调频脉冲速度测速就不再适用。因为目标在同一距离上，产生相同频率的IF信号，通过距离-傅里叶变换会产生距离相同目标的合并信号单个峰值^[62]。

此时要实现测速，需要多个线性调频脉冲。如图2-7所示，雷达系统必须发射一组 N 个等间隔的线性调频脉冲，称为帧（Frame）。

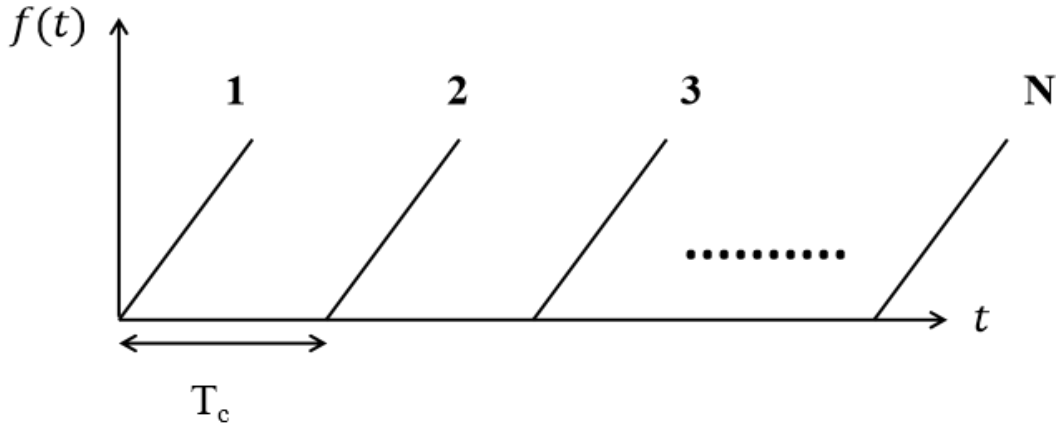


图 2-7 chirp 帧频率作为时间的函数

在同一帧中，当两个等距不等速的目标 v_1 v_2 ，利用多普勒-傅里叶变换计算出目标的不同相位，产生两个具有不同的峰值，如图 2-8 所示，其对应的横坐标为两个目标的相位差。

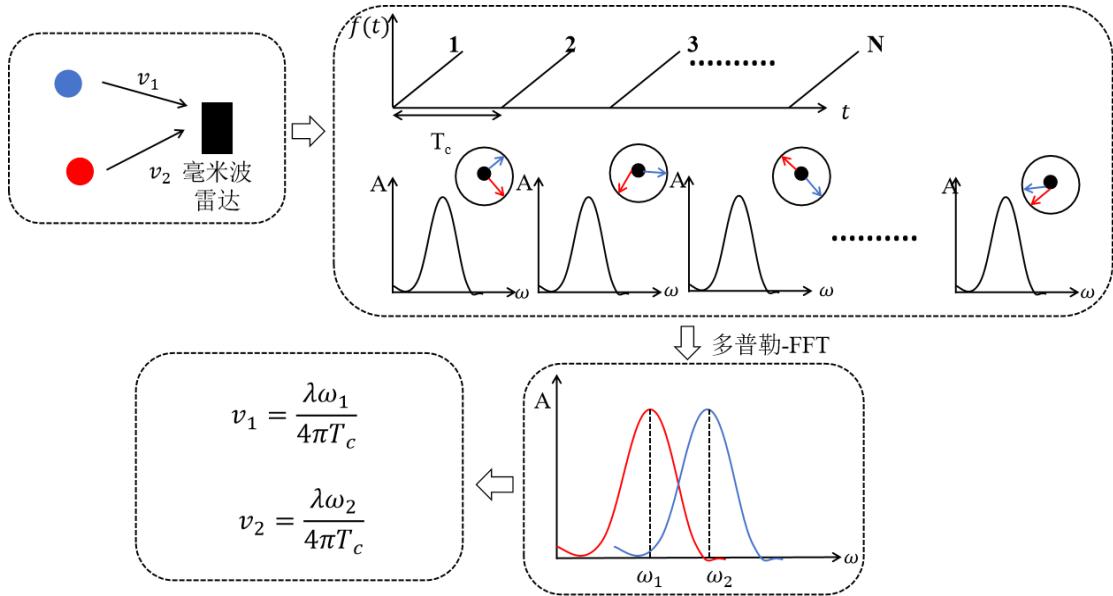


图 2-8 同一距离处多个物体的速度估算

同距不同速的两个目标的速度分别为：

$$v_1 = \frac{\lambda \omega_1}{4\pi T_c}, \quad v_2 = \frac{\lambda \omega_2}{4\pi T_c} \quad (2-17)$$

ω_1 和 ω_2 对应于各个目标的连续线性调频脉冲之间的相位差

（3）速度分辨率

当多个目标将线性调频脉冲反射回雷达时，根据离散傅里叶变换理论，两

个离散频率 ω_1 和 ω_2 在 $\Delta\omega = \omega_1 - \omega_2 > 2\pi/N$ 是可以分辨的。

由于 $\Delta\omega$ 也是由公式 $\Delta\phi = \frac{4\pi v T_c}{\lambda}$ 定义的，因而，当帧周期为 $(v_{res}) T_f = NT_c$ ，则可以用数学方法推导出速度分辨率 (v_{res}) ：

$$v > v_{res} = \frac{\lambda}{2T_f} \quad (2-18)$$

雷达的速度分辨率与帧时间 (T_f) 成反比。

2.3.3 毫米波雷达角度估算

毫米波雷达角度估算是一种利用目标距离细微变化引起的相位变化进行的技术，相位变化 $\Delta\phi = \frac{2\pi\Delta d}{\lambda}$ 通过幅度-傅里叶变换或多普勒-傅里叶变换的峰值来衡量。这种相位变化结果被用于进行角度估计，通常需要至少两个接收天线。

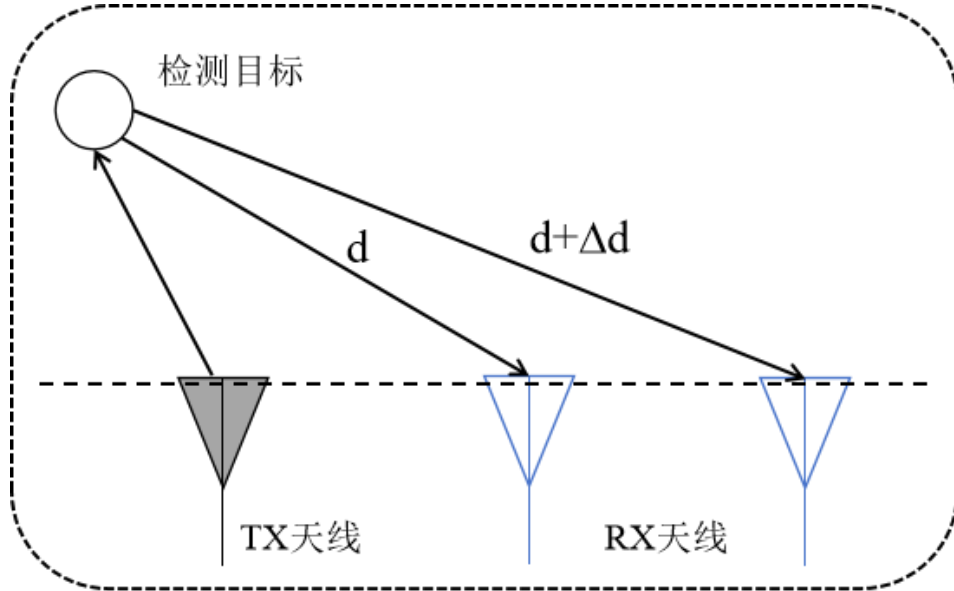


图 2-9 估算到达角需要两个 RX 天线

该方法利用两根接收天线之间的距离差 Δd ，使信号峰值的相位发生变化，为角度估计提供了必要信息。当目标距离远大于天线距离 l 时，从目标到达接收天线的信号和从发射天线到目标的信号是平行的。根据平面波的基本几何假设，可以推导出 $\Delta d = l \sin(\theta)$ ，其中 θ 是到达角。利用两个间隔 l 的接收天线进行角度估计的场景如图 2-9 所示，到达相邻双接收天线的信号所经历的相对延迟可以确定为：

$$\Delta\tau = \frac{l \sin(\theta)}{c} \quad (2-19)$$

其中， c 为光速。然后，基于公式（2-9）中的中频信号模型，在两个相邻的接收天线上接收到的中频信号之间的相位差为

$$\Delta\phi = 2\pi f_c \Delta\tau = \frac{2\pi l \sin(\theta)}{\lambda} \quad (2-20)$$

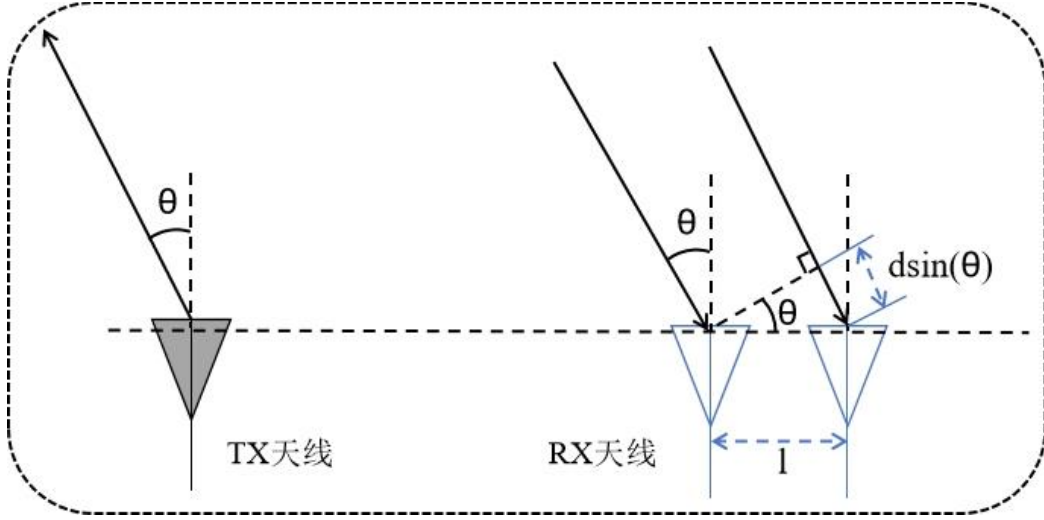


图 2-10 远场的角度估计

到达角(θ)可以通过测量的测向方向计算，表达式为：

$$\theta = \sin^{-1} \left(\frac{\lambda \Delta\phi}{2\pi l} \right) \quad (2-21)$$

其中， $\Delta\phi$ 取决于 $\sin(\theta)$ 。这被称为非线性依赖关系。仅当 θ 很小时， $\sin(\theta)$ 才能近似为线性函数，即 $\sin(\theta) \sim \theta$ 。因此，估计精度依赖于到达角（AOA），当 θ 值较小时，估计精度更准确。

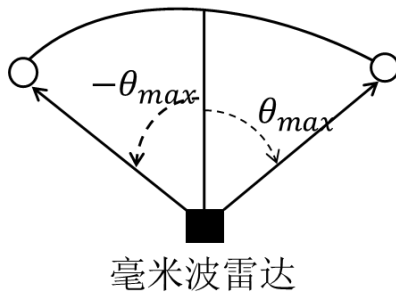


图 2-11 最大角视野

如图 2-11 所示，雷达的最大角视场由雷达能够估计的最大到达角度来定义。当两个接收天线的相位差 $|\Delta\omega| < 180^\circ$ 时，

$$\omega = \frac{2\pi l \sin(\theta)}{\lambda} < \pi \quad (2-22)$$

其中，天线之间的间距 $l = \lambda/2$ 导致最大角视场 $\pm 90^\circ$ 。公式（2-23）表示了

两个间隔为 l 的天线所能服务的最大视野

$$\theta_{max} = \sin^{-1}\left(\frac{\lambda}{2l}\right) \quad (2-23)$$

2.4 毫米波雷达数据集

目前，自动驾驶领域已经涌现出一些流行的数据集，如 KITTI^[63]或 Waymo^[64]。然而，这些数据集大多不包含带注释的毫米波雷达数据，而毫米波雷达数据在自动驾驶系统中至关重要。为了解决这一问题，一些研究者开始关注面向毫米波雷达的数据集，如表 2-1 可见，例如 Dense^[65]提供了与高分辨率视频相配对的数据，但其覆盖范围有限，且缺乏实时性。Zendar^[66]数据集以高帧率采集数据，提供了更多的时间序列信息，但缺乏丰富的标注信息，需要用户进行额外的标注或数据处理。虽然这些数据集为研究提供了宝贵的资源，但都存在数据集的覆盖范围相对较小的限制和缺陷。本文所提出方法实验分别在 NuScenes^[67]数据集和 RadarScenes^[68]上进行了训练和评估。

表 2-1 具有毫米波雷达的数据集

数据集	发布年份	RGB 图像	激光雷达	毫米波雷达	雷达传感器数量
KITTI ^[63]	2012	√	√	×	-
Waymo ^[64]	2019	√	√	×	-
NuScenes ^[67]	2019	√	√	√	5
Dense ^[65]	2020	√	√	√	1
Zendar ^[66]	2020	-	√	√	1
RadarScenes ^[68]	2021	√	-	√	4

