·重庆交通大学前瞻技术综述专题·

# 基于激光雷达的3D目标检测研究综述

余杭

（重庆交通大学机电与车辆工程学院，重庆 400074）

【欢迎引用】余杭. 基于激光雷达的3D目标检测研究综述[J]. 汽车文摘, 2024(2): 18-27.

【Cite this paper】YU H. A Review on LiDAR-Based 3D Target Detection Research[J]. Automotive Digest (Chinese), 2024(2): 18-27.

【摘要】近年来，随着自动驾驶技术的快速发展，智能汽车对于环境感知技术的需求也越来越高，由于激光雷达数据具有较高的精度，能够更好的获取环境中的三维信息，已经成为了3D目标检测领域研究的热点。为了给智能汽车提供更加准确的环境信息，对激光雷达3D目标检测领域主要研究内容进行综述。首先，分析了自动驾驶车辆各种环境感知传感器的优缺点；其次，根据3D目标检测算法中数据处理方式的不同，综述了基于点云的检测算法和图像与点云融合的检测算法；然后，梳理了主流自动驾驶数据集及其3D目标检测评估方法；最后对当前点云3D目标检测算法进行总结和展望，结果表明当前研究中2D视图法和多模态融合法对自动驾驶技术发展的重要性。

**关键词：机器视觉；激光雷达；自动驾驶；3D目标检测；雷达点云**

**中图分类号：**U469.79　　**文献标志码：**A　　**DOI:** 10.19822/j.cnki.1671-6329.20230082

# A Review on LiDAR-Based 3D Target Detection Research

Yu Hang

（School of Mechatronics and Vehicle Engineering, Chongqing Jiaotong University, Chongqing 400074）

【Abstract】In recent years, with the rapid development of autonomous driving technologies, the demand of intelligent vehicles for environment perception technology is also higher and higher. Due to the high accuracy of LiDAR data that can better obtain the 3D information in the environment, it has become a research hotspot in the field of 3D target detection. In order to provide more accurate environmental information for intelligent vehicles, the main research contents in the field of 3D target detection by LiDAR are summarized. Firstly, the advantages and disadvantages of various environment sensing sensors for self-driving vehicles are analyzed; secondly, according to the different data processing methods in 3D target detection algorithms, the detection algorithms based on point cloud and the detection algorithms fused with image and point cloud are reviewed; then, the mainstream self-driving datasets and their evaluation methods for 3D target detection are sorted out; and finally, the current 3D target detection algorithms for point cloud are summarized and outlooked. The results show the importance of the 2D view method and the multimodal fusion method in the current research for the development of autonomous driving technologies.

**Key words: Machine vision, LiDAR , Autonomous driving, 3D object detection, Radar point cloud**

## 0　引言

随着自动驾驶技术的发展，2D物体检测方法的性能已经大幅提高，在KITTI物体检测数据集[1]上实现了90%以上的平均精度。2D方法用于检测图像平面上的对象，而3D方法在2D方法的基础上，将第三维的深度信息引入到定位和回归任务中。然而，在自动驾驶车辆的背景下，2D目标检测和3D目标检测方法之间的性能差距仍然巨大[2]。因此需要进一步研究3D目标检测算法来提升检测精度和效率。

近几年来，各大自动驾驶公司开源大型自动驾驶数据集，推动了深度学习在3D场景下的应用。深度学习模型可以通过卷积神经网络提取学习道路目标特征，提升检测能力。研究人员通常将点云处理方法分为将点云投影到二维平面和直接进行点云处理。投影方法是指将三维空间下的点云特征通过坐标变换将其投影到二维平面中，这种方法是当前自动驾驶车辆3D目标检测中最常用的方法，可运用成熟的2D

目标检测网络进行特征提取，最后再将结果重新映射到三维空间中。投影法因其使用2D检测网络，具有较高的检测效率，但其压缩了空间信息，在检测精度上具有一定的局限性。直接点云处理方法是Qi等[3]在2017年首次提出的，直接将点云作为深度学习神经网络的输入，在大型三维场景下的验证此方法具有较好的表现，因此逐渐受到了研究人员的青睐。本文根据激光雷达点云处理方式的不同将3D目标检测算法分为4大类别：基于体素的方法、基于点的方法、基于体素-点的方法和基于图像与点云融合的方法。图1依照时间顺序，梳理近几年经典的3D目标检测算法，并将其分为单阶段检测和两阶段检测。

