

激光雷达感知障碍物检测 深度学习算法介绍

一汽解放商用车开发院AI实验室

吴再霖

2023年4月

一 CNN Segmentation

二 Point Pillars

三 Mask Pillars

四 Center Points

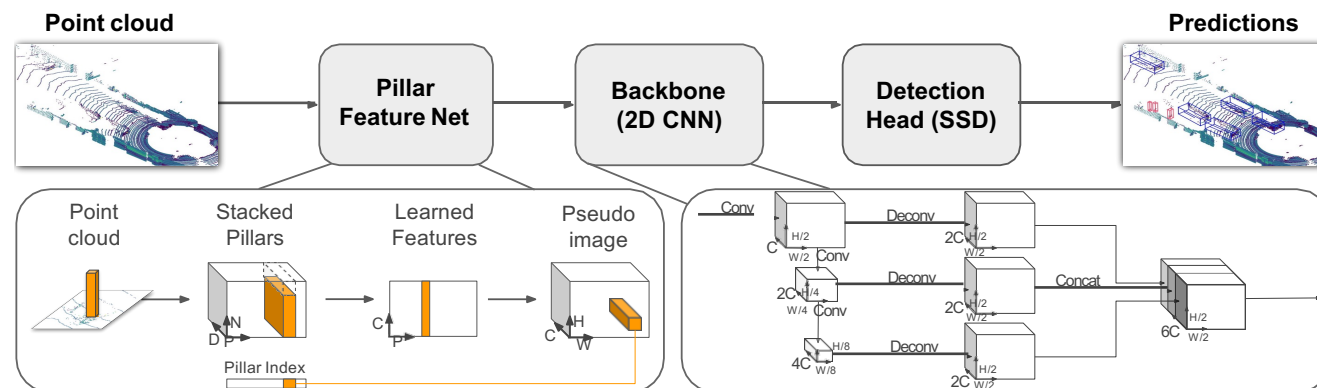
一、CNN Segmentation

□ 算法背景与概述

二、Point Pillars

□ 算法背景与概述

- PointPillars是一个用于对激光雷达点云进行3D目标检测的端到端算法模型，其由创业公司nuTonomy提出并进行效果验证；
- PointPillars算法提出一种新的点云编码方式用于给PointNet提取点云特征，并将提取特征映射为2D伪图像以便通过2D目标检测的方式进行目标分类并提取目标检测框；
- 相较于VoxelNet等3D目标检测方法，PointPillars省去计算量复杂且庞大的三维卷积神经网络(3DCNN)推理过程，以达到检测精度与检测效率间的平衡；
- 在KITTI训练集上，PointPillars算法以较快速度实现较高的检测性能。
- 百度Apollo平台于6.0版本引入PointPillars算法以完成智能驾驶中的激光雷达感知任务



cite: PointPillars: Fast Encoders for Object Detection from Point Clouds (CVPR2019), <https://arxiv.org/abs/1812.05784>

□ Part1: Pillar Feature Net (PFN) —点云转换伪图像 (编码)

- Input: 点云图像, 包括点云坐标、强度等;
- 1. 柱形图单元分割: 每个单元 (pillar) 为在X-Y (笛卡尔坐标) 上以某固定步长对点云进行划分得到的一个3维柱状体, 其在Z轴上无限;
- 2. 生成张量 (Tensor) : (D, P, N)

D: (x,y,z,r,xc,yc,zc,xp,yp)