NuScenes 数据集利用 4 个角位雷达和 1 个前置雷达获得毫米波雷达点云，是首个集成毫米波雷达数据的自动驾驶数据集，与车载相机同步采集不同场景帧的图片和点云信息。此数据集覆盖了多个城市场景，并包含了大量的目标实例，NuScenes 提供了 23 个类别的 3D 检测框以描述道路目标。

RadarScenes 数据集在超过 4 个小时的驾驶过程中利用 4 个汽车毫米波雷达传感器，采集到了涵盖 7000 多个移动道路目标的雷达数据，这些目标被分为 11 个不同类别，其中包括 5 个主要类别，如汽车、卡车、行人等。该数据集提供了雷达点的空间坐标（ x 、 y ）、目标速度（ v_x 、 v_y ）、雷达横截面（RCS）和时间戳(t)等信息，以支持目标检测、跟踪和运动预测等任务。

2.5 数据关联

数据关联在多模态数据融合中具有重要意义，在自车环境中将不同传感器提供的同一目标相关数据组合，由于毫米波雷达点云和相机图像都属于异构数据，两者之间的关联具有挑战性。本节介绍了两种数据关联方法，一种是对相机二维图像和毫米波雷达点云基本坐标进行转换，使二者位于统一坐标系中表示同一目标。另一种是将毫米波雷达点云投影到图像平面上，使数据在同一位置中进行融合。

2.5.1 点云坐标转换

为确保不同模态数据关联后在同一个坐标系中表示，首先，将点云转换为相机的传感器坐标系，以避免数据之间存在空间偏差。表示某个点的齐次坐标矩阵如下：

$$\hat{P}_R = [\hat{P}_{1,R}, \hat{P}_{2,R}, \dots, \hat{P}_{N,R}] \in \mathbb{R}^{4 \times N} \quad (2-24)$$

其中，R表示雷达的坐标系，而 $\hat{P}_{i,R}$ 的第i列表示第i点的坐标向量：

$$\hat{P}_{i,R} = \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \in \mathbb{R}^4 \quad (2-25)$$

转化后表示为

$$\hat{P}_C = T_{CR} \hat{P}_R \quad (2-26)$$

其中，C是相机的坐标系。转换矩阵 T_{CR} 定义如下：

$$T_{CR} = \begin{bmatrix} R & t \\ 0_{1 \times 3} & 1 \end{bmatrix} \in SE(3) \quad (2-27)$$

公式（2-27）描述了雷达和相机坐标系之间的转换关系。 T_{CR} 由旋转矩阵 $R \in SO(3)$ 和平移矩阵 $t = [t_x, t_y, t_z]^T \in \mathbb{R}^3$ 组成。 T_{CR} 是特殊欧几里群 $SE(3)$ 中的一个 4×4 矩阵，而R是特殊正交群 $SO(3)$ 中的一个 3×3 矩阵^[69]。其中，毫米波雷达点云中的点由附加信息组成，如雷达截面积 ρ 值不需要进行坐标转换，大部分雷达传感器只输出一个平面点云；再者，点云中的径向速度不需要平移，只需要使用旋转矩阵R将其旋转到正确的坐标系中。

2.5.2 点云的二维投影

另一种常见的方法是将毫米波雷达数据投影到图像平面上，随后在相同的

位置对数据进行匹配。由于雷达点云缺乏高度信息，因此，在将单个点投影时，其位置不会精确地映射图像平面上。对此，Nobis^[70]为了完成相机图像和毫米波雷达数据的关联，从鸟瞰图地面开始，假定为每个雷达点分配了一个固定的目标高度，并将雷达点云特征尤其是雷达截面积和径向速度投影到图像平面的二维线段上，所得到的特征图像与原始输入图像具有相同的分辨率，并且在投影的二维线段上包含点云的特征值，而在其他位置该值为零。最后将雷达特征图像与原始图像的三个 RGB 通道连接起来，从而形成毫米波雷达和相机之间的数据关联。

2.6 本章小结

本章首先系统地阐述了毫米波雷达的基本工作原理，涵盖了雷达信号的波形特点、工作机制以及发射与接收过程；推导距离测量，速度测量和角度测量的相关公式，确定目标的位置和运动状态，证明了毫米波雷达在目标检测领域中的优势。接着深入分析了包含毫米波雷达点云的各类数据集的特征，这些数据集为训练和验证雷达系统的性能提供了必要的样本，为后续章节实验中数据集的选择进行了铺垫，最后详细介绍了毫米波雷达点云和相机图像数据关联的方法，包括点云坐标的转换和二维投影到图像上的技术，通过融合毫米波雷达数据与视觉信息，可以更精确、高效的完成目标检测和分类任务。

第3章 基于毫米波雷达点云目标检测方法 RPCNET

3.1 引言

近年来，自动驾驶汽车和高级驾驶员辅助系统（ADAS）致力于最大限度地减少人为错误，以更果断有效地做出关键决策。目前，大多数环境感知算法和目标特征提取技术依赖于高分辨率、多通道的激光雷达来为动态目标构建精确的3D检测框。激光雷达作为一种基于光的传感器，通过测量光信号的反射来感知场景的几何形状，从而产生密集的三维点云^[71]。然而，激光雷达受限于不利天气条件，无法穿透雾和灰尘，这对传感器检测性能产生了一定影响。

相比之下，毫米波雷达是一种全天候的传感选择方案，与激光雷达不同的是，毫米波雷达可以不受恶劣天气条件影响的传输毫米波，其波长可以很容易的穿透雾、灰尘和其他微观粒子^[72]。毫米波雷达生成的数据是稀疏的，其点云主要包含来自动态目标（比如车辆，行人等）和其他静态目标（如路障，指引牌等）。然而，毫米波雷达在点云生成方面与激光雷达相比无法像其产生密集的点云。为了在不利的天气条件下实现高质量的感知性能，必须充分利用仅有的稀疏点云信息，然而，由于目标在场景中的数量和位置未知，仅凭稀疏点云作为输入是一大挑战。

为了解决这一挑战，本章提出了一种基于毫米波雷达点云检测的神经网络RPCNET，旨在有效检测多个动态目标，并提供在场景中的精确位置、方向和3D维度，通过三维检测框提供高保真度的感知。为达成此目标，在本章定义了一组允许在稀疏点云中迭代所有边界框的独特的锚框，这些大小固定的锚框被用作最终3D检测框的初始估计，由此，对于网络检测而言不再使用整个场景的点云，而是在每个锚框上单独进行操作。对于每个锚框而言任务被简化为生成一个置信度，指示该锚框内的点是否属于一个目标。接着采用DGCNN网络提取毫米波雷达点云特征，从而确保稀疏雷达点云中的所有信息都能被有效利用，避免丢失点云分布结构信息。值得注意的是，RPCNET即使在恶劣天气下也能实现这种感知，为雷达成为自主感知的主流传感器提供了思路。

本章具体组织内容结构如下：3.1节引言由毫米波雷达的检测性能引出本章

所提出的 RPCNET 网络；3.2 节详细介绍 RPCNET 网络总体框架，该网络从点云无序性特点出发，利用 DGCNN 网络提取点云特征，生成 3D 检测框；3.3 节分别介绍了实验所需的 NuScenes 和 RadarScenes 数据集的评价指标；3.4 节对 RPCNET 网络进行了实验和结果分析，从定性和定量分析的角度评测本算法的同时并给出了与已有方法的对比结果；3.5 节对本章研究进行总结。

3.2 RPCNET 网络总体架构

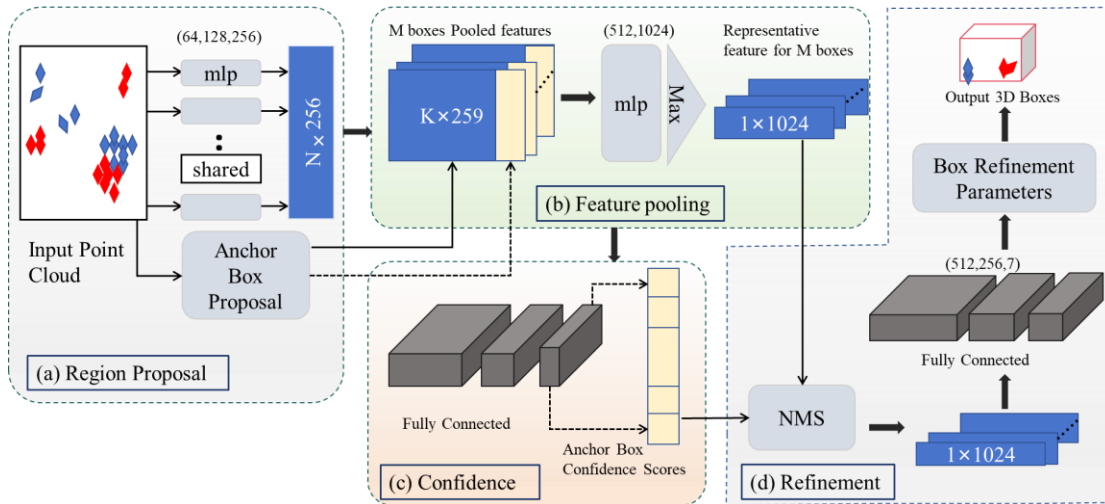


图 3-1 RPCNET 的网络框架

本章介绍了一项称为 RPCNET 的创新深度学习架构，旨在根据场景中原始毫米波点云数据精准估测的 3D 检测框。RPCNET 包括四个阶段，如图 3-1 所示：

（1）候选目标区域生成，该阶段利用 RRPN 网络实时生成候选区域建议；（2）点云特征提取池化，本阶段采用 DGCNN 有效地从点云中提取信息；（3）检测框置信度的预测，旨在选择高置信度的锚框，其中包含属于目标的所有点；（4）确定最终的 3D 检测框。

3.2.1 候选目标区域生成

由于传统方法无法解决场景中目标数量和位置的不确定性，因此，通常情况下，目标检测器使用一组固定的初始区域建议来应对这个问题。Nabati 等人提出的 RRPN 网络，根据相对位置信息即目标与其周围环境的关系，实时生成区域建议（即锚框），通过利用雷达检测结果生成目标建议。RRPN 网络的优势在于利用了这种相对性的考虑，使得网络能够更好地理解目标在空间中的位置关系，从而提高了目标建议的准确性。

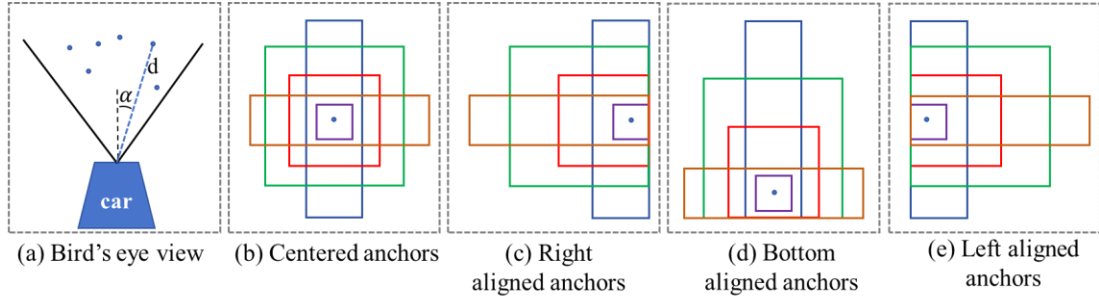


图 3-2 以雷达点为中心的不同尺寸和纵横比的边界框

其中蓝色点为每个雷达探测生成的不同形状和大小的锚点。

在检测过程中一旦雷达探测结果被映射到图像坐标，便能获得每个被检测目标在图像中的近似位置。这些映射的雷达探测结果，以下统称为感兴趣点，提供了关于每幅图像中的目标的价值信息。在给定感兴趣点的情况下，可以使用五种不同位置的锚框，而不需要对图像本身进行任何处理，来提出感兴趣区域。雷达映射过程请参见图 3-2。本节使用 Adam 优化器，学习率为 0.0002，动量为 0.8。在网络的每一层之后都有一个批量处理归一化层。本节使用初始化锚框的尺寸 (2m, 3m, 7m) (宽度，高度，长度)。点云中随机采样 50 个点，以提供网络输入。该网络需要 6 个通道的输入，包括 xyz 坐标、目标速度、雷达横截面和置信度估计。

3.2.2 点云特征提取池化

本节旨在通过学习点云数据的有意义特征表示，以进行分类和三维边界框参数回归。为实现此目标，RPCNET 在锚框生成之前和生成之后的两阶段进行特征提取。

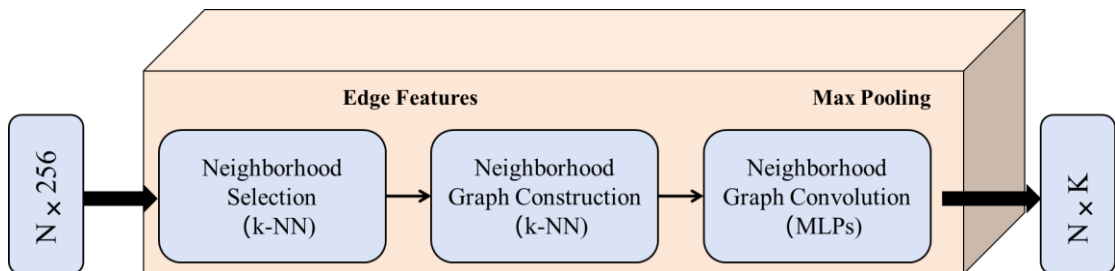


图 3-3 边缘卷积结构

在第一阶段，本算法使用 DGCNN 网络架构从整个点云中提取特征，如图 3-3 所示。DGCNN 融合了 PointNet 和图卷积神经网络方法，作为其主干，接收整个点集作为输入，并为每个点返回一个语义标签。DGCNN 将点云描述为使