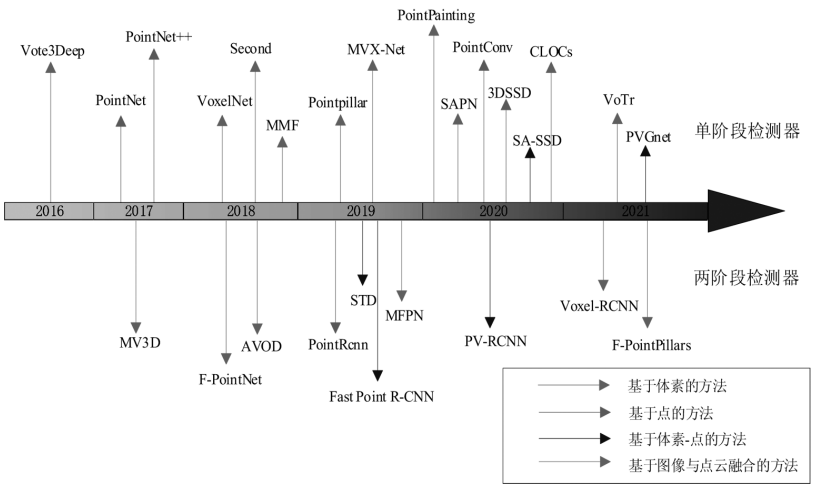



**图1　3D目标检测算法发展进展**

基于以上分析，本文对当前已开源的自动驾驶数据集、3D目标检测算法以及各类评价指标进行综述总结。

## 1　车载感知系统传感器

为了更加全面地理解感知系统传感器数据采集及其处理原理，本节主要对比不同车载传感器的优、劣势。

自动驾驶车辆通过车载传感器感知车辆周围行驶环境，这些传感器主要包括相机、激光雷达及毫米波雷达。不同传感器的数据采集功能和优缺点如表1所示。

**表1　各传感器功能及优缺点对比**

| 传感器 | 功能 | 优点 | 缺点 |
|---|---|---|---|
| 相机 | 2D目标检测 | 成本低，能够提供颜色和纹理信息 | 容易受到环境光照的影响 |
| 激光雷达 | 3D目标检测 | 能够获取360°环境信息，精度高 | 点云稀疏性高，识别难度大，成本高，没有颜色信息和纹理信息 |
| 毫米波雷达 | 测距、测速、目标检测 | 探测距离远，不受天气的影响 | 目标识别精度低 |

### 1.1　相机

相机作为自动驾驶车辆中成本低且最常见的传感器，已经被各大自动驾驶企业部署在车辆上。相机具有较高的分辨率，能够识别所见物体的颜色、形状和纹理等，利用采集的信息，通过算法处理可以使自动驾驶车辆识别道路上的障碍物，旨在了解行驶环境。由于相机出色的识别性能，它能够对道路上的红绿灯和交通标志进行精确地识别，因此在自动驾驶车辆中应用广泛。目前在自动驾驶车辆上使用较多的相机组合形式有以下2种。

（1）单目相机通过将三维空间下的物体转变到二维平面，利用二维视图展示物体的形状和纹理等信息，研究人员利用这类信息完成目标检测、分类等任务。但是，单目相机不能提供深度信息，测距性能较差。

（2）多目相机拥有单目相机的所有功能，在测距和三维物体检测定位上，由于多目相机具有多个摄像头，可以通过匹配算法对摄像头进行融合并得到稠密的深度图，这弥补了单目相机测距性能差的缺点，但是其计算量大，实时性较低。

### 1.2　雷达

雷达通过发射无线电波去检测目标并对其进行定位。雷达可分为激光雷达、毫米波雷达等，是自动驾驶车辆主要的3D检测传感器。

#### 1.2.1　激光雷达

激光雷达（LiDAR）作为自动驾驶汽车主要的传感器之一，主要用于物体的定位感知，根据扫描形式可分为机械式激光雷达、固态激光雷达和混合式激光雷达3大类。

（1）机械式激光雷达

在垂直方向上，发射器能够以一定频率发射多组激光光束，这些光束在接触到物体后，经过漫反射返回到接收器，并且通过发射器不停地旋转可以实时扫描周围360°的环境信息。因此，机械式激光雷达具有信息扫描快和视野范围广的优点。但是，其复杂的机械式旋转结构长时间工作会导致其精度降低，并且存在价格昂贵和体积大等缺点。

（2）混合式固态激光雷达

机械式激光雷达利用发射器旋转的方式来实现360°扫描，而混合式固态激光雷达则是利用驱动转镜或棱镜进行扫描。如MEMS扫描镜，它是由半导体器件组成，在硅基芯片上集成了体积十分微小的微振镜，其内部主要结构是尺寸微小的悬臂梁，反射镜悬挂在扭杆之间以一个固定的谐波频率振荡，通过微振器的旋转来反射激光的光束，扫描周围环境。硅基MEMS微振镜可控性好，可实现快速扫描，可媲美高线束雷达。因此，在相同的点云密度下，混合式固态激光雷达与传统机械式激光雷达相比所需激光发射器更少、体积更小、可靠性更好。

（3）固态激光雷达

与机械式激光雷达相比，固态激光雷达没有机械式激光雷达的内部旋转件，外形尺寸大幅减小，成本相对较低。使用寿命和可靠性较高，符合当前自动驾驶车辆对于雷达的需求。固态激光雷达主要有2种技术路线，分别为光学相控阵（Optical Parametric Amplification, OPA）和快闪（Flash）。OPA激光雷达通过光学相控阵技术，用多个光源组成激光束的发射阵列，通过调节发射阵列中每个发射单元的相位差，来控制输出激光束的方向以达到对不同方向的扫描，具有效率高、体积小和易控制等优点。但是，其存在制造难度高和探测距离短的缺点。Flash固态激光雷达采用类似相机的工作原理，瞬时发射一片覆盖整个区域的激光，通过高灵敏接收器记录场景信息，具有集成度高、扫描速度快和生产量大等优点。但是，其探测距离短、抗干扰能力差、分辨率低。

1.2.2　毫米波雷达

毫米波雷达是指以1~10 mm为波段，30~300 GHz为工作频率的毫米波探测雷达，通过发射和接收毫米波来采集物体距离和速度信息，常见的毫米波雷达有以下3种[32]。

（1）短距毫米波雷达，主要以24 GHz为工作频率，感知距离小于30 m，但是其探测角度广、成本低，可以实现车身全覆盖，是当前使用最多的毫米波雷达。

（2）中距毫米波雷达，主要是以77 GHz为工作频率，感知距离1~100 m，相比于短距毫米波雷达可以实现更高的精度，探测距离更远，但是成本也更高，视角较小。适用于自车与前车的测速和测距等功能。

（3）长距毫米波雷达，主要是以77 GHz为工作频率，感知距离大于200 m，针对高速行驶的车辆，长距毫米波雷达能够很快地检测前车信息，做到提前预警，为自动驾驶车辆或驾驶员预留足够的时间制动或避让。

### 1.3　传感器应用分析

自动驾驶车辆作为一个复杂的系统，选择合适的传感器组合能够有效提高环境感知能力。目前有以下2种主流的传感器组合方式：基于纯视觉和基于激光雷达、毫米波雷达以及视觉融合的方案。