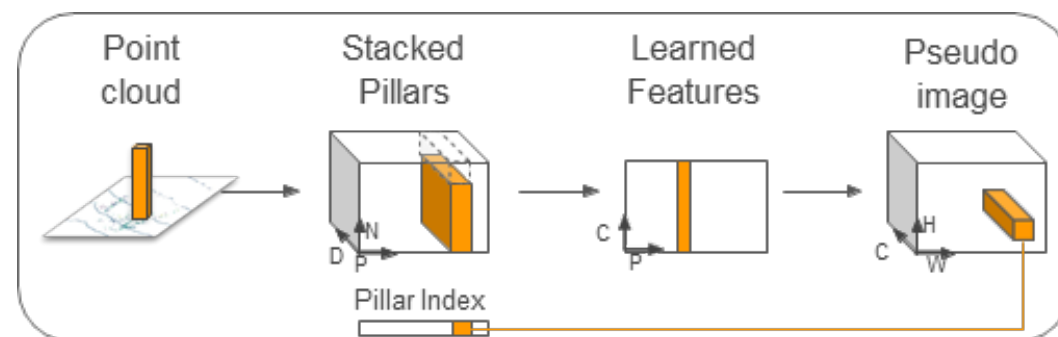
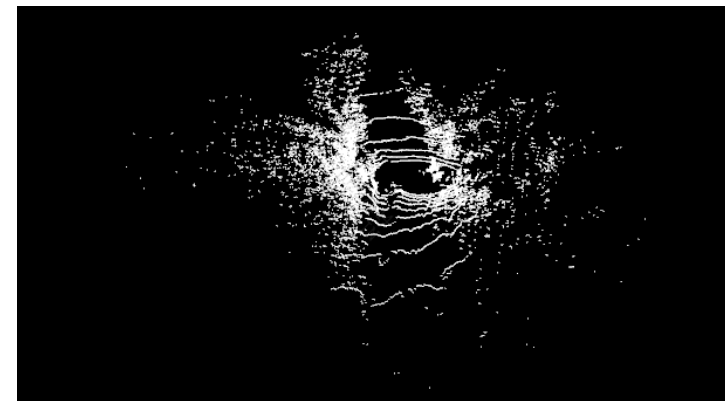
- x,y,z,r: 单元中每个点在三维空间中的坐标(x,y,z)与反射强度r
- xc,yc,zc: 单元中每个点到该单元所有点的算术平均值点各坐标的距离;
- xp,yp: 单元中每个点到该单元X-Y坐标中心的偏移量;

P: 单元格数量(Pillars Num)及索引 (Index)

- 由于计算复杂度的限制, 模型会滤掉不包含点云或包含少量点云的单元格, 即对单元格数量做出限制, 最多处理P个单元格;

N: 单个单元格内点数(Points Num per Pillar)

- 每个单元中最多包含N个表示点特征的向量, 若点数量大于N, 则采用随机采样的方法从中选取N个, 反之, 若点数量少于N, 则用零填充的方法填充到N个。



二、Point Pillars

□ Part1: Pillar Feature Net (PFN) —点云转换伪图像 (伪图像生成)

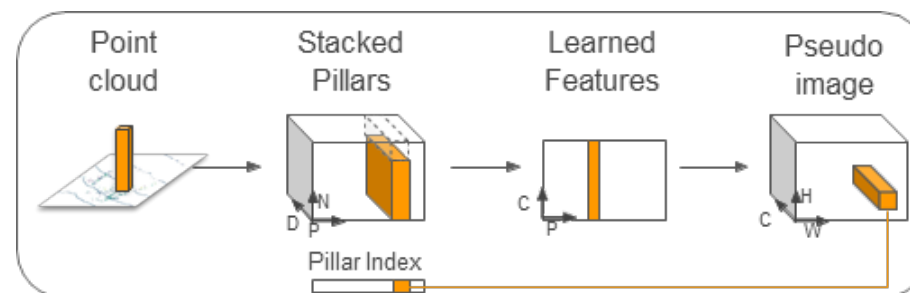
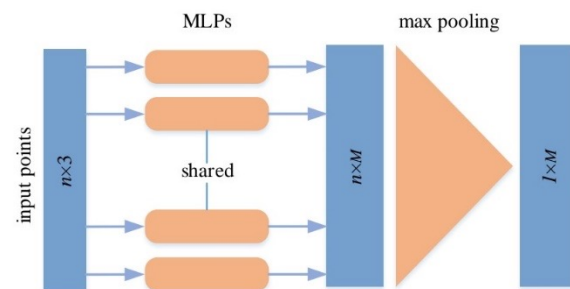
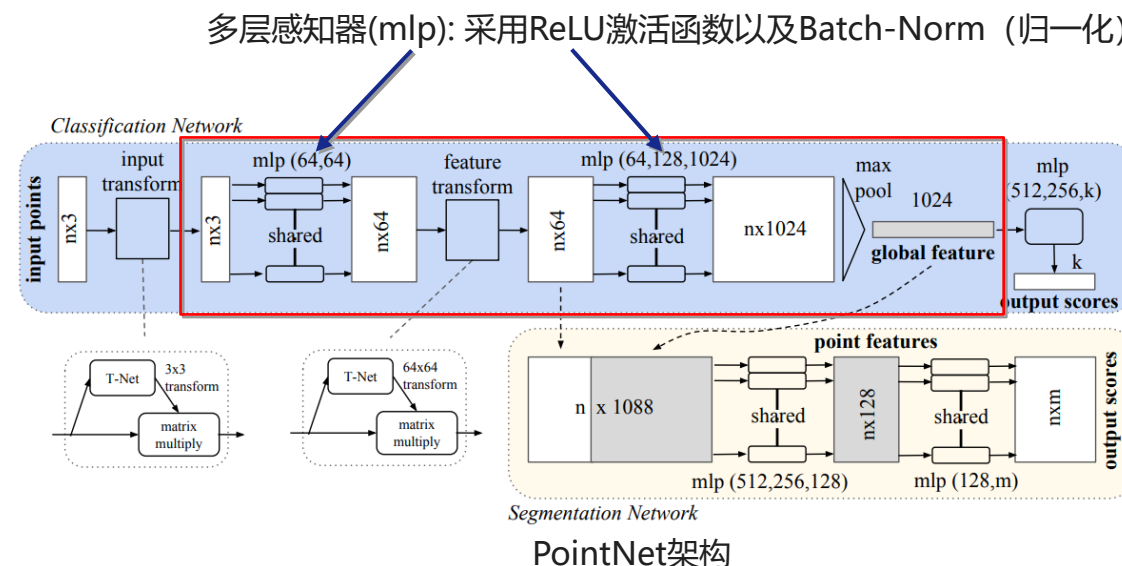
3. 特征学习：利用简化PointNet模型对张量中每一个点特征进行处理，并对每个单元格中的点进行最大池化(max pool)

