分类号：TP391　　单位代码：10363
密　级：公开　　学　号：2210330108

题　目　激光雷达与视觉融合的车辆环境感知方法研究

题　目　**Research on vehicle environment perception method Based on LiDAR and vision fusion**

学 生 姓 名：李梦楠
校 内 导 师：徐晓光教授
校 外 导 师：郝旭耀工程师
专业学位类别：电子信息
研究方向(领域)：人工智能与系统集成

# 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：

日期：　　年　　月　　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：　　指导教师签名：

日期：　　年　　月　　日　　日期：　　年　　月　　日

# 激光雷达与视觉融合的车辆环境感知方法研究

## 摘 要

随着人工智能和机器人技术的快速发展，智能驾驶技术已经成为全球研究的热点。智能驾驶车辆旨在通过搭载激光雷达、毫米波雷达、相机等传感器实时、高效、可靠的感知车辆周围的环境，为车辆的控制和决策提供依据，实现车辆的自主驾驶。目标检测是车辆感知周围环境的重要手段，是确保智能驾驶车辆安全稳定运行的前提。传统的目标检测的算法大多都是基于单一传感器进行的，在复杂的路况或恶劣的天气条件下很难完成对车辆周围环境的感知，进而影响智能驾驶车辆的安全运行。为此，本文采用激光雷达和相机融合检测的方案，利用不同传感器采集的信号的互补性，对多模态数据融合的目标检测算法进行研究，改进了一种多模态焦点稀疏卷积目标检测算法。本文主要工作内容如下：

(1) 针对不同类型传感器之间的坐标系转换与对齐问题，推导了多种坐标系之间相互转换的关系。基于矩形棋盘格标定方法，获取相机内外参数。在此基础上，进行激光雷达与相机的联合标定，为传感器数据融合奠定基础。

(2) 针对多模态焦点稀疏卷积目标检测算法对小尺寸目标检测精度不足的问题，对多模态焦点稀疏卷积目标检测算法进行改进。首先，对算法的 2D主干网络部分进行优化，提出了一种多尺度特征提取网络，使算法能够关注小尺寸目标。其次，分析了传统非极大值抑制NMS算法的原理并针对多模态焦点稀疏卷积目标检测算法目标框筛选精度不足的问题，引入了Soft-NMS算法，并利用高斯加权系数动态调整置信度。然后，针对传统的目标筛选算法判别标准单一

的问题，结合 3D目标检测框的特性，引入了 3D-IoU姿态估计损失函数提升 3D目标框的筛选精度和效率。最后，在KITTI数据集中对改进前后的算法进行训练和验证，并可视化了实验结果。

(3) 进行实验验证。本次实验在校园的主干道上进行激光雷达和相机数据包的录制。在实验室内解析数据包，获取激光雷达点云数据和图像数据。将处理后的数据输入改进的多模态焦点稀疏卷积目标检测算法，可视化了实验结果，并对结果展开分析。改进后的多模态焦点稀疏卷积目标检测算法在对距离较远的小尺寸目标以及遮挡严重的目标均有良好的检测效果。

**关键词：**激光雷达，相机，多模态焦点稀疏卷积，非极大值抑制，目标框筛选

# RESEARCH ON VEHICLE ENVIRONMENT PERCEPTION METHOD BASED ON LIDAR AND VISION FUSION

## ABSTRACT

With the rapid development of artificial intelligence and robotics, intelligent driving technology has become a global research hotspot. Intelligent driving vehicles aim at real-time, efficient and reliable perception of the environment around the vehicle by carrying sensors such as LiDAR, millimeter wave radar, cameras, etc., which provides the basis for vehicle control and decision-making and realizes autonomous driving of the vehicle. Target detection is an important means for vehicles to perceive the surrounding environment, which is a prerequisite to ensure the safe and stable operation of intelligent driving vehicles. Most of the traditional target detection algorithms are based on a single sensor, and it is difficult to complete the perception of the environment around the vehicle in complex road conditions or bad weather conditions, which in turn affects the safe operation of intelligent driving vehicles. For this reason, this paper adopts the scheme of fusion detection of LiDAR and camera, and utilizes the complementary nature of signals collected by different sensors to study the target detection algorithm of multimodal data fusion, and improves a multimodal focus sparse convolution target detection algorithm. The main work of this paper is as follows:

(1) Aiming at the problem of coordinate system conversion and alignment between different types of sensors, the relationship of mutual conversion between multiple coordinate systems is derived. Based on the rectangular checkerboard grid calibration method, the internal and external parameters of the camera are obtained. On this basis, the joint calibration of LiDAR and camera is carried out to lay the foundation for sensor data fusion.

(2) Aiming at the problem that the multimodal focus sparse convolution target detection algorithm has insufficient detection accuracy for small-size targets, the multimodal focus sparse convolution target detection algorithm is improved. First, the 2D backbone network part of the algorithm is optimized, and a multi-scale feature extraction network is proposed so that the algorithm can focus on small-size targets.

Second, the principle of the traditional non-maximal suppression NMS algorithm is analyzed and the Soft-NMS algorithm is introduced for the problem of insufficient target frame screening accuracy of the multimodal focus sparse convolutional target detection algorithm, and the confidence level is dynamically adjusted by using Gaussian weighting coefficients. Then, to address the problem of a single discriminative criterion of traditional target screening algorithms, combined with the characteristics of 3D target detection frames, the 3D-IoU pose estimation loss function is introduced to improve the screening accuracy and efficiency of 3D target frames. Finally, the algorithm before and after the improvement is trained and verified in KITTI dataset, and the experimental results are visualized.

(3) Perform experimental validation. This experiment was conducted on the main road of the campus to record LIDAR and camera packets. The packets are parsed in the lab to obtain the LIDAR point cloud data and image data. The processed data were fed into the improved multimodal focal sparse convolutional target detection algorithm, the experimental results were visualized, and the results were analyzed. The improved multimodal focus sparse convolution target detection algorithm has good detection effect on small-sized targets at a long distance and heavily occluded targets.

**KEY WORDS:**LiDAR, camera, Multimodal Focused Sparse Convolution, non-maximal suppression, target frame screening

# 目录

# 第1章　绪论

## 1.1 研究背景和意义

根据中国汽车工业协会的相关数据，截至 2023 年，全国汽车的生产量和销售量分别为 3016.1 万台和 3009.4 万台，我国汽车保有量 4.35 亿台[1]。随着我国城市化快速发展，机动车的数量不断增加，这不仅造成城市交通拥堵和环境污染等问题，还给人们的生命、健康和财产安全带来了损失。根据国家统计局发布的数据，2021 年全国道路交通事故共发生 256409 起，死亡人数 60676 人，其中绝大多数交通事故都是由机动车驾驶人违规操作引起的，如车辆超速、超载、酒驾和分散注意力等[2]。

为了减少交通事故所造成的人员伤亡和财产损失，各国都在进行相关科学研究，以期通过技术进步减少车祸发生率，缓解道路交通拥挤等问题。当前备受关注的智能驾驶系统相较于机动车驾驶人具有运行更稳定、反应速度更快等优点，可以有效减少交通事故的发生。由于事故率降低，智能驾驶车辆可以同车道内的其他车辆保持安全距离，更好地适应道路交通管理的需要，减少交通拥堵，提高道路的通行效率[3]。

通常情况下，搭载智能驾驶系统的车辆通过感知自身周围的环境，规划出最佳行驶路线，从而实现安全驾驶[4]。换而言之就是汽车通过智能驾驶系统获取并分析车辆周围的环境信息，为用户提供最佳驾驶路径和行驶策略，进而达到智能驾驶的目的。

在恶劣天气以及驾驶场景复杂的情况下，单个传感器难以满足环境感知需求。汽车、行人、自行车是车辆行驶过程中最常遇到的障碍物，也是最容易引发交通事故的不确定因素[5]。因此，对车辆、行人以及自行车进行检测是智能驾驶环境感知的重点[6]。当前，为了在实际道路场景中检测车辆周围的障碍物，完成车辆周围的环境感知，通常采用多传感器融合技术[7]。

多源传感器数据融合技术是提高智能驾驶汽车行驶安全性和稳定性的关键技术。不同传感器采集的数据具有多样性，多源传感器数据融合能够利用信息

间互补的特性，提升车辆感知周围环境的能力[8]。

在环境感知领域中，常见的传感器有超声波雷达、激光雷达、毫米波雷达、红外线和视觉相机等，每种传感器都有其独特的优势[9]。在众多的多源传感器配置方案中，相机与激光雷达的组合能够提供包括空间位置信息、颜色信息等多个维度的信息，从而显著增强车辆感知周围环境的能力，并受到广泛的关注[10]。因此，基于激光雷达和视觉融合的车辆环境感知研究对智能驾驶汽车具有重要意义[11]。

## 1.2 国内外研究现状

### 1.2.1 基于单一传感器目标检测研究现状

相机作为智能驾驶感知系统中最常见的的传感器，通常会在车辆上配置多个相机，以满足多样化的任务需求[12]。在众多的检测技术中，二维图像检测技术已相对成熟，能有效识别车辆周围的障碍物、可行驶区域以及交通标识等[13]。二维图像检测算法主要分为传统图像处理方法和基于深度学习的方法[14]。传统图像处理方法侧重于从图像中提取关键特征点，如边缘点或角点[15]。但这类方法往往在分类任务中表现不佳，产生大量无关的检测结果。而基于深度学习的方法，通过简化流程实现了从输入到输出的端到端检测[16]。这种方法具有很高的泛化性能，是当前的研究热点。

基于深度学习的方法分为单阶段目标检测和多阶段目标检测[17]。其中，单阶段目标检测仅通过一次特征提取就能完成目标检测任务[18]。而多阶段目标检测算法需要先进行特征提取，然后对提取的目标特征进行框选以完成检测[19]，相比单阶段目标检测算法在检测精度上有明显的优势，但检测速度较慢[20]。Faster-RCNN[21]算法是典型的多阶段目标检测算法，该算法首先通过特征提取网络提取目标特征，然后通过区域建议网络生成目标的候选区域，最后将这些目标的候选区域进行分类和检测框回归。然而，这种算法的缺点在于需要区域建议网络，导致推理速度较慢[22]。以SSD[23]和YOLO[24]为代表的单阶段目标检测算法则可以不通过生成目标候选区域，直接从图像中生成物体的坐标和类别。因此单阶段的目标检测算法具有检测速度快，但相对精度较低等特点[25]。

2015 年，He[26]等人为了解决深度学习过程中容易出现的梯度消失问题提出了ResNet网络模型，该网络模型一经提出就在学术界和工业界得到了广泛的应用，至今仍然广泛应用于计算机视觉领域。2017 年，Howard等人[27]提出了一种适用于移动终端和嵌入式设备的轻量级网络MobileNet。该网络通过引入深度可分离模块，有效的减少了网络参数量[28]。2020 年，Zhang等人[29]提出了不同规模大小的ResNeSt网络，其中具有代表性的有ResNeSt-50、ResNeSt-101、ResNeSt-256。ResNeSt将信道式注意力引入不同的网络分支中，充分利用特征图的注意力和多路径表示的互补优势[30]。

点云是由三维空间中的点所组成的，一种方法是利用点云的三维特性将其投影至二维平面进行目标检测[31]。M.Simon等人[32]正是利用点云的三维特性提出了一种快速的 3D目标检测网络Complex-YOLO，通过对原始点云数据进行预处理，并生成点云的鸟瞰图，然后利用卷积网络从点云生成的鸟瞰图中提取目标特征，最后通过区域建议网络生成目标候选区域。Yang Bin等人[33]提出了一种基于点云的单阶段检测网络PIXOR，通过骨干网络提取点云俯视图的特征并使用预测分支进行目标框选。另一种方法是对三维点云进行体素化，将得到的点云体素网格进行编码得到点云 4 维稀疏特征，然后利用稀疏卷积网络进行点云特征提取[34]。Zhou Yin等人[35]提出的VoxelNet网络首先将稀疏不规则的点云进行体素化，并将其进行特征编码，并通过全连接层和最大池化层将点云的局部特征与原始特征进行拼接，得到聚合后的特征，最后使用区域建议网络生成目标候选区域。Yan Yan等人[36]在VoxelNet网络的基础之上提出了SECOND网络，旨在将VoxelNet中的点云特征提取部分的卷积替换为 3D稀疏卷积，提升网络的检测能力。

### 1.2.2 多传感器融合研究现状

虽然基于传感器检测算法的性能不断提升以及硬件可靠性不断增强，但是基于单一传感器的目标检测仍然难以应对复杂条件下车辆环境感知的要求[37]。例如，视觉传感器可以精确、丰富地捕捉目标的二维信息，但是却不能精确地获取目标的坐标及尺寸等三维信息。与之相比，激光雷达在获取目标物体的坐标及尺寸等三维信息方面具有显著优势，但其在目标分类以及目标颜色的获取

等方面却难以与视觉传感器相提并论[38]。为了全面、精确地获取目标的信息，研究人员将多个传感器进行组合，利用多模态数据融合技术增加信息获取的维度[39]。

多源传感器融合的方式按照传感器的结构布局可以划分为分布式、集中式和混合式三大类[40]。在分布式结构中，每个传感器对应一个小型处理器[41]，负责对原始信息进行处理，然后将处理后的信息送入中央处理器进行融合；在集中式结构中，将原始数据直接输入到中央处理器进行处理[42]；在混合式结构中，部分传感器会对原始数据进行预处理，其余部分直接输入到中央处理器与预处理信息进行融合[43]。智能驾驶汽车受到整车体积以及供电等诸多因素限制，绝大多数都采用集中式结构进行信息融合[44]。按照传感器数据融合的具体实现方式和层次，又可细分为决策级融合、特征级融合以及数据级融合。

Cvejic等人[45]通过将可见光与红外监控摄像机视频进行像素级融合，研究其对目标跟踪能力的影响。Chongjian Yuan[46]等人提出了一种新的相机与雷达校准的方法，通过将两个传感器捕捉的自然边缘特征进行对齐来实现像素级精度，进而使高分辨率的激光雷达和RGB相机在在无目标的环境中进行自动外部校准。Shuo Gu等人[47]提出了一种将三维点云投影到相机坐标系中利用距离和颜色信息，由点云反深度诱导的道路检测框架。Xiangmo Zhao等人将两种传感器采集的信息进行融合，利用点云数据定位目标的候选区域，然后将目标候选区域映射至图像空间，并从中选取感兴趣区域，利用卷积神经网络进行目标识别[48]。

## 1.3 研究内容及章节安排

本文基于多源传感器数据优势互补特性，将多传感器捕捉的目标信息进行融合，为车辆准确地感知周围环境提供信息支持。为解决单一传感器在对障碍物检测过程中因存在信息缺失而导致小尺寸目标检测精度较低的问题。首先，以多模态焦点稀疏卷积目标检测算法为主体框架，结合多模态数据融合技术，提出了一种多尺度特征提取网络；其次，分析了传统非极大值抑制NMS算法的原理并针对多模态焦点稀疏卷积目标检测算法目标框筛选精度不足的问题，引入了高斯加权系数调整置信度，并根据重叠部分的面积动态确定高斯加权系数；

然后，针对传统的目标筛选算法判别标准单一的问题，结合 3D目标检测框的特性，引入了 3D-IoU姿态估计损失函数提升 3D目标框的筛选精度和效率。此外，本次实验采用激光雷达、相机、计算机以及相关软件，搭建实验平台，并利用KITTI数据集和实车采集的数据分别对算法进行验证。

根据以上研究内容，本文的内容章节安排如下：

第 1 章：绪论。首先通过分析中国的道路交通安全情况，阐述智能驾驶系统的研究背景和意义；然后分别介绍了智能驾驶系统中的两种目标检测方案的研究现状；最后，阐述了本文研究的主要内容及工作安排。

第 2 章：传感器标定与融合方法。首先对本次实验的多源传感器配置方案进行了描述，介绍了激光雷达与相机的型号、参数以及工作原理；然后对激光雷达坐标系、相机坐标系、世界坐标系、图像坐标系以及像素坐标系之间相互转换的关系进行了推导；最后，介绍了相机的标定的过程，在此基础上描述了激光雷达与相机的联合标定。

第 3 章：基于多模态焦点稀疏卷积的目标检测算法。阐述了常规稀疏卷积、子流型稀疏卷积和焦点稀疏卷积的概念，并说明了非极大抑制算法的原理及其存在的问题、2D目标边框的融合方法和衡量边界框重叠的标准。对点云体素化的规则、编码方式进行了详细的阐述。本文基于多模态焦点稀疏卷积的目标检测算法在小尺寸目标检测中出现的漏检问题，提出了一种多尺度特征提取网络。针对传统的非极大抑制算法处理边界框时出现的漏检和误检问题，在 2D IoU边界框融合方法的基础上，引入了高斯加权系数，调整IoU的分数，并通过重叠部分面积动态调整高斯加权系数。针对传统的目标筛选算法判别标准单一的问题，结合 3D目标检测框的特性，引入了 3D-IoU姿态估计损失函数提升 3D目标框的筛选精度和效率。基于KITTI数据集进行目标框的筛选网络的搭建与训练，实验结果表明，改进算法在对“car”、“Pedestrians”和“Cyclist”这三个目标类别的mAP(mean Average Precision，平均精度)分别提升了 1%、1.41%、1.73%。

第 4 章：实验验证与结果分析。介绍了实验所用的设备及平台，设计了联合标定的节点以及环境感知节点，并在校园主干道进行数据包的采集并提取数据。利用改进的多模态焦点稀疏卷积目标检测算法实现点云数据与图像数据的

融合，以完成车辆周围的环境感知，并对结果进行可视化和探讨。改进后的多模态焦点稀疏卷积目标检测算法在对距离较远的小尺寸目标以及遮挡严重的目标均有良好的检测效果。

第 5 章：总结与展望。对本文的研究的工作进行总结，同时对以后的研究方向进行展望。

# 第2章　传感器标定与融合方法

对于智能驾驶技术中的多模态数据融合任务而言，传感器的选型、不同传感器特点的互补性是无人驾驶技术的硬件基础。只有充分了解传感器的特点，发挥传感器之间的优势互补作用，才能更好的提升感知系统的性能。此外，传感器的联合标定能有效解决两种不同类型传感器之间坐标转换和对齐的问题，确保它们能在同一坐标系下提供的数据可以准确的进行融合和分析。

## 2.1 多源传感器配置方案

在智能辅助驾驶技术领域，传感器的选择对系统性能至关重要。目前主流的传感器有激光雷达、摄像头、超声波雷达、红外线传感器和毫米波雷达等。虽然激光雷达具有高精度的优势，但是其成本较高，大多数辅助驾驶系统更倾向于采用摄像头、毫米波雷达和超声波雷达。在一些高端汽车型号中，如智界S7、星际元ES、蔚来ET9 等，通常会采用激光雷达与摄像头相结合的配置方案，这为更高级别的驾驶辅助功能提供了可能性。

事实上，不同雷达传感器产生的点云数据具有一定的相似性。相比之下，摄像头采集的RGB数据与点云数据之间存在较大的差异。因此，研究激光点云数据与视觉RGB数据之间的融合，能够更好地整合两种传感器的优势，提高系统在复杂驾驶场景下的感知和决策能力，从而推动无人驾驶技术的发展。

### 2.1.1 激光雷达传感器

雷达传感器是人工智能技术中最常用到的传感器之一，在自动驾驶、机器人、测绘和环境监测等领域广泛应用。激光雷达能够提供准确的距离信息，高分辨率的地图和对周围环境的详细感知，是实现精准导航和避障的关键技术之一。常见的雷达传感器有毫米波雷达、超声波雷达、激光雷达(Light Detection And Ranging，LiDAR)等。激光雷达有多种类型，包括机械扫描雷达、固态激光雷达。机械扫描雷达通过旋转或倾斜的机械部件扫描激光束，而固态激光雷达则利用电子控制并调整激光束的发射[49]。激光雷达通常由激光发射器、接收