（1）特斯拉自动驾驶采用纯视觉方案，通过多相机融合的方式来实现自动驾驶车辆的定位感知功能，它在一定程度上规避了激光雷达硬件成本高、计算量大的缺点，但是其纯视觉的环境感知系统，易受到环境变化的影响，在强光和昏暗条件下，会损失感知系统的鲁棒性。

（2）谷歌的Waymo与百度的Apollo等公司采用激光雷达、毫米波雷达与视觉融合的方案，利用不同传感器的优势可以降低环境变化带来的影响，具有较高的环境感知能力，但是其硬件成本也随之提高，对于计算量的需求增大。

## 2　三维目标检测

### 2.1　基于点云的检测方法

基于点云的三维目标检测技术可分为基于体素的方法（Voxel-base）、基于点的方法（Point-base）和基于体素-点的方法（Voxel-point base）。

2.1.1　基于体素的方法

采用体素化思想处理点云数据是常用的点云数据处理方法，是通过输入的点云数据创建一个三维体素栅格，每个体素内用体素中所有点的重心来近似显示体素中其他点，这样该体素内所有点都用一个中心点最终表示，减少了原始点云的数据量。

基于体素的方法可以利用深度学习中卷积神经网络有效进行特征提取并进行3D检测，具有很高的计算效率，但其离散化点云的过程使得部分数据丢失，这导致了部分情况下检测精度降低。

Engelcke等[4]提出了Vote3Deep算法，首先通过构建一种有效的卷积层，采用中心对称的投票机制去处理输入点云中存在的稀疏问题，然后经过修正的线性单元

和$L_1$正则去解决CNN堆叠过程中的中间层特征稀疏的问题。由于其在特征提取过程中采用了手工特征的方法使得局部信息不能够有效的提取。因此，Yin等[5]在2018年引入了VoxelNet改善这种情况，如图2所示。所提出的模型是一个通用的3D检测网络，它将特征提取和边界盒预测结合到一个单级、端到端可训练的深度网络中，以增强高稀疏点结构的状态。为了提取逐点特征以将数据区域划分为相等的体素，使用了具有体素特征编码（Voxel Feature Encoding，VFE）层的特征学习网络，但是其使用3D卷积使得计算复杂度提高。

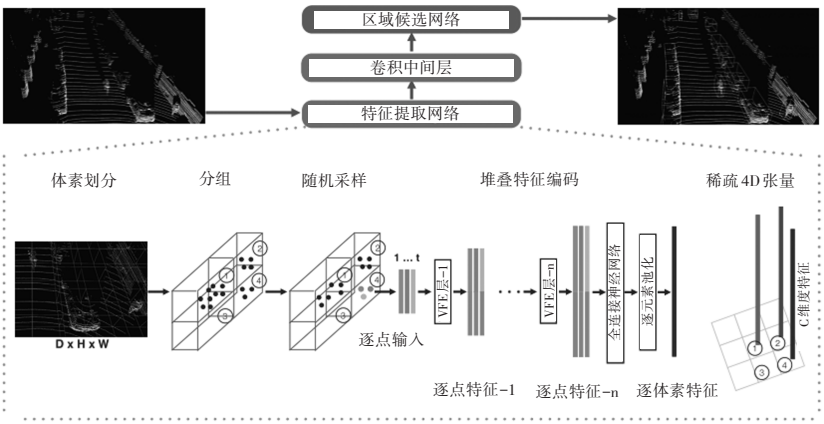



**图2　VoxelNet网络结构[5]**

Yan等[6]在2018年提出了second目标检测网络，通过利用3D稀疏卷积来解决VoxelNet[5]中计算复杂度高的缺点，极大地提高了计算效率。Deng等[7]提出了Voxel-RCNN利用两阶段检测思想，通过进一步提取Proposal中的特征进行Proposal的refinement，解决了体素化过程中信息丢失的问题，提高了检测精度，但是其检测速度较低。Alex等[8]提出了Pointpillar，它利用PointNet[3]来学习按垂直列柱组织的点云。然后利用2D卷积网络进行特征提取，具有极高的运行速度。针对于体素大小的不同会导致信息丢失的问题，Xiang等[9]提出了SAPN网络，从点云中提取多分辨率支柱级特征，使检测方法更具尺度意识。其次，使用空间注意力机制来突出特征图中的对象激活。Mao等[10-11]提出的VoTr是一种基于体素的Transformer网络，利用稀疏体素模块和子流形体素模块，可以有效地对空体素和非空体素位置进行操作，解决了传统体素3D检测器无法捕获上下文信息与感受野不足的问题。

2.1.2　基于点的方法

点云格式的数据通常是不规则的，研究人员通常将其转换为规则的3D体素或者二维图像。这仍然需要对数据进行分类，导致数据过于庞大，并导致部分点云信息消失。

为了直接从未处理的点云中的点特征中学习，Qi等[3]首先提出了PointNet模型，由2个网络组成：一个分类网络，通过仿射变换矩阵的输入和特征变换来处理数据，并将该变换直接应用于点的坐标，然后通过最大池化层进行聚合，获得全局特征。一个分割网络，将全局特征与局部特征进行拼接，得到点分割并得到评分结果。PointNet++[12]基于PointNet因采样点不均匀而缺失局部特征问题，通过添加扩展结构对模型进行了改进，它结合了不同规模区域的特征，以响应输入样本密度的变化。Wu等[13]提出的PointConv具有与PointNet++相似的结构，但用PointConv层取代了PointNet中的结构，它使用多层感知机（Multilayer Perception，MLP）为每个卷积滤波器近似一个权重函数，然后使用密度尺度重新加权学习的权重函数。Shi等人提出了两阶段3D目标检测网络PointRCNN[14]（见图3），第一阶段将点云分割为前景点和背景，第二阶段结合第一阶段每个点的语义特征，实现了精确的预测，但是其实时性相对较差。Yang等[15]提出了3DSSD网络，它移除了Point-base方法中必须的FP层和细化模块，提出了一种新的基于特征距离的融合采样策略F-FPS，用来保留各类前景实例中的内部点，以此来实现分类和回归任务信息的丰富性，并且相比于最先进的基于点的方法快了2倍。

