

**论文地址：**https://arxiv.org/abs/2007.13970

**代码地址：**https://github.com/Zengyi-Qin/Weakly-Supervised-3D-Object-Detection

**论文信息：**ACM MM 2020；麻省理工学院；微软研究院；

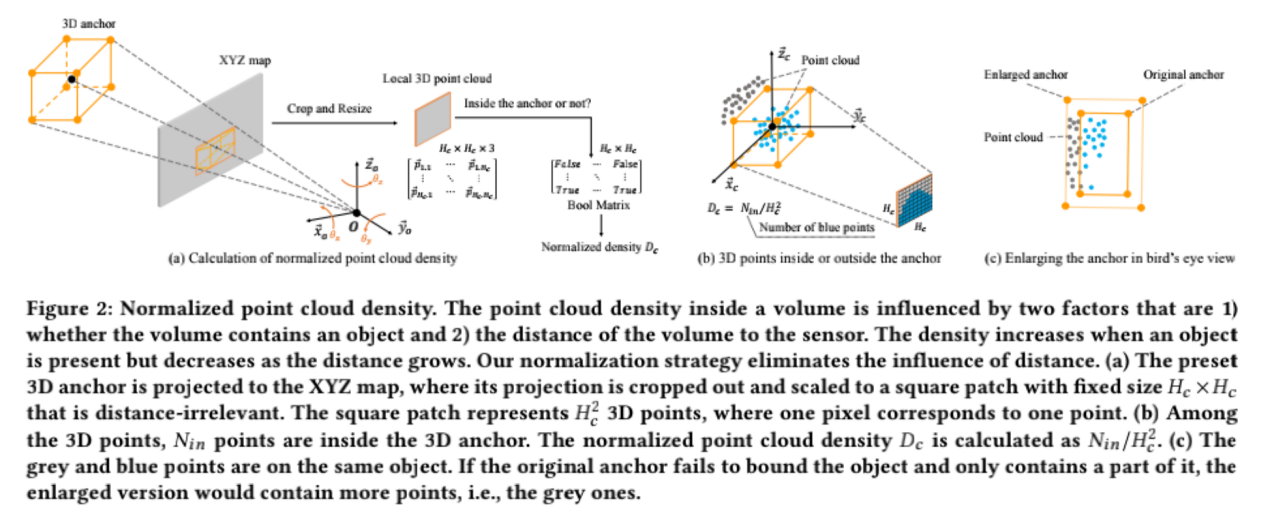
**动机：**使用弱监督学习规避3D检测任务中，标签信息较难获得且不准确的问题。整体方案分为两步：第一步根据归一化点云密度无监督产生多个候选框，第二步，利用跨模态的知识蒸馏策略，将在图像上学习到的检测知识用在3D检测问题上，生成最终的候选框。

**如何无监督提候选框：**

论文提出根据点云分布密度和是否存在物体是相关的思路，先初步找到目标物体位置。为了避免anchor中点云密度受深度影响，提出了一种归一化密度的方式。

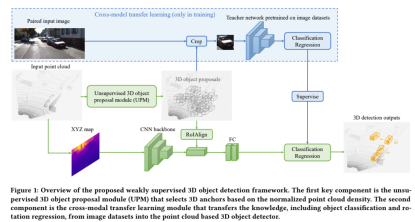
作者将点云投影至前视图，得到pixel-wise map，每个pixel对应三维坐标有三个特征值。在map上以间隔0.2m截取大小为32\*32的patch，每个patch是一个四棱台范围内的数据投影得来的，所以每个patch在深度上可以对应到多个3D box，同时随着深度增加，3D box的体积越大，密度=落在box内的点数量/32\*32，这样就规避了随着深度加深，密度降低的问题。这部分因为没用GPU并行化计算，所以耗时作者是没有算进整体网络的耗时的。