器、扫描机构和数据处理单元组成。激光雷达是通过激光光束测量距离和创建高分辨率三维地图的感知技术。激光雷达的工作原理是基于激光脉冲的发射和接收，利用光的飞行时间来计算目标的距离。激光雷达的工作流程如下：激光发射器向目标发送激光脉冲，激光束照射在目标表面后反射回来，接收器接收反射的激光，并通过记录光脉冲信号从发射到接受反射所需时间(即介质飞行时间，ToF)，来计算目标距离。通过控制激光束的方向，激光雷达可以在水平和垂直方向上扫描周围环境，生成点云数据。因此，本文选用镭神C-16 机械式激光雷达进行实验，激光雷达如图 2-1 所示，主要参数如表 2-1 所示。

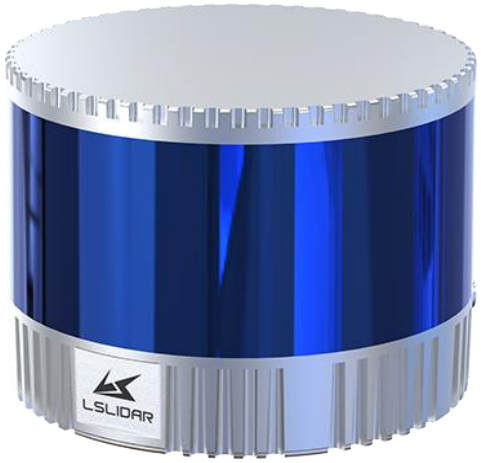



图 2-1 镭神 C-16 机械式激光雷达

表 2-1 镭神 C-16 激光雷达参数

| 型号 | C-16 |
|---|---|
| 测距方式 | 脉冲式 |
| 激光波段 | 905nm |
| 激光等级 | 1 级(人眼安全) |
| 激光通道 | 16 路 |
| 信号传输方式 | 无线功率与信号传输 |
| 最大测程 | 50 米 |
| 最小测程 | 0.5 米 |
| 测距精度 | ±10cm |
| 数据获取速度 | 最高 32 万点每秒 |
| 垂直视场角 | ±15º |
| 水平视场角 | 360º |

激光雷达在对目标进行扫描时，由于光的传播速度非常快，因此测量目标

用时比较短。通过控制设备，使激光发射装置和激光接收装置以一定频率旋转，进而可以得出激光雷达扫描角度范围内目标的距离[50]。通过这种扫描方式获取空间中目标的信息，由许多三维空间中的点所组成的集合称为点云。

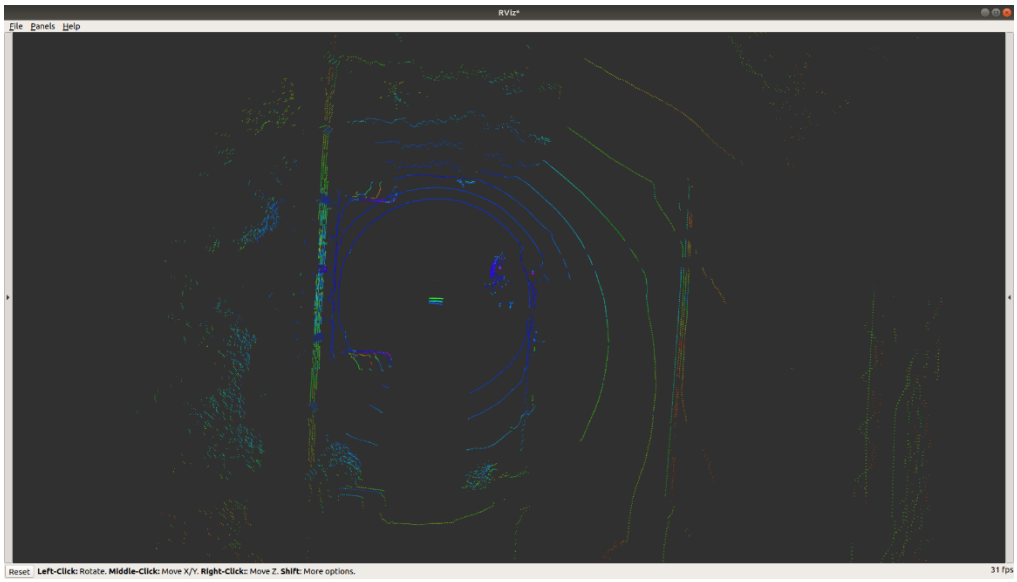



图 2-2 激光雷达采集点云

### 2.1.2 视觉相机

视觉相机是一种通过光学传感器捕捉视觉信息的设备，广泛应用于各种领域，包括智能驾驶、监控系统、机器视觉和虚拟现实等。车规级相机是自动驾驶技术中至关重要的传感器之一，是一种用于采集道路和车辆周围环境的信息的关键感知设备，通常采用计算机视觉技术与其他传感器协同工作，帮助自动驾驶系统实现环境感知、目标检测、车道保持等功能。在自动驾驶车辆传感器配置中，车规级相机已成为主流选择。一方面，这些相机不仅在前视方向应用广泛，还涵盖了环视等多种用途，力图实现对车辆周围环境的全方位 360°覆盖，这种全面的视野对于自动驾驶系统的感知和决策至关重要；另一方面，与激光雷达相比，视觉传感器可以获取更过的细节信息，如纹理、色彩等。视觉传感器对信号灯和交通标识的检测更为敏感，这使得车辆对目标进行准确的识别与分类成为可能。

本次实验选用的是奥比中光的Astra Pro相机。Astra Pro是一款多功能用途的相机，它既能通过光学传感器捕捉场景中的RGB信息，又能通过发射光条并

测量其在场景中的形变获取场景的深度信息。表 2-2 为Astra Pro相机的主要参数。图 2-3 为相机采集图像。

表 2-2 Astra Pro 相机的主要参数

| 参数 | 数据 |
| --- | --- |
| 型号 | 奥比中光Astra Pro |
| 帧率 | 30 帧/秒 |
| 分辨率 | 640*480 |

图 2-3 相机采集图像

## 2.2 坐标系转换及传感器标定

无人驾驶技术多采用“雷达+相机”的解决方案，激光雷达能提供点云数据，相机能提供高分辨率的RGB信息，两种类型数据融合之后能够更加准确、高效的实现环境感知，而激光雷达与相机的联合标定是二者融合的基础。激光雷达与相机联合标定能够较好的解决两种不同类型传感器之间坐标转换和对齐的问题，确保它们在同一坐标系下提供的数据可以准确的进行融合和分析。

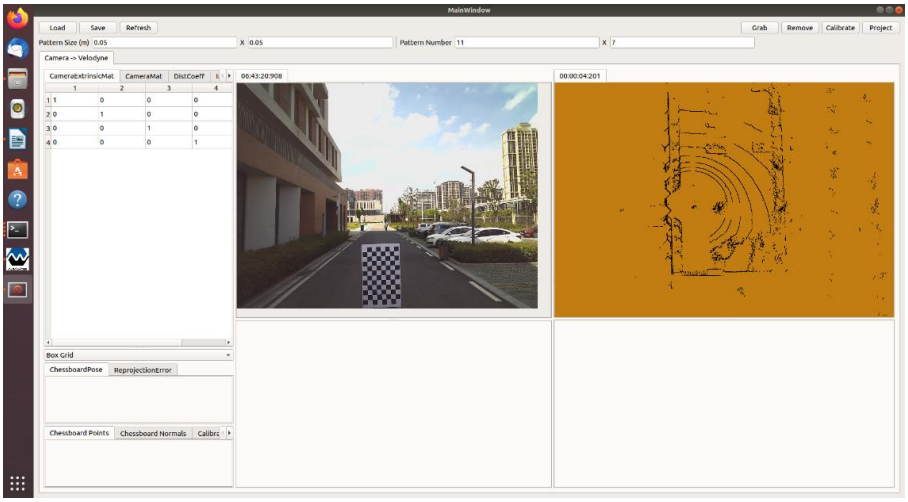



图 2-4 传感器标定过程

### 2.2.1 坐标系选取

为了将激光雷达采集的点云数据与相机采集的RGB数据进行对应，首先需要解决坐标系之间相互转换的问题，获取二者空间中的相对关系。其中涉及的坐标系包括激光雷达坐标系 $O_L-X_LY_LZ_L$ 、相机坐标系 $O_C-X_CY_CZ_C$ 、图像坐标系 $O_T-X_TY_T$ 、像素坐标系 $O_P-u_Pv_P$ 。融合坐标系转换图如图 2-5 所示。

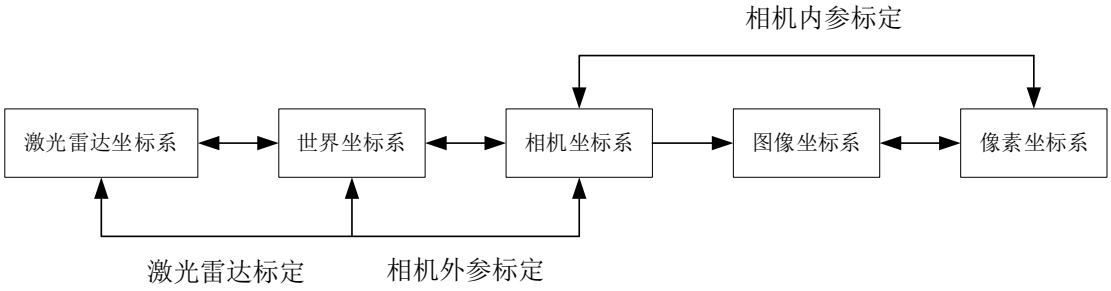



图 2-5 融合坐标系转换图

由图 2-5 可知，为确保激光雷达与相机坐标系在时空中的一致性，需要对 $O_L-X_LY_LZ_L$ 和 $O_C-X_CY_CZ_C$ 进行相互转化以实现信息匹配。转换流程如下：

(1) 世界坐标系是基于真实的空间测量数据建立的，用来准确的反应传感器与目标障碍物之间的相对位置关系，用*O-XYZ*表示。

(2) 激光雷达坐标系是指在三维空间中采集数据时所使用的坐标系，以激光雷达自身为原点，选取车辆正前方为激光雷达 $X_L$ 轴的正方向；垂直于车辆前方且水平向左的为 $Y_L$ 轴正方向；垂直车辆向上为 $Z_L$ 轴正方向。

(3) 相机坐标系是用来描述相机位置和方向的坐标系，相机坐标系的原点为

相机的光学中心，$X_C$ 轴正方向水平向右；$Y_C$ 轴正方向垂直向下；$Z_C$ 轴为车辆前进方向。

(4) 图像坐标系是用来描述数字图像中像素位置的坐标系，通常以图像的几何中心为坐标原点，沿着图像的水平中心线向右为 $X_T$ 轴正方向；沿着图像的垂直中心线向下且垂直于地面为 $Y_T$ 轴正方向。

(5) 像素坐标系同样是用来描述数字图像中像素位置的坐标系，通常选取图像左上角的顶点为坐标原点，水平向右为$u$轴正方向；垂直向下为$v$轴正方向。

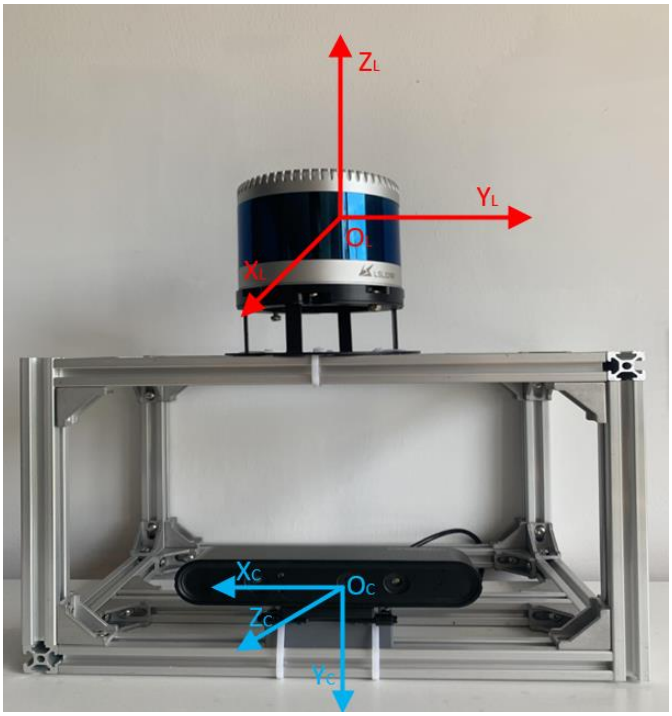



图 2-6 坐标系建立

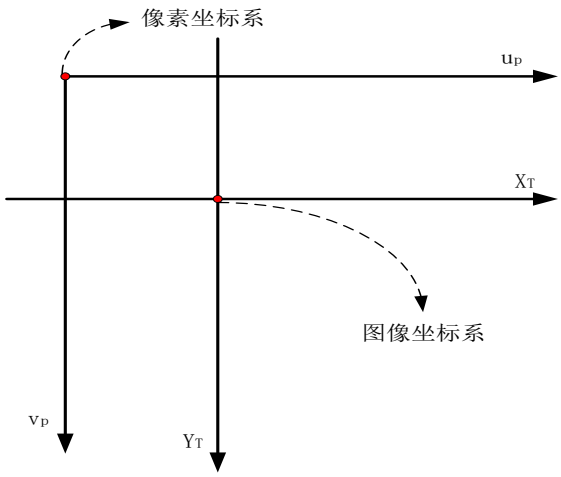



图 2-7 图像与像素坐标系

### 2.2.2 激光雷达坐标系与世界坐标系之间的转换

由于激光雷达通过激光发射器向目标发送激光脉冲，激光束照射在目标表面后反射回来，接收器接收反射的激光，获取三维空间中的信息，而世界坐标系也是三维空间坐标系。因此需在三维空间中对坐标系进行平移与旋转等变换实现激光雷达坐标系与世界坐标系间的相互转化。

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R \begin{bmatrix} x_L \\ y_L \\ z_L \end{bmatrix} + t \tag{2-1}$$

其中，$R$为转换过程中的旋转矩阵，t为转换过程中的平移向量。由式(2-1)可知，当激光雷达坐标系$O_L$-$X_LY_LZ_L$中存在一点$L(x_L,y_L,z_L)$，其在世界坐标系$O$-$XYZ$中的坐标可表示为(x,y,z)。

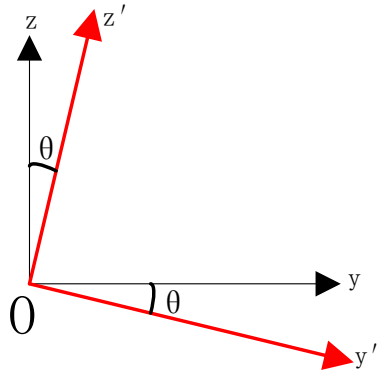



图 2-8 绕 x 旋转过程示意图

$$Y_L^{'} = Y_L \cos\theta - Z_L \sin\theta \tag{2-2}$$

$$Z_L^{'} = Y_L \sin\theta + Z_L \cos\theta \tag{2-3}$$

其矩阵形式为：

$$\begin{bmatrix} X_L' \\ Y_L' \\ Z_L' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} X_L \\ Y_L \\ Z_L \end{bmatrix} \tag{2-4}$$

同理，绕y轴和z轴之后矩阵表示为：

$$\begin{bmatrix} X_L' \\ Y_l' \\ Z_L' \end{bmatrix} = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \begin{bmatrix} X_L \\ Y_L \\ Z_L \end{bmatrix} \tag{2-5}$$

$$\begin{bmatrix} X_L' \\ Y_L' \\ Z_L' \end{bmatrix} = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_L \\ Y_L \\ Z_L \end{bmatrix} \tag{2-6}$$

将绕三个方向旋转得到的分量相乘可得到两个三维坐标系之间的旋转矩阵：

$$R = \begin{bmatrix} \cos\beta\cos\theta & -\sin\theta\cos\beta & \sin\beta \\ \sin\theta\sin\beta\cos\alpha + \cos\theta\sin\alpha & \cos\theta\cos\alpha - \sin\theta\sin\beta\sin\alpha & -\sin\theta\cos\beta \\ \sin\theta\sin\alpha + \cos\theta\cos\alpha\cos\beta & \cos\theta\sin\beta\sin\alpha & \cos\beta\cos\theta \end{bmatrix} \tag{2-7}$$

式(2-7)的齐次坐标表示形式如式(2-8)所示，其中 $0 = [0 \quad 0 \quad 0]^T$ 。

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{\mathrm{L}} \\ y_{\mathrm{L}} \\ z_{\mathrm{L}} \\ 1 \end{bmatrix} \tag{2-8}$$

### 2.2.3 相机内参标定

在对激光雷达和相机进行联合标定时，首先要获取相机的内部参数矩阵。本文利用激光雷达坐标系和世界坐标系转换相类似的旋转、平移等原理，推导相机坐标系与世界坐标系之间的转换关系，具体转换关系如下：

(1) 世界坐标系 $O-XYZ$ 与相机坐标系 $O_C-X_CY_CZ_C$ 的转换关系

相机坐标系和世界坐标系之间的转换关系与激光雷达坐标系和世界坐标系之间的转换关系采用的是相类似的旋转、平移转换原理。假定在世界坐标系中的一点 $p$ 的齐次坐标为 $(x,y,z,1)$，在相机坐标系中有与之相对应的一点 $p_0$ 的齐次坐标为 $(x_c,y_c,z_c,1)$。于是，可得相机坐标系与世界坐标系之间的空间转换关系如式所示。

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} \mathrm{R}_1 & t_1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{2-9}$$

式(2-9)中，$R_1$ 为 3×3 的单位正交旋转矩阵，$t_1$ 为平移向量。

(2) 世界坐标系 $O-XYZ$ 与图像坐标系 $O_T-X_TY_T$ 的转换关系

透视投影关系实际上是指相机所拍摄图像的二维平面坐标系与相机本身三维空间坐标系之间的转换关系，也被称为小孔成像的投影变换关系。假定空间中有任一点 $A$，其在相机所在三维空间坐标系中的坐标可表示为 $A(X_C,Y_C,Z_C)$，

其在图像坐标系 $O_T-X_TY_T$ 中的坐标表示为 $B(X_T,Y_T)$，小孔成像的投影原理如图2-9所示，图中 $f$ 为相机的焦距。

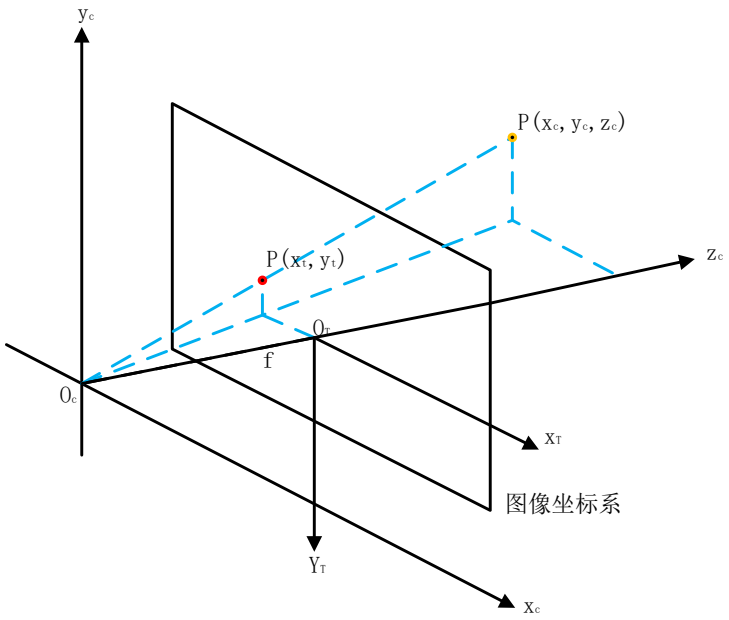



图 2-9 小孔成像原理

由相似三角形定理可得式(2-10)：

$$\begin{cases} X_t = f\dfrac{X_t}{Z_v} \\ Y_t = f\dfrac{Y_t}{Z_c} \end{cases} \tag{2-10}$$

式(2-10)齐次矩阵的表达形式为：

$$Z_c\begin{bmatrix} X_T \\ Y_T \\ 1 \end{bmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \tag{2-11}$$