2.1.3　基于体素-点的方法

通常，基于体素的方法在计算方面具有很高的效率，但是体素划分过程中物体划分不全使得局部信息丢失，导致检测精度降低。基于点的方法计算更为复杂，但是其能获得更大的感受野，检测精度相对较高。有学者结合二者的优点提出了基于体素-点的方法。

Chen等[16]提出的Fast Point R-CNN是一个两阶段检测模型（如图4）。第一阶段通采用了体素化思想使用VFE网络将点云进行编码并作为输入完成3D目标

预测。第二阶段，将与原始点云和上下文特征提取合并，并融入注意力机制以获取更好的定位信息。Yang等[17]提出了STD两阶段检测模型，它使用原始点云作为输入，计算每一个点并使用球形锚框来生成精确的候选框，与基于体素化思想的候选框栅格特征提取方法相比，它使用较少的计算量实现了更高的精度。在第二阶段使用并行交叉IoU分支，使得定位精度提高，从而进一步提升性能。

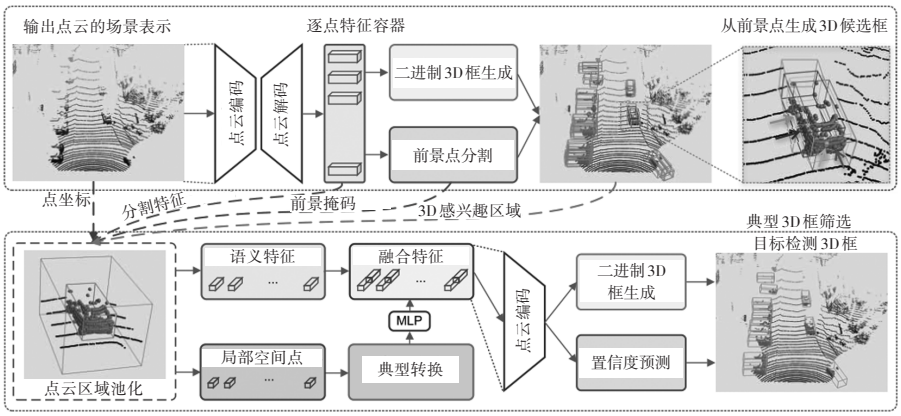



**图3　PointRCNN网络结构[14]**

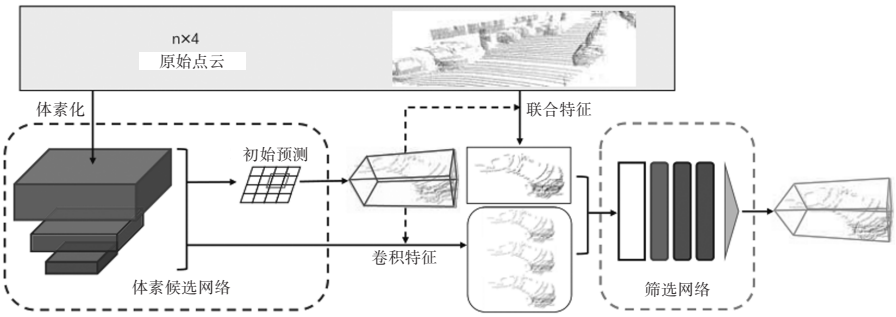



**图4　Fast Point R-CNN网络结构[16]**

Shi等人提出了PV-RCNN[18]网络，利用体素到关键点场景编码与点到网格RoI特征提取，利用Voxel-based操作进行有效的多尺度信息编码，生成高质量的3D候选框；同时利用改进的SA模块操作保留精确的位置信息和灵活的感受野。He等[19]提出了SA-SSD网络，通过预处理对点云进行体素化，基于backbone学习体素特征，并在主干网络外通过点监督网络将各体素特征转换为点特征，通过增加2个点级的任务让学习来的特征能更好地感知位置信息。Miao等[20]提出了一个基于点云、体素以及网格特征融合的单阶段3D目标网络PVGNet。该网络使用一个网络来对提取点云、体素和网络特征，通过融合不同层的特征可以更好的挖掘点云信息。

### 2.2　基于图像与点云融合的方法

基于图像与点云融合的检测方法融合了图像检测中丰富的纹理信息与点云检测中的深度信息，纹理信息对于识别和分类起着至关重要的作用，而深度信息可以准确地定位物体的大小以及位置关系。通过两者检测信息互补，理论上可以达到更好的检测效果。基于融合的方法主要分为顺序融合与并行融合2类。

#### 2.2.1　顺序融合

这种方法是以顺序的方式对图像和点云进行融合，首先提取图像特征，然后将图像特征投影或映射到点云上，最后通过检测网络输出检测结果，流程如图5所示。

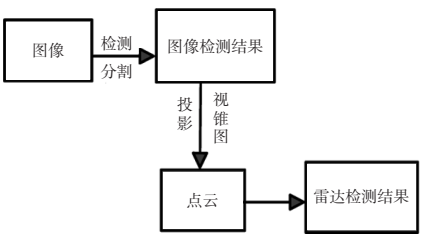



**图5　顺序融合原理**

Qi等[21]提出了F-PointNet网络（见图6），该网络利用2D CNN对象检测器来提出2D区域并对其内容进行分类。然后将2D区域提升到3D，从而成为平截头体方案。最后，框估计网络估计对象的amodal 3D边界框，这在一定程度上提升了检测精度，但3D对象检测预测结果容易受到从2D图像获得的外部依赖性的影响。针对这一问题，Pei等[21]提出了混合多种特征金字塔网络（Mutiple Feature Pyramid Network，MFPN），