张量处理： $(D, P, N) \rightarrow (C, P, N)$

最大池化： $(C, P, N) \rightarrow (C, P)$

4. 伪图像生成：通过每个单元格索引值P将上步生成的 (C,P) 张量转换回其原始坐标来创建大小为 (C,H,W) 的伪图像，H表示伪图像的高，W表示伪图像的宽，以上将原有三维空间点云转换为伪图像，伪图像中的每一个像素表示对应的单元(Pillar)。

- Output: 伪图像(Pseudo-image)

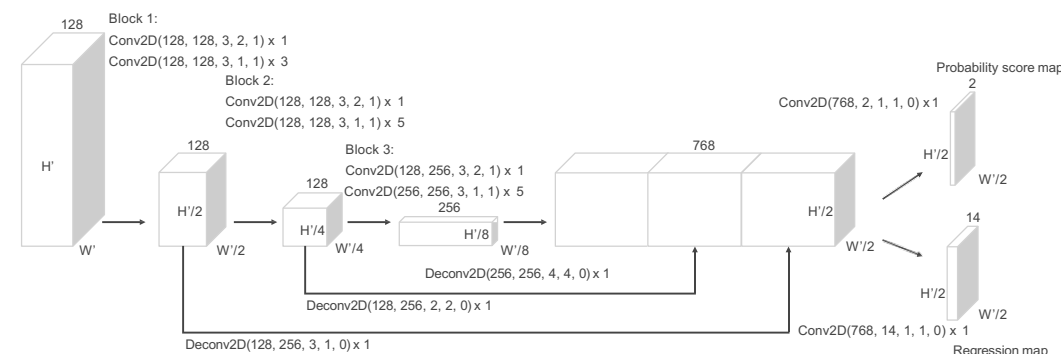
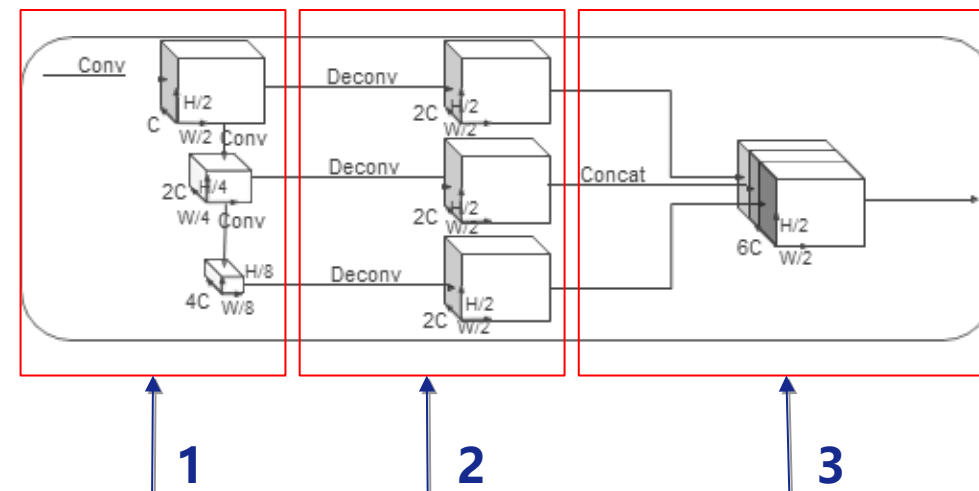


cite: PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation (CVPR2017), <https://arxiv.org/abs/1612.00593>