(3) 像素坐标系 $O_P-u_Pv_P$ 和图像坐标系 $O_T-X_TY_T$ 的转换关系

在理想状态下，像素坐标系 $O_P-u_Pv_P$ 和图像坐标系 $O_T-X_TY_T$ 是重合的，但在实际过程中，由于安装过程中会产生误差，导致两坐标系不会完全重合。因此需要计算图像坐标系到像素坐标系的坐标平移矩阵。

$$\begin{cases} u_p = \dfrac{X_i}{dx} + u_0 \\ v_p = \dfrac{Y_i}{dy} + v_0 \end{cases} \tag{2-12}$$

式(2-12)中，图像坐标系的原点在像素坐标系中的坐标为$(u_0,v_0)$，$dx$与$dy$为单个像素在$X_iY_i$坐标轴上的大小。将其转换成齐次矩阵可表示为：

$$\begin{bmatrix} u_p \\ v_p \\ 1 \end{bmatrix} = \begin{pmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \tag{2-13}$$

(4) 像素坐标系$O_P-u_Pv_P$和世界坐标系$O-XYZ$的转换关系

将式(2-11)代入式(2-13)可得相机坐标系与像素坐标系的转换公式为：

$$Z_C \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_1 \\ 0 & \frac{1}{dy} & v_1 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{bmatrix} x_C \\ y_C \\ z_C \\ 1 \end{bmatrix} \tag{2-14}$$

即

$$Z_C \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{f}{dx} & 0 & u_1 & 0 \\ 0 & \frac{f}{dy} & v_1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_C \\ y_C \\ z_C \\ 1 \end{bmatrix} \tag{2-15}$$

在式(2-15)中，矩阵$A=\begin{bmatrix} \frac{f}{d_x} & 0 & u_1 \\ 0 & \frac{f}{d_y} & v_1 \\ 0 & 0 & 1 \end{bmatrix}$为相机的内参矩阵。将式(2-8)代入式(2-15)得到世界坐标系与像素坐标系之间的转换关系：

$$Z_C \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{f}{dx} & 0 & u_1 & 0 \\ 0 & \frac{f}{dy} & v_1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ y \\ z \end{bmatrix} \tag{2-16}$$

相机内部参数能够确定三维空间坐标到二维图像坐标的投影关系，同时，相机在世界坐标系中的具体位置由其外部参数决定。相机的内部与外部参数，

均可通过相机标定过程获得。

张正友相机标定法[51](Zhang's Camera Calibration Method)是最常见的相机内参标定法，用相机拍摄不同角度和位置下的矩形棋盘标定板，从拍摄的图像中提取特征点信息，通常为棋盘格的交叉点或标志的角点，根据棋盘格的实际大小和图像中提取的特征点信息建立三维空间坐标系与图像坐标系之间的对应关系，通过最小化图像坐标系中检测到的角点和世界坐标系中对应的真实坐标之间的重投影误差，求解相机的内部参数和外部参数。

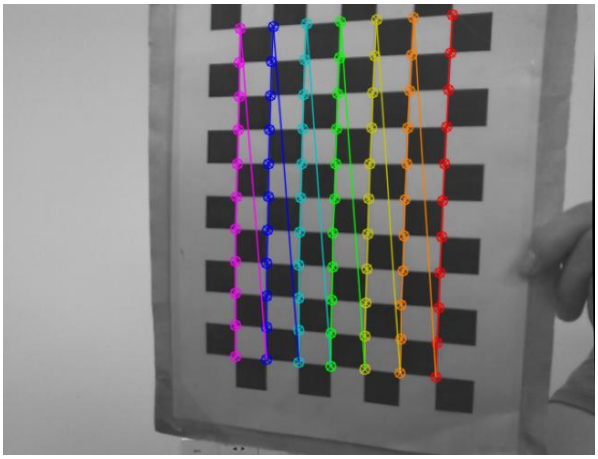


图 2-10 相机内参标定

```
width
640

height
480

[autoware_camera_calibration]

camera matrix
520.528666 0.000000 315.779702
0.000000 519.256271 243.125909
0.000000 0.000000 1.000000

distortion
0.081500 -0.364745 -0.002034 -0.002670 0.290581

rectification
1.000000 0.000000 0.000000
0.000000 1.000000 0.000000
0.000000 0.000000 1.000000

projection
513.128357 0.000000 313.731025 0.000000
0.000000 513.916626 242.237678 0.000000
0.000000 0.000000 1.000000 0.000000
```

图 2-11 相机的内部参数

本次实验正是采用上述的张正友相机标定法，通过基于Ubuntu18.04 开源的Autoware14.0 平台进行相机内外参标定以及联合标定。在实际的标定过程中，

手持标定板距离相机在 1-2 米位置，调整标定板的方向与角度保证标定的准确性 ，分别将标定板向上倾斜、向下倾斜、向左倾斜、向右倾斜以及正面面对相机，直至完成相机内参标定。

```
%YAML:1.0
---
CameraExtrinsicMat: !!opencv-matrix
   rows: 4
   cols: 4
   dt: d
   data: [ 6.9576710497410699e-01, 3.5868986121855795e-02,
       -7.1737127867690653e-01, -2.9575799578697864e-01,
       7.1812018473029382e-01, -1.4514392878282933e-02,
       6.9576772897447858e-01, 2.4075311383048165e-01,
       1.4544274436307791e-02, -9.9925109368664922e-01,
       -3.5856880053295015e-02, -1.7977480510191240e-01, 0., 0., 0., 1. ]
CameraMat: !!opencv-matrix
   rows: 3
   cols: 3
   dt: d
   data: [ 5.4440723976555569e+02, 0., 3.2653751827942790e+02, 0.,
       5.4573816187229045e+02, 2.3976638758617850e+02, 0., 0., 1. ]
DistCoeff: !!opencv-matrix
   rows: 5
   cols: 1
   dt: d
   data: [ 1.4651644277141793e-01, -3.8404544036889604e-01,
       -9.7166232890489999e-05, 4.0055255836540689e-03,
       3.1941618047184644e-01 ]
ImageSize: [ 640, 480 ]
ReprojectionError: 2.9199188918481905e-01
```

图 2-12 相机的外参标定文件

## 2.3 激光雷达与相机的联合标定

激光雷达与相机的联合标定是多模态数据进行融合的基础，通过联合标定能够获得激光雷达与相机之间的空间对应关系，从而能够将捕捉到的信息进行有效的融合和协同处理。

对于空间中的任意一点，在相机坐标系与激光雷达坐标系中均有与之对应的数据点。通过矩形棋盘标定板对其进行标定时，首先需要提取相机坐标系和激光雷达坐标系内相匹配的棋盘格的角点，进行二者之间的外参标定。然后，根据相机内部参数标定所得的内参矩阵，可以计算出点云数据在像素坐标系内所对应的点的位置。实际采集的目标点在像素坐标系中的坐标可以表示为$(u,v)$，其次坐标可表示为$(u,v,1)$，在激光雷达坐标系中存在其对应的点的坐标为$(X_L,Y_L,Z_L)$，其次坐标表示为$(X_L,Y_L,Z_L,1)$，由式(2-17)表示为：

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = A\begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix}\begin{bmatrix} X_L \\ Y_L \\ Z_L \\ 1 \end{bmatrix} \tag{2-17}$$

式(2-17)中 $A$ 为相机的内参矩阵， $R$ 、 $T$ 分别为旋转矩阵和平移矩阵。

本次实验采用基于Ubuntu18.04 的开源Autoware 14.0 软件和Calibration Tool Kit标定工具。启动激光雷达与相机，打开Autoware 14.0 软件，通过ROSBAG录制激光雷达与相机的标定数据包并保存；

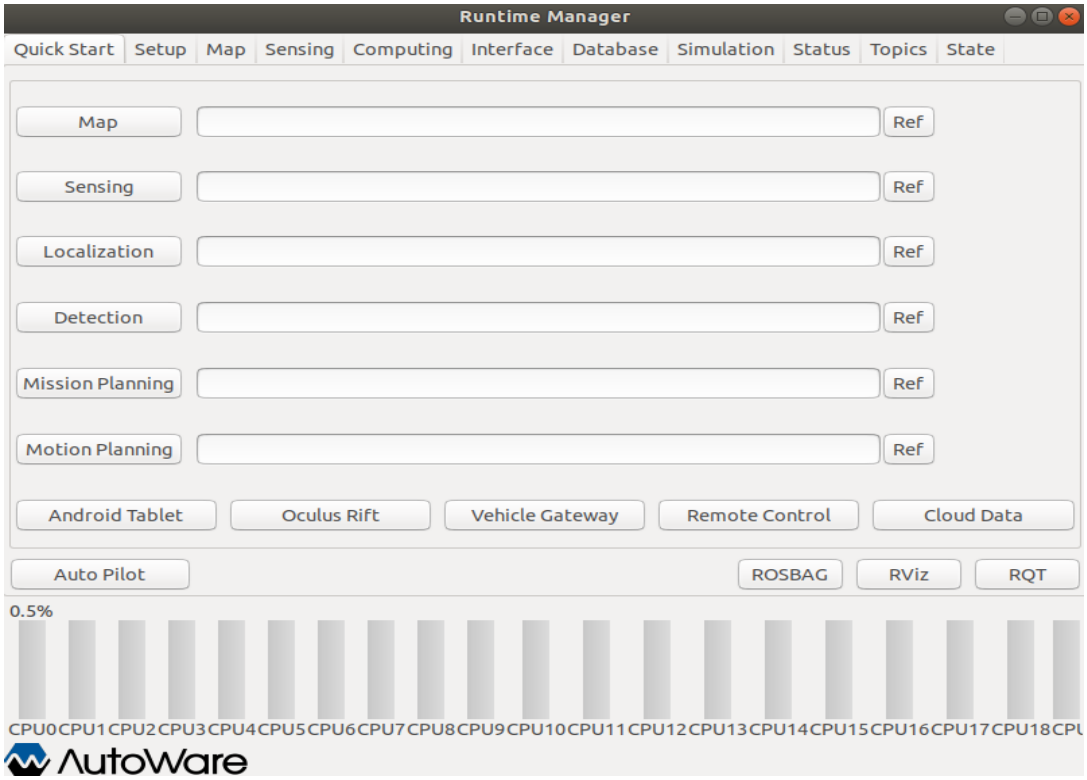



图 2-13 Autoware 14.0

(1) 打开Calibration Tool Kit标定工具，配置Calibration Tool Kit，依次选取/image_raw和Camera>Velodyne作为话题和标定类型，设置矩形棋盘标定板的整体尺寸 11*7 以及单个正方形的边长 50mm，导入相机标定的内参文件，重启配置；

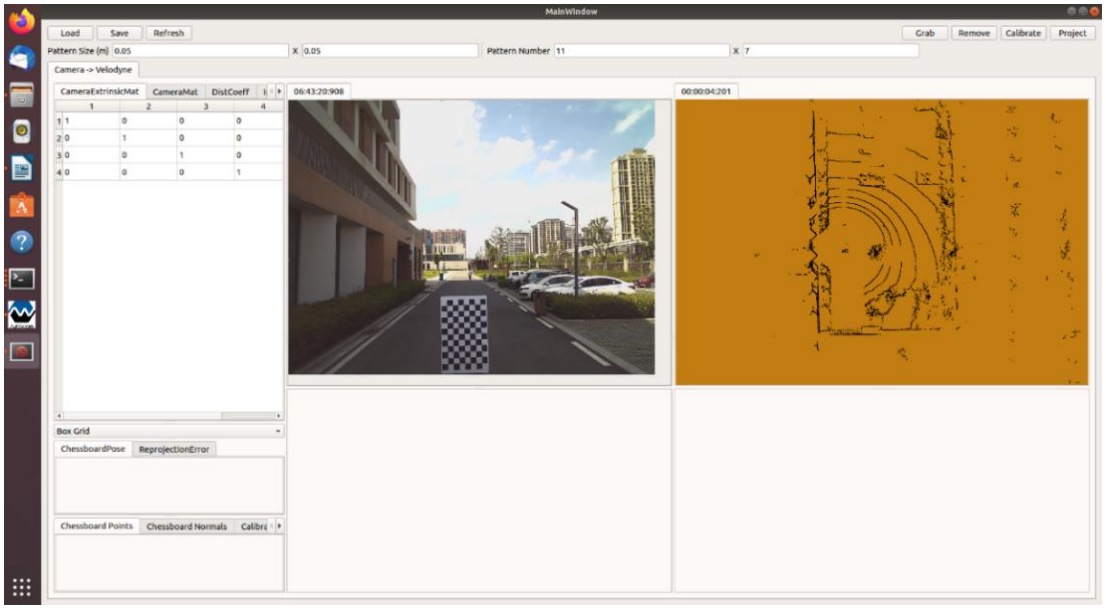



图 2-14 Calibration Tool Kit 标定工具

(2) 播放录制的数据包，选取点云数据中标定板的中心位置进行标记，重复多次进行直至完成标定；

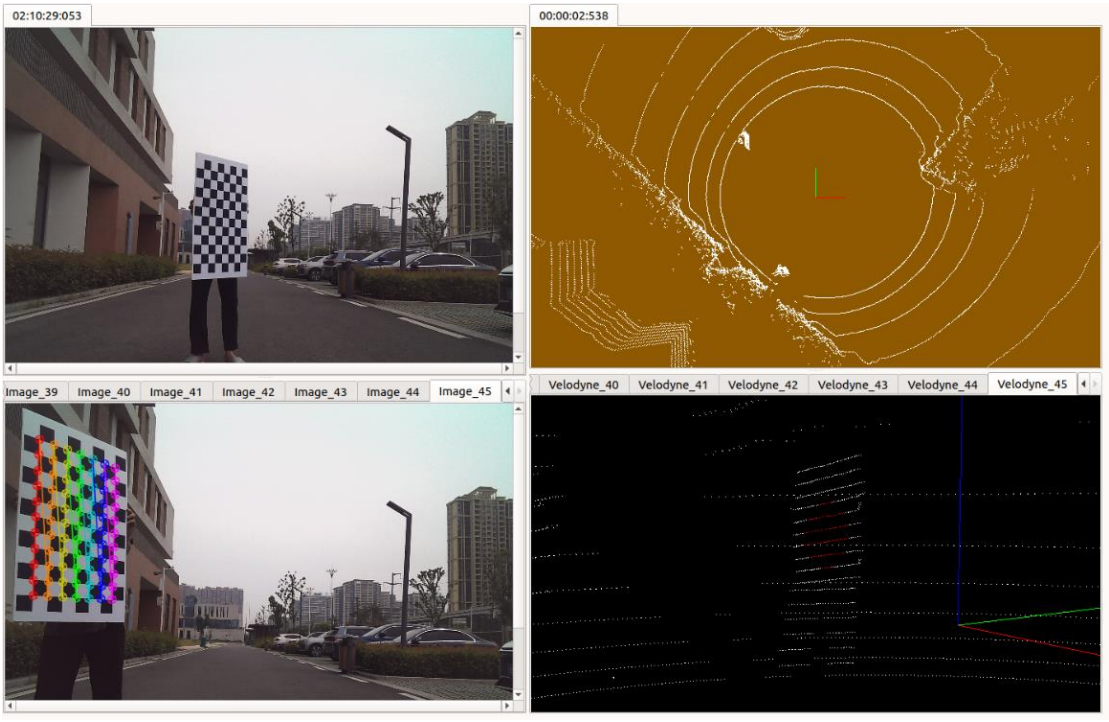



图 2-15 激光雷达与相机联合标定

(3) 标定结果

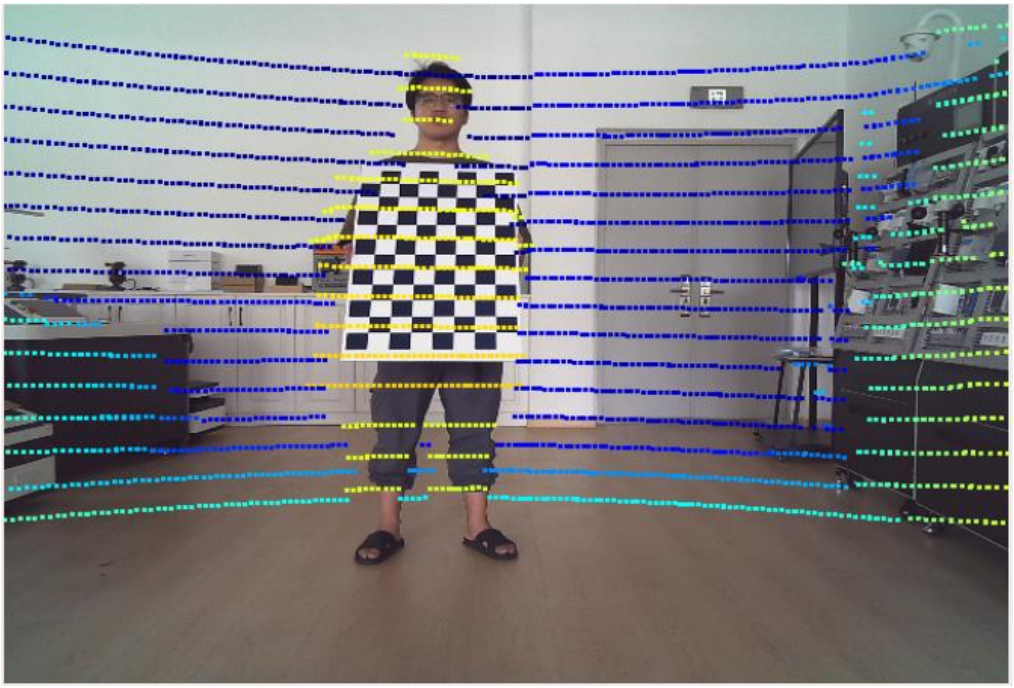


图 2-16 激光雷达与相机联合标定结果

## 2.4 多模态融合方案

多模态融合技术具有整合不同结构数据的能力，使各个传感器之间的优势相辅相成，得到充分的综合利用。多模态数据融合技术的核心在于对数据进行处理和融合，而不同的融合阶段也对整体效果产生重要影响。在深度学习领域，多源传感器融合通常被划分为三个关键层次。

首先，数据级融合发生在原始数据层面，旨在整合不同传感器采集的原始信息。通过这种方式，系统能够建立起一个更为丰富、全面的数据集，为后续处理提供更为有力的基础。

其次，特征级融合发生在特征提取之后，将每个传感器提取的特征进行融合。这个阶段的关键在于通过学习更高级别的抽象特征，使系统能够更深刻地理解多模态数据之间的内在关系。

最后，决策级融合发生在各部分结果产生后，通过整合这些结果做出最终的决策。这个层次的融合使得系统能够充分考虑每个传感器对任务的贡献，生成更为准确、鲁棒的综合决策。

### 2.4.1 数据级融合

数据级融合也被称为像素级融合，属于底层数据融合阶段。数据级融合要求多个传感器是同质的，即它们观测的是同一物理量，否则需要进行尺度校准。在数据级融合过程中，来自多个传感器的数据直接被融合；然后，对融合后的数据进行特征提取，进而实现目标识别。这种融合方式能够有效的避免数据丢失问题，融合的结果通常也很准确。不过，由于直接处理原始数据，会导致融合过程中的计算量较大，对系统通信带宽要求较高。

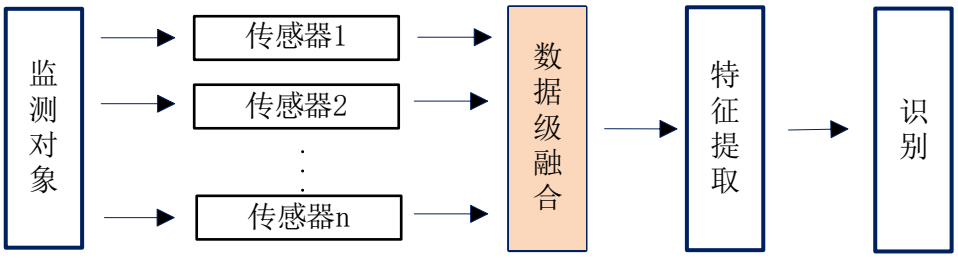



图 2-17 多传感器数据级融合示意图

### 2.4.2 特征级融合

特征级融合就是在多传感器融合过程中的一种中间层次的策略。在这个层次上，首先从提取每个传感器采集的原始数据中的特征；然后，将这些特征整合成向量，而选择哪些特征进行整合成为决策的关键。