根据每个3D box算出来的归一化密度，和预先设置的阈值0.5做比较，低于阈值的认为不包含物体，进一步，将保留下的3D box向外扩张0.2m，避免候选框中只有一部分的物体。



**图像和点云之间的迁移学习：**

作者提了一种跨模态的知识蒸馏策略，点云做3D检测的网络作为student，在大型图像数据集上训练得到的模型作为teacher。



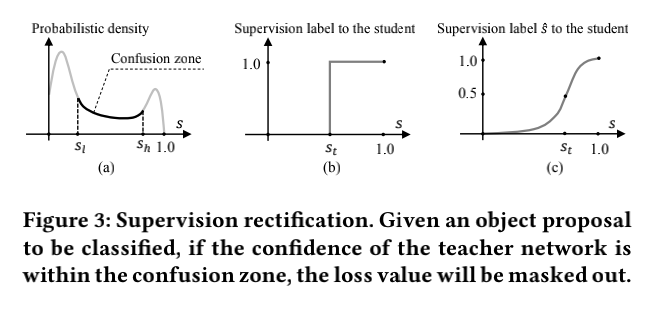
Teacher(蓝色部分)：模型解决两个任务，第一个是输入一个物体，输出该物体种类；第二个是朝向预测，将360°分为16个bin，将朝向问题转换为multi-bin classification的问题。

Student(绿色部分)：由backbone，RoIAlign和FC组成，输入backbone的是pixel-wise map，将由无监督学习产生的候选框对应到pixel-wise map和图像数据（作者选用kitti数据中提供了和点云数据对应的图像数据）中，裁剪出候选框覆盖的区域，将裁剪出的图像数据送入teacher网络中做分类预测，将student backbone提取的候选框特征利用RoIAlign得到固定维度，后送入FC做分类预测，通过交叉熵损失函数保证student给出的预测结果和teacher给出的结果是一致的。

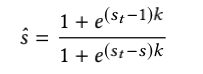
跨模态的知识蒸馏会有以下两点问题

1 如下图a，teacher给出某些样本的预测结果在sl=0.4和sh=0.6区间中，表示他对这些样本分类不是很有把握，这种时候，student是学习不到什么知识的。

2 对于teacher预测的结果，往往会设置一个阈值，将结果二分类，这样就会导致student学习没有区分度，比方说阈值设置为0.6时，teacher对正样本预测0.9，0.7，转为student监督标签时都是1，如下图b。



根据以上两点，做以下修改：

1 现利用公式将teacher输出结果曲线平滑为上图c

2 将student和teacher之间的损失函数改为，即不计算落在sl和sh区间中的样本。其中^是平滑化后的teacher输出，~是student输出。

截屏2020-08-01 上午12.46.43

**实验结果：**

实验对比挑选了三个SOTA的弱监督学习方法：PCL[Peng Tang, Xinggang Wang, Song Bai, Wei Shen, Xiang Bai, Wenyu Liu, and Alan Loddon Yuille. 2018. Pcl: Proposal cluster learning for weakly supervised object detection. *IEEE transactions on pattern analysis and machine intelligence* (2018).], OICR[Peng Tang, Xinggang Wang, Xiang Bai, and Wenyu Liu. 2017. Multiple Instance Detection Network with Online Instance Classifier Refinement. In *CVPR*. ]和MELM[Fang Wan, Pengxu Wei, Jianbin Jiao, Zhenjun Han, and Qixiang Ye. 2018. MinEntropy Latent Model for Weakly Supervised Object Detection. In *Proceedings of*

*the IEEE Conference on Computer Vision and Pattern Recognition*. 1297–1306. ],因为该论文提出的方法是不能预测3D检测框的，所以比较是在2维检测结果上进行的，结果显示在recall、AP都是有很大的提升。

**个人总结：**

提供了一种跨模态知识蒸馏的思路，但是限定需要是成对的数据，还是有些不容易的。论文中关于如何用点云密度无监督筛选候选框还是有借鉴意义的，可以用在一些后处理中。