用 k 近邻 (k -Nearest Neighbors, k -NN) 在潜在空间中动态构造的链图。主要过程分为图的初始化, 图的构建, 图的边缘化生成以及确定锚框。

(1) 图的初始化

本节定义了一组 N 个无序排列的点云, 其集合表示为 $P = \{p_1, \dots, p_N\}$, 每个点 $p_i = (c_i, m_i)$ 由三维笛卡尔坐标向量 $c_i = (x_i, y_i, z_i)$ 表示, 以及点附加的额外属性状态向量 $m_i \in \mathbb{R}^k$, 其中 m 可以包含传感器测量特性, 如雷达横截面, 速度以及新生成的特征, 例如连接到每个节点的边数。特征向量越丰富, 对模型性能的影响就越好。

(2) 图的构建

在构造图时, 本节利用这些点的坐标来直接生成特征向量。将点表示为图上的一个节点, 因此 p_i 可以看作是该节点的信号, 因此使用 $c_i = (x_i, y_i, z_i)$ 来构造该图。首先, 计算每个中心点的 k 个最近邻点, 即 k -NN。第 i 点和第 j 点的距离表示为 $A_{i,j} = \|c_i - c_j\|^2 = d(i,j)^2$, 这构成了邻接矩阵 $A \in \mathbb{R}^{k \times k}$ 。给定一个点云 P , 构造全连通图 $G = (\mathcal{V}, E, \mathcal{W})$ 如式(3-1)所示:

$$E = \{(i,j) | d(i,j) < r\} \quad (3-1)$$

本节中取 $r = 1.5m$, 并使用高斯函数作为顶点 i 和顶点 j 之间的权值函数如式(3-2)所示:

$$w_{ij} = \begin{cases} \exp\left(\frac{-d(i,j)^2}{2\sigma^2}\right), & (i,j) \in E \\ 0, & otherwise \end{cases} \quad (3-2)$$

其中, $d(i,j)$ 是点 i 和点 j 之间的距离。需要指出的是, 本节提出的算法可支持根据局部几何点云结构而变化的图拓扑。

(3) 图的边缘化生成

DGCNN 使用边缘卷积技术 (EdgeConv), 该模块在局部邻域图中的特征描述了点与其相邻点的连接特征, 以处理点云的局部结构。这确保了其结构的交换和旋转不变性, 有效地从点云数据中提取特征信息。

基本上, G 是 $\mathcal{K}^N$ 中点 $P = \{p_1, \dots, p_N\}$ 的 k -NN 图, 首先, 每个点 p_i 的 k 边特征 $E_{i,j}$ 如式(3-3):

$$E_{i,j} = h_{\theta}(p_i, p_j) \quad (3-3)$$

其中 p_j 是 p_i 的一个邻域点之一， h_Θ 是一个具有可学习参数 Θ 的非线性函数。然后，利用非对称函数式 h_Θ 如式(3-4):

$$h_\Theta(p_i, p_j) = \bar{h}_\Theta(p_i, p_j - p_i) \quad (3-4)$$

表示全局形状结构和局部邻域信息。最后，边缘卷积通过对每个顶点执行通道对称操作，通过聚合所有与顶点 p_i 相关的边缘特征来计算输出的 p'_i :

$$p'_i = \max_{j:(i,j) \in \mathcal{E}} \text{ReLU}(\Theta_M \cdot (p_j - p_i) + \phi_M \cdot p_i) \quad (3-5)$$

其中， $\Theta = \{\theta_1, \theta_2, \dots, \theta_M, \phi_1, \phi_2, \dots, \phi_M\}$ 和 $\cdot$ 表示欧几里得内积。

如图 3-3 边缘卷积结构所示，有三个主要的模块进行图的生成和卷积的发生：邻域选择、邻域图构造和邻域图卷积。邻域选择模块的目的是通过 k-NN 图算法为中心点选择 k 个最近点。在邻域图构造模块中，边缘卷积使用动态图构造来生成定义点与其邻点之间连接的边特征。为了更新特征向量的中心，利用多层感知机（MLP）实现了图的卷积运算。

（4）确定锚框

在第二阶段，针对每个点确定了锚框。网络中的感兴趣区域（RoI）特征池化模块对位于每个锚框内的所有点特征进行了池化。在特征池化阶段，从每个锚框内的所有点中随机抽取 42 个点。若点数小于 42，则进行点的重复取样，以确保在整个网络操作中的一致性。这些特征通过另一个卷积层处理，然后通过最大池化操作得到每个场景中定义的锚框代表性特征。

在这两阶段的设计使得 RPCNET 能够高效地从点云数据中提取丰富的特征，为后续的分类和边界框回归任务提供高质量的输入。

3.2.3 检测框置信度预测

在前一个步骤后，通过全连接层组成的分类网络对锚框的整体代表性特征集进行处理。这些全连接层学习了从锚框的代表性特征到每个检测框的置信度值的映射。通过对基于感兴趣区域（ROI）的最大池化特征进行分类，确保来自锚框内所有邻域点的上下文信息，从而提高了分类结果的准确性。直接在锚框上进行分类解决了分割的问题。网络的目标是高可信度地选择相应的锚框，其中包含所有属于一个目标的点。

3.2.4 确定检测框

在先前的步骤中，使用固定大小的锚框进行了对最终 3D 边界框的尺寸、中心和方向的粗略估计。然而，进一步细化这些参数对于准确估计检测框的尺寸和位置至关重要。在这个过程中，针对点云内的每个点进行了边界框的预测，并得到了相应的置信度分数。通过采用非极大值抑制（NMS）的方法^[73]选择高置信度框，并删除与同一目标的其他高度重叠的检测框，以优化结果，这确保了 RPCNET 算法最终输出的边界框具有代表性。

为了训练分类网络，基于锚框对地面真实值边界盒的交集联合（IoU）值生成了地面真实值。使用锚框的鸟瞰图来计算二维 IoU。根据 IoU 匹配选择前 100 个锚框进行分类。将二维 IoU 重叠大于 0.2 的边界框被认为是分类的正例。对于细化网络，使用预测置信分数来消除二维 IoU 重叠超过 0.5 的框，并采用非极大值抑制，同时对区域建议和细化网络同时进行训练。然而，在训练的初级阶段，由于分类头无法提供正确的置信度分数；本节使用具有地面真实框的二维 IoU 分数作为非极大值的置信值。

考虑细化过程，通过三个完全连接的层输出一个元组 $[h', w', l', x', y', z', \theta']$ ，该元组分别对应于边界框的长度、宽度、高度、中心坐标和方向角。这些细化后的特征被添加到锚定框的参数中，从而生成最终的三维边界框预测。这一系列改进确保了检测结果在尺寸和位置上更为准确。

3.2.5 损失函数

在基于毫米波雷达点云进行目标检测领域，损失函数的设计至关重要。损失函数的目标是在训练过程中有效地引导模型学习正确的特征表示和边界框细化，以实现准确的目标检测结果。

首先，考虑到雷达点云特征提取阶段的分类和回归任务。对于分类任务，采用交叉熵损失函数 $\mathcal{L}_{cls}$ ，用于衡量目标类别的预测准确性。具体而言，将每个点云按照标注的边界框进行分类，将标注类别的点予以正确分类，而将其他点分类为背景类。因此，分类损失 $\mathcal{L}_{cls}$ 包括了杂波预测任务和类别预测任务的交叉熵损失之和：

$$\mathcal{L}_{refinement}(r, r') = \begin{cases} \frac{1}{2}(r, r')^2, & \text{for } |r - r'| < 1 \\ \delta|r - r'| - \frac{1}{2}, & \text{otherwise.} \end{cases} \quad (3-6)$$

其中， r 和 r' 分别为每个参数的地面真实值和回归细化值 $h', w', l', x', y', z', \theta'$ 。

其次，针对边界框参数的细化任务，视为回归问题，本节采用 Smooth- L_1 损失来衡量预测边界框参数和真实值之间的差异。Smooth- L_1 损失函数在减少异常值的影响同时保持了较好的平顺性，这有助于提高模型的稳定性。同时以防止模型过拟合^[74]引入了 L2 正则化项 $\mathcal{L}_{reg}$ ，总体损失函数表达式如下：

$$\mathcal{L} = \mathcal{L}_{cls} + \mathcal{L}_{refinement} + \gamma \mathcal{L}_{reg} \quad (3-7)$$

3.3 实验环境及评价指标

在本节中，对所提出的RPCNET方法分别在 NuScenes 数据集和 RadarScenes 数据集上进行了评估，并进行了广泛的实验，以验证所提出方法的有效性。

3.3.1 实验环境

本文的实验环境见表 3-2，使用 Python3.8+Pytorch1.10.0+Cuda11.3 版本进行实验。

表 3-2 实验环境

开发环境	环境配置
CPU	Platinum 8255C
GPU	RTX 2080 Ti
Cuda	11.3
运行系统	Ubuntu 20.04
开发语言	Python 3.8
机器学习库	Pytorch 1.10.0

3.3.2 NuScenes 数据集评价指标

(1) 平均精度指标(Average Precision, AP)

AP 通过计算超过 10%的精度-召回率曲线下的归一化面积^[67]。在计算 AP 时，去除召回率或精度小于 10%的操作点，以减少精度和召回率区域常见的噪声影响。若某个目标类别在特定的距离范围内没有监测点，则该类的 AP 值为零，对

不同距离 $\mathbb{D} = \{0.5m, 1m, 2m, 4m\}$ 和 $\mathbb{C}$ 类别的匹配阈值进行平均，得到 mAP 公式表示为：

$$\text{mAP} = \frac{1}{|\mathbb{C}||\mathbb{D}|} \sum_{c \in \mathbb{C}} \sum_{d \in \mathbb{D}} \text{AP}_{c,d} \quad (3-8)$$

(2) 真正度量指标(True Positive metrics TP)

TP 衡量了与地面真实框匹配预测的准确性，在在匹配过程中使用 $d = 2m$ 中心距离计算的，并被设计为正标量。若某个特定类别的召回率未达到 10%，则该类的所有 TP 错误都设置为 1。TP 错误包括：平均平移误差（ATE），平均比例误差（ASE），平均方向误差（AOE），平均速度误差（AVE）和平均属性。对于每个 TP，计算所有类的平均 TP（mTP），如式(3-9)：

$$\text{mTP} = \frac{1}{|\mathbb{C}|} \sum_{c \in \mathbb{C}} \text{TP}_c \quad (3-9)$$

(3) NuScenes 检测分数(NuScenes Detection Score, NDS)

NDS 综合考虑目标检测的各个方面，包括位置、大小、方向、属性和速度的估计，其计算公式(3-10)：

$$\text{NDS} = \frac{1}{10} \left[5\text{mAP} + \sum_{\text{mTP} \in \text{TP}} (1 - \min(1, \text{mTP})) \right] \quad (3-10)$$

NDS 通过结合平均精度指标（mAP）和真正度量（TP）来量化目标检测的质量。

3.3.3 RadarScenes 数据集评价指标

RadarScenes 数据集中的目标检测质量是在其并集（Intersection over Union, IOU）阈值为 0.4 的 mAP 值测量的。另一个常用的评价指标是对数平均错过率（Log-Average Miss Rate, LAMR）^[75]，尤其适用于小目标。LAMR 旨在引导检测器将更多的关注放在召回率上而非精确度。在雷达领域另一常用的度量标准是类别平均 F1 分数，其计算方式类似于 mAP，但是在计算单个类别 F1 分数之后进行宏观平均。此外，本实验报告了。考虑到使用所有 11 个目标类别可能导致训练和验证集的不平衡，建议采用映射的标签，即主要涵盖 5 个主要目标类别的标签，包括行人、人群、自行车、汽车和大型车辆类别。

3.4 实验结果与分析

3.4.1 在 NuScenes 数据集上的实验结果与分析

(1) 定量分析

在构建 RPCNET 网络架构时，特别注重于利用 DGCNN 网络来提取毫米波雷达点云数据的关键特征，这种独特网络的设计不仅提高了特征提取的效率和准确性，而且在 NuScenes 数据集的广泛基准测试中，显示出了比现有的毫米波雷达点云检测方法更优的性能。

对比分析显示，相较于最新的 Radar-PointGNN 方法，本章所提的 RPCNET 方法在多项误差指标上表现更为出色，并且在 NDS 方面有所提升。具体而言，如表 3-3 所示的 RPCNET 方法在 NuScenes 测试集上不同方法的性能指标，其中本章所提的方法在 NDS 为 0.37，mATE，mASE，mAOE 和 mAVE 分别为 0.89，0.74，0.84 和 0.92，相比于另外两种方法，误差更小。

表 3-3 在 NuScenes 测试集上评估的性能指标

Methods	NDS(10^{-1})	mATE	mASE	mAOE	mAVE
PointNet ^[7]	0.20	1.34	1.09	1.22	1.34
Radar-PointGNN ^[28]	0.34	1.02	0.86	0.90	1.02
RPCNET(ours)	0.37	0.89	0.74	0.84	0.92

根据表 3-4 数据显示，本章所提出的网络在 NuScenes 数据集上的实验相比于 PointNet 和 Radar-PointGNN 网络表现更为出色，实验结果表明，RPCNET 在各种目标类别的平均检测精度方面都有显著提升，特别汽车的 AP 达到了 58.8%，相较于 PointNet 网络，行人，卡车，公交车和栅栏的 AP 值分别提升了 3.1%，9.6%，6%和 9.5%。