特征级融合分为目标状态融合与目标特征融合两种类型。目标状态融合主要用于多传感器的目标跟踪，其过程包括对数据进行预处理以实现配准，然后着重于参数的关联和状态的估计。目标特征融合也称特征层融合识别，其本质上是一个模式识别的问题。在融合之前需要将特征进行关联，然后将这些特征向量进行分类以形成有意义的组合。

特征融合技术已相对发展成熟，得益于特征层的高效特征关联技术，能够确保信息融合的一致性。特征级融合在实际应用中取得了广泛的成功，尤其在深度分析和理解不同传感器提供的场景中发挥着至关重要的作用。

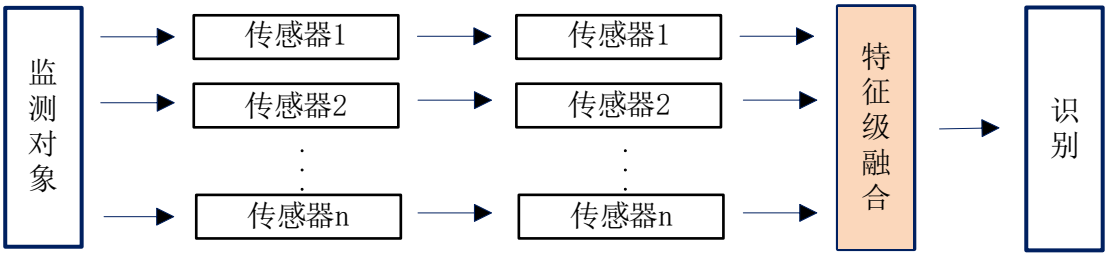



图 2-18 多传感器特征级融合示意图

### 2.4.3 决策级融合

决策级融合属于多传感器信息融合的高层次策略，其主要特点在于对数据进行高层次的抽象，产生一个更为全面、精确的联合决策结果。决策级融合在面对环境和目标的随时间的变化而发生改变、先验知识获取困难、知识库的信息量巨大等特性以及面向对象的系统设计需求等方面，具有优越的性能。联合决策的结果比任何单传感器决策更为精确。

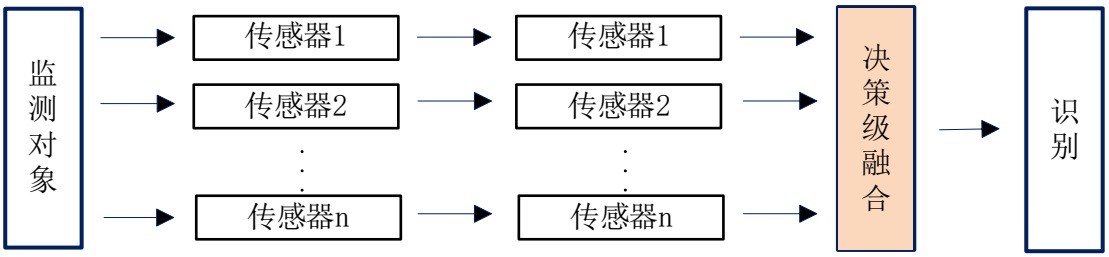



图 2-19 多传感器决策级融合示意图

## 2.5 本章小结

本章首先阐述了激光雷达和相机的原理、参数以及多传感器数据融合在智能驾驶环境感知领域中的优势。其次通过选取坐标系，推导了激光雷达坐标系、相机坐标系、图像坐标系、像素坐标系以及世界坐标系之间的转换关系。然后利用开源Autoware14.0 软件对相机进行了标定，获取了相机的内部参数矩阵，在此基础上进行了激光雷达和相机进行了联合标定，得到了联合标定的结果。最后阐述了多源数据融合的三种方式，数据级融合、特征级融合以及决策级融合。

# 第3章　基于多模态焦点稀疏卷积的目标检测算法

对于智能驾驶车辆准确的感知车辆周围所有的障碍物是保障车辆安全平稳运行的前提。本章首先针对多模态焦点稀疏卷积目标检测算法在对小尺寸目标检测精度不足的问题，在 2D主干网络部分提出了一种多尺度的特征提取网络。其次针对传统的非极大抑制算法处理边界框时出现的漏检和误检问题，在 2D IoU边界框融合方法的基础上，引入了高斯加权系数，调整IoU的分数，并通过重叠部分面积动态调整高斯加权系数。然后针对传统的目标筛选算法判别标准单一的问题，结合 3D目标检测框的特性，引入了 3D-IoU姿态估计损失函数提升 3D目标框的筛选精度和效率。最后基于KITTI数据集进行目标框的筛选网络的搭建与训练，并根据可视化结果和数据进行分析。

## 3.1 多模态焦点稀疏卷积目标检测算法相关原理

### 3.1.1 稀疏卷积

稀疏卷积是一种卷积操作，常用于处理具有稀疏性质的输入数据，在传统的卷积操作中，输入数据通常是密集的，即每个位置都有非零值。传统的卷积操作通常在规则的网格结构上进行，例如图像上的像素网格或文本上的字符序列。然而，对处理非规则的集合数据，如点云、曲面或三维模型，传统卷积并不适用。

常规稀疏卷积是一种应用于卷积神经网络的稀疏卷积操作，通过一些约束来促使卷积层学习到稀疏的特征表示。常规稀疏卷积的目标是在保持模型准确性的同时减少神经网络中的冗余连接和参数数量。有助于提高模型的计算效率和泛化能力，并减少过拟合的风险。常规稀疏卷方法可以在不损失模型性能的前提下，减少参数数量和计算的复杂度，提高模型的解释性和可解释性，增强对特定特征的感知能力。下图 3-1 为常规稀疏卷积的示意图。

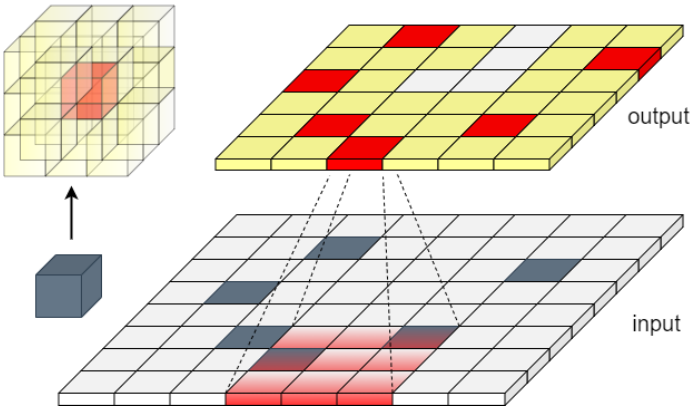



图 3-1 常规稀疏卷积示意图

在 $n$ 维空间中的位置 $p$，其输入的特征为 $X^p$，通道数量为 $c_{in}$，卷积核的权重为 $W \in R^{K^d \times C_{in} \times C_{out}}$，在点云所处的三维空间中，$K^d = 3^3$，其卷积过程可以表示为：

$$y^p = \sum_{k \in 3^3} W_k \bullet X_{\bar{p}k} \tag{3-1}$$

$k$ 为卷积核空间 $K^d$ 中所有的离散位置，$\bar{P}_k = p + k$ 为对应的特征位置，$k$ 为对应位置 $p$ 的偏移距离。对于稀疏数据，$K^d(p, P_{in}) = \{k \mid p + k \in P_{in}, k \in K^d\}$ 输入和输出的特征空间分别为 $P_{in}$ 和 $P_{out}$，则对应稀疏数据的卷积过程可表示为：

$$y_P \in P_{out} = \sum_{k \in K_d(p, p_{in})} W_k \cdot X_{\bar{p}k} \tag{3-2}$$

其中 $K^d(p, P_{in})$ 是 $K^d$ 的子集，取决于 $n$ 维空间中的位置 $p$ 和输入的特征空间 $P_{in}$，通常被表示为：

$$K^d(p, P_{in}) = \{k \mid p + k \in P_{in}, k \in K^d\} \tag{3-3}$$

如果 $P_{out}$ 包含 $P_{in}$ 周围的 $K^d$ 内的所有扩展位置，则该过程被表示为：

$$P_{out} = \bigcup_{p \in P_{in}} P(p, K^d) \tag{3-4}$$

其中，当 $P(p, K^d) = \{p + k \mid k \in K^d\}$ 时，式(3-3)即为常规稀疏卷积。

当 $P_{in} = P_{out}$ 时，式(3-4)可表示子流形卷积，子流形卷积通过在子流形上定义局部坐标系，并在该坐标系上进行卷积操作，来实现对几何数据的卷积。子流形卷积将输出位置固定为与输入位置相同，它保持了效率，禁用了断开连接的功能之间的信息流。可以应用于各种几何数据分析任务，如点云分割、点云分类、曲面重建等。它在处理非规则几何数据时能够保持对称性、局部性和旋转不变性。图 3-2 所示为子流形卷积示意图。

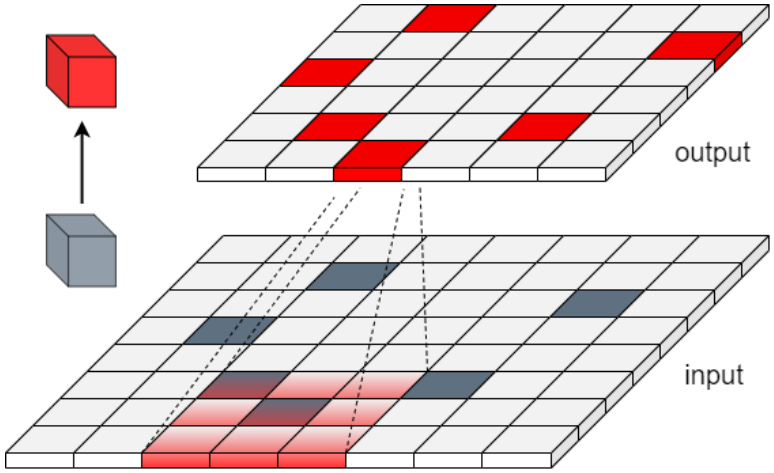



图 3-2 子流型卷积示意图

焦点稀疏卷积作为一种新型的稀疏卷积，能够根据输入数据不同部分的重要性或相关性，有选择的关注或者聚焦于重要的点云区域，动态的确定输出位置，并利用稀疏性质减少计算和存储开销。陈玉康等人[52]在论文中提出了焦点稀疏卷积网络用于 3D物体检测任务的创新方法，实现了在处理 3D点云数据时高效准确的物体检测。图 3-3 为焦点稀疏卷积示意图。其公式可表示为：

$$P_{out}=(\bigcup_{p\in P_{im}}P(p,K_{im}^{d}(p)))\cup P_{in/im} \tag{3-5}$$

焦点稀疏卷积过程包含 3 个重要的步骤，立方体重要性预测、重要位置输入选择、动态输出形状。

立方体重要性图中包含点云输入特征中位置点$P$处周围的候选输出特征的重要性，其通过子流形稀疏卷积和Sigmoid函数实现。

重要位置输入选择中$P_{im}$为：

$$P_{im}=\{p\,|\,I_0^p\geq\tau,p\in P_{in}\} \tag{3-6}$$

其中，重要性特征图中心 $I_0^p$ 要大于等于某个特定的阈值 $\tau$，当 $\tau$ 为 0 或 1 时，就变为常规稀疏卷积和子流形稀疏卷积。动态输出表示为：

$$K_{im}^{p}(p)=\{k\,|\,p+k\in P_{in},I_k^p\leq\tau,k\in K^d\} \tag{3-7}$$

动态输出形状在原始膨胀内被修剪，并满足 $I_k^p\leq\tau$。

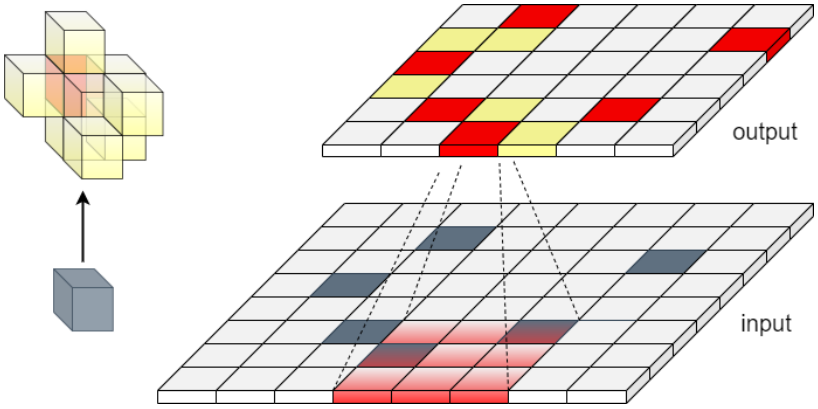



图 3-3 焦点稀疏卷积示意图

### 3.1.2 激活函数

激活函数在神经网络中定义了输入与输出间的非线性映射关系。通过引入非线性变换，提升网络学习复杂特征的能力。焦点稀疏卷积融合目标检测算法的主干网络通过将稀疏卷积的输出连接上激活函数，对卷积后的结果进行非线性变换，提升网络的非线性拟合能力，增强特征的表达能力，并且帮助控制神经元的激活程度，从而提升模型的性能。在神经网络中常用的激活函数有ReLU、Sigmoid等。

(1) Sigmoid 函数

Sigmoid函数是一种常用的激活函数，其主要功能是将神经元的输出映射到[0,1]，实现对输出非线性转换，进而更好的拟合非线性的数据分布，使神经元的输出更容易进行梯度计算和反向传播。作为一种连续并具有可导性质的函数，Sigmoid函数的梯度变化是渐变的，这有利于防止训练过程中出现的梯度突变，从而避免输出跳跃的数值。

Sigmoid函数的公式如式(3-8)所示：

$$f(x)=\frac{1}{1+e^{-x}} \tag{3-8}$$

Sigmoid函数图像如图 3-4 所示：

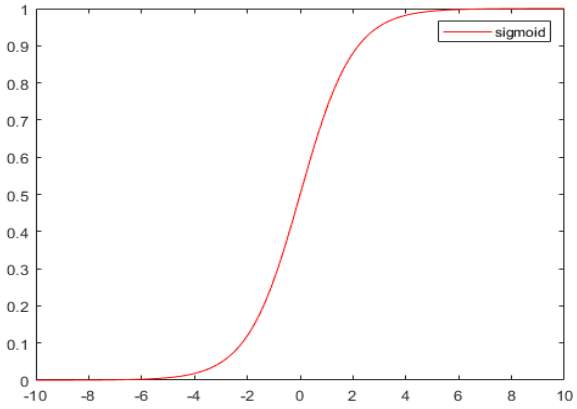



图 3-4 Sigmoid 函数图像

(2) ReLU 函数

ReLU (Rectified Linear Unit)函数能将所有负数映射为 0，正数则保持不变。由于ReLU函数能将所有负数输入变为 0 的特性，因此通过使用ReLU函数能够实现稀疏激活，并减少过拟合的风险。如公式(3-9)所示，当输入值大于或等于 0 时，ReLU函数的梯度为 1，使得神经网络在进行反向传播过程中能够更好的传递梯度，有效避免了梯度相乘引起的梯度消失问题。需要注意的是，当输入为负数时的梯度为 0，表明神经元的权重不再更新，使得后续训练过程中无法再被激活。

ReLU激活函数的公式如式(3-9)所示：

$$f(x)=\begin{cases} x, & x\geq 0 \\ 0 & x<0 \end{cases} \tag{3-9}$$

ReLU激活函数图像如图 3-5 所示：

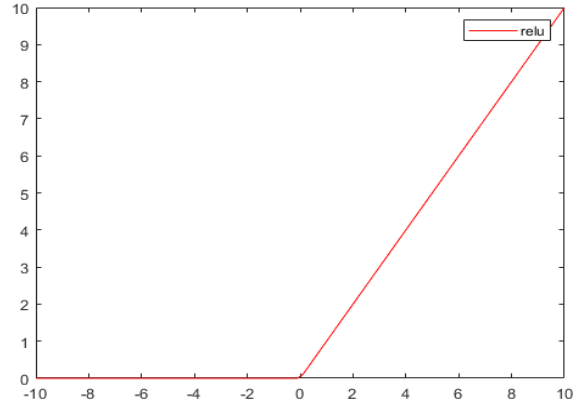



图 3-5 ReLU函数图像

## 3.2 点云体素化特征编码与图像特征提取

### 3.2.1 点云体素特征编码

提取点云和图像数据的特征是 3D目标检测的首要任务。在处理激光点云和视觉图像特征时，关键在于如何有效地将原始数据抽象化，以精确地捕捉目标特征。同时，还需要考虑如何对这些特征进行编码，确保它们之间融合的结果是有效的。

点云体素化将点云空间离散化为均匀的体素网格，每个体素单元可以视为原始点云数据的一个局部代表，可以较好的保留点云数据的原始结构。更重要的一点是，体素网格本质上是一个三维张量，与传统的图像数据结构相类似。通过对每个体素单元的特征进行编码提取点云数据的高级特征能使点云特征与视觉图像特征融合相对容易。

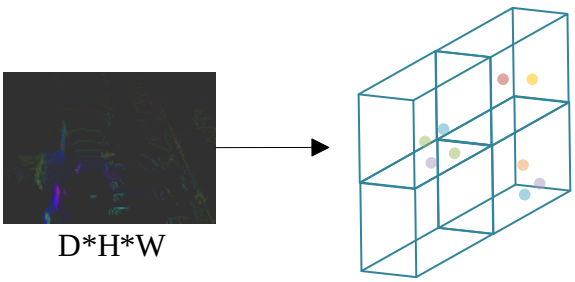



图 3-6 点云体素网格

在体素特征编码过程中，首先需要对指定范围内的点云进行体素划分，即将点云数据转换为规则的体素网格。其次通过对体素网格进行抽象，将每个体素格子表示为一个特征向量，包含了该体素网格内点云数据的特征信息。然后可以通过 3D卷积进行计算得到俯视角度的特征图，特征图包含了体素网格的主要特征，以支持后续的任务处理。

(1) 点云体素划分

在划分体素网格时，需要明确点云在三维空间中的构造范围$(D,H,W)$，划分的体素网格的尺寸为$(v_D,v_H,v_W)$，体素网格的维度为$(D',H',W')$。其中，$D'=D/v_D$，$H'$=H/$v_H$，$W'$=W/$v_W$。

(2) 分组和采样

点云数据是三维空间中点的集合，将空间中的点装入划分好的voxel内进行点云体素分组。由于点云数据具有庞大的规模且分布不均匀的特点，经过划分和分组后的体素网格分为空体素网格和非空体素网格。针对非空体素网格，设定非空体素网格的最大数量，同时规定了非空体素网格内点的数量最多为$n$个。当实际点数超过$n$时，通过随机采样保留$n$个点。

(3) 体素特征编码

首先通过计算每个体素内部所有点的局部平均位置作为质心，并利用质心偏移生成输入特征集。其次通过全连接层将特征映射到特征空间中，通过点的特征聚集信息编码包含在体素内部曲面形状。全连接层由线性层、批量归一化层以及ReLU激活函数所构成。在特征表示形成后，通过对所有与体素相关的元素通过最大池化获取局部聚合特征；然后将局部聚合特征与点的特征进行拼接，形成最终的特征。对所有的非空体素采用相同的编码方法，并共享全连接层的网络参数。

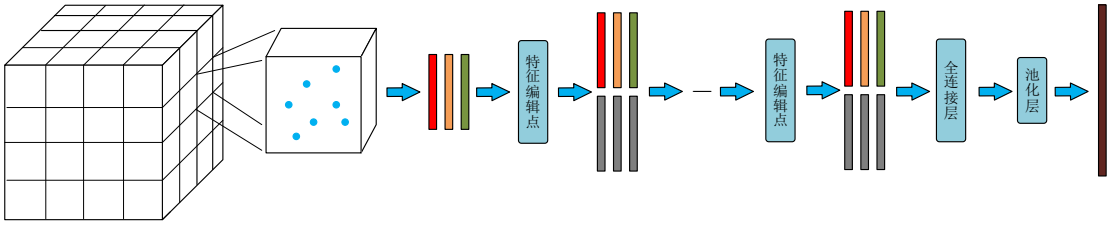



图 3-7 体素特征编码

(4) 稀疏张量表示

通过处理点云数据中的非空体素，构建一个体素特征列表，其中体素的特征与其对应体素内的空间位置有关。这些体素特征可以用一个 4D稀疏张量 $C\times D^{'}\times H^{'}\times W^{'}$ 表示。虽然点云数据包含大量的点，但对点云数据进行体素划分后所得的体素大多都是空的。因此，使用稀疏张量来表示非空体素可以减少在反向传播过程中的消耗。