二、Point Pillars

□ Part2: Backbone (2D CNN) —神经网络特征提取

- 此部分由特征候选网络(Region Proposal Network)构成:
 - 1. **下采样层**: 由图中自上而下的网络[1]以越来越小的空间分辨率产生特征, 用以检测不同大小的目标, 如小分辨率检测诸如车辆等较大目标, 大分辨率检测诸如行人、单车等较小目标;
 - 2. **上采样层**: 由图中的平行网络[2]构成, 将下采样层得到的特征图进行上采样(反卷积), 使其具有同样的分辨率;
 - 3. **串联层**: 图末[3], 不同步骤得到的特征图进行串联(特征融合), 使得最后得到的特征图具有之前各个不同步骤的所有特征。
- 最后得到的特征图用以对点云进行目标检测及分类。

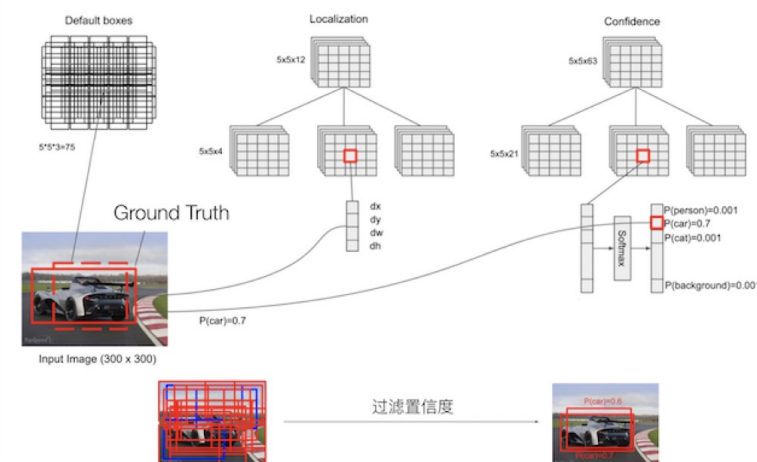


Region proposal network architecture.

cite: VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection (CVPR2018), <https://arxiv.org/abs/1711.06396>

□ Part3: Detection Head —检测头推理目标检测框与类别 (Single Shot Detector, SSD)

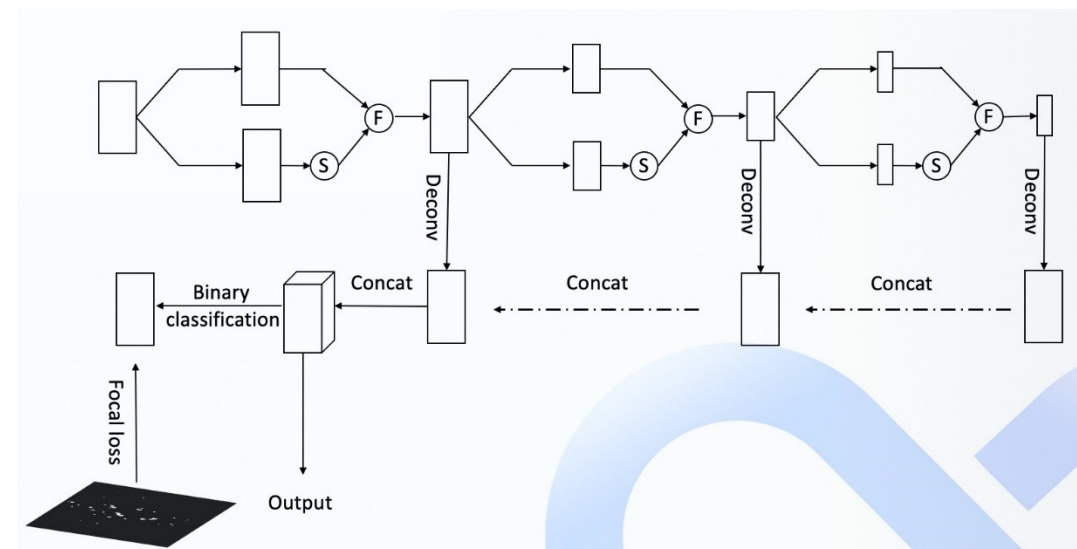
- 由SSD(Single Shot Detector)算法目标检测部分 (检测头) 构成:
 - 先验框(Default Box)生成:** 设计不同大小的先验框, 具体为以特征图上每个点为中心, 生成一系列同心的先验框, 这些先验框具有不同的长、宽, 具体的长宽由具体公式严格定义;
 - 置信度、偏移量与类别推理:** 通过预测, 生成目标类别分数 (不同属性的概率) 与边界框回归参数 (偏移量: dx, dy, dw, dl), 方向 (朝向: $d\theta$);
 - 损失计算:** 分类损失, 朝向损失—Softmax; 定位损失—SmoothL1
 - 检测框筛选:** 由设定置信度 (概率) 阈值, 非极大值抑制方法 (NMS), 选择该目标唯一检测框。
- 检测框的高度与中心点z轴坐标无法通过学习得到, 需单独回归。
- 最后, 可以得到检测结果类别与检测框 (x, y, z, w, l, h, θ), 其中 x, y, z 表示检测框中心点坐标; w, l, h 代表长宽高; θ 表示朝向。



