

HEDNet: A Hierarchical Encoder-Decoder Network

for 3D Object Detection in Point Clouds

概览

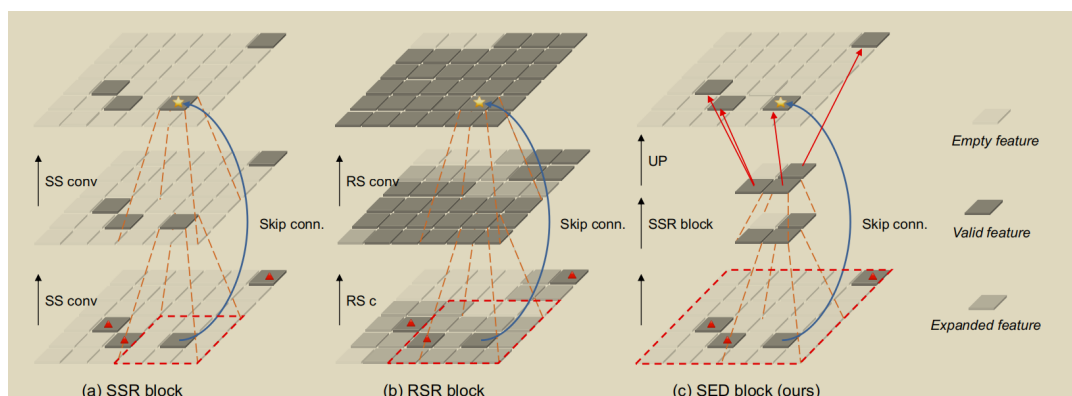
1. 现存的基于点云的3D Object Detection主要的挑战是3D场景内点云的稀疏分布（目标具体太远，点云数据太少了；目标与背景混淆，得到的目标点云数据过少；目标点云占整个场景的点云的比例过低）。
2. 现存的高性能方法通常采用具有小卷积核的3D稀疏卷积神经网络来提取特征，为了降低计算成本，这些方法采用子流型稀疏卷积，但这种方法使3D空间上不连续的特征之间的信息交换。
3. 本文提出了HEDnet，一种用于3D物体检测的分层编码器-解码器网络，它利用编码器-解码器模块来捕获空间特征之间的远程依赖性，特别是对大型和遥远的目标。在Waymo Open和nuScenes两个数据集上都实现了更高的检测精度。

介绍

1. 现存方法始终在精度和计算开销上权衡，目前基于体素的3D 物体检测是效果最高的算法。
2. 作者发现编码器-解码器架构能够高效的捕捉长距离的依赖关系。论文提出了一种稀疏编码器-解码器块（SED），一方面，编码器通过特征下采样提取多尺度特征，从而促进空间上不连续的特征之间的信息交换；另一方面，解码器融合多尺度的特征以恢复丢失的细节。再由DED块进一步改善得到了SED块，DED块将稀疏特征的提取扩展到了物体的中心（目前领先的3D 物体检测通常都更加关注物体中心）。利用SED和DED块，论文构建了HEDNET层次化编码器-解码器网络，用于点云输入的3D 物体检测，用于对大型和遥远目标的检测。
3. 在Waymo数据集上L2 的mAPH为75%，nuScenes数据集上的NDS为72%。超过了先前利用稀疏大卷积核的卷积神经网络以及transformer等提取主干的方法，且效率更高。

方法

稀疏卷积网络



SS卷积层和SSR模块

- (1) SS卷积层只作用于有效特征
- (2) SSR模块的输入输出特征图稀疏度相等
- (3) 影响了空间上无连接的特征的信息交流（如五角星得不到任何关于三角形的信息）

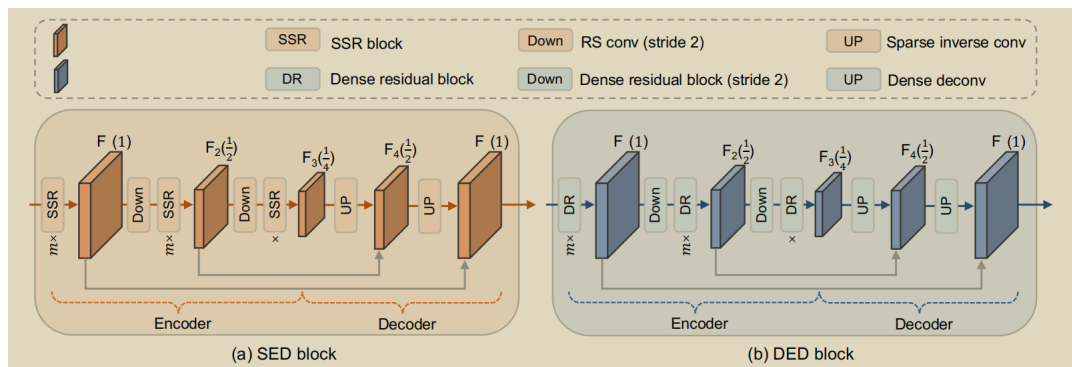
RS卷积层和RSR模块

- (1) RS卷积层同时作用于有效特征和扩展特征，扩展特征是有效特征附近的特征。
- (2) SSR模块输出特征图的稀疏度低于输入特征图。
- (3) 稀疏度极具降低，计算开销急速上升。