### 3.2.2 图像特征提取

图像特征提取在计算机视觉领域起着至关重要的作用。特征提取的目的是从原始图像中提取出有用的特征信息，忽略不必要的冗余信息，进而执行高级别的视觉任务，如分类、检测或分割。

本文对原始图像进行特征提取，并与点云稀疏特征进行融合。因此本文仅采用轻量化的残差网络对输入的图像进行特征提取和尺度调整，图像特征提取网络如图 3-8 所示。具体来说，输入的图像经过一个卷积核大小为 7*7，步长为 2 的初始卷积层，捕捉图像的全局信息，并将特征图的尺寸缩小到原来的一半。在初始卷积层之后，图像特征会经过一个最大池化层，将图像下采样到 1/4 分辨率并逐步提取图像的特征。接着经过三个相同的残差块，残差块内部由两个 3*3 的卷积层组成，同时在卷积层后面添加一个批量归一化层和激活函数，用于加速收敛和提升网络的表达能力。这些残差块通过残差连接使网络可以轻松地学习残差函数，避免了梯度消失问题。最后经过最大池化层将特征通道数减小到与点云稀疏特征一致，以便后续融合。

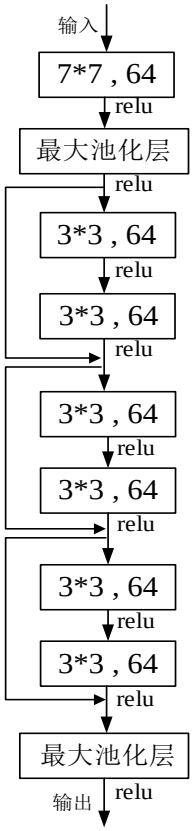



图 3-8 图像特征提取网络

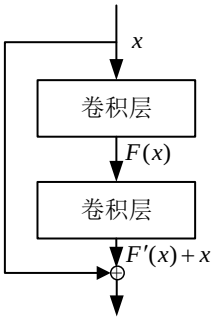



图 3-9 残差结构

图像特征提取部分残差块的结构示意图如图 3-9 所示。残差网络是一种以基于深度学习为基础的神经网络，常被用于解决深层网络在训练中遇到的梯度消失和梯度爆炸等问题。传统的深度学习神经网络会在每层后面添加非线性转换模块，从而导致信息丢失等问题。而残差网络将输入特征与块的输出特征相

加，在保留原始输入信息的同时，还能使网络学习输入和输出之间的差异，从而学习到复杂的映射关系。

## 3.3 主干网络

在基于多模态的焦点稀疏卷积目标检测算法中，3D主干网络和 2D主干网络共同发挥作用，提升检测效率和精度。3D主干网络将点云体素化特征编码后的稀疏特征与图像特征进行融合，利用稀疏卷积进行深层次特征提取，然后将输出的特征进行压缩，得到鸟瞰图(Bird Eye View, BEV)特征,为二维特征提取奠定基础；2D主干网络将BEV特征进行多尺度特征提取，加强对不同尺度目标的描述能力。

### 3.3.1 3D主干网络

在 3D目标检测算法中，特别是在处理多模态数据时，3D主干网络对多模态特征进行融合并进行特征提取，以获取包含丰富空间信息的特征表示，并为算法提供了目标检测所必需的基础信息。

基于多模态的焦点稀疏卷积目标检测算法的 3D主干网络旨在将点云稀疏特征与图像特征融合，并根据融合特征的稀疏性采用多个不同的稀疏卷积提取目标特征的深层次信息。具体来说，3D主干网络将融合后的多模态特征转化为高维度的特征表示，利用稀疏卷积计算的特点，仅在点存在的空间位置进行计算，捕捉了空间中目标对象的尺寸、形状以及其他关键信息，并大大提高了计算的效率。

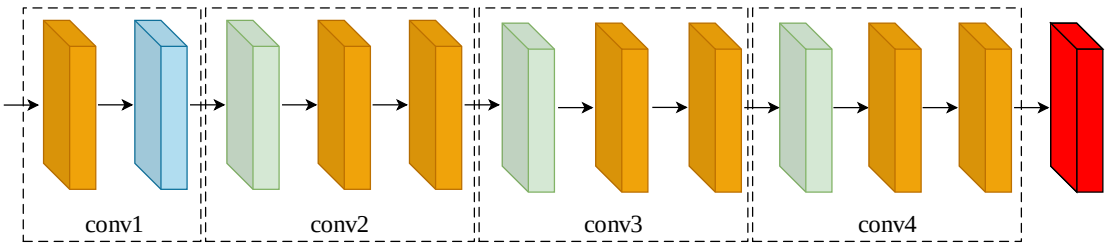



图 3-10 3D 主干网络结构图

如图 3-10 所示，3D主干网络结构如网络共有 4 个卷积块，每个卷积块包含若干个卷积层。在网络结构图中，不同颜色代表不同的卷积模块，其中黄色代表子流形卷积模块，绿色代表常规稀疏卷积模块，蓝色表示多模态焦点稀疏卷

积融合模块，红色表示稀疏-稠密转换模块，每一个卷积模块都由“卷积层-批量归一化层-激活函数层”组成。网络的输入为 3×1600×1408×41，在第一个卷积块的作用下输出 16×1600×1408×41 维度的特征，经过第二个卷积块输出特征为 32×800×704×21，经过第三个卷积块后输出的特征为 64×400×352×11，经过第四个卷积块后输出的特征为 64×200×176×5，最后经过稀疏-稠密转换模块得到主干网络的输出为 128×200×150×2。

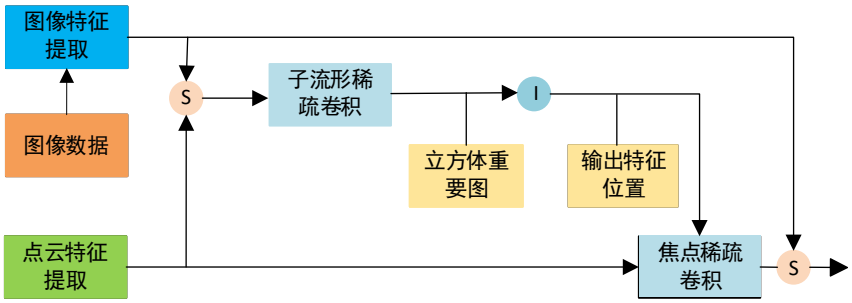



图 3-11 多模态焦点稀疏卷积融合模块示意图

图 3-11 为多模态焦点稀疏卷积融合模块示意图，图中s表示图像特征与点云特征进行加权特征融合，I表示索引。如图 3-11 所示焦点稀疏卷积融合算法从前往后有 3 条路径。第 1 条路径是输入的点云特征经过焦点稀疏卷积模块得到关于点云稀疏特征信息；第 2 条路径是图像数据经过特征提取模块得到RGB特征，并与点云特征进行特征对齐，二者具有相同的数据维度，通过加权求和的方式对二者进行融合，融合后的特征经过子流形稀疏卷积获得融合后的目标重要性特征位置，再经过焦点稀疏卷积得到融合后的特征信息；第 3 条路径是图像数据经过特征提取后的RGB特征。最后通过加权求和的方式将 3 条路径的特征进行融合，得到点云与图像融合后的特征。

多模态的焦点稀疏卷积算法采用了一种加权特征融合的方法将点云稀疏张量与图像特征进行融合，充分利用点云数据和图像数据的优势，使 3D目标检测算法能够更好的理解和描述三维场景。具体来说，在保持两种数据特征的通道数一致的前提下，将点云体素化特征编码后的 4D稀疏张量与图像特征提取的输出结果进行通道级融合，以获取更加全面的特征表示。

假设 $n$ 个通道的特征分别为 $F_1$，$F_2$，……，$F_n$，其中每个通道的权重由参数 $W_1$，$W_2$，……，$W_n$ 决定。则特征通道级融合的公式可以表示为：

$$F_{fused} = \sum_{i=1}^{n} W_i \cdot F_i \tag{3-10}$$

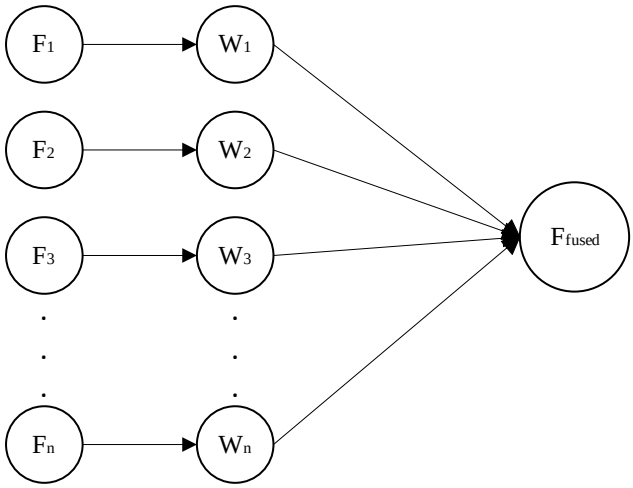



图 3-12 特征通道级融合示意图

### 3.3.2 2D主干网络优化

在许多应用场合，算法不仅需要从 3D数据中提取特征，还要能高效地处理这些特征。为此，在 3D主干网络对多模态数据进行特征提取后，对特征的高度信息进行压缩，然后利用卷积神经网络进行 2D特征提取。

基于多模态的焦点稀疏卷积目标检测算法是陈玉康博士等人[52]提出的一种3D目标检测算法，该算法在检测汽车这类大尺寸目标时能取得较好的结果。但在定位行人和自行车这两类小尺寸目标时存在漏检的情况。因此，本文针对原始多模态焦点稀疏卷积目标检测算法进行改进，在原始特征提取模块基础上，提出了多尺度特征提取网络，并将其与特征学习网络和卷积中间层相结合，形成端到端的可训练流程。图 3-13 为多尺度特征提取网络结构图。

如图 3-13 所示，提出的多尺度特征提取网络包含 4 个卷积块，其中每个卷积块包含多个卷积层，并且在每个卷积块中还包含了多次重复的二维卷积。利用重复的二维卷积进行下采样可以提取输入数据中的特征，同时减小特征图的尺寸，降低计算的复杂度。随着网络层数的增加，重复的二维卷积和下采样操作增加了网络对输入特征的抽象能力，同时增加了神经元的感受野，使网络能够捕捉更广泛的特征，能够帮助网络理解抽象的场景，从而使网络能够学习到更复杂的特征。此外，在进行上采样操作时，首先将前 3 个卷积块的输出特征

用 0 进行填充，然后通过二维反卷积进行上采样来增大特征图的尺寸。上采样的过程能有效缓解模型在连续卷积过程中丢失的细节信息，帮助模型提升对不同尺寸目标的检测精度。对这些上采样后的特征进行跳跃连接操作，将其与网络的输出特征相结合。通过这样的连接，使其能够直接影响网络的输出，不仅增强了模型输出的分辨率，同时还包含了多尺度的特征信息。

提出的多尺度特征提取网络随着网络深度的增加会带来梯度消失问题，进而影响网络有效提取特征。为此，在每一层卷积之后通过批量归一化层和ReLU激活函数缓解网络深度所带来的梯度消失问题，同时也能提升网络的非线性表达能力。

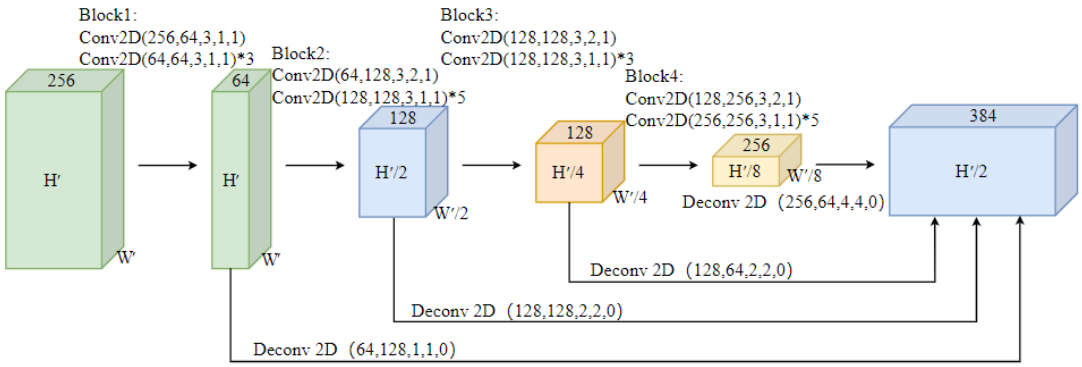



图 3-13 多尺度特征提取网络

提出的多尺度特征提取网络通过对输入特征进行逐级下采样至 1/8 分辨率，使网络能够从不同尺度上提取特征，同时网络也可以获取特征局部细节信息。此外，通过二维反卷积进行的上采样操作，并将输出跳跃连接至网络的末端，从而使网络浅层的特征信息与深层的特征信息相结合，增强了对整体特征的表达能力。在实际场景中，目标可能因距离、角度和遮挡等因素出现不同的尺度变化。多尺度特征提取网络能够适应这种变化，提升检测算法对不同尺度大小物体的识别能力。对于目标检测任务来说，多尺度特征提取能提升整体算法在复杂场景中检测小尺寸目标的能力。

## 3.4 非极大值抑制NMS

在传统的区域建议网络中，非极大值抑制过程的参数设置往往是不固定的，尤其是在训练和推理阶段。在推理阶段，参数的设置可能会根据评估和可视化的需求发生变化，这意味着用于计算平均精度的参数设置可能与实际绘制边界框时的参数不同。NMS在目标检测中起着关键作用，直接影响到检测的效果[53]。目标检测任务中，通过动态提取边界框并对其进行分类及边界框赋值。目标检测结果图片上有时会出现多个边界框，边界框与边界框之间可能会相交。NMS的作用就是将较高分数的边界框保留，对分数较低的边界框进行抑制，不会在最终的检测结果中显示。NMS的伪代码如下：

*Input:* $B = \{(B_n, S_n)\}_{n=1\ to\ N}$, *where* $S_n$ *is the score of* $B_n$, $D = \emptyset$

*Step 1:Select the maximum score box D in B*

*Step 2:Add G and its score to D and remove G and its score for B*

*Step 3:*

*For* $B_i$ *in B*

*If* $IoU(G,B) \geq NMS\_threshold$

*Then*

*Remove* $B_i$ *and* $B'$ *score in B*

*End if*

*End for*

*Step 4:*

*Repeat Step 1-3 until* $B = \emptyset$

*Output:D*

其中，$S$为初始识别置信度；$B$为所有初始边界框的集合；$G$为得分最高的边界框；$D$为最终的输出结果。通过NMS筛选出得分最高的边界框$G$，而与$G$重叠的边界框通过IoU衡量。当IoU的值大于预设的阈值时，与$G$重叠的边界框会被抑制，反之则会被保留。传统的NMS每次都会选择$G$，并将$G$添加至$D$，然后将$G$从$B$中删除；在$B$中，与$G$的IoU大于预设的阈值的边界框也会被抑制；NMS会一直重复上述步骤直至$B$变为空集合，最后输出结果$D$。根据传统的NMS算法，在实际应用过程中常会出现如下问题：

(1) 物体重叠：当一张图像中出现两个外观类似的物体重叠时，传统的

NMS算法难以识别。如图 3-14 所示，两只外观相似的马重叠时，会得到两个边界框。红色边界框的置信度为 0.9，绿色边界框的置信度为 0.8。当两个边界框的重叠面积大于预设的阈值时，传统的NMS算法会删除置信度较小的绿色边界框，从而导致识别精度降低。

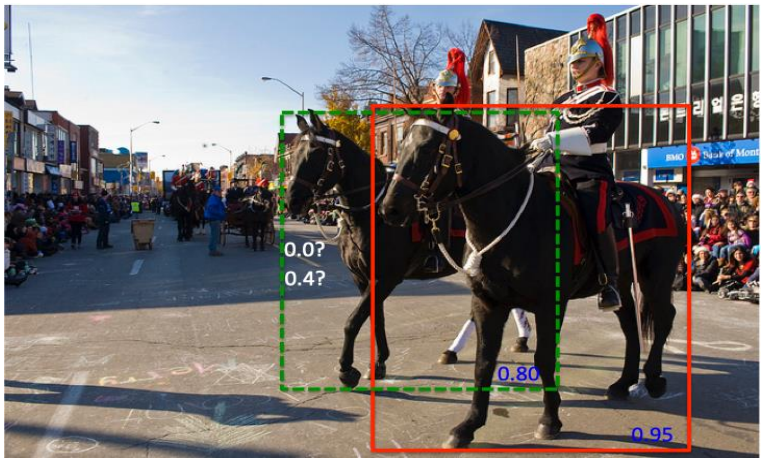



图 3-14 非极大值抑制算法物体重叠情况

(2) 预测不准：在复杂的环境条件下进行目标识别时，目标边界框会出现识别不够准确的情况，出现两个边界框重叠在一个目标之上。如图 3-15 所示。

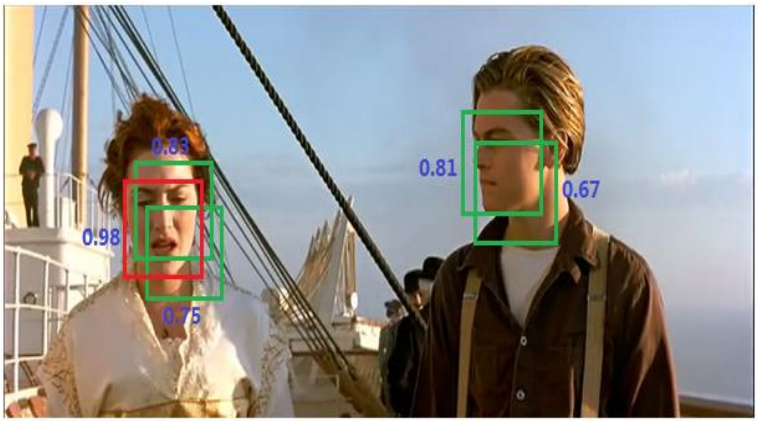



图 3-15 非极大值抑制算法预测不准

(3) 判别标准单一：传统的NMS算法以置信度为基础的，只会保留置信度分数较高的边界框。通常来说，识别框的置信度与识别框之间重叠的面积之间并没有密切的关系。如图 3-16 所示，单纯以边界框的置信度作为判断标准是不准确的。

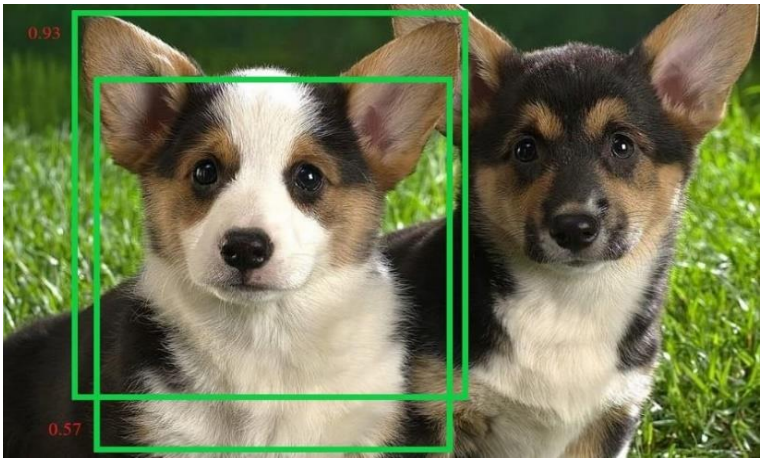



图 3-16 非极大值抑制算法判断标准单一情况

## 3.5 目标框筛选优化

### 3.5.1 2D-IoU目标边界框融合方法

在目标检测任务中，通常采用矩形边界框来框选目标的位置，每个边界框包含了目标的尺寸、位置和可能的类别标签信息。因此，准确的框选目标对目标检测任务是非常重要的。交并比(Intersection over Union，IoU)是目标检测中最常用到的衡量两个边界框重叠程度的评价指标，通过计算两个边界框交集的面积与两个边界框并集的面积的比值得到。IoU的计算方法如图 3-17 所示。