表 3-4 在 NuScenes 测试集上实验各类平均检测精度

Methods	AP(%)					
	Car	Truck	Bus	Pedestrian	Barrier	Motorcycle
PointNet ^[7]	48.2	15.0	10.3	8.4	33.7	8.5
Radar-PointGNN ^[28]	55.7	18.9	14.5	10.9	41.3	10.4
RPCNET(ours)	58.8	24.6	16.3	16.1	43.2	12.2

这些结果进一步验证了 RPCNET 方法在目标检测任务上的优越性，尤其是

在识别不同类别目标方面具有更强的性能。在 NuScenes 数据集上的实验结果突显了该方法在处理点云数据上的优越性，为各种目标类别的准确检测提供了有力支持。

（2）定性分析

在 NuScenes 测试数据集上的实验结果如图 3-4 所示，图中展示了本章所提出的方法在实际场景中生成的 3D 检测框。蓝色框代表地面真实存在的目标数据，而橙色框则是所提出模型预测出的结果。本节特意选择了三个典型场景进行实验分析，前两列描绘的是在晴朗天气情况下的主干道上的情况，第三列则聚焦于光线昏暗的弯道驾驶场景。这种情况下，对于传统视觉系统可能会面临挑战，因为光线不足会严重影响目标检测的准确性，然而，RPCNET 网络展现出了良好的适应性和鲁棒性，通过仔细比较算法的预测框与地面真实检测框的差异，可以发现即便在不利条件下（例如光线较弱），RPCNET 模型也能有效地完成任务，并且预测的误差较小。这些实验结果进一步证明了本章所提出的方法在处理复杂环境时的可靠性，并且揭示了其在特定条件下的优越性能。

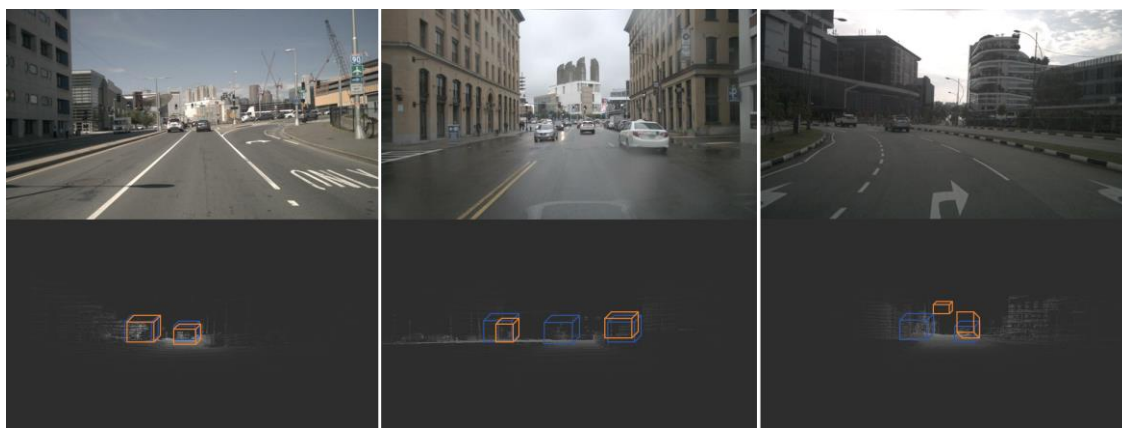


图 3-4 在 NuScenes 数据集实验结果

其中，蓝色框是地面真实数据，橙色框是模型预测结果。

3.4.2 在 RadarScenes 数据集上的实验结果与分析

（1）定量分析

在根据 RadarScenes 数据集上的实验结果，RPCNET 方法再验证集上取得了显著的成绩，其 mAP 值达到了 62.1%，见表 3-5。具体到不同目标类别的表现，自行车、汽车和大型车辆类别的 AP 值表现最为突出，分别为 67.6%，74.5%和

72.7%。尽管行人类别的 AP 值相对较低，为 35.7%，然而与 YOLOV3 和 RadarGNN^[76]相比，RPCNET 在各个类别的精度均呈现出明显的提升。

表 3-5 在 RadarScenes 数据集上的目标检测结果

方法	AP _{行人}	AP _{人群}	AP _{自行车}	AP _{汽车}	AP _{大型车}	mAP
YOLOv3 ^[23]	34.5	54.2	55.4	68.2	65.9	55.6
RadarGNN ^[76]	34.0	57.2	64.5	71.4	68.9	59.2
RPCNET	35.7	60.1	67.6	74.5	72.7	62.1

根据表 3-5 表明 RPCNET 方法在目标检测任务中取得了较为全面和优越的性能，特别是在自行车、汽车和大型车辆的识别方面表现卓越。相较于传统方法 YOLOV3 和 RadarGNN，RPCNET 在各类目标上都取得了更为令人满意的结果，突显了其在处理雷达场景数据中的有效性和优越性。

(2) 定性分析

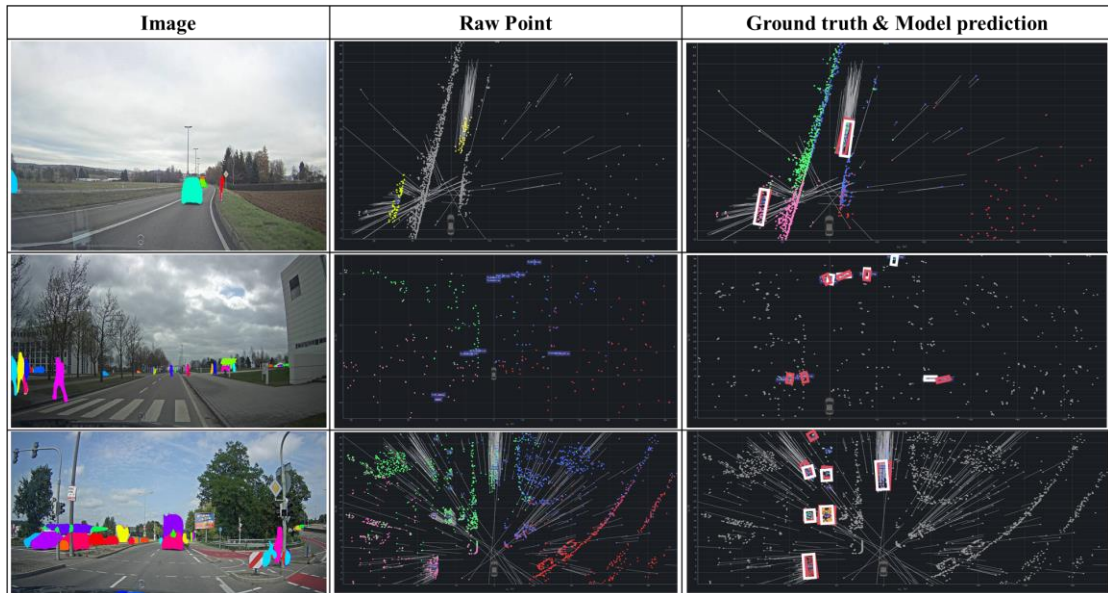


图 3-5 RPCNET 定性分析。左边显示的是 RadarScenes 图像，数据集使用不同颜色来表示被检测到的目标。中间为鸟瞰图中对应的可视化原始雷达点云数据，并用不同颜色表示不同目标类别并有标签。最右侧显示的是标注的地面真实数据如红色框所示，而白色框是表示目标类别和边界框的模型预测。

针对图 3-5 和图 3-6 的定性分析，为了更进一步的证明本文所提网络的有效性，本节对 RPCNET 模型的 3D 目标检测结果进行了可视化，如图 3-5 展示了在乡村道路、园区内以及交通枢纽等不同场景下的检测效果。图 3-5 中展示了 RadarScenes 数据集上的雷达点云数据，并在鸟瞰图中展示了本章所提出的方法检测到的地面真实数据（最右侧图中的红色框表示）和模型预测的目标类别和

边界框。

具体而言，从图 3-5 的第一行在乡村小道上的场景中可见，存在行人和车辆等多样化目标，RPCNET 显著提高了三维目标检测的精度，展现了对复杂场景的强大的适应性。第二行在园区内的场景中，主要关注行人和人群类别检测，结果显示本章提出的方法在这类别上取得了令人满意的结果。最后，从图 3-5 的第三行在繁华的交通枢纽场景中，包括了公交车、卡车、汽车和行人等多样化目标，RPCNET 方法同样表现出高度的检测准确性。这些不同场景的测试进一步证明了本章所提方法的鲁棒性和普适性。

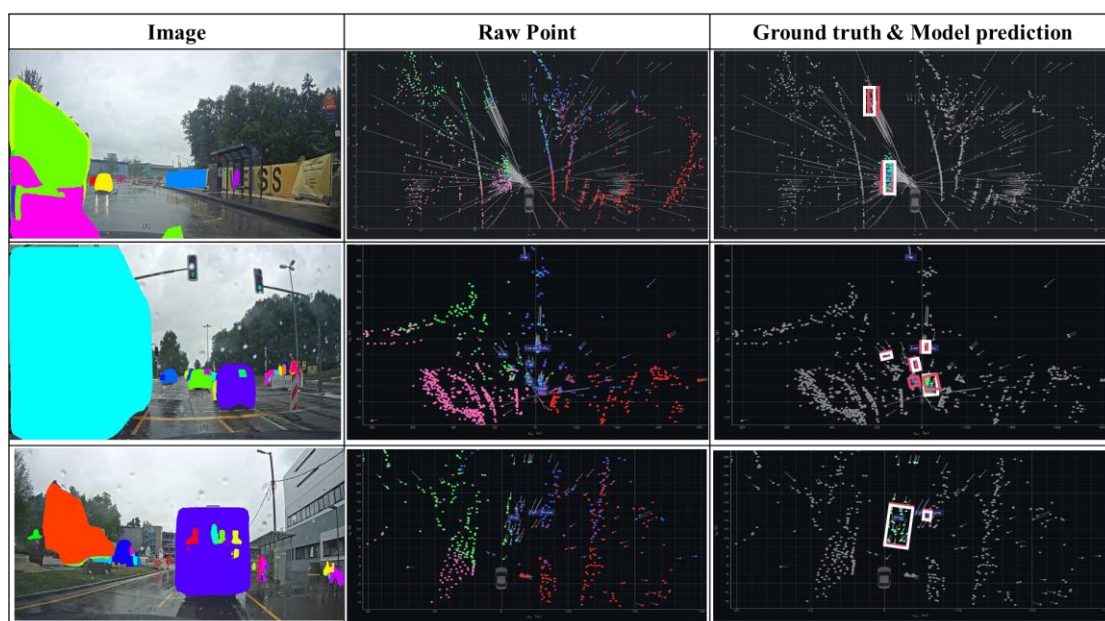


图 3-6 雨天情况下使用 RPCNET

图 3-6 展示了 RPCNET 模型在雨天场景下的 3D 目标检测结果，包括十字路口和城市街道等雨天复杂情况，结果显示，在雨天环境中，本网络仍然取得良好的检测效果。图 3-6 中第一行和第二行是在高流量和高密度车辆的十字路口和城市街道场景中，本章所提出的方法展现了较好的检测效果，进一步证明了其在雨天复杂环境中的稳健性和有效性。从图 3-6 的第三行中可以看出，本章所提的方法在十字路口和城市街道这种人流量大、高密度车辆下仍具有较好的检测效果，进一步证明了其在雨天环境中的稳健性和有效性。这个结果有助于解决雨天缺乏光照的复杂交通背景下的三维目标检测问题，提高了车辆目标检测的识别精度高并且鲁棒性强。

3.5 本章小结

本章提出了一种基于毫米波雷达点云的动态目标 3D 目标检测，即 RPCNET 网络结构，充分利用了毫米波雷达能够获得目标位置、方向和角度信息的优势，并且不受恶劣天气光线影响。通过在具有挑战性的 NuScenes 和 RadarScenes 数据集上对进行评估，实验证明，RPCNET 方法在 RadarScenes 数据集上的 mAP 值达到了 62.1%。在自行车、汽车和大型车辆这些类别中，AP 值表现尤为突出，分别为 67.6%，74.5%和 72.7%。尽管行人类别的 AP 值相对较低，为 35.7%，然而与 YOLOV3 和 RadarGNN 相比，RPCNET 在各个类别的精度均呈现出明显的提升。在 NuScenes 数据集中汽车的 AP 达到了 58.8%，即使在光线较弱的环境下，本章的方法仍能够准确预测，与地面实际检测框的误差相对较小。这表明本章所提的方法在复杂环境中表现出色，尤其在光照不足的情况下具有强大的鲁棒性。这些观察结果进一步支持了本章所提的方法在不同场景下的可靠性和适应性。

第 4 章 视觉和毫米波雷达融合的 3D 目标检测算法

CenRadfusion

视觉-毫米波雷达数据融合已成为当前三维目标精准检测的主要解决方法。本章专注于相机图像和毫米波雷达点云的融合，提出了一种改进的融合方法——CenRadfusion。该方法在经典的 CenterFusion 网络基础上进行了改进和优化，充分利用了毫米波雷达和相机数据的融合特征，以实现三维目标的精准检测。首先，为了保证融合体系结构的完整性，在系统检测初期将毫米波雷达点云投影到图像平面上，并将其作为附加通道添加到 CenterNet 图像检测网络的输入中，从而初步形成了三维检测框。其次，针对毫米波雷达点云稀疏且不均匀，采用毫米波雷达点云密度聚类技术，对毫米波雷达点云进行聚类，以获得标签，过滤掉无关点云和白噪声，从而提高数据质量和目标检测的可靠性。接着，为了增强特征的方向性，添加一个通道注意力模块（Squeeze-and-Excitation Networks, SeNet）对每个特征通道进行加权，以增强网络的重要特征。最后，在 NuScenes 数据集上进行了实验，实验结果表明，与原始 CenterFusion 算法相比较，对汽车，卡车和摩托车的检测 AP 值分别提高了 7.8%，5.5%和 5.4%。