通过2D目标检测网络识别目标在RGB图像中的位置，然后利用视锥图将图像映射到点云中，通过改变视椎体(Frustum)的建议框，将结果与BEV物体检测进行比较，并惩罚由于条件造成的漏点以提高准确性。

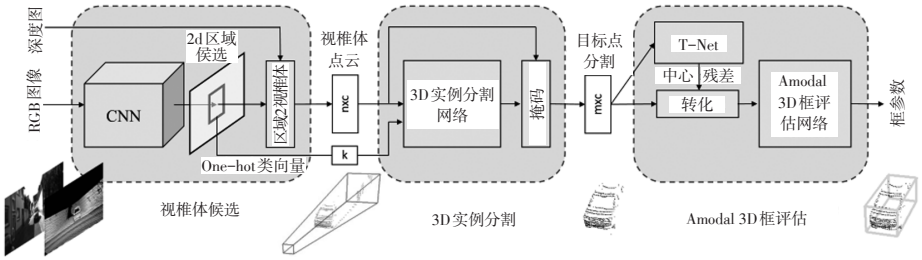
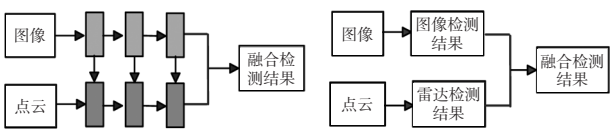



**图6　F-PointNet网络结构[23]**

Anshul等[23]基于F-PointNet的思想提出了F-PointPillars网络，首先将二维检测映射到三维边界截锥体中，并去除截锥体外的点。其次对于每个2D检测，使用高斯函数创建一个掩码，表示像素属于对象的可能性。可能性值被投影到点云上，并将整个3D空间离散化为一个2D网格，形成一组支柱。在每个非空支柱内使用PointNet提取支柱特征，然后将这些特征散回到一个2D伪图像中。使用一组卷积和反卷积提取多个分辨率下的空间特征。最后采用边界框回归进行检测。Vora等人提出了一种通用的顺序融合检测方法PointPainting[24]，该网络通过对图像进行语义分割得到各类别障碍物分割分数，然后将点云投影到分割图像上融合分割结果以达到语义增强的效果，该网络对于小目标检测有较大的提升。Sindagi等人提出了MVX-Net[25]融合网络，使用2D检测网络提取图像语义编码特征，分别融入到体素点特征上和进过VFE编码后的体素特征上最后得出3D检测结果。

2.2.2　并行融合

并行融合是通过对图像和点云分别进行特征提取，然后对图像特征和点云特征进行融合，主要有特征融合和目标融合2种，如图7所示。




(a)特征融合原理　　(b)目标融合原理

**图7　并行融合原理**

Chen等[26]提出了MV3D网络(见图8)，通过点云的鸟瞰图生成3D候选框，再将候选框投影到鸟瞰图、点云前视图和图像上以获取区域特征，再将不同的模态信息进行融合得到融合特征，最后用于分类和边界框回归，由于其使用了下采样导致小目标信息丢失，使得小目标检测精度低。针对这一问题，Ku等[28]提出了AVOD算法，利用FPN[28]网络对特征进行提取得到图像和BEV视角下全尺寸特征图，然后利用1×1卷积和crop&resize操作处理并融合特征图，这在一定程度上改善了小目标检测的效果，但是其裁剪操作可能使得特征之间存在不对应关系。针对这一问题，Liang等[29]提出了MMF网络有2个支流，一个是通过ResNet18[30]提取图像特征并融合多尺度图像特征，另一个支流是通过连续融合层将多尺度图像特征融入点云鸟瞰(Bird's Eye View，BEV)特征提取网络，实现了多尺度的传感器融合，最终在BEV空间下生成检测结果。Pang等人提出了一种高效的低复杂度融合模型CLOCs[31]，该模型首先利用2D和3D目标检测网络分别提出各自的候选框，然后通过编码网络将各自的候选框编码为稀疏张量，最后利用2D卷积对非空元素进行特征融合并输出检测结果[32]。

## 3　数据集及评估

在自动驾驶中安全性是最重要的要求，所以对环境感知算法的研究需要考虑各种各样的道路环境，并且在深度学习中无论是模型训练还是试验验证都离不开数据集，基于这一问题部分科研机构开源了大型自动驾驶数据集，常用的自动驾驶数据集如表2所示。

### 3.1　自动驾驶数据集

(1)KITTI数据集

KITTI数据集是由德国卡尔斯鲁厄理工学院和丰田美国技术研究院联合创立。它是最早开源的自动驾驶数据集，使用64线激光雷达、2个灰度相机和2个彩色相机采集道路信息，可用于2D/3D检测。该数据集主要包括城市、乡村和高速等场景信息。由7 481帧标注图片组成训练集和验证集，7 518张图片组成测试集，共计有超过20万个3D标注对象。主要标注物体为人、汽车和骑行者，然后依据遮挡、远近等因素分为简单、中等、困难3个不同等级供研究人员验证自己的网络。