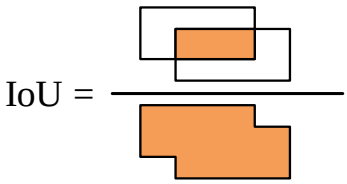



图 3-17 交并比的计算方法示意图

根据IoU值的大小，可以对两个边界框之间的重叠关系进行判断。当计算的IoU的值小于设定阈值空间的最小值时，认为检测结果有效，不进行融合处理；当IoU的值在给定的阈值区间时，认为边界框检测到的是同一个目标，此时，对两个边界框进行融合，保留两个边界框重合的部分作为新的检测结果；当IoU的值大于设定阈值空间的最大值时，认为两者检测的也是同一个目标，此时将两个边界框的并集作为新的检测结果输出。图 3-18 为 2D边界框融合示意图。

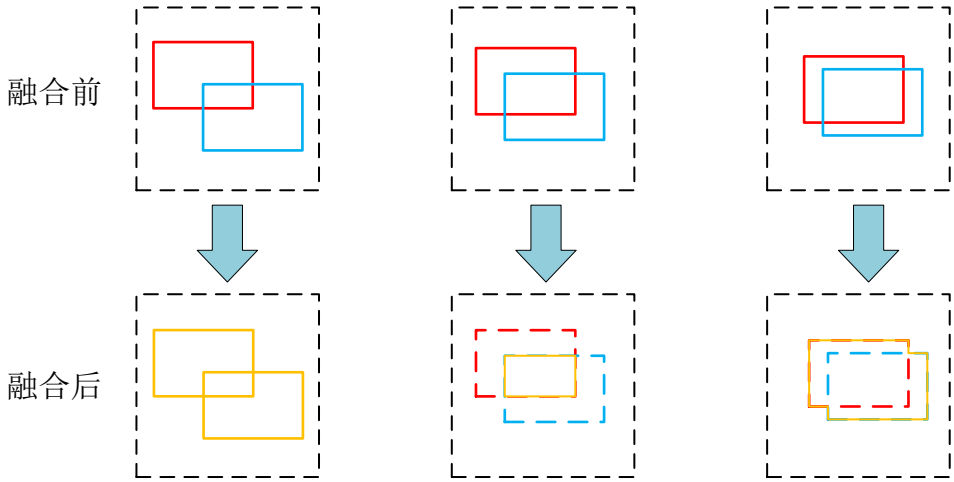



图 3-18 2D 边界框融合示意图

### 3.5.2 3D-IoU损失函数

传统的NMS算法在处理重叠边界框时存在一些问题，如在智能驾驶汽车应用过程中，即便使用两阶段目标检测器，也可能出现一些漏检和误检现象。针对传统的NMS算法的判断标准和阈值单一可能会将一些重合度较高的边界框删除的问题，Qiu S等人[54]通过降低检测框的阈值并引入额外的惩罚机制来解决这一问题。虽然这种方法提高了边界框的精度，但卷积神经网络可能会将检测分数传递至相邻的区域，增加了边界框的数量，使边界框更加难以区分。忽视类内差异不仅降低了对密集目标的识别能力，还延长了优化的过程。尽管已经提出了一些针对 2D目标框的改进方案，但是 3D目标检测框通常以高度和方向作为其重要的检测标准，在使用上述改进方案时还是有一定的局限性。整理和阅读相关文献后，发现在复杂环境下，3D目标检测中目标框与目标不匹配、中心偏移的消除和目标框比例偏差的减少是智能驾驶汽车准确感知其周围环境的主要挑战。与 2D边界框的评价标准一样，3D边界框同样采用IoU作为目标框的评价标准。本节基于提出的传统的目标筛选算法判别标准单一的问题，结合 3D目标检测框的特性，从传统 2D-IoU的角度进行优化。传统 2D-IoU的公式如下：

$$IoU = \frac{|B^a \cap B^c|}{|B^a \cup B^c|} \tag{3-11}$$

$$L_{IoU} = 1 - \frac{|B^a \cap B^c|}{|B^a \cup B^c|} \tag{3-12}$$

以上两式中，$B^a$ 为检测框，$B^c$ 为真实框，LIoU为损失函数。当IOU的值为

0 时，意味着检测框与真实框之间没有重叠部分，即它们之间没有交集。在这种情况下，IoU通常表示检测框未能正确地定位目标对象。此时，常见的距离算法，如欧氏距离、曼哈顿距离等，也无法提供有效的反馈信息，同时梯度传导会受到影响，导致模型在优化过程中无法有效地学习。此外，IoU评价标准的单一性也会导致一些问题，尤其是在处理角度问题时。例如，即使两个框在角度上有所偏差，但它们的IoU值仍可能相同，这会导致检测精度下降，因为IoU无法捕捉到框之间的角度偏移。如图 3-19 所示。

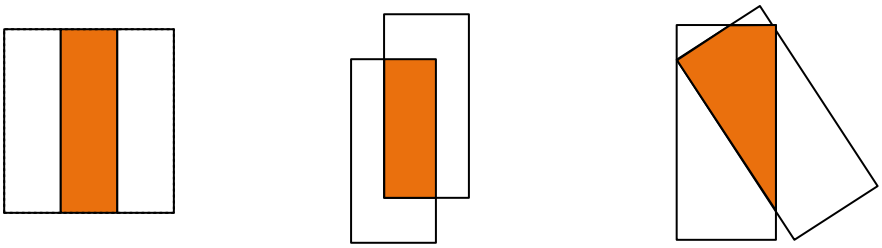


图 3-19 IoU 值相同的多种目标框重叠情况

如图 3-19 所示三张目标框的IoU值相同，但它们GIoU的值却不同，最低的是最后一张图，它的总体重合度最小。GIoU(Generalized Intersection over Union)是一种扩展了IoU的目标检测评价指标，它能够更准确地反映目标框之间的重叠情况。特别是在目标框不完全重合的情况下，GIoU考虑了目标框的形状以及位置之间的关系，相比于传统的IoU，具有更好的稳定性和鲁棒性。因此，GIoU能够准确地反映出两个矩形框在存在夹角情况下的重叠程度，从而在一定程度上减少目标框不重合的情况。文献[55]中提到的GIoU的损失函数，能够在某种的程度上降低 2D目标框不重合的问题。

假定存在一个预测框 $B^a$ 和一个真实框 $B^c$ ，通过计算求出预测框与真实框的最小闭包，即将两个框包围起来的最小矩形框Q。进一步计算Q中不包含预测框与真实框的部分占Q的面积的比值，然后用预测框与真实框的IoU减去这个比值，得到下面的计算公式：

$$IoU = \frac{|B^a \cap B^c|}{|B^a \cup B^c|} \tag{3-13}$$

$$GIoU = IoU - \frac{|C \setminus (B^a U B^c)|}{|C|} \tag{3-14}$$

$$L_{GIoU} = 1 - GIoU \tag{3-15}$$

在 2D目标检测过程中，设预测框 $B^a$ 的坐标为 $B^a = (x_1^a, y_1^a, x_2^a, y_2^a)$ ，真实框 $B^c$

的坐标为 $B^c=(x_1^c,y_1^c,x_2^c,y_2^c)$，且 $x_2^a>x_1^a,y_2^a>y_1^a$。则可以求出预测框 $B^a$ 与真实框 $B^c$ 的面积分别为：

$$A^a=(x_2^a-x_1^a)\times(y_2^a-y_1^a) \tag{3-16}$$

$$A^c=(x_2^c-x_1^c)\times(y_2^c-y_1^c) \tag{3-17}$$

根据上述公式可以求出预测框与真实框的重叠部分面积如下所示：

$$x_1^I=\max(\hat{x}_1^a,x_1^c),x_2^I=\min(\hat{x}_2^a,x_2^c) \tag{3-18}$$

$$y_1^I=\max(\hat{y}_1^a,y_1^c),y_2^I=\min(\hat{y}_2^a,y_2^c) \tag{3-19}$$

$$I=\begin{cases}(x_2^I-x_1^I)\times(y_2^I-y_1^I) & x_2^I>x_1^I,y_2^I>y_1^I\\ 0\end{cases} \tag{3-20}$$

因此可以计算出预测框与真实框的最小闭包 $B^Q$ 的坐标为：

$$x_1^Q=\max(\hat{x}_1^a,x_1^c),x_2^Q=\max(\hat{x}_2^a,x_2^c) \tag{3-21}$$

$$y_1^Q=\max(\hat{y}_1^a,y_1^c),y_2^Q=\max(\hat{y}_2^a,y_2^c) \tag{3-22}$$

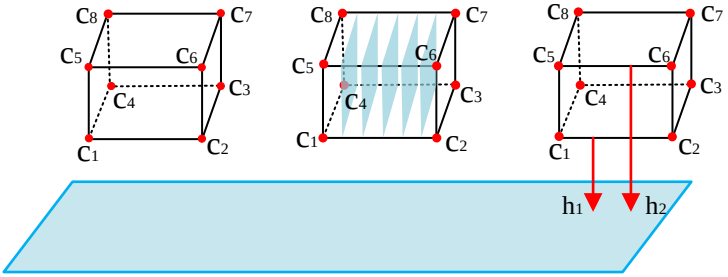
最小闭包 $B^Q$ 的面积为：

$$A^Q=(x_2^Q-x_1^Q)\times(y_2^Q-y_1^Q) \tag{3-23}$$

可以求出最终的损失为：

$$L_{GloU}=1-\frac{I}{A^a+A^c-I}+\frac{A^Q-A^a+A^c-I}{A^Q} \tag{3-24}$$

与 2D目标检测相比，3D目标检测回归点的数量通常更多。由于 3D检测框的编码方式与角度定义的影响，$L_{GloU}$ 中回归参数的定义也会受到影响。立方体3D检测框是利用立方体空间中心点的长宽高进行表示的。然而，这种中心点回归方法在精度上有所不足，并且不提供方向角信息。如图 3-20 展示了三种典型的立方体检测框编码方法。




(a) (b) (c)

图 3-20 三种典型的立方体检测框编码方法

在图 3-20(a)中，通过立方体的每个顶点回归检测框的位置，并以最长边的

方向向量方向确定车辆的朝向角。然而，此方法忽略了立方体的几何约束，导致输出较多冗余信息。在图 3-20(b)中，通过将立方体分割成多个与坐标轴平行的区域，并在每个区域内使用一个简单的回归模型(如线性回归)进行预测。这种方法虽然能最大程度减少冗余参数，但无法捕捉复杂的结构。对于斜向的立方体，这种方式的回归表现不佳。综合上述方法并结合立方体的结构特性，本文引入 3D-IoU算法[56]。如图 3-20(c)所示，利用立方体底面 4 个顶点的坐标以及立方体的高度来表示具有角度的 3D检测框。

根据上述 2D目标框的推导过程和 3D目标框的编码方式，可假设 3D真实框的为 $B_{3D}^{c}$，那么 3D预测框 $B_{3D}^{a}$ 的坐标可以表示为：

$$B_{3D}^{a}=(c_1^a,c_2^a,c_3^a,c_4^a,h_1^a,h_2^a),B_{3D}^{c}=(c_1^c,c_2^c,c_3^c,c_4^c,h_1^c,h_2^c) \tag{3-25}$$

根据上述坐标可以求出 3D预测框 $B_{3D}^{a}$ 的体积 $A_{3D}^{a}$ 与 3D真实框 $B_{3D}^{c}$ 的体积 $A_{3D}^{c}$ 可表示为：

$$A_{3D}^{a}=(c_3^a-c_1^a)\times(c_4^a-c_c^a)\times(h_2^a-h_1^a) \tag{3-26}$$

$$A_{3D}^{c}=(c_3^c-c_1^c)\times(c_4^c-c_c^c)\times(h_2^c-h_1^c) \tag{3-27}$$

根据上述公式可以求出 3D预测框 $B_{3D}^{a}$ 与 3D真实框 $B_{3D}^{c}$ 的重叠部分体积可表示为：

$$c_1^I=\max(\hat{c}_1^a,c_1^c),c_3^I=\min(\hat{c}_3^a,c_3^c) \tag{3-28}$$

$$c_2^I=\max(\hat{c}_1^a,\hat{c}_1^c),c_3^I=\min(\hat{c}_3^a,\hat{c}_3^c) \tag{3-29}$$

$$I=\begin{cases}(c_3^I-c_1^I)\times(c_4^I-c_2^I)\times(h_2^g-h_1^g) & c_3^I>x_1^I,c_1^I>c_2^I\\ 0\end{cases} \tag{3-30}$$

则可计算出包含 3D预测框和 3D真实框的最小 3D框的体积 $A_{3D}^{Q}$，用公式可表示为：

$$c_1^I=\min(\hat{c}_1^a,c_1^c),c_3^I=\max(\hat{c}_3^a,c_3^c) \tag{3-31}$$

$$c_2^I=\max(\hat{c}_2^a,\hat{c}_2^c),c_4^I=\max(\hat{c}_4^a,\hat{c}_4^c) \tag{3-32}$$

$$A_{3D}^{Q}=(x_2^Q-x_1^Q)\times(y_2^Q-y_1^Q)\times(h_2^c-h_1^c) \tag{3-33}$$

可以求出多传感器融合的 3D目标框检测算法的为损失函数为：

$$L_{3D-G1oU}=1-\frac{I}{A_{3D}^{a}+A_{3D}^{c}-I}+\frac{A_{3D}^{Q}-A_{3D}^{a}+A_{3D}^{c}-I}{A_{3D}^{Q}} \tag{3-34}$$

### 3.5.3 高斯加权的Soft-NMS

在目标检测算法中，常通过区域建议网络(Region Proposal Network,RPN)生成大量的候选框，然而其中大部分可能与目标无关，导致计算量大大增加，从而降低了整体的计算速度。在阅读有关文献后，本文基于NMS方法，引入Soft-NMS算法，并通过高斯加权系数对置信度进行调整，根据重叠的面积动态确定高斯加权系数，更好的反馈检测框之间重叠的情况[57]。其伪代码如下所示：

*Input: B = {$b_1$,$b_2$,$b_3$……,bn},S = {$s_1$,$s_2$,$s_3$……$s_n$}, $N_t$,D = Ø*

*Step1: Screening the detection frame by vanishing point detection method;*

*Step2: Calculate the area of all detection frames in set B;*

*Step3: Merge all bounding boxes and sort them from low to high according to the score S;*

*Step4:*

*For j = 1,2,……,n-1*

$$iou(G,B_i^{'})=\frac{S_{B_i\cap B}}{S_{B_i\cup B}}$$

$$s_i \leftarrow iou(G,B_i')$$

*End*

*Output: B′,S′*

其中*B*为所有的初始检测框集合，*S*为检测框对应的置信度，*Nt*为NMS阈值。

Step1，基于几何学原理和相机参数来计算预测框的位置和大小，并根据消失点的位置及检测结果*B*的中心位置，推断预测框的位置；

Step2，所有初始检测框的面积可以表示为：

$$S=\frac{x_2-x_1+1}{y_2-y_1+1} \tag{3-35}$$

Step3，通过高斯加权的方式对置信度进行调整，更好反映检测框之间的重叠情况，并根据重叠的面积，动态的确定高斯加权系数，将所有检测框组成新的集合$B^{'}$，对集合中的检测框按照分数排序：

$$G=sort(B^{'},S) \tag{3-36}$$

Step4，将$B^{'}$中的检测框按照得分高低的顺序进行排列，并计算每个检测框与其他检测框的$iou(G,B_i')$：

$$S_{B_i \cap B} = (max[x, x_i] - min[x, x_i]) \times (max[y, y_i] - min[y, y_i]) \tag{3-37}$$

$$iou(G, B_i^{'}) = \frac{S_{B_i \cap B}}{S_{B_i \cup B}} = \frac{S_{B_i \cap B}}{S_{B_i} + S_B - S_{B_i \cap B}} \tag{3-38}$$

高斯加权系数可以用 $iou(G, B_i')$ 进行代替，在高斯加权系数的作用下，检测框的得分可以表示为：

$$S_i = S_i * e^{-\frac{iou(G, b_i)^2}{\sigma}} \tag{3-39}$$

通过设定的阈值对检测框的分数进行筛选，输出符合要求的检测框作为最终的目标框。

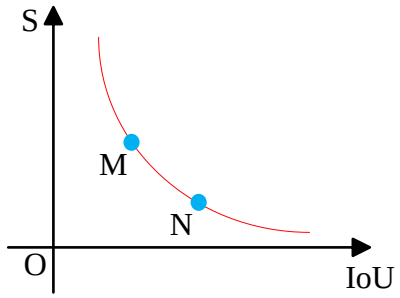



图 3-21 置信度分数与 IoU 函数曲线

## 3.6 目标框筛选网络的搭建与训练

在OpenPCDet框架内配置实验环境，实验平台的配置情况如下表 3-1 所示。

表 3-1 实验平台配置

| 软件与硬件 | 配置 |
|---|---|
| 操作系统 | Ubuntu 18.04 |
| 内存 | 16 GB |
| GPU | NVIDIA RTX3090 |
| GPU加速器 | CUDA 10.2 CUDNN8.0 |
| 深度学习框架 | PyTorch 1.10 |
| 编程语言 | Python 3.6 |

将提出的多尺度特征提取网络替换至原始目标算法的 2D主干网络中，并在此基础上引入了 3D-IoU与高斯加权的Soft-NMS算法，搭建了最终的目标检测网络。改进的多模态焦点稀疏卷积目标检测算法框架如图 3-22 所示。

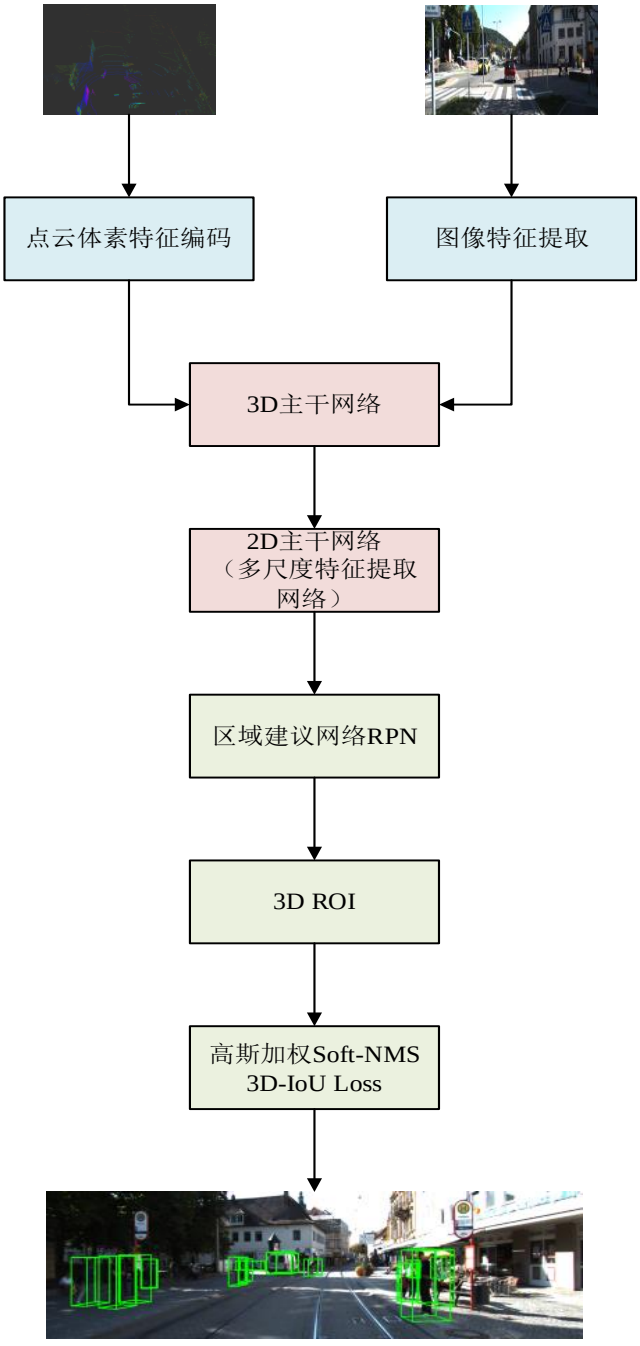
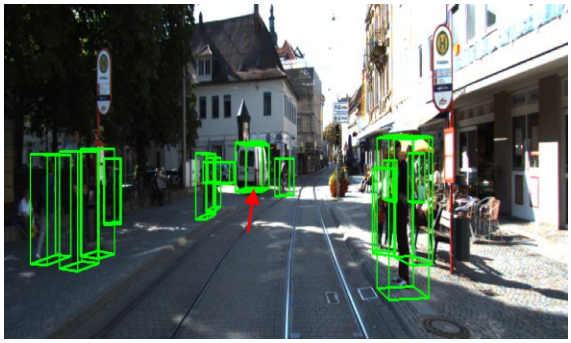


