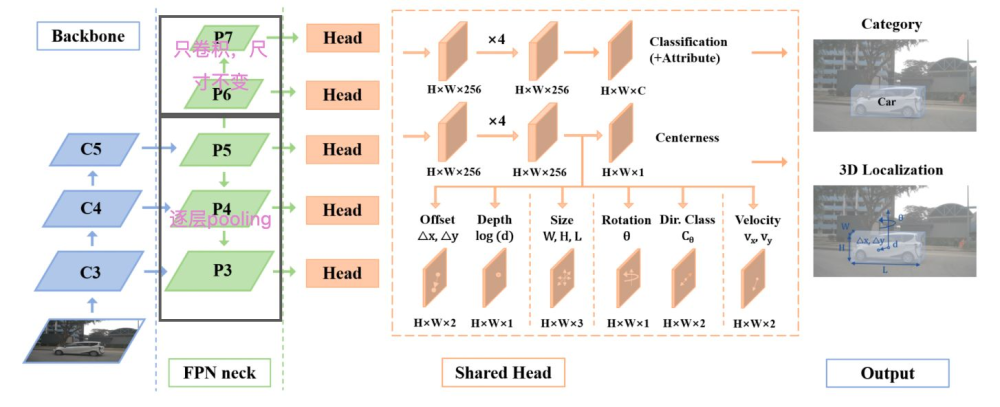
# FCOS3D

论文地址：

<https://arxiv.org/pdf/2104.10956.pdf>

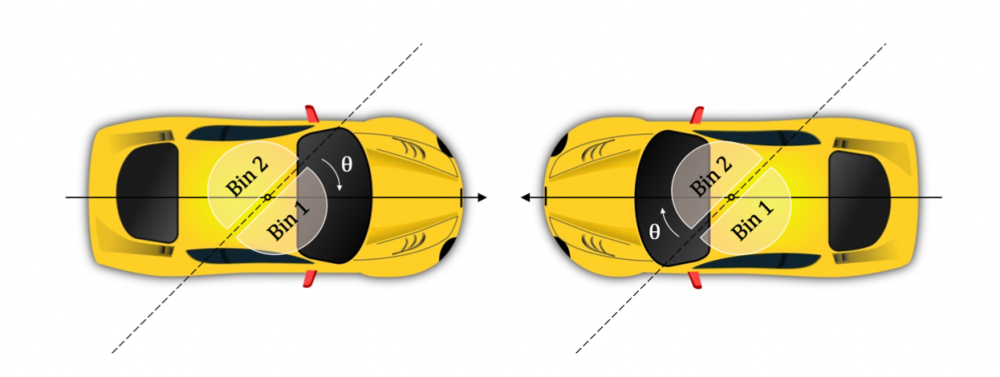
## 网络结构

* Backbone： 使用 resnet101，固定第一层的卷积参数从而减少显存消耗

* FPN： 生成特征层 P3-P7，每个特征层用于检测不同尺度的目标

* Detection Head：包含4个共享参数的卷积层和 small heads 用于不同的 targets 预测；回归分支需要较高的解耦程度，即每个子 targets 都设置一个 heads，所有回归的目标如上图所示。

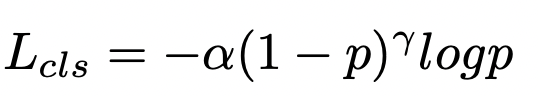
从网络结构对比来看，FCOS3D的backbone和neck部分沿用FCOS的结构，检测头head部分因regression targets不同有所差异:FCOS的分类分支有两个子分支，共享4个卷积层，分别预测目标类别和center-ness，回归分支预测feature map上的点到bounding box的left，top，right，bottom边界的距离  ；FCOS3D在检测头部分对分类分支中目标类别预测和center-ness预测进行了解耦，center-ness预测分支和回归分支共享4个卷积层，在回归分支中每个特征点预测3Dbox的在2D图上3D-center的偏置  ，3Dbox中心点深度信息  ，3Dbox的长宽高  ，3Dbox的航向角  以及x，y方向上的速度  ，其中在预测航向角时还添加了方向辨别任务(二分类);FCOS和FCOS3D检测头都是参数共享的。



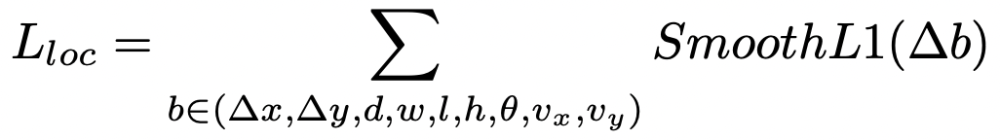
fcos3d 设计了一种新的旋转编码方案，即将360°角度回归解耦为方向的二分类和180°角度回归，在角度正确性上超过了 CenterNet

## loss 设计

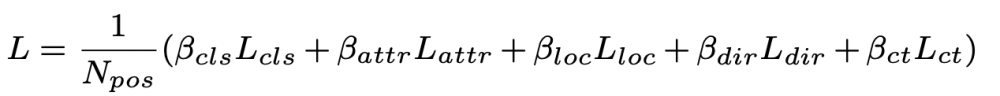
* + 分类使用 focal loss



* + 回归基本上使用 smoothL1 loss，方向分类和 center-ness 使用 binary cross entropy 损失



* + 总损失



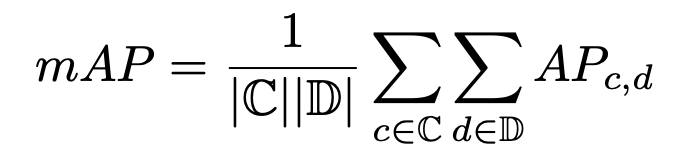
## 测试阶段

给定一张图片，经过该框架推理，得到检测框的class score，attibute score及center-ness结果。class score 和 center-ness 相乘作为预测的置信度，然后在 bird view 中进行 rotated NMS 得到最终的的结果。

## 评测指标

### mAP

* 使用在地平面上的 2D center 与 gt 的距离作为 threshold，避免使用 IoU 作为 threshold 对物体size 和朝向敏感的问题。最终通过匹配threshold来计算mAP，D代表4个阈值{0.5, 1, 2, 4}米，C表示所有的类别：



### True Positive metrics

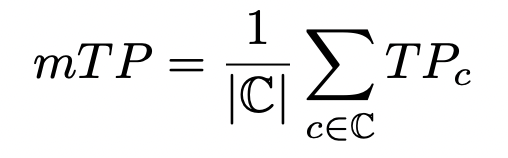
* Average Translation Error (ATE)： 2d 下的中心距离差距 （m）

* Average Scale Error (ASE)： 1-IoU，IoU为对齐 translation 和 orientation 后计算的值

* Average Orientation Error (AOE)：smallest yaw angle difference（radians）

* Average Velocity Error (AVE)： 速度差异的 L2-Norm (m/s)

* Average Attribute Error (AAE)：1−acc，其中 acc 指代属性分类准确度



### NuScenes Detection Score（NDS）

nuScenes官方提供的指标



## 总结

FCOS3D是具有代表性的基于单目相机进行深度估计的方法，其他刷榜的基于单目的3D检测方法大多是基于该方法进行的改进，如[PGD\_Camera](https://github.com/open-mmlab/mmdetection3d/tree/master/configs/pgd)，基于FCOS3D框架，在深度估计头上加入了概率和几何深度学习来进一步提升深度估计的准确性。