4.1 CenterFusion 目标检测网络概述

Nabati 等人提出了 CenterFusion 算法，如图 4-1 展示了该方法的网络体系结构。在第一阶段，首先利用基于相机的目标检测方法 CenterNet 来预测 2D 边界框和一个初步的 3D 边界框。然后，将这两个边界框都用于在整个雷达点云中对应的雷达点与目标相关联。在第二阶段相关的雷达点被用作网络的第二阶段的附加输入，在第二阶段根据相关雷达点信息的引入调整 3D 边界框生成。利用相机参数创建视锥。视锥关联方法用于关联来自这两种传感器模式的数据。最初，使用基于图像的检测网络体系结构进行目标预测。该网络的输出包括 2D 和 3D 边界框，但 3D 边界框包含目标的深度和方向信息。由于雷达点云缺乏高度信息，CenterFusion 网络将雷达点扩展成线，以补充雷达点的高度信息。

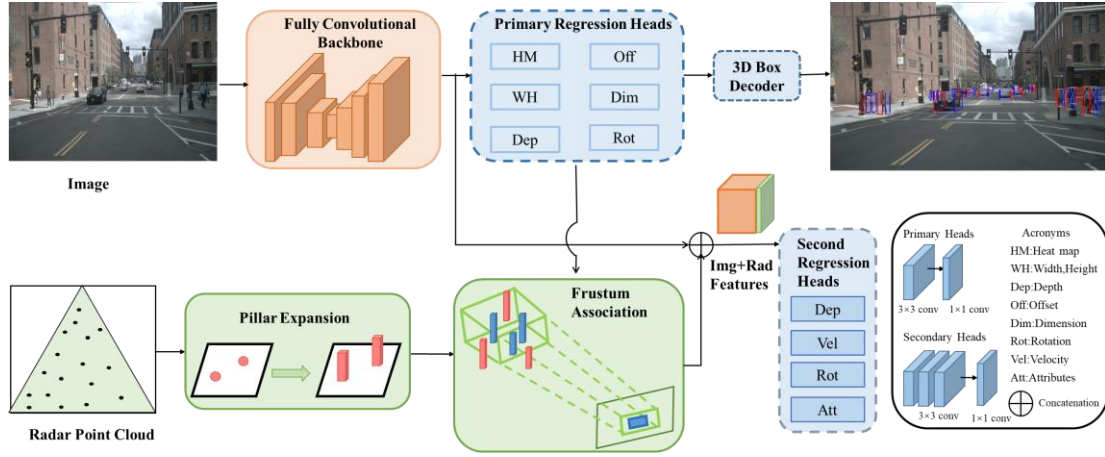


图 4-1 CenterFusion 网络框架

本节阐述了 CenterFusion 中利用中心点检测目标的 CenterNet 网络以及基于视锥的雷达点关联两种主要思想。首先，在第 4.1.1 节中介绍了 CenterNet 的方法，该方法用来创建图像的特征图，并进行初级头检测中 2D 边界框的生成和初步 3D 边界框参数的回归。随后，在第 4.1.2 节中，基于在点云和图像数据视锥关联基础，解释了如何将相关雷达点与特征图的融合。

4.1.1 CenterNet 网络

CenterNet 是一种无锚框（anchor-free）的目标检测网络，相比于依靠大量锚框的检测网络如 YOLO^[77]，SSD^[78]和 Fast R-CNN^[79]，CenterNet 网络通过创建图像特征并且解释第一阶段检测初步回归数据来实现基于相机的目标检测，其目的是为了输出关键点的热力图，将目标表示为单个点来检测，每一张热力图上的坐标若存在目标中心点，则在该坐标处产生一个关键点。通过回归到所有目标的宽度，高度，三维位置，中心偏移量等其他目标属性来得到实际的目标检测框。

对于 2D 和 3D 目标检测训练网络，CenterNet 首先将图像 $I \in \mathbb{R}^{W \times H \times 3}$ 输入网络，以获得关键点的热力图 $\hat{Y} \in [0,1]^{\frac{W}{R} \times \frac{H}{R} \times C}$ ，其中 C 是关键点类型的数量，R 是输出步幅，输出步幅通过因子 R 向下采样输出预测，而 W，H 分别是输入图像的宽度和高度。当预测值 $\hat{Y}_{x,y,c} = 0$ 时，则表示该位置为背景，当预测值 $\hat{Y}_{x,y,c} = 1$ 代表检测到图像上 (x,y) 为中心点的 C 类目标的关键点。接着，CenterNet 利用不同的全卷积编解码器网络从图像 I 中预测 $\hat{Y}$ ，包括堆叠沙漏网络（Stacked Hourglass Network）^[80]，上卷积残差网络（Up-convolutional Residual Networks,

ResNet) 以及深度聚合网络 (Deep Layer Aggregation, DLA)^[81]。在 CenterFusion 中以及在本章工作中, 只有 DLA 网络架构被用于实验。

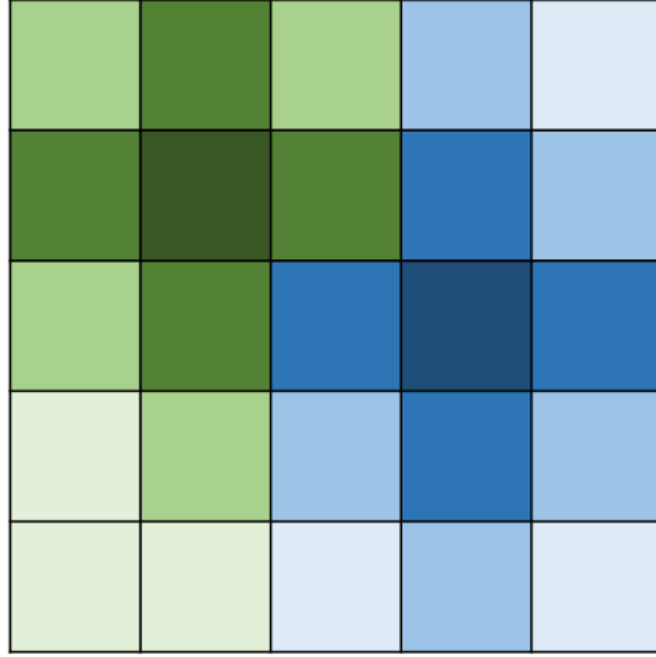


图 4-2 高斯核像素网格图

如图 4-2 所示, 地面真实热力图 Y 显示了一个像素网格, 其中一个类别的两个中心点分别是深绿色和深蓝色, 像素网格的其余部分以中心点的颜色显示, 其对应的高斯核在每个像素处具有最大值。热力图中的每个像素都有一个确定的目标中心点值。

对于图像上的属于 C 类的地面真实值关键点 $p \in R^2$, 计算一个低分辨率的等效值 $\tilde{p} = \left\lfloor \frac{p}{R} \right\rfloor$, 使用高斯核公式(4-1)将所有地面真实关键点分割到一个热力图 $Y \in [0,1]^{\frac{W}{R} \times \frac{H}{R} \times C}$ 。

$$Y_{xyz} = \exp \left(-\frac{(x-\tilde{p}_x)^2 + (y-\tilde{p}_y)^2}{2\sigma_p^2} \right) \quad (4-1)$$

其中, σ_p 是目标大小自适应标准差, 根据目标大小来控制每个目标的热力图大小, 然后采用全卷积编码器网络来预测 $\hat{Y}$ 。如果同一类别的两个高斯分布重叠, 取元素级最大值。高斯核受到相应关键点附近的可变半径的限制, 若在一个图像中有多个类别目标, 则关键点的高斯核重叠。损失函数定义如公式 (4-2) 所示:

$$L_k = \frac{1}{N} \sum_{\text{xyz}} \begin{cases} (1 - \hat{Y}_{\text{xyz}})^\alpha \log(\hat{Y}_{\text{xyz}}) & Y_{\text{xyz}} = 1 \\ (1 - Y_{\text{xyz}})^\beta (\hat{Y}_{\text{xyz}})^\alpha \log(1 - \hat{Y}_{\text{xyz}}) & \text{otherwise} \end{cases} \quad (4-2)$$

其中 N 是目标的数量， $Y \in [0,1]^{\frac{W}{R} \times \frac{H}{R} \times C}$ 是标定对象的真实热力图， α 和 β 是损失超参数。

为了生成三维边界框，每个中心点需要三个额外的目标属性：深度、3D 尺寸和方向。作者为三个属性分别添加了一个单独的头检测，其中深度是每个中心点的单个标量，计算深度作为关键点的额外输出通道 $\hat{D} \in [0,1]^{\frac{W}{R} \times \frac{H}{R}}$ 。目标的 3D 尺寸[8]是由三个标量表示，使用 $\hat{f} \in [0,1]^{\frac{W}{R} \times \frac{H}{R} \times 3}$ 直接回归到以米为单位的绝对值。方向是单个标量，由于难以回归，根据 Mousavian 等人^[82]的方法，在两个框内进行方向的表示。具体来说，该方向使用 8 个标量进行编码，每个框有 4 个标量。对于每个框，两个标量用于 softmax 分类，其余的两个标量回归到每个框内的一个角度。

4.1.2 基于视锥的关联

本节描述了基于视锥的数据关联技术，视锥是通过将相机投影线投射到 3D 空间中来创建的，使得 3D 检测框与投影线相切。视锥平面沿着投影线由检测框的最小和最大深度定义的，视锥的底部和顶部与图像平面平行。视锥关联的输入包括 2D 边界框、3D 边界框和点云，在视锥关联中，将与目标关联的感兴趣区域确定在 3D 边界框周围，如果存在多个关联点，则仅选择最近的点进行融合。接下来将详细解释这一融合过程。

所选的雷达点 r 具有深度特征 d_r ，和X轴，Y轴速度分别为 v_x^r 和 v_y^r ，这些特征被映射为检测框热力图 $r \in \mathbb{R}^{\frac{W}{R} \times \frac{H}{R} \times C_r}$ 到图像平面，由于选择了雷达的三个特征，因此 $C_r = 3$ 。检测到的目标在图像中的深度 d_r 通过颜色映射表示，其中浅色对应较大的深度值。当相机未检测到目标时，相应区域的热力图值被设置为零。每个特征都被填充到检测框热力图 r 的不同通道中，目标 S 的特征 $f = [d_r \ v_x^r \ v_y^r]^T$ 映射公式为：

$$r_{x,y,i}^S = \frac{1}{M_i} \begin{cases} f_i, & |x - c_x^S| \leq \eta \omega^S \text{ and } |y - c_y^S| \leq \eta h^S \\ 0, & \text{otherwise} \end{cases} \quad (4-3)$$

其中， $[x \ y]^T$ 是图像检测框热力图 r 中的像素位置， ω^S 和 h^S 是目标 S 的宽度和

高度, η 是缩放因子用于缩小边界框覆盖的区域, $[c_x^S \ c_y^S]^T$ 是 2D 边界框的中心, 参数 M_i 是每个通道 $i \in [1,2,3]$ 的归一因子。当多个检测框覆盖一个像素时, 深度最小的检测框具有优先权。检测框热力图 r 将被传到第二阶段的检测头中。

正如前文所述, 毫米波雷达数据具有稀疏性、不准确性和噪声, 这使得在数据级融合中的关联变得困难。Nabati 和 Qi 为了弥补雷达点云的高度信息的缺失, 不仅将这些雷达点扩展成线, 还将其扩展成柱状。这些柱状的正方形横截面为 $20\text{cm} \times 20\text{cm}$, 高度为 1.5 m。径向速度和雷达截面积的测量值被相等地分配给整个拓展柱。由于雷达点测量的不确定性, 这些点被扩展到柱状上而不仅仅是线上。对柱状的扩展使得视锥对这种不准确更加可靠, 通过视锥关联的集成, 将检测框和柱状扩展进行了映射, 从而允许将雷达探测映射到目标中心, 并减少了重叠的问题。

如果点云拓展后柱状的任何部分在视锥区域内, 则这些柱状被认为与视锥有关。因此, 与支柱对应的雷达点与相同的初步目标探测相关联。最后, CenterFusion 的作者建议在所有相关的支柱中选择最近的支柱用于融合步骤。这由深度决定, 而不是由欧氏距离决定。

该网络训练从相机数据和相机-毫米波雷达融合的数据的回归关键参数起着重要作用, 主要分为两个阶段初级回归和第二次回归。初级回归仅从图像中提取 2D 检测框和初步 3D 检测框的参数, 第二次回归利用雷达中的丰富信息进行调整。

4.2 CenRadfusion

根据上节对 CenterFusion 网络框架的介绍可知, CenterFusion 网络主要依赖于相机采集到的图像特征, 若相机在第一阶段因为故障或者天气环境原因没有检测到任何物体, 则 CenterFusion 算法无法有效利用毫米波雷达, 因为毫米波雷达只有在图像初步检测后才参与预测, 所以当相机无法提供初级预测时, 毫米波雷达单独起不到作用。为了解决这些问题, 在本节中提出了一种有效的基于相机和毫米波雷达融合的改进方法, 称为 CenRadfusion 网络框架, 以提高三维目标检测的精度。如图 4-3 中呈现了该框架, 其主要包括三个关键阶段:

(I) 图像初步检测阶段: 通过将雷达点云早期投影到相机采集的图像上, 并利用 CenterNet (DLA) 网络作为主干网络, 从提取的图像特征中预测出目标

的中心点，以获取物体的二维尺寸（宽度和高度）、中心偏移量、三维尺寸、深度和旋转，其可为检测到的目标提供精确的 2D 边界框以及初步的 3D 边界框。这一阶段的主要目标是基于图像特征提取初步的三维检测框。

（II）雷达数据预处理与关联：首先，将雷达点云数据应用 DBSCAN^[83]聚类算法进行聚类之后，过滤点云中的白噪音。接着，通过视锥关联模块，将初步检测框与扩展后的雷达点云柱进行关联，实现雷达数据与检测目标的关联，并生成雷达特征图。这一步骤是为了确保不同传感器数据之间的有效协调。

（III）多模态数据融合与增强：图像和雷达特征图被连接起来后，为了强调重要特征，增加 SeNet^[84]注意力机制，通过重新计算深度、旋转角度以及估计物体的速度和属性来改进初步检测值，再形成最终的三维目标检测框。

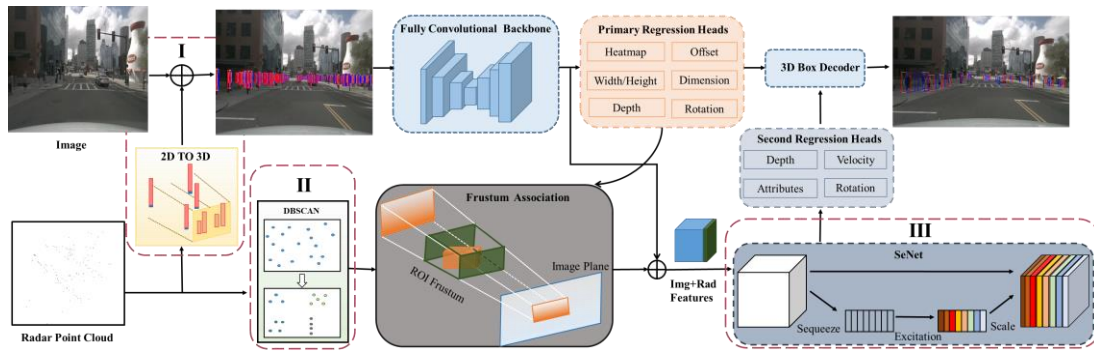


图 4-3 CenRadfusion 的网络架构

4.2.1 空间扩展投影

CenterFusion 的三维检测仅依赖于使用相机输入的初步检测。因此，雷达模式中所包含的互补信息并不用于初级探测，而只是用于调整相机的预测。虽然 CenterNet 是一个可靠的基于图像的检测网络，但由于雨或雾等天气条件，使得依赖于相机的检测结果不那么可靠。为了解决缺乏鲁棒性的问题，可以在特征融合的早期阶段额外引入雷达信息。为了提高目标检测的精度，可以将雷达点云中包含的信息可以投影到图像平面上，并作为附加通道添加到 CenterNet 的输入中。通过这种方法，在环境条件或传感器故障方面为目标检测的融合体系结构提供了一个更精准的主干，最大程度地保留了数据的初始完整性。

根据文献^[85]的启发，本节实施了类似的空间投影方法。与相机图像相比，雷达点云的数据非常稀疏。为了部分弥补这个问题，本研究采取了几个措施。首先，点云由若干个雷达扫描组成，其次，在投影出三维雷达点云投影进入二

维图像平面之后，这些点在维度、高度和宽度上都进一步扩展。



(a)雷达点云投影到图像上

(b)雷达点云拓展为柱

图 4-4 投影到 2D 图像平面上雷达点的可视化

如图 4-4 显示了投影雷达点扩展到三维柱上。在图 4-4(a)中，雷达点投影到二维图像中的单个点，为了解决高度信息的不准确问题，每个雷达点都被扩展为一个固定尺寸的柱，如图 4-4 (b) 所示。由于这些探测现在与 3D 空间中的一个维度相关联，所以这些柱状创建了一个更好的表示被雷达探测到的物理目标。如图 4-4 所示，通过这些点投影到垂直线上并拓展成柱，雷达图像的特征在数据中变得更密集，同时在图像的x坐标中保持精度。柱的尺寸以米为单位指定，这意味着，柱状的体积就会使雷达点远离相机的远点而减小。在图 4-4(b)中，雷达点被扩展为柱，以部分补偿雷达点云中的稀疏性，随着雷达点远离相机原点越远，三维柱状体积就会越小。

4.2.2 雷达数据预处理

为了提高融合后的特征质量和准确性，本节引入了毫米波雷达数据预处理步骤，如图 4-3 中 II 所示，并采用 DBSCAN 聚类方法对毫米波雷达点云进行过滤。该步骤主要包括以下关键步骤。

雷达数据预处理。在对雷达数据检测前，首先预处理时进行一次聚类，产生分类标签。由于毫米波雷达采集到的每个雷达点都对应着一个潜在的反射目标，而毫米波雷达通常捕获大量的白噪声。这些白色的噪音往往会以非聚集的方式出现，噪声点之间的密度相对较低，因此，考虑密度聚类进行噪声滤波。使用 DBSCAN 进行聚类，进而从聚类结果中过滤离群点。

DBSCAN 在发现任意形状的团簇方面明显更加有效，DBSCAN 算法的优点包括，其可以不需要人为假设的情况下，自动决定类别数量。和 K-means 只能

发现圆形簇不同，该算法可以发现任意形状的簇类，同时它可以识别噪声点方面的鲁棒性较强。

通过增加毫米波雷达数据预处理步骤和点云过滤，并将过滤后的点云投影到图像上，可以提高 CenRadfusion 网络的输入数据质量和准确性。这样的处理步骤能够减少噪声干扰、提取关键点和区域，并将毫米波雷达数据与相机数据更好地融合，为后续的目标检测和定位提供更可靠的输入。在实验中，可以评估增加预处理步骤和点云过滤对目标检测性能的影响，从而验证其有效性和优越性。

4.2.3 增加注意力机制

当将毫米波雷达特征和视觉检测特征融合后，为了进一步提高融合特征的表达能力和区分性，可以引入 SeNet (Squeeze-and-Excitation Network) 注意力机制。SeNet 注意力机制是一种轻量级的注意力模块，可以自适应地学习特征通道之间的重要性权重。

首先，从毫米波雷达和视觉检测两个传感器获得的特征图中提取特征后，其通道数分别 67, 112, 200。当通道数为 200 时，特征数量最高，此时，SeNet 的加入会得到特征映射区域的最大值，为对于每个特征图，通过引入 SeNet 模块来计算通道权重。SeNet 模块由两个关键部分组成：Squeeze 操作和 Excitation 操作。Squeeze 操作通过全局平均池化将特征图的空间维度压缩成一个通道维度，然后通过两个全连接层进行特征压缩和激活操作。Excitation 操作通过学习到的权重向量对通道特征进行重新加权。特征融合：将两个传感器的特征图分别与对应的通道权重相乘，以实现特征的重加权融合。进一步处理：将融合后的特征图送入后续的网络模块进行进一步的处理。

SeNet 模块的结构如图 4-5 所示，输入特征层大小为 $67 \times 112 \times 200$ ，在 Squeeze 操作中，通过全局平均池化操作，候选特征大小变为 1×1 ，在 Excitation 操作中，第一个全连接层用于减少 $\frac{200}{r}$ 的通道数， r 时比例因子，用于减少通道数，从而减少计算量，在本研究中，通过多次实验，当 $r = 25$ 时效果最好，当使用 Relu 激活函数时，特征维数保持不变，当使用第二个全连接层升级维数时，此时通道回到 200；使用 S 型函数输出每个通道的权值，将权重与原

始对应的通道相乘，得到加权特征层。

通过引入 SeNet 注意力机制，可以自适应地学习不同特征通道的重要性，使得融合后的特征更加具有区分性和表达能力。这有助于提高目标检测和定位的准确性和鲁棒性，特别是在复杂场景下的应用。

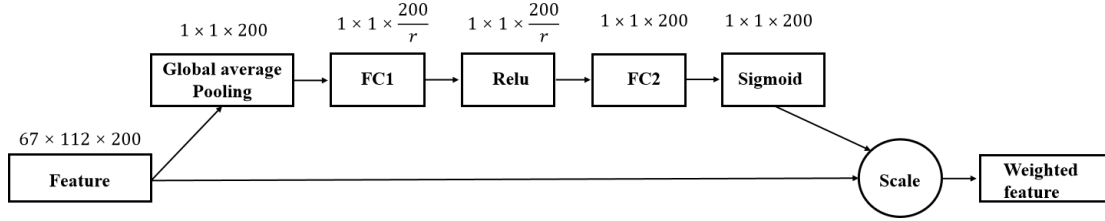


图 4-5 SeNet 框架图

在实验中，可以通过对比使用和不使用 SeNet 注意力机制的性能差异来评估其有效性。通过引入 SeNet 注意力机制，可以进一步提升 CenRadfusion 算法在三维目标检测任务中的性能，并丰富论文中对于特征融合和注意力机制的研究内容。

4.3 实验与结果

4.3.1 数据准备

自动驾驶数据集通常围绕视觉图像和激光雷达原始数据设置的，由于公共数据集中缺乏毫米波雷达数据，在本章的所有实验中，所提出的方法均在 NuScenes 数据集上进行了训练和评估，该数据集是第一个包含雷达数据的自动驾驶数据集，选择不同场景帧的图片和点云数据作为数据集，对于目标检测，NuScenes 提供了 23 个类别的 3D 边界框。

NuScenes 数据集包括来自车辆前部，侧面和后部的 6 个不同摄像头的图像。雷达探测是由四个角雷达和一个前雷达获得。本章使用数据集中可用的样本的两个子集来进行实验，第一个是子集包含来自前摄像头和前雷达的数据，第二个子集是来自后置相机和后置两个雷达数据。

4.3.2 消融实验

为了验证雷达点云数据预先处理，增加雷达点云聚类 and 通道注意模块 SeNet 增加的有效性，本章针对不同的模块设计了消融实验，所有的消融实验以 CenterFusion 为基础，在 NuScenes 数据集上进行验证。表 4-1 显示了消融研究的

总体检测结果。

表 4-1 对 NuScenes 验证集的整体消融研究结果

Method	空间扩展投影	DBS CAN	Se Net	NDS	mAP	mATE	mASE	mAOE	mAVE	mAAE
Center-Fusion [36]	-	-	-	0.449	0.326	0.631	0.261	0.516	0.614	0.145
Ours	√	-	-	0.448	0.315	0.629	0.262	0.498	0.689	0.142
Ours	-	√	-	0.454	0.324	0.625	0.256	0.483	0.574	0.140
Ours	√	√	-	0.458	0.335	0.623	0.254	0.475	0.570	0.131
Ours	√	√	√	0.460	0.342	0.619	0.263	0.472	0.563	0.130

在第一个实验中如表 4-1，本节只增加了空间扩展投影，将雷达点投影至图像上，将雷达特征作为额外输入添加到图像检测。第二个实验中仅加入 DBSCAN 点云聚类的方式，过滤雷达噪声点，这使得 NDS 评分相对于提高了 6%，同时加入空间扩展投影和聚类方式时，NDS 和 mAP 得分分别相对提高了 9%和 0.9%，当 SeNet 注意力机制加入进去后，与最开始的 CenterFusion 网络相比较，NDS 和 mAP 得分分别 11%和 1.6%。

表 4-2 在 NuScenes 验证集上的基于不同类的消融研究结果

Method	空间扩展投影	DBS CAN	Se Net	Car	Truck	Bus	Pedestrian	Motor	Bicycle
Center-Fusion [36]	-	-	-	0.509	0.258	0.234	0.37	0.314	0.201
Ours	√	-	-	0.512	0.265	0.239	0.371	0.321	0.218
Ours	-	√	-	0.52	0.271	0.251	0.37	0.318	0.226
Ours	√	√	-	0.532	0.284	0.26	0.379	0.323	0.23
Ours	√	√	√	0.543	0.301	0.274	0.386	0.323	0.239

如表 4-2，在本节消融实验过程中，每种目标类别的 AP 值均有提升，同时加入空间扩展投影和聚类方式时，汽车和自行车的 AP 得分分别相对提高了 2.3%和 2.9%，当 SeNet 注意力机制加入进去后，与最开始的 CenterFusion 网络相比较，汽车和行人的 AP 得分分别提高了 3.4%和 1.6%。

4.3.3 与现有方法的对比

目前，有许多类型的 3D 目标检测网络。为了验证本研究中该网络的有效性，本章选择了 NuScenes 数据集列表中各种流行的三维目标检测网络进行实验比较。在单模态检测中，选择了 PointPillars_DSA^[86]，CenterNet，FCOS3D^[87]，其中 PointPillars_DSA 是激光雷达进行三维目标检测的算法，而在多模态融合网络中，选择了 CenterFusion 和 CRAFT^[88]网络。