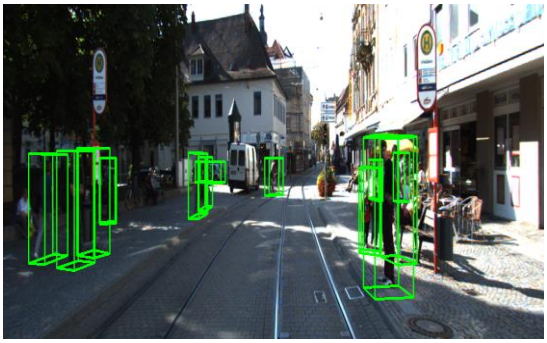

图 3-22 改进的多模态焦点稀疏卷积目标检测算法框架

## 3.7 检测结果与分析

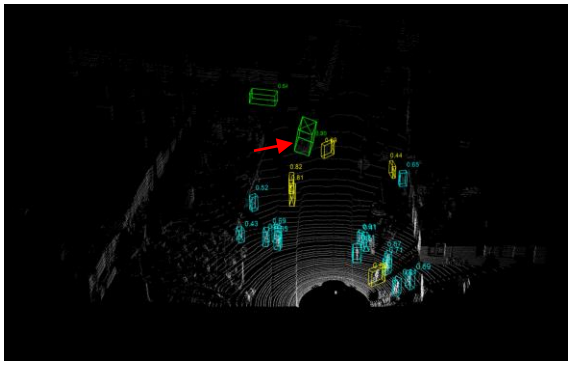
为了验证提出的改进融合算法，在KITTI验证集上对训练后的模型进行了测试，并通过可视化呈现了实验结果。图 3-23 展示了在KITTI数据集上，对点云 3D图和BEV图进行框选的可视化检测结果。



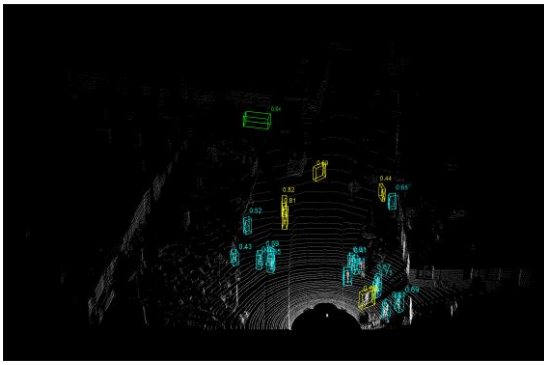
(a)改进后场景 1 图像检测结果



(b)改进前场景 1 图像检测结果

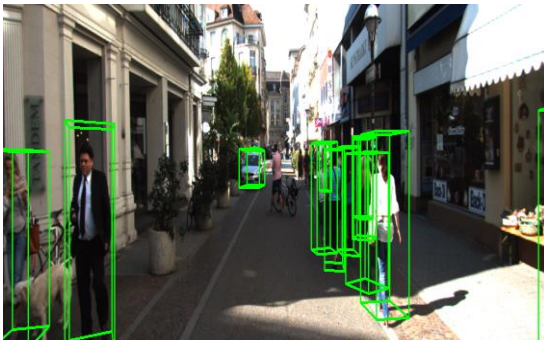


(c)改进后场景 1 点云检测结果

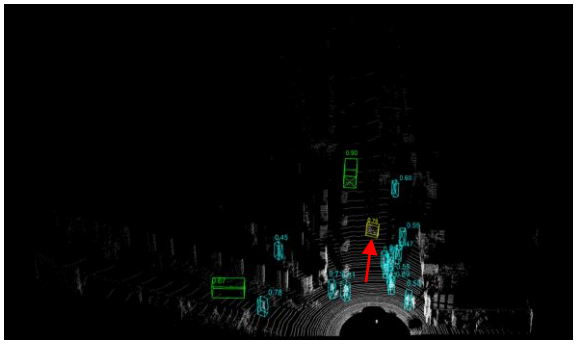


(d)改进前场景 1 点云检测结果

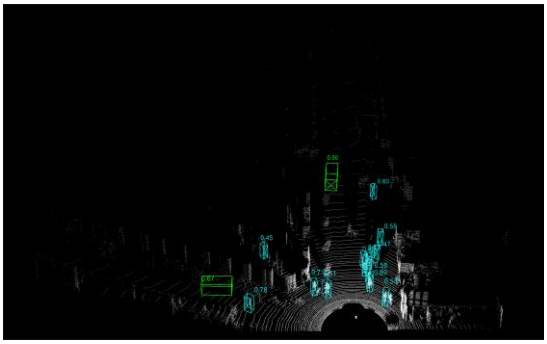
(e)改进后场景 2 图像检测结果



(f)改进前场景 2 图像检测结果



(g)改进后场景 2 点云检测结果



(h)改进前场景 2 点云检测结果

图 3-23 改进前后检测结果对比图

如图 3-23 所示，(a)、(b)、(c)、(d)和(e)、(f)、(g)、(h)分别为两种场景下点

云与图像改进前后检测的结果对比图。从图 3-23(a)和(c)中红色箭头指示的识别框为改进前的目标检测算法漏检的汽车被成功检测；从图 3-23(e)和(f)中红色箭头指示的识别框为改进前的目标检测算法漏检的行人被成功检测。改进后的目标检测算法所检测出的汽车和行人均为视野中较远处的小目标，由此证明了改进后的目标检测算法能够提升模型对小尺寸目标检测的精度。

表 3-2 对比了典型多模态融合算法与我们改进后算法在“car”类别检测中的平均精度(Average Precision,AP)值，表 3-3 比较了改进算法与其他算法在“Pedestrians”类别检测中的AP值，而表 3-4 则对比了改进算法与其他算法在“Cyclist”类别检测中的AP值。

表 3-2 典型多模态融合算法在KITTI数据集上对“car”类别检测结果AP值对比

| Method | 3D | | | BEV | | |
|---|---|---|---|---|---|---|
| | Easy | Mod | Hard | Easy | Mod | Hard |
| MV3D | 71.09 | 62.35 | 55.12 | 86.02 | 76.90 | 68.49 |
| F-PointNet | 81.20 | 70.39 | 62.19 | 88.70 | 84.00 | 75.33 |
| AVOD | 73.59 | 65.78 | 58.38 | 86.80 | 85.44 | 77.73 |
| AVOD-FPN | 81.94 | 71.88 | 66.38 | 88.53 | 83.79 | 77.90 |
| UberATG-ContFuse | 82.54 | 66.22 | 64.04 | 88.81 | 85.83 | 77.33 |
| RoarNet | 83.71 | 73.04 | 59.16 | N/A | N/A | N/A |
| 3D-CVF | 89.20 | 80.05 | 73.11 | N/A | N/A | N/A |
| Focals Conv-F | 92.26 | 85.32 | 82.95 | N/A | N/A | N/A |
| Ours | 92.78 | 85.45 | 85.30 | 95.82 | 91.11 | 91.00 |

通过表 3-2 中的数据分析可知，相比改进前的多模态焦点稀疏卷积目标检测算法，改进后的算法对“car”类检测时，在“Easy”和“Mod”难度上AP仅有微弱提升，但在“Hard”难度上提升了 2.35%。与MV3D算法相比，本文所改进的算法对“car”类进行检测，结果表明，在三维平均精度($AP_{3D}$)和鸟瞰图平均精度($AP_{Bev}$)上均有明显的提升，平均提升精度分别为 24.99%、15.5%。与UberATG-ContFuse算法相比，本文所改进的算法对“car”类进行检测，结果表明，在$AP_{3D}$、$AP_{Bev}$上同样有明显的提升，平均提升精度分别为 16.9%、15.5%。

表 3-3 改进算法在 KITTI 数据集上对“Pedestrians”类别的检测结果 AP 值对比

| Method | 3D | | | BEV | | |
|---|---|---|---|---|---|---|
| | Easy | Mod | Hard | Easy | Mod | Hard |
| PointPillars | 50.42 | 44.57 | 40.73 | 54.70 | 49.67 | 46.03 |
| PartA2 | 60.75 | 53.75 | 49.43 | 63.36 | 57.93 | 53.15 |
| PointRCNN | 64.50 | 55.41 | 51.55 | 65.82 | 60.43 | 53.84 |
| PointRCNN2 | 63.33 | 56.24 | 51.24 | 64.90 | 60.66 | 54.10 |
| Focals Conv-F | 68.93 | 62.15 | 59.25 | 70.32 | 66.20 | 64.35 |
| Ours | 69.50 | 64.71 | 60.51 | 73.31 | 67.46 | 64.58 |

在表 3-3 中，我们观察到改进前后的算法对“Pedestrians”类别的检测精度均有一定程度的提升。与改进前的算法相比，改进后的算法$AP_{3D}$在三种难度上的提升分别为 0.57%、2.56%、1.26%，平均提升了 1.46%；在三种难度上的$AP_{Bev}$提升分别为 2.99%、1.26%、0.23%，平均提升了 1.49%。与PointRCNN2算法相比，本文改进的算法对“Pedestrians”类检测，结果显示在$AP_{3D}$和$AP_{Bev}$上均有明显的提升，平均提升精度分别为 7.97%、8.56%。

表 3-4 改进算法在 KITTI 数据集上对“Cyclist”类别的检测结果 AP 值对比

| Method | 3D | | | BEV | | |
|---|---|---|---|---|---|---|
| | Easy | Mod | Hard | Easy | Mod | Hard |
| F-PointNet | 71.96 | 56.77 | 50.39 | 75.38 | 61.96 | 54.68 |
| AVOD | 60.11 | 44.90 | 38.80 | 63.66 | 47.74 | 46.55 |
| AVOD-FPN | 59.97 | 46.12 | 42.36 | 62.39 | 52.02 | 47.87 |
| VoxelNet (LiDAR) | 61.22 | 48.36 | 44.37 | 66.70 | 54.76 | 50.55 |
| SECOND | 70.51 | 53.85 | 46.90 | 73.67 | 56.04 | 48.78 |
| Focals Conv-F | 86.62 | 70.87 | 68.22 | 87.47 | 72.83 | 70.51 |
| Ours | 87.11 | 73.02 | 70.93 | 88.18 | 76.96 | 73.12 |

对表 3-4 的数据分析，可以看出改进前后的算法在对“Cyclist”类别的检测精度也有一定程度的提升。改进后的算法$AP_{3D}$在三种难度上的提升分别为 0.49%、2.15%、2.71%，平均提升了 1.78%；$AP_{Bev}$在三种难度上的提升分别为 0.71%、4.13%、2.61%，平均提升了提升了 2.48%。与SECOND算法相对比，本文改进的算法在对“Cyclist”类进行检测，结果表明，在$AP_{3D}$和$AP_{Bev}$上均有明显的提升，平均提升精度均为 19.9%。

通过以上的数据对比可以看出，本文改进的算法在KITTI数据集上相较于其他多模态融合算法和单一的点云检测算法都取得了较好的检测结果，尤其是对“Pedestrians”类和“Cyclist”类的检测精度上的提升，表明本文改进的基于多模态的焦点稀疏卷积目标检测算法能有效提升对小尺寸目标的检测精度。

表 3-5 改进前后算法的 mAP(3D)比较

| Method | Car | Pedestrians | Cyclist |
|---|---|---|---|
| Focals Conv-F | 86.84 | 63.44 | 75.24 |
| Ours | 87.84 | 64.91 | 77.02 |

通过对表 3-5 中的对比实验数据分析，发现改进算法在对“car”类别的检测平均精度与原始算法相比mAP提升了 1%，在对“Pedestrians”和“Cyclist”这两个小目标类别的mAP分别提升了 1.41%、1.73%。这表明改进后的算法不仅能提升大目标类别的检测精度，同时有效提升了对小目标类别的检测精度。

## 3.8 本章小结

本章首先阐述了基于多模态焦点稀疏卷积目标检测算法的相关原理。其次，详细介绍了点云体素化、特征编码以及图像特征提取模块。基于多模态焦点稀疏卷积目标检测算法，并针对小目标检测精度不足的问题，提出了一种多尺度特征提取网络。然后针对目标框筛选精度不足的问题，通过引入高斯加权稀疏调整置信度，并根据重叠区域面积动态确定高斯加权系数。再然后针对目标框筛选算法判别标准单一的问题，引入了一种 3D-IoU姿态估计损失函数提升 3D 目标框筛选的精度和效率。最后，在KITTI数据集中对改进前后的算法进行分析验证，并可视化了检测结果。将改进后的算法与改进前的算法和其他类型的检测算法进行了数据对比，实验结果表明，改进算法在对“car”、“Pedestrians”和“Cyclist”这三个目标类别的mAP分别提升了 1%、1.41%、1.73%，证明改进后的目标检测算法能有效提升对小尺寸目标的检测精度。

# 第4章　实验验证与结果分析

本章主要阐述了实验所用的软件和硬件，据此设计并进行了实验。在校园主干道上通过上位机实时采集激光雷达和相机数据，并在实验室内对数据进行解析。将解析后的数据用于验证改进后的多模态焦点稀疏卷积目标检测算法，对不同场景下的检测结果进行了可视化，并根据可视化的结果进行了分析。

## 4.1 实验的软硬件平台

### 4.1.1 实验平台介绍

如图 4-1 所示为环境感知系统实验车，实验所用的设备包括镭神C-16 激光雷达、奥比中光相机、户外电源、上位机等。

将激光雷达与相机通过刚性连接安装到车顶支架，利用户外电源对设备进行供电。在上位机内启动ROS，并通过Autoware 14.0 软件利用激光雷达和相机进行实时数据采集。

图 4-1 环境感知系统实验车

### 4.1.2 ROS节点设计

ROS(Robot Operating System)是一个灵活的软件框架，旨在简化机器人编程。它由通讯协议、工具集与应用模块及生态系统构成[58]。在本文中，利用ROS框架设计了一个点云数据和图像数据的环境感知系统。

ROS环境感知系统架构包含两大部分: (1)传感器联合标定节点；(2)环境感知节点。由于点云与图像在分辨率和帧率上存在差异，本文实验所用到的激光雷达获取频率为每秒 10 帧，相机获取频率为每秒 30 帧。为了实现两者数据的融合，需要对数据进行处理。

(1) 传感器联合标定节点

由于环境感知系统是基于点云和图像进行设计的，因此，需要进行联和校准，使点云数据能够通过旋转矩阵和平移矩阵与相机坐标系对齐；然后利用时间戳同步Rosbsg中的点云数据和图像数据。

(2) 环境感知节点

为确保点云数据能精确地投影到图像上，需要实现检测结果和校准参数的相互传递。将多模态数据融合目标检测算法封装成独立的ROS节点，通过ROS消息传递机制将点云数据和图像数据传入融合网络中，实现激光雷达与视觉融合的车辆环境感知。

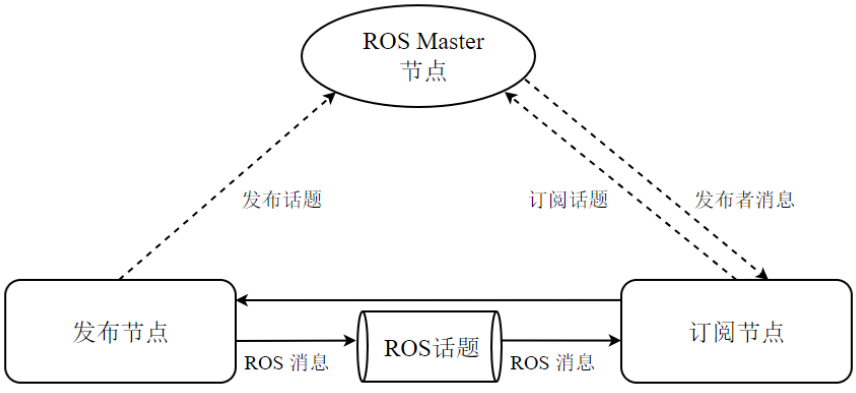



图 4-2 ROS 发布订阅模式示意图

## 4.2 数据集

由德国卡尔斯鲁厄理工学院联手丰田美国技术研究所共同发布的KITTI数据集，目前被认为是国际上为数不多针对自动驾驶技术评估而设的大型数据集

之一[59]。这个数据集是在一辆特别改装的大众汽车上收集的，该汽车配备了多个传感器，包括GPS/INS组合的导航系统、一个 64 线激光雷达扫描仪，以及两套黑白摄像头和彩色摄像头。如图 4-3 所示。这些传感器可以提供丰富的数据，例如激光雷达可以提供高精度的三维点云数据，相机可以提供图像数据，GPS/INS系统可以提供车辆的位置和姿态信息。通过这些传感器，KITTI数据集可以提供多种类型的数据，包括图像、点云、GPS位置信息等。官方提供了一套车载计算机用于记录各传感器的数据，并将其打包为二进制文件供用户下载。同时，官方还提供了工具，可以将这些二进制文件转换为通用的bag文件格式，便于用户使用。此外，官方还提供了相机内参和各传感器之间的外参标定结果，方便用户进行数据处理和算法开发。

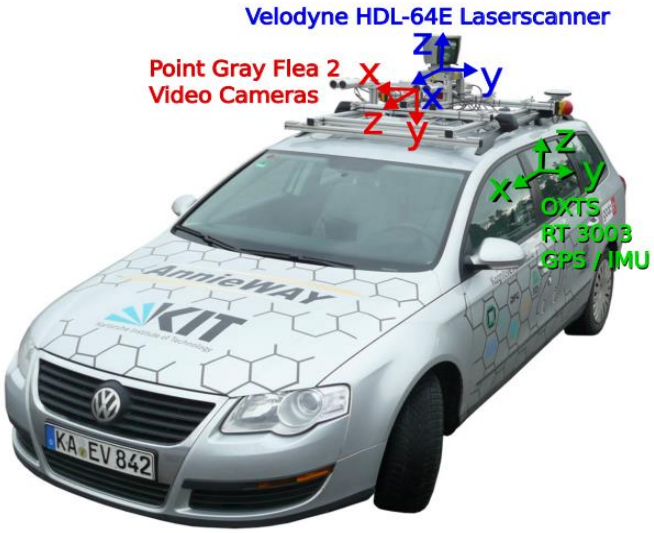



图 4-3 KITTI 数据集采集实验平台

## 4.3 评价指标

性能评价指标如准确率(Precision，P)、召回率(Recall，R)和平均精度(Average Precision，AP)是衡量检测算法优劣的关键标准，为算法性能提供了客观评估[60]。这些指标能够客观地评估算法的优劣。

P表示在所有检测结果中，判断正确的样本数所占的比例。如果使用表格内符号表示对应类别的样本数，则P可以定义为：

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{4-1}$$

式(4-1)中，TP表示真阳性(True Positives)，即被正确识别为正例的样本数；FP 表示假阳性(False Positives)，即被错误识别为正例的样本数。R是指正确识别的目标数占目标总数的百分比。R用公式可以表示为：

$$recall = \frac{TP}{TP + FN} \times 100\% \tag{4-2}$$

其中，FN表示假阴性(False Negatives)，即被错误判断为负例的样本数。

在评估 3D目标检测的性能时，常用的度量包括$AP_{3D}$、$AP_{Bev}$。这两个指标都是基于物体边界框的预测结果与真实框之间的匹配情况来计算的。这些评价指标在KITTI等数据集中被广泛使用，它们提供了一种量化检测器性能的方式，并且考虑了目标在三维空间中的准确性以及在平面上的准确性。