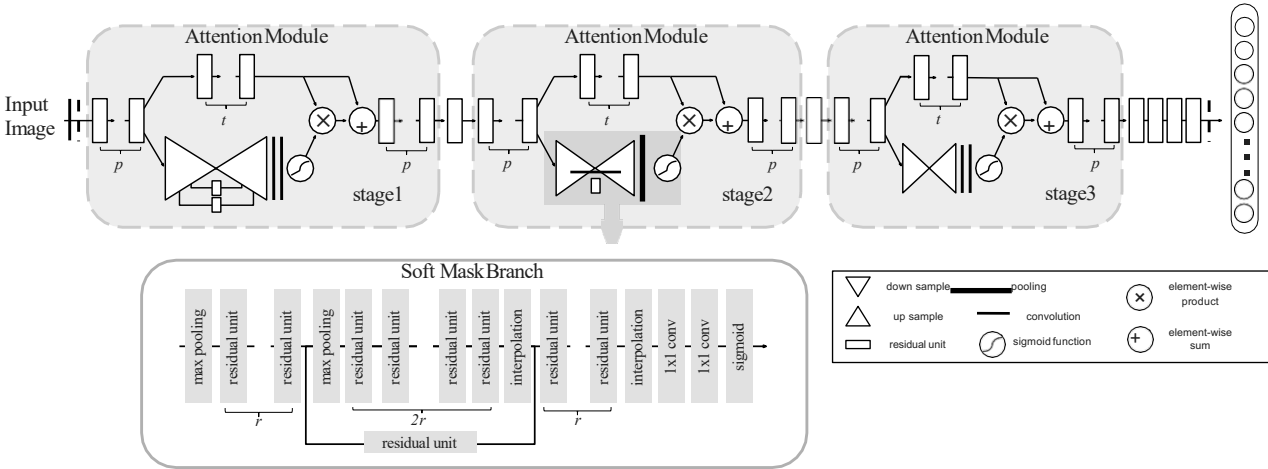
cite: SSD: Single Shot MultiBox Detector (ECCV2016), <https://arxiv.org/abs/1512.02325>

三、Mask Pillars

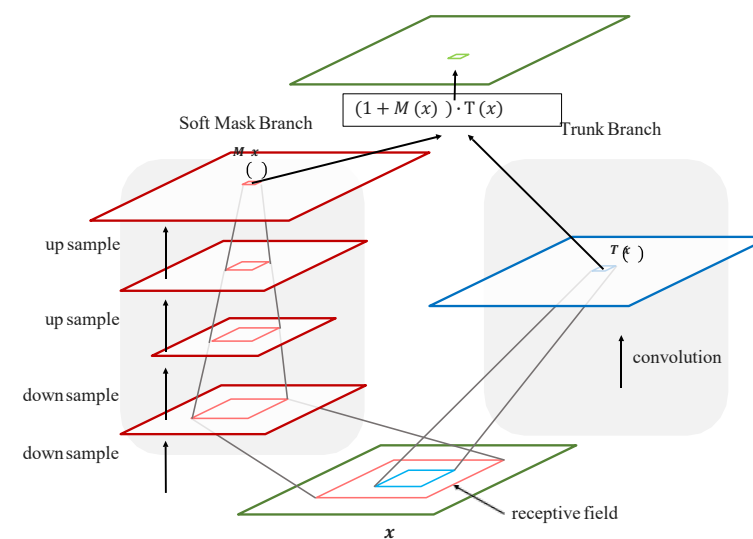
□ 算法背景与概述



三、Mask Pillars



三、Mask Pillars



N: 单个单元格内点数(Points Num per Pillar)

- 每个单元中最多包含N个表示点特征的向量，若点数量大于N，则采用随机采样的方法从中选取N个，反之，若点数量少于N，则用零填充的方法填充到N个。；

中国一汽 创领未来