在 NuScenes 数据集上的评价结果如表 4-3 所示。本章所提的方法在目标检测上的平均准确度略低于目前最先进的 CRAFT；然而，值得注意的是，该方法优于仅使用相机进行三维目标检测的算法，与 CenterNet 和 FCOS3D 相比，汽车的平均准确度提高了 0.8%和 1.9%。与其他方法相比，基于激光雷达的方法 PointPillars_DSA 虽然在大多数类别上超过本算法，但其对自行车，摩托车的检测上存在短板，本章所提的算法在自行车的平均准确度上是优于激光雷达算法 15.6%。

表 4-3 在 NuScenes 验证集上的算法性能比较

Method	Modalit										
	y							AP			
	C	R	L	Car	Truck	Bus	Trailer	Pede strian	Motor	Bicycle	Barrier
PointPillars _DSA ^[86]			√	0.812	0.438	0.572	0.478	0.733	0.321	0.079	0.553
CenterNet ^[37]	√			0.536	0.27	0.248	0.251	0.375	0.291	0.207	0.533
FCOS3D ^[87]	√			0.524	0.27	0.277	0.255	0.397	0.345	0.298	0.538
CenterFusion ^[36]	√	√		0.524	0.265	0.362	0.235	0.389	0.314	0.229	0.484
CRAFT ^[88]	√	√		0.681	0.391	0.356	0.324	0.437	0.429	0.241	0.51
Ours	√	√		0.543	0.301	0.274	0.26	0.386	0.323	0.235	0.534

4.3.4 检测结果显示

对该算法的检测结果进行了可视化测试。使用本研究中提出的网络对 NuScenes 数据集的图像和毫米波雷达点云数据进行检测。检测结果如图 4-5 和图 4-6 所示。在图 4-5 中，是白天天气较好的情况下，在街道和交通路口处可以观察到，该网络可以准确地检测到不同距离的车辆和行人。图 4-6 显示了在乌云天气时，光线较弱的情况下，在停车场附近时，可以观察到，该网络能够正确

地检测到行人，摩托车以及被遮挡的物体。因此，本研究提出的 CenRadfusion 网络是一种具有良好鲁棒性的多模态融合三维目标检测网络。



图 4-5 在白天天气较好情况下，最终检测效果图



图 4-6 在阴天天气光线较差情况下，最终检测效果图

4.4 本章小结

本章在具有挑战性的 NuScenes 3D 的检测基准上评估了本章提出的方法 CenRadfusion 网络结构，一种用于自动驾驶应用中的目标检测的雷达和视觉融合的三维目标检测算法，充分利用毫米波雷达信息和视觉图像检测。CenRadfusion 方法对 CenterFusion 先前的融合网络框架进行了多种改进。

(1) 本章实现了早期融合，避免因环境问题或者视觉传感器故障导致图像输入偏差，提高了三维目标检测精度。

(2) 本章使用了毫米波雷达点云密度聚类 DBSCAN 进行噪声滤波，滤除毫米波雷达反射目标点中的白噪声，这些白噪声以非聚合的方式存在，噪声点之间的密度相对较小，密度聚类进行噪声滤波。

(3) 本章添加了基于 SeNet 的特征提取网络，自适应地调整每个特征信道

的重要性，为每个特征信道分配不同的权值，增强网络的重要特征，可以显著提高网络的运行效率和学习能力。

（4）实验表明，在三维目标检测中，改进后的 CenterFusion 算法在车辆、行人、摩托车三种检测类别中的评价准确率可以达到 54.3%，38.6%和 23.9%，NDS 比原来的 CenterFusion 算法提高了 1.1%，mAP 提高了 1.6%。

第 5 章 总结与展望

毫米波雷达技术在学术界和工业界日益受到关注，其相较于其他传感器模式具备的天气鲁棒性和运动测量等优势，使其在自动驾驶领域的环境感知中发挥了重要作用；在目标检测和分类能力上，视觉可以提供详细的目标轮廓信息。本研究针对毫米波雷达点云目标检测，处理技术以及毫米波雷达和相机图像特征融合展开研究，提出了相应的网络模型框架和算法。本章主要对论文中已经完成的工作进行总结，并对后续研究的内容进行分析和展望。

5.1 本文总结

本研究对毫米波雷达和视觉融合的三维目标检测研究的工作主要集中在以下方面：

(1) 首先详细从两个整理汇总毫米波雷达的相关知识：毫米波雷达的信号传播理论和毫米波雷达测速测距测角度的基本工作原理。其次介绍了包含毫米波雷达信息的数据集以及点云和图像数据关联技术的常用方法，为后续网络架构的提出提供了理论支撑。

(2) 构建了基于毫米波雷达点云的三维目标检测网络模型 **RPCNET**。该模型充分利用了毫米波雷达获取目标精确位置，速度和角度信息以及其对恶劣天气的抗干扰能力。**RPCNET** 通过四个关键步骤实现目标检测：首先，使用 **RRPN** 网络根据毫米波雷达点云分布提出初始的区域建议；其次，利用 **DGCNN** 网络从点云中提取特征，并为每个点确定锚框，通过卷积层定义锚框代表特征。接着，通过置信度预测固定大小的锚框作为三维边界框的初始估计；最后，通过全连接层分类网络对锚框进行分类，进一步细化边界框以获得更准确的检测结果。**RPCNET** 的创新之处在于直接在原始的毫米波雷达点云上进行分类，通过非极大值抑制优化最终输出，从而实现对目标的检测。

(3) 构建了一种基于毫米波雷达点云和相机图像特征融合的三维目标检测网络模型 **CenRadfusion**。该模型实现了点云数据和相机图像的早期融合，提高了检测精度。针对毫米波雷达点云的密度聚类进行噪声滤波，滤除了反射目标

点中的白噪声，最后采用基于 SeNet 的特征提取网络，自适应地调整每个特征信息通道的重要性，显著增强了网络的学习能力和运行效率。实验结果表明，CenRadfusion 模型在公开的 NuScenes 数据集上表现优于之前的 CenterFusion 网络和单模态目标检测网络。

5.2 展望

本文主要研究了相机和毫米波雷达特征信息融合的三维目标检测技术，尽管所提出的网络在算法层面上取得了一定的创新和改进，但就当前多模态信息融合感知技术的发展现状而言，有必要在以下方面进行进一步研究，这也将是后续的努力方向：

（1）数据融合坐标系的转化

将数据转换为鸟瞰图（BEV）视角，是一个非常具有前瞻性的方向。这种转换不仅可以简化数据处理流程，还有助于提高目标检测的准确性和效率。在未来的研究中，可以探索如何更好地将不同传感器获得的信息映射到共同的坐标系下，并利用这种转换来实现更精确的目标检测与跟踪。

（2）毫米波雷达点云噪声处理

针对毫米波雷达数据中存在的背景杂波和多路径效应引起的结构化噪声问题，提出有效的噪声处理方法是至关重要的。未来的研究可以集中于开发能够准确识别并过滤这些噪声的算法，以提高目标检测系统的鲁棒性和可靠性。此外，还可以考虑如何结合深度学习技术和传统信号处理方法来处理这些噪声，从而进一步优化目标检测的性能。

（3）4D 毫米波成像雷达的应用

随着 4D 毫米波成像雷达技术的不断发展^[89]，它已经成为自动驾驶领域一个备受关注的新兴趋势。新型的 4D 毫米波成像雷达不仅保留了传统毫米波雷达的优点，而且还增加了俯仰角的测量信息，并且提供了更密集的点云和更高的分辨率，能够通过输出大量的测量点清晰地呈现目标障碍物的轮廓。在未来的研究中，可以进一步探索如何充分利用 4D 毫米波成像雷达的优势，设计出更高效、更可靠的自动驾驶系统，这可能涉及到改进目标检测算法，优化传感器融合策略，以及开发适用于不同场景和环境的 4D 毫米波成像雷达的数据处理技术。

参考文献

- [1] 王云鹏,吴琼,宋德王等.自动驾驶数据集及 3D 目标感知方法综述[J].人工智能,2023,(05):31-47.DOI:10.16453/j.2096-5036.202347.
- [2] 彭育辉,江铭,马中原等.汽车自动驾驶关键技术研究进展[J].福州大学学报(自然科学版),2021,49(05):691-703.
- [3] Othman K. Public acceptance and perception of autonomous vehicles: a comprehensive review[J]. AI and Ethics, 2021, 1(3): 355-387.
- [4] Cho M. A study on the obstacle recognition for autonomous driving RC car using LiDAR and thermal infrared camera[C]//2019 Eleventh international conference on ubiquitous and future networks (ICUFN). IEEE, 2019: 544-546.
- [5] Song Z, Liu L, Jia F, et al. Robustness-Aware 3D Object Detection in Autonomous Driving: A Review and Outlook[J]. arXiv preprint arXiv:2401.06542, 2024.
- [6] Cao C, Preda M, Zaharia T. 3D point cloud compression: A survey[C]//Proceedings of the 24th International Conference on 3D Web Technology. 2019: 1-9.
- [7] 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.
- [8] Qi C R, Yi L, Su H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space[J]. Advances in neural information processing systems, 2017, 30.
- [9] 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.
- [10] Danzer A, Griebel T, Bach M, et al. 2d car detection in radar data with pointnets[C]//2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, 2019: 61-66.

- [11]Feng Z, Zhang S, Kunert M, et al. Point cloud segmentation with a high-resolution automotive radar[C]//AmE 2019-Automotive meets Electronics; 10th GMM-Symposium. VDE, 2019: 1-5.
- [12]Tilly J F, Haag S, Schumann O, et al. Detection and tracking on automotive radar data with deep learning[C]//2020 IEEE 23rd International Conference on Information Fusion (FUSION). IEEE, 2020: 1-7.
- [13]Scheiner N, Kraus F, Wei F, et al. Seeing around street corners: Non-line-of-sight detection and tracking in-the-wild using doppler radar[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 2068-2077.
- [14]Palffy A, Dong J, Kooij J F P, et al. CNN based road user detection using the 3D radar cube[J]. IEEE Robotics and Automation Letters, 2020, 5(2): 1263-1270.
- [15]Yang G, Mentasti S, Bersani M, et al. Lidar point-cloud processing based on projection methods: a comparison[C]//2020 AEIT International Conference of Electrical and Electronic Technologies for Automotive (AEIT AUTOMOTIVE). IEEE, 2020: 1-6.
- [16]Werber K, Rapp M, Klappstein J, et al. Automotive radar gridmap representations[C]//2015 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM). IEEE, 2015: 1-4.
- [17]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.
- [18]Lombacher J, Hahn M, Dickmann J, et al. Detection of arbitrarily rotated parked cars based on radar sensors[C]//2015 16th International Radar Symposium (IRS). IEEE, 2015: 180-185.
- [19]Lombacher J, Laudt K, Hahn M, et al. Semantic radar grids[C]//2017 IEEE intelligent vehicles symposium (IV). IEEE, 2017: 1170-1175.
- [20]Mescheder L, Oechsle M, Niemeyer M, et al. Occupancy networks: Learning 3d reconstruction in function space[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 4460-4470.

- [21]Roddick T, Cipolla R. Predicting semantic map representations from images using pyramid occupancy networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 11138-11147.
- [22]Dreher M, Erçelik E, Bänziger T, et al. Radar-based 2D car detection using deep neural networks[C]//2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2020: 1-8.
- [23]Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
- [24]Sless L, El Shlomo B, Cohen G, et al. Road scene understanding by occupancy grid learning from sparse radar clusters using semantic segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019: 0-0.
- [25]Scheiner N, Kraus F, Appenrodt N, et al. Object detection for automotive radar point clouds—a comparison[J]. AI Perspectives, 2021, 3(1): 6.
- [26]李文静,白静,彭斌等.图卷积神经网络及其在图像识别领域的应用综述[J].计算机工程与应用,2023,59(22):15-35.
- [27]Shi W, Rajkumar R. Point-gnn: Graph neural network for 3d object detection in a point cloud[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 1711-1719.
- [28]Svenningsson P, Fioranelli F, Yarovoy A. Radar-pointgnn: Graph based object recognition for unstructured radar point-cloud data[C]//2021 IEEE Radar Conference (RadarConf21). IEEE, 2021: 1-6.
- [29]Gong P, Wang C, Zhang L. Mmpoint-gnn: Graph neural network with dynamic edges for human activity recognition through a millimeter-wave radar[C]//2021 International Joint Conference on Neural Networks (IJCNN). IEEE, 2021: 1-7.
- [30]Phan A V, Le Nguyen M, Nguyen Y L H, et al. Dgcnn: A convolutional neural network over large-scale labeled graphs[J]. Neural Networks, 2018, 108: 533-543.