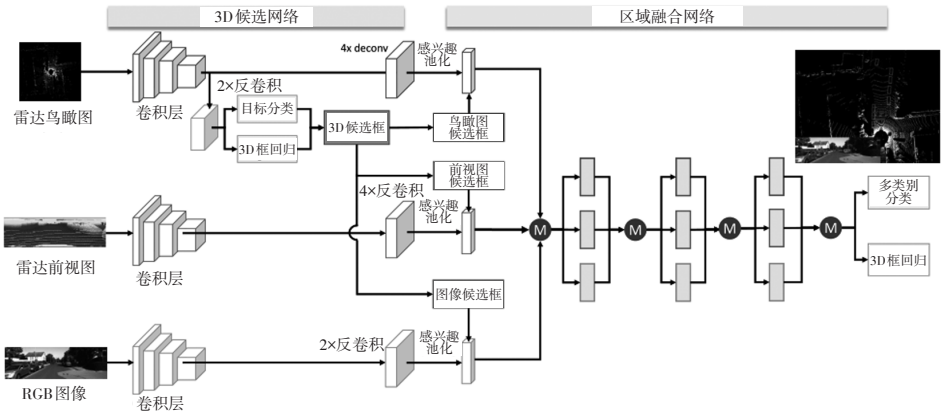



**图8　MV3D网络结构[26]**

**表2　开源数据集对比**

| 数据集 | 发布时间/年 | 视角范围/° | 场景/个 | 总时长/h | 环境 |
|---|---|---|---|---|---|
| KITTI | 2012 | 90 | 22 | 6 | 白天 |
| Waymo | 2019 | 360 | 1 000 | 16.7 | 白天、夜晚、雨天 |
| NuScenes | 2019 | 360 | 1 000 | 15 | 白天、夜晚、雨天 |
| ApolloScape | 2018 | 360 | | > 100 | 白天、夜晚 |
| Lyft | 2020 | 360 | 170 000 | > 1 000 | 白天、夜晚、雨天 |

该数据集是使用最广泛的数据集，但是数据集存在局限性，其标注信息依照相机视角只标注了正向90°区域的目标，并且全是在视野良好的白天工况，其中大多数标注对象为汽车，其标注信息缺乏多样性。

（2）Waymo数据集

Waymo数据集[33]是谷歌自动驾驶公司公布的开源数据集，它由5个雷达和5个相机采集而成，整个数据集分为1 000个训练集和150个测试集，每个场景有20 s标注数据，总计有超过1 200万个标注信息，其中包含了行人、车辆和路标等目标，并且在每帧之间使用一致的标识符，可以为跟踪任务提供基线。

该数据集限制激光雷达数据的范围，并为每个激光脉冲的前2次返回提供数据。相机图像是通过滚动快门扫描拍摄的，精确的扫描模式可能会因场景而异。所有相机图像都被下采样并从原始图像中裁剪，这样可以获得更加精确的环境信息。

（3）NuScenes数据集

NuScenes[34]是由Motional团队公布的开源数据集。由6个相机、1个激光雷达采集而成，它包括了新加坡和波士顿2个城市中1 000个不同的驾驶场景，整个数据集分为850个训练集和150个测试集，每个场景有20 s标注数据，包括不同天气情况以及道路条件。该数据集的标注信息包括了汽车、行人、卡车、公交以及交通标注等23种标注类别总计超过140万个标注对象。

相比于KITTI数据集，NuScenes的数据规模更大，实现了360°标注，包括不同的天气和光照等场景，其标注信息更具多样性，并且还提供了人类注释语义地图。但是其主要针对3D目标检测任务，缺少2D包围框的标注。

（4）ApolloScape数据集

ApolloScape数据集[35]是由百度公司开源的大型数据集。为了刻画高细粒度的静态3D世界，ApolloScape使用Reigl移动三维激光扫描仪收集点云。这种方法生成的三维点云要比Velodyne激光雷达生成的点云更精确、更稠密。在采集车车顶上安装有标定好的高分辨率相机，以30帧/s的速率同步记录采集车周围的场景。该数据集是目前行业内环境最复杂、标注最精准、数据量最大的自动驾驶公开数据集。ApolloScape的标注精细度超过同类型的KITTI、Cityscapes数据集。并且Apollo Scape还使用仿真环境来标注数据集，通过模拟虚拟驾驶场景来实现对真实道路的还原，并记录相关环境信息。

该数据集是由图像和稠密点云组成，包含了超过14万张高清图像。该数据集标注了25种类别，包括汽车、行人和交通标注等，相比于传统标注信息，该数据集标注了不同类型的车道线，做到对场景的全面分析。

（5）Lyft数据集

Lyft[36]是由美国自动驾驶车队公布的开源数据集，由20辆搭载了7个摄像头和5个激光雷达的自动驾驶汽车组成的车队在加利福尼亚州帕洛阿尔托的一条固定路线上收集的。该数据集由170 000个场景组成，每个场景长25 s，总计超过1 000 h，捕捉自动驾驶系统的感知输出，该系统对附近车辆、骑车者和行

人随时间变化的精确位置和运动进行编码。除此之外，数据集还包含一张高清语义图，其中包含15 242个标记元素和该地区的高清鸟瞰图。该数据集是可用于训练预测和规划解决方案的最大、最详细的数据集。它比目前的最佳替代方案大3倍，而且更具描述性。这种差异会显著提高轨迹预测和运动规划任务的性能。