在目标检测任务中，置信度阈值和交并比阈值是两个决定性的参数，用于评估预测的边界框是否正确。置信度阈值衡量了模型对边界框内存在目标的相信程度，只有当边界框的分数高于预设的阈值时，才会被视为正例；反之，则视为负例。然后需要测量预测边界框和实际目标边界框的重叠程度，只有当边界框达到所需的分数和重叠标准时，才会被保留。在评估的过程中，主要考虑四种情况：真正例是正确确定的边界框；假正例是不正确确定的边界框；在背景样本中，假负例是错误地识别为背景的边界框，而真负例是正确识别为背景的边界框。

表 4-4 检测框的判断标准

| | 结果预测 | 真实结果 |
|---|---|---|
| TP | 真 | 真 |
| FP | 真 | 假 |
| TN | 假 | 假 |
| FN | 假 | 真 |

## 4.4 实验结果可视化

实车实验数据是在校园主干道路上进行采集的。在实际采集数据的过程中，尽可能多的采集距离较远、有遮挡场景。实验数据采集场景如图 4-4 所示。

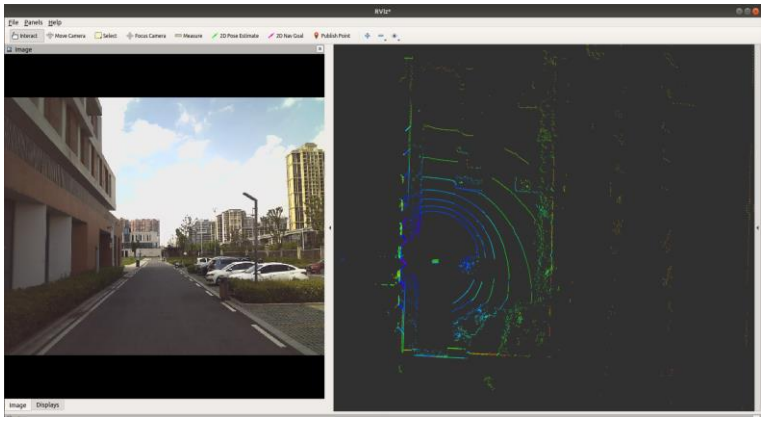


图 4-4 实验数据采集场景

为了更好地评估算法的性能，在数据录制完成之后，在实验室内对数据进行解析，如图 4-5 所示为Rosbag解析后的图像数据，如图 4-6 所示为Rosbag解析后的点云数据。

图 4-5 Rosbag 解析后的图像数据

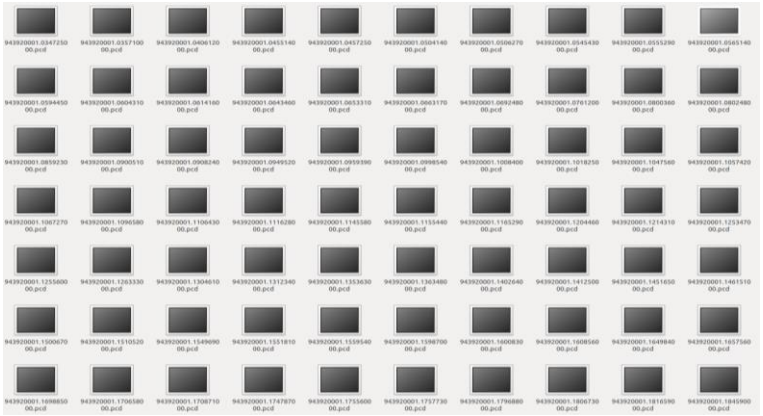
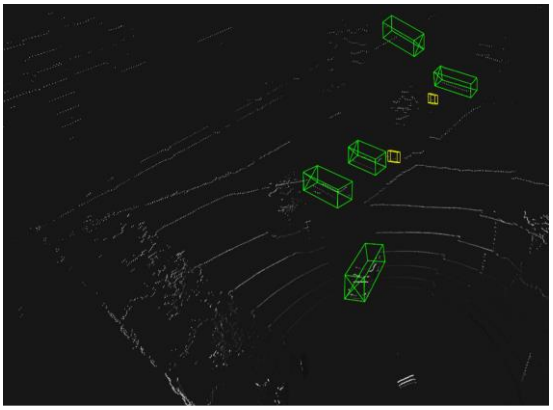


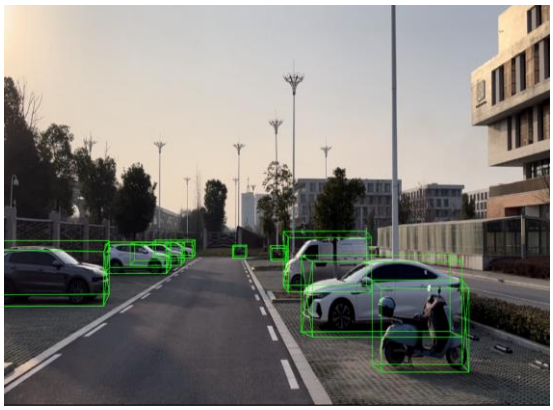
图 4-6 Rosbag 解析后的点云数据

将解析后的数据输入改进后的多模态焦点稀疏卷积检测算法中进行测试并对结果进行了可视化。实验可视化的结果如图 4-7 所示。

(a)场景 1 图像检测结果



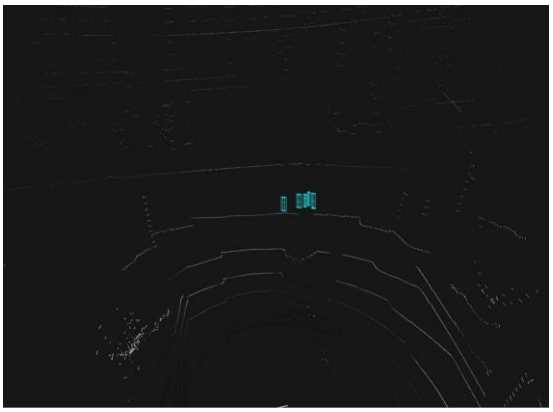
(b)场景 1 点云检测结果



(c)场景 2 图像检测结果

(d)场景 2 点云检测结果

(e)场景 3 图像检测结果



(f)场景 3 点云检测结果

图 4-7 实车实验检测结果

如图 4-7 所示，(a)和(b)、(c)和(d)、(e)和(f)分别为三种不同场景下图像与点云的可视化结果。从图 4-7(a)和(b)中可以看出，汽车和电动车在被部分遮挡的情况下，本文改进的多模态焦点稀疏卷积目标检测算法能够准确的定位出目标的位置；从图 4-7(c)和(d)中可以看出,远处的电动车和被遮挡严重的汽车，也能被本文改进的多模态焦点稀疏卷积目标检测算法准确的定位；从图 4-7(e)和(f)中可以看出，中间有个人在被严重遮挡的情况下，也被改进的多模态焦点稀疏

卷积目标检测算法成功定位。但是在图 4-7 (e)中可以看出，较远处有两个人没有被成功检测，这主要是由于超出了激光雷达使用的有效范围，因此在表 4-2 中没有对该数据进行统计。

表 4-2 检测准确率、漏检误检率

| 组数 | 实际目标数 | 漏检、误检数 | 正确检测数 | 漏检、误检率 | 检测准确率 |
|---|---|---|---|---|---|
| 1 | 7 | 0 | 7 | 0% | 100% |
| 2 | 9 | 0 | 9 | 0% | 100% |
| 3 | 5 | 0 | 5 | 0% | 100% |

从以上的可视化结果分析中可以得出，改进后的多模态焦点稀疏卷积目标检测算法在对距离较远的小尺寸目标以及遮挡严重的目标均有良好的检测效果。

## 4.5 本章小结

本章对实验所用的软件和硬件进行了详细的阐述，其次介绍了性能评价指标。然后通过环境感知实验平台进行雷达和相机的数据录制，并在实验室对数据进行解析。最后用解析后的数据对改进后的多模态焦点稀疏卷积目标检测算法进行验证，可视化了实验结果，并对不同场景下的检测结果进行了分析。

# 第5章 总结与展望

## 5.1 总结

作为智能驾驶的核心技术之一，环境感知的重要性不言而喻。目标检测是智能驾驶环境感知的重要手段，传统的基于单一视觉传感器的目标检测算法在明亮、且无遮挡的情况下展现出较好的性能，但是在复杂的路况和恶劣的天气条件下检测效果大打折扣。为了克服基于单一传感器检测系统的局限性，基于多模态数据融合的检测算法应运而生。多模态数据融合算法集成了单个传感器的优势，可以提供更加全面和准确的环境信息，进而提升智能驾驶车辆环境感知的准确性和可靠性。本文采用激光雷达和相机融合检测的方案，利用不同传感器采集的信号的互补性，对多模态数据融合的目标检测算法进行研究，改进了一种多模态焦点稀疏卷积目标检测算法。本文主要工作内容如下：

(1) 首先研究并确定了多源传感器的配置方案，并推导了传感器坐标系间的转换关系。在理论的基础上，采用张正友标定法对相机进行标定，获取相机的内外参数。其次，利用开源Autoware 14.0 软件对相机和激光雷达进行联合标定，获得了激光雷达与相机标定后的旋转和平移矩阵，保存联合标定后的文件。

(2) 首先详细的阐述了多模态焦点稀疏卷积目标检测算法的相关原理。其次针对多模态焦点稀疏卷积目标检测算法对小尺寸目标检测存在漏检问题，在 2D 主干网络的基础上提出了一种多尺度的特征提取网络。然后针对原算法目标框筛选精度较低的问题，对传统非极大值抑制算法NMS进行优化，引入了高斯加权系数调整置信度，并根据重叠部分的面积动态地确定高斯加权系数。再然后针对传统的目标筛选算法判别标准单一的问题，结合 3D目标检测框的特性，引入了 3D-IoU姿态估计损失函数提升 3D目标框的筛选精度和效率。最后在KITTI数据集上对改进后的方法进行训练和验证，可视化了实验结果。实验结果表明，改进算法在对“car”、“Pedestrians”和“Cyclist”这三个目标类别的mAP分别提升了 1%、1.41%、1.73%。

(3) 在校园主干道上进行数据采集，通过Autoware软件录制数据包，在实验室中解析图像和点云数据，并传入改进后的网络进行验证，得出实验结果。结

果表明，改进后的多模态焦点稀疏卷积目标检测算法在对距离较远的小尺寸目标以及遮挡严重的目标均有良好的检测效果。

## 5.2 展望

本文虽然对基于焦点稀疏卷积融合目标检测算法进行了改进，在对小目标检测时漏检的情况较之前有提升。但仍然存在部分问题需要改进：

(1) 本文改进的多模态焦点稀疏卷积目标检测算法现阶段仅仅在KITTI数据集上进行了验证，今后可以利用其他数据集对改进的多模态焦点稀疏卷积目标检测算法进行验证，以提升算法的普适性。

(2) 本文基于改进的多模态焦点稀疏卷积目标检测算法仅实现了对车辆周围障碍物的定位，没有实现对障碍物距离的检测，在今后的研究中需要进行改进。

(3) 本文在实车实验环节没有进行障碍物的实时检测，仅仅采集了点云和图像数据，在离线的状态下对改进的多模态焦点稀疏卷积进行了验证，在今后的研究过程中，需努力实现车载实时的环境感知。

# 参考文献

[1] 姚兰.2023 年乘用车产销量均创历史新高[J].汽车纵横,2024,(2): 102-103.

[2] 戴帅，刘金广，褚昭明编.中国大城市道路交通发展研究报告 2021[M].人民交通出版社股份有限公司，2022.

[3] 武京利,米泉硕,纪磊.无人驾驶汽车环境感知技术分析[J].现代国企研究,2019,(12):455.

[4] 樊贵全.浅谈新能源汽车电机驱动电磁干扰产生机理[J].时代汽车,2023,(06):112-114.

[5] 梁浩.现代有轨电车运行控制技术发展和展望[J].现代城市轨道交通,2021,(04):33-37.

[6] 梁丽丽,杨圣清,吴铂涵,等.自动驾驶汽车目标检测技术应用研究[J].汽车测试报告,2023,(13):45-47.

[7] 刘晋成,唐伦,陈前斌.基于数据特征的多传感器融合实时目标检测[J].计算机应用研究,2023,40(11):3456-3461.

[8] 伊笑莹,芮一康,冉斌,等.车路协同感知技术研究进展及展望[J].中国工程科学,2024,26(01):178-189.

[9] 陈小凤,史亚锋,江和耀,等.地面无人车系统架构及自主行驶关键技术[J].指挥信息系统与技术,2023,14(06):61-65.

[10] 李佳迪,汤会增,相黎阳,等.基于 5G与AR技术的特高压GIS数字化运检平台设计与研究[J].高压电器,2023,59(12):83-93.

[11] 王亚波,靳玉良,张亚,等.基于激光雷达的结构化道路障碍物检测方法[J].中国惯性技术学报,2023,31(06):593-600+619.

[12] 李学军,权林霏,刘冬梅,等.基于Faster-RCNN改进的交通标志检测算法[J/OL].吉林大学学报(工学版),1-10[2024-04-07].

[13] 张新宇,徐子贤,闫冬梅,等.基于深度学习的 3D目标检测算法综述[J].控制工程,2024,31(03):526-534.

[14] 李晓雯,李海涛,高树静,等.基于Co-PSPNet的轻量级水下鱼体图像分割算法[J].计算机测量与控制,2024,32(02):268-275+308.

[15] 林俊杰,刘昊洋.基于GPU加速的多源点云融合算法研究[J].大众科技,2023,25(12):11-14.

[16] 刘毅,李彬,张向阳.基于改进VGG-19 的井下视觉指纹匹配定位方法[J].华中科技大学学报(自然科学版),2023,51(09):68-73+102.

[17] 龙邹荣,蔡林峰,叶彬强,等.改进YOLOv5s的轻量化目标检测算法研究[J].重庆理工大学学报(自然科学),2023,37(12):244-251.

[18] 刘庆.基于yolov5s的街道场景检测的实现[J].电脑知识与技术,2022,18(09):96-98.

[19] 史时喜.基于视频检测的高风险区域门禁防尾随预警[J].机械设计与制造工程,2021,50(11):91-97.

[20] 刘紫燕,袁磊,朱明成,等.融合SPP和改进FPN的YOLOv3 交通标志检测[J].计算机工程与应用,2021,57(07):164-170.

[21] Jiang H , Learned-Miller E .Face Detection with the Faster R-CNN[J].IEEE, 2017:650-657.

[22] 黄治翔,张艺骞.基于深度学习的目标检测综述[J].科技资讯,2023,21(24):13-16.

[23] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//European conference on computer vision. Springer, Cham, 2016: 21-37.

[24] Tian Y, Yang G, Wang Z, et al. Apple detection during different growth stages in orchards using the improved YOLO-V3 model[J]. Computers and electronics in agriculture, 2019, 157: 417-426.

[25] 宋忠浩,谷雨,陈旭,等.基于加权策略的高分辨率遥感图像目标检测[J].计算机工程与应用,2021,57(13):199-206.

[26] He K, Zhang X, Ren S, et al. Identity mappings in deep residual networks[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14. Springer International Publishing, 2016: 630-645.

[27] Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861, 2017.

[28] 张晓勐,朱德利,余茂生.无人机遥感图像中的玉米雄穗轻量化检测模型[J].江西农业大学学报,2022,44(02):461-472.

[29] Zhang H, Wu C, Zhang Z, et al. Resnest: Split-attention networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 2736-2746.

[30] 王贺,张震.基于ResNeSt和改进Transformer的多标签图像分类算法[J].测试技术学报,2024,38(01):48-53.

[31] 刘晋,杨容浩,文文,等.一种车载LiDAR点云道路提取深度神经网络模型[J].测绘通报,2023,(12):8-12+18.

[32] Simony M, Milzy S, Amendey K, et al. Complex-yolo: An euler-region-proposal for real-time 3d object detection on point clouds[C]//Proceedings of the European conference on computer vision (ECCV) workshops. 2018: 0-0.

[33] Yang B, Luo W, Urtasun R. Pixor: Real-time 3d object detection from point clouds[C]//Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2018: 7652-7660.

[34] 张冬冬,郭杰,陈阳.基于原始点云的三维目标检测算法[J].计算机工程与应用,2023,59(03):209-217.

[35] Zhou Y, Tuzel O. Voxelnet: End-to-end learning for point cloud based 3d object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4490-4499.

[36] Yan Y, Mao Y, Li B. Second: Sparsely embedded convolutional detection[J]. Sensors, 2018, 18(10): 3337.

[37] 李玉萍.基于视觉的车辆检测技术现状[J].国外电子测量技术,2007,(10):21-23+32.

[38] 陈睿韵,田文斌,鲍海波,等.农业轮式机器人三维环境感知技术研究进展[J].智慧农业(中英文),2023,5(04):16-32.

[39] 刘瑞佳,周怡.基于多传感器融合的跌倒检测研究综述[J].现代计算机,2020,(34):55-58+82.

[40] 熊姝涵,赵彬.数字孪生应用中多对象定位与通讯问题研究[J].通信与信息技术,2023,(06):36-40.

[41] 杨明川,朱敬华,李元婧,等.一种考虑隐私保护的深度强化学习任务分配模型[J].计算机研究与发展,2023,60(11):2650-2659.

[42] 李程,夏丹,董世运,等.复杂陆战场环境下的智能感知理论现状与发展[J].国防

科技,2021,42(03):42-48.

[43] 刘蜀,侯国超,吉玉洁.舰载指控系统雷达目标航迹信息融合算法研究[J].计算机仿真,2016,33(03):9-12.

[44] 况博裕,李雨泽,顾芳铭,等.车联网安全研究综述：威胁、对策与未来展望[J].计算机研究与发展,2023,60(10):2304-2321.

[45] N. C ,S.G. N ,H.D. K , et al.The effect of pixel-level fusion on object tracking in multi-sensor surveillance video[J].Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition,2007:1-7

[46] Chongjian Y ,Xiyuan L ,Xiaoping H , et al.Pixel-Level Extrinsic Self Calibration of High Resolution LiDAR and Camera in Targetless Environments[J].IEEE Robotics and Automation Letters,2021,6(4):7517-7524.

[47] Gu S ,Lu T ,Zhang Y , et al.3-D LiDAR + Monocular Camera: An Inverse-Depth-Induced Fusion Framework for Urban Road Detection.[J].IEEE Trans. Intelligent Vehicles,2018,3(3):351-360.

[48] Xiangmo Z ,Pengpeng S ,Zhigang X , et al.Fusion of 3D LIDAR and Camera Data for Object Detection in Autonomous Vehicle Applications[J].IEEE Sensors Journal,2020,1-1.

[49] 高丽伟. 基于激光雷达的车辆目标检测算法研究[D].东北大学,2020.

[50] 倪志康. 基于三维激光的移动机器人SLAM算法研究[D].苏州大学,2021.

[51] Zhang,Zhengyou.A Flexible New Technique for Camera Calibration.[J].IEEE Transactions on Pattern Analysis & Machine Intelligence, 2000.

[52] Chen Y, Li Y, Zhang X, et al. Focal sparse convolutional networks for 3d object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 5428-5437.

[53] Hosang J , Benenson R , Schiele B .Learning non-maximum suppression[J].IEEE Computer Society, 2017.

[54] Qiu S , Wen G , Deng Z ,et al.Accurate non-maximum suppression for object detection in high-resolution remote sensing images[J].Remote Sensing Letters, 2018, 9(3):238-247.

[55] 周兴凯.可自主跟随智能物流机器人设计与试验研究[D].合肥工业大学,2022.

[56] Li J , Luo S , Zhu Z ,et al.3D IoU-Net: IoU Guided 3D Object Detector for Point Clouds[J]. 2020.

[57] 于小川.基于卷积神经网络的交通图像目标检测研究[D].哈尔滨工程大学,2018.

[58] Quigley M , Gerkey B P , Conley K ,et al.ROS: An open-source Robot Operating System[J]. 2009.

[59] Geiger A , Lenz P , Stiller C ,et al.Vision meets robotics: The KITTI dataset[J].International Journal of Robotics Research, 2013, 32(11):1231-1237.

[60] 贾志成,段棋峰,汪东.基于无人机多光谱遥感的林火监测模型[J].中南林业科技大学学报,2024,44(03):22-32.