- [31]Meyer M, Kusch G, Tomforde S. Graph convolutional networks for 3d object detection on radar data[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 3060-3069.
- [32]Wu Y, Wang Y, Zhang S, et al. Deep 3D object detection networks using LiDAR data: A review[J]. IEEE Sensors Journal, 2020, 21(2): 1152-1171.
- [33]Yadav R, Vierling A, Berns K. Radar+ RGB fusion for robust object detection in autonomous vehicle[C]//2020 IEEE International Conference on Image Processing (ICIP). IEEE, 2020: 1986-1990.
- [34]Nabati R, Qi H. Radar-camera sensor fusion for joint object detection and distance estimation in autonomous vehicles[J]. arXiv preprint arXiv:2009.08428, 2020.
- [35]Nabati R, Qi H. Rrpn: Radar region proposal network for object detection in autonomous vehicles[C]//2019 IEEE International Conference on Image Processing (ICIP). IEEE, 2019: 3093-3097.
- [36]Nabati R, Qi H. Centerfusion: Center-based radar and camera fusion for 3d object detection[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021: 1527-1536.
- [37]Zhou X, Wang D, Krähenbühl P. Objects as points[J]. arXiv preprint arXiv:1904.07850, 2019.
- [38]Chadwick S, Maddern W, Newman P. Distant vehicle detection using radar and vision[C]//2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019: 8311-8317.
- [39]Targ S, Almeida D, Lyman K. Resnet in resnet: Generalizing residual architectures[J]. arXiv preprint arXiv:1603.08029, 2016.
- [40]Kowol K, Rottmann M, Bracke S, et al. Yodar: Uncertainty-based sensor fusion for vehicle detection with camera and radar sensors[J]. arXiv preprint arXiv:2010.03320, 2020.
- [41]John V, Mita S. RVNet: Deep sensor fusion of monocular camera and radar for image-based obstacle detection in challenging environments[C]//Image and Video Technology: 9th Pacific-Rim Symposium, PSIVT 2019, Sydney, NSW, Australia,

- November 18–22, 2019, Proceedings 9. Springer International Publishing, 2019: 351-364.
- [42]John V, Nithilan M K, Mita S, et al. So-net: Joint semantic segmentation and obstacle detection using deep fusion of monocular camera and radar[C]//Image and Video Technology: PSIVT 2019 International Workshops, Sydney, NSW, Australia, November 18–22, 2019, Revised Selected Papers 9. Springer International Publishing, 2020: 138-148.
- [43]Dong X, Zhuang B, Mao Y, et al. Radar camera fusion via representation learning in autonomous driving[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 1672-1681.
- [44]Chunmei M, Yinong L, Ling Z, et al. Obstacles detection based on millimetre-wave radar and image fusion techniques[J]. 2016.
- [45]Meyer M, Kusch G. Deep learning based 3d object detection for automotive radar and camera[C]//2019 16th European Radar Conference (EuRAD). IEEE, 2019: 133-136.
- [46]Cui H, Wu J, Zhang J, et al. 3D detection and tracking for on-road vehicles with a monovision camera and dual low-cost 4D mmWave radars[C]//2021 IEEE International Intelligent Transportation Systems Conference (ITSC). IEEE, 2021: 2931-2937.
- [47]Bansal K, Rungta K, Bharadia D. Radsegnet: A reliable approach to radar camera fusion[J]. arXiv preprint arXiv:2208.03849, 2022.
- [48]Harley A W, Fang Z, Li J, et al. A simple baseline for bev perception without lidar[J]. arXiv preprint arXiv:2206.07959, 2022, 3(4).
- [49]Lim T Y, Ansari A, Major B, et al. Radar and camera early fusion for vehicle detection in advanced driver assistance systems[C]//Machine learning for autonomous driving workshop at the 33rd conference on neural information processing systems. 2019, 2(7).

-
- [50] Ehler L E. Integrated pest management (IPM): definition, historical development and implementation, and the other IPM[J]. Pest management science, 2006, 62(9): 787-789.
 - [51] Feng D, Rosenbaum L, Glaeser C, et al. Can we trust you? on calibration of a probabilistic object detector for autonomous driving[J]. arXiv preprint arXiv:1909.12358, 2019.
 - [52] 张科遥,林福江,白雪飞. 77 GHz FMCW 车载雷达系统设计[J]. 信息技术与网络安全, 2020, 39(04): 53-57+72. DOI:10.19358/j.issn.2096-5133.2020.04.010.
 - [53] Eaves J, Reedy E. Principles of modern radar[M]. Springer Science & Business Media, 2012.
 - [54] Jain V, Heydari P, Jain V, et al. Radar fundamentals[J]. Automotive radar sensors in silicon technologies, 2013: 5-11.
 - [55] Wang W Q. Overview of frequency diverse array in radar and navigation applications[J]. IET Radar, Sonar & Navigation, 2016, 10(6): 1001-1012.
 - [56] Wang G, Munoz-Ferreras J M, Gu C, et al. Application of linear-frequency-modulated continuous-wave (LFMCW) radars for tracking of vital signs[J]. IEEE transactions on microwave theory and techniques, 2014, 62(6): 1387-1399.
 - [57] Winkler V. Range Doppler detection for automotive FMCW radars[C]//2007 European Radar Conference. IEEE, 2007: 166-169.
 - [58] Richards M A. Fundamentals of radar signal processing[M]. New York: Mcgraw-hill, 2005.
 - [59] Xu S, Xu D, Luo H. Information theory of detection in radar systems[C]//2017 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT). IEEE, 2017: 249-254.
 - [60] Heckbert P. Fourier transforms and the fast Fourier transform (FFT) algorithm[J]. Computer Graphics, 1995, 2(1995): 15-463.
 - [61] Ozaktas H M, Kutay M A, Mendlovic D. Introduction to the fractional Fourier transform and its applications[M]//Advances in imaging and electron physics. Elsevier, 1999, 106: 239-291.

- [62]Mahafza B R. Radar systems analysis and design using MATLAB[M]. Chapman and Hall/CRC, 2005.
- [63]Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? the kitti vision benchmark suite[C]//2012 IEEE conference on computer vision and pattern recognition. IEEE, 2012: 3354-3361.
- [64]Sun P, Kretzschmar H, Dotiwalla X, et al. Scalability in perception for autonomous driving: Waymo open dataset[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 2446-2454.
- [65]Bijelic M, Gruber T, Mannan F, et al. Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 11682-11692.
- [66]Mostajabi M, Wang C M, Ranjan D, et al. High-resolution radar dataset for semi-supervised learning of dynamic objects[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020: 100-101.
- [67]Caesar H, Bankiti V, Lang A H, et al. nuscenes: A multimodal dataset for autonomous driving[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 11621-11631.
- [68]Schumann O, Hahn M, Scheiner N, et al. Radarscenes: A real-world radar point cloud data set for automotive applications[C]//2021 IEEE 24th International Conference on Information Fusion (FUSION). IEEE, 2021: 1-8.
- [69]Frank C. Modern robotics-mechanics, planning, and control[M]. Cambridge University Press, 2017.
- [70]Nobis F, Shafiei E, Karle P, et al. Radar voxel fusion for 3D object detection[J]. Applied Sciences, 2021, 11(12): 5598.
- [71]Wei Z, Zhang F, Chang S, et al. Mmwave radar and vision fusion for object detection in autonomous driving: A review[J]. Sensors, 2022, 22(7): 2542.

- [72]Stock K. Self-driving cars can handle neither rain nor sleet nor snow[J]. Bloomberg Businessweek, 2018, 17.
- [73]Liu S, Huang D, Wang Y. Adaptive nms: Refining pedestrian detection in a crowd[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 6459-6468.
- [74]Huber P J. Robust estimation of a location parameter[M]//Breakthroughs in statistics: Methodology and distribution. New York, NY: Springer New York, 1992: 492-518.
- [75]Jahanshahi M, Dehghan M, Meybodi M R. LAMR: learning automata based multicast routing protocol for multi-channel multi-radio wireless mesh networks[J]. Applied intelligence, 2013, 38: 58-77.
- [76]Fent F, Bauerschmidt P, Lienkamp M. Radargnn: Transformation invariant graph neural network for radar-based perception[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 182-191.
- [77]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.
- [78]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.
- [79]Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.
- [80]Newell A, Yang K, Deng J. Stacked hourglass networks for human pose estimation[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII 14. Springer International Publishing, 2016: 483-499.
- [81]Yu F, Wang D, Shelhamer E, et al. Deep layer aggregation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 2403-2412.

- [82]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.
- [83]Hahsler M, Piekenbrock M, Doran D. dbscan: Fast density-based clustering with R[J]. Journal of Statistical Software, 2019, 91: 1-30.
- [84]Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.
- [85]Wang Y, Guan Y, Li S, et al. Fusion perception of vision and millimeter wave radar for autonomous driving[C]//Proceedings of the 8th International Conference on Computing and Artificial Intelligence. 2022: 767-772.
- [86]Bhattacharyya P, Huang C, Czarnecki K. Sa-det3d: Self-attention based context-aware 3d object detection[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 3022-3031.
- [87]Wang T, Zhu X, Pang J, et al. Fcos3d: Fully convolutional one-stage monocular 3d object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 913-922.
- [88]Kim Y, Kim S, Choi J W, et al. Craft: Camera-radar 3d object detection with spatio-contextual fusion transformer[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2023, 37(1): 1160-1168.
- [89]张远,肖宝华,杨大林.基于 4D 毫米波雷达点云的多目标跟踪算法[J].科技资讯,2024,22(01):38-42.DOI:10.16661/j.cnki.1672-3791.2306-5042-9247.

攻读硕士学位期间发表学术论文及参与项目

一、发表学术论文

- [1] Peicheng Shi, **Tong Jiang**, Zhiqiang Liu and Aixi Yang. CenRadfusion: Fusing Image Center Detection and Millimeter Wave Radar for 3D Object Detection[J]. Signal Image and Video Processing. (中科院四区, 除导师外一作, 录用)
- [2] Peicheng Shi, **Tong Jiang** and Aixi Yang. RPCNET: Radar Point Cloud Network for 3D Object Detection[J]. Arabian Journal for Science and Engineering(中科院四区, 除导师外一作, 终审)

二、申请及以获得的专利

- [1] 时培成, **江彤**, 董心龙等. 一种用于电驱动桥控制器的减震装置[P]. 中国: CN115111302B, 2024-03-01. (发明专利, 除导师外一作, 授权)
- [2] 时培成, **江彤**, 张建国等. 一种车辆自适应转向雷达系统及其工作方法[P]. 中国: CN115871567A, 2023-03-31. (发明专利, 除导师外一作, 实审)
- [3] 时培成, **江彤**, 李新桥等. 一种新能源汽车驱动电机的检测装置[P]. 中国: CN114814581A, 2022-07-29. (发明专利, 除导师外一作, 实审)

三、参研项目

- [1] 安徽省自然科学基金, (2022, 2208085MF173), 基于多模态融合的智能车辆多任务感知技术研究, 参与;
- [2] 安徽省重点研究与开发计划项目, 202104a05020003, 特定场景无人车通用化全线控底盘开发与产业化应用, 参与;

四、省级比赛获奖情况

- [1] 2023年4月, 作品《我眼中的冬奥精神》, 获得安徽省第六届校园读书创作活动(“喜迎二十大, 奋进新征程”)主题征文比赛三等奖;
- [2] 2023年7月, 作品《基于多模态信息融合的三维感知系统及其在智能车上的应用》获得第十届“挑战杯·华安证券”安徽省大学生课外学术科技作品竞赛二等奖;
- [3] 2024年3月, 获得第十七届安徽省大学生职业规划大赛二等奖。

致谢

三年时光转瞬即逝，坐在陪伴自己三年的工位电脑前敲下了这段文字，感恩这三年的读研旅程中遇见的每一个人，发生的每一件事与见证的每一处风景。

春晖寸草，难以回报。二十余年，正是家庭的幸福给予了我最好的滋养。感谢父母和爷爷奶奶外婆在我求学之路上的默默支持和鼓励，给我无尽的疼爱和呵护，家人们尽最好的一切供我求学，暖意饱食，无后顾之忧，养育之恩，无以为报，愿成长脚步能一路生花，为您们挡下岁月的风沙。

一朝沐杏雨，一朝念恩师。感谢我的导师时培成教授，依然记得三年前考研成绩出来后给您发的第一封自荐信，您的一句欢迎加入我的团队真的让我激动了很久，在之后的三年里从选题开题，论文撰写和小论文投稿各个方面您都给予了许多鼓励和指导，也是在您的鼓励下，我参加了一系列的比赛锻炼了自己的表达能力。感谢老师三年来的辛苦付出，愿老师身体健康，学术长青。

一鸣从此始，相望青云端。感谢齐恒师兄，刘志强师兄，张程辉师兄，陈新禾师兄，李龙师兄，刘糠继师兄对我这个师妹的指导和鼓励。特别感谢从我研一入学以来就一直耐心教导关照我的齐恒师兄，刘志强师兄，还记得我刚报道那天，是二位师兄开着我们课题组实验小车去接我帮我运行李，那天走在校园里很是骄傲。毕业前你们会仔细的帮我完善实验细节，润色论文语句，比赛过程中照顾我的情绪，愿师兄们前程似锦，步步高升。感谢图书馆一楼并肩作战的各位战友们：张建国，毛飞，张庆，李屹，有幸遇见，感谢你们就像哥哥一样对我这个女生格外照顾，会在比赛筹备阶段，担心我害怕，陪我到工位锁门；会在我手足无措时，伸出援手；这段共同学习，互帮互助，相约聚餐的时光是我读研生活中浓墨重彩的一笔，接下来的日子，一起努力顶峰相见吧！感谢董心龙师弟，周梦如师妹，许柳柳师妹，李帅师弟，杨礼师弟，潘艺鑫师弟，吴文超师弟在生活中的帮助，相会是缘，同行是乐，共事是福！

山水一程，三生有幸。感谢我的室友卢娜，赵亚玲，陈清清，即使各奔前程，也会互相记挂！感谢我亲爱的好朋友们，遇到可爱又有趣的你们是我宝贵的财富，谢谢你们给我带来了许多快乐。有幸遇到汪涌，对我宠爱有加，教我人情世故，护我天真如初，陪我共同成长。来日方长，我们一起携手去做更多有趣的事情，去看更多美丽的风景吧！

读研期间，经常劝慰自己的一句话：“关关难过关关过”，但现在我想说的是“前路漫漫亦灿灿”最后，向在百忙中审阅论文的各位专家、教授、学者致以崇高的敬意和最诚挚的感谢！

江彤

尚模敦学 唯实惟新