### 3.2 评估

为了对一个模型检测性能进行判断，常用的评估方法有模型检测速度、目标定位精度、目标检测精度、平均方向相似性4种。

（1）模型检测速度，通常采用每秒检测帧数来评估，通常每秒处理的帧数越多，模型检测的实时性能越高。

（2）目标定位精度，当前常用的方法是通过交并比（IoU）数值的大小来评估定位精度，即通过模型检测生成的预测框与真实框之间重合度的比值大小，如图8所示。*IoU*变化范围为[0,1]，越接近1定位精度越高，计算公式如式（1）。

$$IoU = \frac{A \cap B}{A \cup B} \tag{1}$$

式中，$A$为预测框大小；$B$为真实框大小。

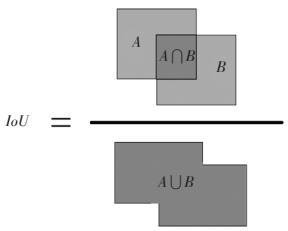



**图8 *IoU*计算原理**

（3）目标检测精度，通常采用查准率（precision）与查全率（recall）来评估检测精度，计算如式（2）、式（3）。

$$precision = \frac{TP}{TP + FP} \tag{2}$$

$$recall = \frac{TP}{TP + FN} \tag{3}$$

式中，$TP$为被正确识别的正样本；$FP$为负样本但被识别为正样本；$FN$为正样本但被识别为负样本。

（4）针对3D目标检测任务KITTI数据集定义了平均方向相似性（Average Orientation Similarity，AOS）指标，用于评价目标航向角的预测结果，定义如式（4）。

$$AOS = \frac{1}{11} \sum_{r \in \{0, 0.1, \cdots, 1\}} \max_{\tilde{r}: \tilde{r} \geq r} s(\tilde{r}) \tag{4}$$

式中，$r$代表物体检测的召回率recall。

在因变量$r$下，方向相似性$s \in [0,1]$被定义为所有预测样本与ground truth余弦距离的归一化，如式（5）。

$$s(r) = \frac{1}{|D(r)|} \sum_{i \in D(r)} \frac{1 + \cos \Delta_\theta^{(i)}}{2} \delta_i \tag{5}$$

式中，$D$(r)表示在召回率$r$下所有预测为正样本的集合，$\Delta_\theta^{(i)}$表示检出物体$i$的预测角度与真实值的差。为了惩罚多个检出匹配到同一个真实值，如果检出$i$已经匹配到真实值（$IoU \geq 50\%$）设置$\delta_i = 1$，否则$\delta_i = 0$。

### 3.3 小结

本节主要分析了主流的自动驾驶开源数据集。其中KITTI数据集作为开源最早的自动驾驶数据集，为2D和3D环境感知技术的研究提供了巨大的帮助，但是存在标注信息的局限性。NuScenes数据集作为3D目标检测主要的数据集具有标注多样性，场景丰富等优点，可用于复杂环境的模拟，但是2D标注信息较少，不适用于二维检测任务。Waymo数据集是目前最大的自动驾驶开源数据集，它包含了丰富的2D和3D标注信息，适用于多数自动驾驶场景。ApolloScape是目前为止纹理信息最为精确的数据集，并且标注了车道线信息，可以适用于全方面的检测任务。Lyft包含了语义级别的高清地图，可以更好地进行轨迹跟踪与预测。本节还分析了模型评估方法，利用检测帧数分析模型检测速度，利用交并比IoU分析模型定位精度，利用查准率和查全率分析模型检测精度以及利用AOS分析模型的航向角预测结果。

## 4 结束语

三维物体检测是自动驾驶汽车领域的一项重要任务，本文首先介绍了车载传感器相关知识及应用场景，其次综述了以雷达信息为主要输入的三维目标识别技术和模型，包括基于点云的方法和基于图像与点云融合的方法。基于点云方法是一个具有最佳效果的潜在应用领域，但面临的挑战是最大限度地降低计算资源和实时应用的成本。基于融合的方法在实际应用的实施资源和时间上都有很大的改进潜力，但对该方法的研究仍然有限。最后针对自动驾驶领域开源的大型数据集做了相应的总结分析及3D目标检测评价指标的分析，为后续研究人员提供帮助。

近几年来，随着自动驾驶技术的发展，对于环境感知的能力也随之提高，3D目标检测作为自动驾驶技术中的关键任务，仍然面临着许多难题和挑战。结合本文综述内容，对未来可能的研究趋势进行了分析。

（1）2D视图法

目前，将雷达点云处理为2D鸟瞰图（BEV）的方法是3D目标检测领域研究热点。其主要是通过压缩

空间特征，将3D目标检测任务转换为2D目标检测，使得检测任务更加简单、快速。例如PointPillar[8]基于柱状的思想将点云压缩到二维平面，然后利用2D卷积进行运算，从而提高处理效率。Complexer-YOLO[37]直接将原始点云压缩成为2D鸟瞰图，然后基于2D卷积进行运算，极大地提高了计算效率。

（2）多模态融合法

目前，多模态融合检测是自动驾驶车辆上运用最为广泛的方法。其主要是通过对雷达与图像数据进行对齐投影，构建跨数据特征融合，从而获取更好的检测效果。例如BEVFusion[38]分别将图像特征和雷达特征进行编码，然后通过共享网络进行融合，这很大程度上提高了检测效率与精度。

在未来的一段时间内，自动驾驶技术会逐渐地趋于成熟，无论是基于视图的3D目标检测还是基于多模态的3D目标检测算法，都能为自动驾驶技术带来无限的可能性，促进自动驾驶行业的发展。

## 参 考 文 献


[1] 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.

[2] KU J, MOZIFIAN M, LEE J. Waslander.Joint 3D proposal generation and object detection from view aggregation [C]// 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2018: 1-8.

[3] CHARLES R Q, HAO S, MO K C, et al. PointNet: Deep learning on point sets for 3D classifification and segmentation [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2017: 77-85.

[4] ENGELCKE M, RAO D, WANG D Z, et al. Vote3Deep: Fast object detection in 3D point clouds using effifficient convolutional neural networks [C]//2017 IEEE International Conference on Robotics and Automation. IEEE, 2017: 1355-1361.

[5] ZHOU Y, TUZEL O. VoxelNet: End-to-end learning for point cloud based 3D object detection [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 4490-4499.

[6] YAN Y, MAO Y, LI B. SECOND: Sparsely embedded convolutional detection [J]. Sensors: Basel, 2018, 18(10): E3337.