SED模块

- (1) 经过特征下采样，将底部特征图中的无连接的有效特征集成到中间特征图中的相邻有效特征。
 - (2) 利用一个SSR模块提取中间特征图的特征，建立相邻特征之间的关联交互信息。
 - (3) 最后进行特征上采样，将中间特征图转换为能够匹配输入特征图分辨率的输出特征图。
- (2) SED模块输出特征图的稀疏度等于于输入特征图，故叠加SED模块，计算开销不会过大。

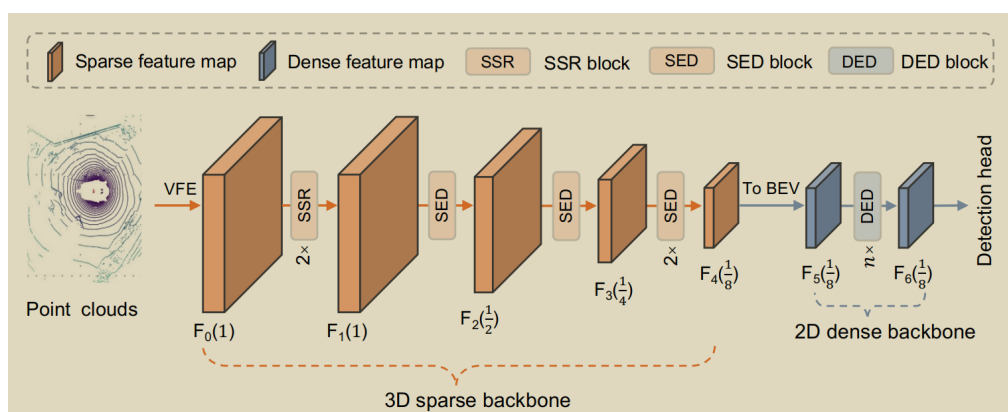
SED和DED



$$\begin{aligned} F_1 &= SSR^m(X) \\ F_2 &= SSR^m(Down_1(F_1)) \\ F_3 &= SSR^m(Down_2(F_2)) \\ F_4 &= UP_2(F_3) + F_2 \\ F_4 &= UP_1(F_4) + F_1 \end{aligned}$$

1. 编码器负责提取多尺度特征，解码器体用跳跃连接负责按顺序融合提取到的多尺度特征。
2. 按照上述顺序进行编码解码。利用RS卷积层进行下采样，稀疏逆卷积进行上采样。
3. 该架构实现了无连接的特征信息的交流。使得模型可以捕捉到远距离的依赖。
4. 高效率的3D物体检测通常依赖于对于物体中心的探测。利用纯的稀疏卷积层提取物体特征可能会使物体中心出现空洞，为了克服这一问题，论文构建了DED模块。与SED模块架构基本相同，只是替换了一些密集卷积块。如SSR替换为DR，下采样的RS替换为步幅为2的DR，上采样的稀疏逆卷积替换为密集的反卷积。

HEDNet



- (1) 原始点云数据通过VFE模型进行体素化，然后生成一个表示为F0的体素网络。
- (2) 一系列3D稀疏卷积特征提取，包括SSR和SED。
- (3) 3维稀疏特征被压缩成2维的BEV向量。
- (4) 利用DED模块进行2维特征提取，将稀疏特征向物体中心靠拢。
- (5) 最后输出特征提取结果，用于物体预测。

实验结果

Figure4 相较于DSVT-Voxel 和 LargeKernel3D两个模型，检测精度更高，检测速度更快。原因是另外两者采用了卷积核较大的卷积网络。

Figure5 汽车比行人精度要高，适用于检测更大的物体；距离越远，改善越好，HEDNET使点数较少的物体之间的联系建立起来。

Tabel3.a RSR相较于SSR检测精度更高但时延更大；论文提出的SED和显著优于RSR和SSR，且时延也仅仅是略低，加入DED模块效果更好。

Tabel3.b DED块数越多效率越高，4到5发生了过拟合。

Tabel3.c 在2维物体检测层面，SED+DED依然是较优的选择。

Tabel3.d 提取特征的尺度越多，最后检测的精度就越高，但时延也就越长。
需要权衡

HEDNET的核心观点就是捕捉距离上离得较远的物体之间的联系。