

## Project4 Reference 整理

### 1. Cylinder3D: An Effective 3D Framework for Driving-scene LiDAR Semantic Segmentation

原文地址: <https://arxiv.org/abs/2011.10033>

Github: <https://github.com/xinge008/Cylinder3D>

对于室外驾驶场景中大规模 LiDAR 点云分割,常用的方法是将点云投影到二维空间,再通过 2D 卷积对其进行处理。这样做虽然有不错的表现,但它丢弃了点云的三维拓扑和几何关系信息。一种减少这种信息丢失的方法是对点云使用立方体划分的体素化再通过 3D 卷积对其进行处理,但这种方法得到的改善相当有限,因为室外点云还具有稀疏性和密度不均匀性。这篇论文提出了 Cylinder3D:

1. 针对点云的稀疏性和密度不均匀性,对点云进行圆柱体划分的体素化
2. 使用不对称的 3D 卷积网络来生成体素级输出
3. 引入逐点细化模块,减少了因体素化带来的标签干扰

### 2. Center-based 3D Object Detection and Tracking

原文地址: <https://arxiv.org/abs/2006.11275>

Github: <https://github.com/tianweiy/CenterPoint>

提出了一种 center-based 的激光雷达点云同步三维目标检测与跟踪框架。我们的方法使用一个标准的 3D 点云编码器,在头部使用一些卷积层来生成鸟瞰图热图和其他密集回归输出。检测是一个简单的局部峰值提取与细化,而跟踪是一个最接近距离匹配。CenterPoint 很简单,接近实时,并在 Waymo 和 nuScenes 基准测试中实现了最先进的性能

### 3. BEVFusion: Multi-Task Multi-Sensor Fusion with Unified Bird's-Eye View Representation

原文地址: <https://arxiv.org/abs/2205.13542>

Github: <https://github.com/mit-han-lab/bevfusion>

BEVFusion 在共享的 BEV 空间下统一 camera feature 和 Lidar 点云 feature, 为了实现 camera-to-BEV 转换的高效,文中提出实现并行计算的 GPU kernel,将速度提升 40 倍。BEVFusion 打破了以往将 point-level fusion 作为最佳解决方案的经验,并在 3D 检测和 BEV segmentation 均实现新 SOTA (小目标、雨天和夜间的效果也非常好)。

### 4. EA-LSS: Edge-aware Lift-splat-shot Framework for 3D BEV Object Detection

原文地址: <https://arxiv.org/abs/2303.17895>

Github: <https://github.com/hht1996ok/EA-LSS>

这篇文章介绍了一种名为 EA-LSS (Edge-aware Lift-splat-shot) 的新型框架,用于改进基于 Lift-Splat-Shot (LSS) 的 3D 目标检测性能。

EA-LSS 框架包含两个关键模块:边缘感知深度融合 (EADF) 模块和细粒度深度 (FGD) 模块。如下图所示,EADF 用于提高深度估计网络在处理深度变化快速的区域(深度跳跃)时的准确性,FGD 则是在训练阶段提供更精细的深度监督。

### 5. PVRCNN: Point-Voxel Feature Set Abstraction for 3D Object Detection

原文地址: <https://arxiv.org/pdf/1912.13192>

Github: <https://github.com/open-mmlab/OpenPCDet>

提出了 PV-RCNN,一种从点云中精确检测三维物体的新方法。该方法使用本文提出的 voxel set abstraction layer 集成了多尺度 3D 体素 CNN 特征和使用 PointNet 采集到

的特征。并且通过关键点的特征，有效的优化了一阶段检测生成的预测框和评分。在 KITTI 和 Waymo Open 数据集上取得了非常优秀的效果。与以前最先进的算法相比，本文提出的 voxel-to-point 和 keypoint-to-grid 结构有效显著的提高了三维物体检测的性能。

## 6. Mask3D: Mask Transformer for 3D Semantic Instance Segmentation

原文地址: <https://ieeexplore.ieee.org/abstract/document/10160590>

Github: <https://github.com/jonasschult/mask3d>

本文介绍了一种基于 Transformer 的 3D 语义实例分割方法，名为 Mask3D。与现有的方法不同，Mask3D 直接预测实例掩码而不需要手动选择的投票方案或手工制作的分组机制。文章详细介绍了 Mask3D 的组成部分，包括稀疏特征骨干、实例查询和掩码模块等。实验结果表明，Mask3D 在多个数据集上取得了最先进的表现。

## 7. OneFormer3D: One Transformer for Unified Point Cloud Segmentation

原文地址: <https://arxiv.org/abs/2311.14405>

GitHub: <https://github.com/filapro/oneformer3d>

3D 点云分割主要包含三类: 语义、实例和全景。语义分割为每个语义类别输出一个掩码，使得点云中的每个点被分配一个语义标签。实例分割返回一组单个对象的掩码。全视角分割为每个前景对象预测一个掩码，并为每个背景点预测一个语义标签。

三个三维分割任务实际上都隐含着预测一组掩码，但它们通常是完全不同的架构。三维语义分割方法依赖于 U-Net 网络。三维实例分割方法将语义分割模型与基于聚类、目标检测或 Transformer 解码器的聚合方案相结合。三维全景分割

在二维图像中执行全景分割，然后将预测的掩码提升到三维空间并逐点聚合。

至此，出现了一个很自然的问题：能否用单一统一的方法共同解决所有三个三维分割任务？

这就是 OneFormer3D 的工作，第一个多任务统一的三维分割框架。使用 SPFormer 基线，在 Transformer 解码器中并行添加语义查询和实例查询，以统一预测语义和实例分割掩码。然后，确定了基于 Transformer 的三维实例分割性能不稳定的原因，并通过新的查询选择机制和新的高效匹配策略解决了问题。最后，提出了一个只需训练一次的单一统一模型，即使它们专门针对每个任务进行了调整，也能优于三维语义、三维实例和三维全视角分割方法。