[7] DENG J J, SHI S S, LI P W, et al. Voxel R-CNN: Towards high performance voxel-based 3D object detection [J]. Proceedings of the AAAI Conference on Artifificial Intelligence, 2021, 35(2): 1201-1209.

[8] LANG A H, VORA S, CAESAR H, et al. Pointpillars: Fast encoders for object detection from point clouds [C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2019: 12689-12697.

[9] SONG X, WEI Q Z, XIAO Y C, et al. Scale-Aware Attention-Based PillarsNet (SAPN) Based 3D Object Detection for Point Cloud [J]. Mathematical Problems in Engineering (10), 2020: 3927365.

[10] MAO J G, XUE Y J, NIU M Z, et al. Voxel transformer for 3D object detection [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2021: 3144-3153.

[11] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017: 5999-6009.

[12] QI C R, YI L, SU H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space [J/OL]. (2017-06) [2023-12-30]. https://www.researchgate.net/publication/317426798_PointNet_Deep_Hierarchical_Feature_Learning_on_Point_Sets_in_a_Metric_Space.

[13] WU W, QI Z, LI F X. Pointconv: Deep convolutional networks on 3d point clouds [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019: 9621-9630.

[14] SHI S S, WANG X G, LI H S. PointRCNN: 3D object proposal generation and detection from point cloud [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019: 770-779.

[15] YANG Z T, SUN Y N, LIU S, et al. 3DSSD: Point-based 3D single stage object detector [C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020: 11037-11045.

[16] CHEN Y L, LIU S, SHEN X Y, et al. Fast point R-CNN [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2019: 9774-9783.

[17] YANG Z T, SUN Y N, LIU S, et al. STD: Sparse-to-dense 3D object detector for point cloud [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2019: 1951-1960.

[18] SHI S S, GUO C X, JIANG L, et al. PV-RCNN1: Point-voxel feature set abstraction for 3D object detection [C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020: 10526-10535.

[19]HE C H, ZENG H, HUANG J Q, et al. Structure aware single-stage 3D object detection from point cloud [C]//2020

IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020: 11870-11879.

[20] MIAO Z, CHEN J, PAN H, et al. PVGNet: a bottom-up one-stage 3D object detector with integrated multi-level features [C]//Proceedings of the 2021 IEEE Conference on Computer Vision and Pattern Recognition, Nashville, Jun 20-25, 2021. Piscataway: IEEE, 2021: 3279-3288.

[21] QI C R, LIU W, WU C X, et al. Frustum PointNets for 3D object detection from RGB-D data [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 918-927.

[22] CAO P, CHEN H, ZHANG Y, et al. Multi-view frustum pointnet for object detection in autonomous driving [C]// 2019 IEEE International Conference on Image Processing (ICIP). IEEE, 2019: 3896-3899.

[23] PAIGWAR A, SIERRA-GONZALEZ D, ERKENT Ö, et al. Frustum-PointPillars: A Multi-Stage Approach for 3D Object Detection using RGB Camera and LiDAR[C]//2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). IEEE, 2021: 2926-2933.

[24] VORA S, LANG A H, HELOU B, et al. PointPainting: Sequential fusion for 3D object detection [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020: 4603-4611.

[25] SINDAGI V A, ZHOU Y, TUZEL O. MVX-net: Multimodal VoxelNet for 3D object detection [C]//2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019: 7276-7282.

[26] CHEN X Z, MA H M, WAN J, et al. Multi-view 3D object detection network for autonomous driving [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2017: 6526-6534.

[27] KU J, MOZIFIFIAN M, LEE J, et al. Joint 3D proposal generation and object detection from view aggregation [C]// 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2018: 1-8.

[28] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature Pyramid Networks for Object Detection [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017: 936-944.

[29] LIANG M, YANG B, CHEN Y, et al. Multi-task multi-sensor fusion for 3D object detection [C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019: 7337-7345.

[30] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016: 770-778.

[31] PANG S, MORRIS D, RADHA H. CLOCs: Camera-LiDAR object candidates fusion for 3D object detection[C]// 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2020: 10386-10393.

[32] AHMAD W A, WESSEL J, NG H J, et al. IoT-ready millimeter-wave radar sensors [C]//2020 IEEE Global Conference on Artifificial Intelligence and Internet of Things. IEEE, 2020: 1-5.

[33] SUN P, KRETZSCHMAR H, DOTIWALLA X, et al. Scalability in perception for autonomous driving: Waymo open dataset [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020: 2443-2451.

[34] CAESAR H, BANKITI V, LANG A H, et al. nuScenes: A multimodal dataset for autonomous driving [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020: 11618-11628.

[35] HUANG X Y, WANG P, CHENG X J, et al. The ApolloScape open dataset for autonomous driving and its application [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(10): 2702-2719.

[36] HOUSTON J , ZUIDHOF G , BERGAMINI L , et al. One Thousand and One Hours: Self-driving Motion Prediction Dataset [J/OL]. (2020-06-25)[2023-12-30]. https://www.researchgate.net/publication/342464189_One_Thousand_and_One_Hours_Self_driving_Motion_Prediction_Dataset.

[37] SIMON M, AMENDE K, KRAUS A, et al. Complexer-YOLO: Real-time 3D object detection and tracking on semantic point clouds[C]//2019 IEEE/CVF Conference On Computer Vision And Pattern Recognition Workshops (CVPRW). IEEE, 2019: 1190-1199.

[38] LIU Z, TANG H, AMINI A, et al. BEVFusion: Multi-Task Multi-Sensor Fusion with Unified Bird's-Eye View Representation[C]//2023 IEEE International Conference on Robotics and Autimation. IEEE, 2023: 2774-2781.

（责任编辑　**明慧**）

**【作者简介】**

余杭(1998—),男,重庆交通大学,硕士研究生,研究方向为智能网联汽车3D目标检测。

E-mail:yh2022_03@163.com