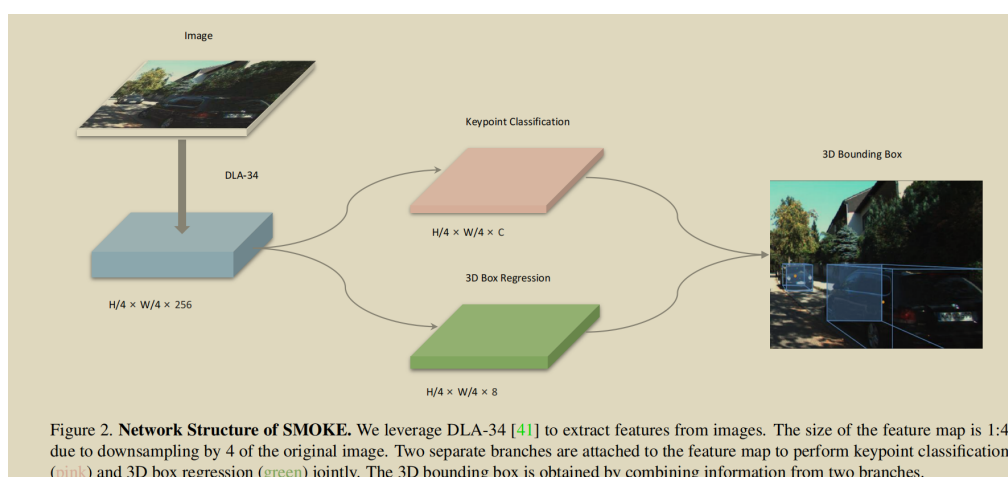


Paper2: Smoke : Single-Stage Monocular 3D Object Detection via Keypoint Estimation

1. 常见的利用单目摄像机进行三维物体检测的方法大致分为两步：（1）向神经网络中输入图像，产生对应物体的2D轮廓边界。（2）利用R-CNN神经网络，根据边界框来预测三维物体的方向和运行姿态。2D探测网络对于3D物体的探测是一种冗余，并且可能带来一定的噪声干扰因素，反而使得3D物体探测不准确。
2. 2D Object Detection已经被证明无论是在精度和速度方面都有着良好的性能，但3D Object Detection却困难的多，由于对物体所处深度无法准确判断，故对于物体的具体位置和具体姿态的确定是一个巨大挑战。目前三维较好的方法是使用激光雷达点云，从点云中学习和分析物体的具体特征。但实际问题是激光雷达较贵并且使用寿命很短。对于自动驾驶汽车来说并不符合其经济要求。而相机则成本较低，且易于安装，预期使用寿命较长，对于轻量级的3D Object Detection 已足够使用。单目摄像机获取的单个RGB图像很难获得全部的空间位置信息，双目摄像机可以解决，但实际情况下某些情况很难实现双目摄像机，本文尝试用单目摄像机解决这个困难的问题。
3. 本文提出了一种Smoke的方法通过关键点结合和三维变量回归来实现对物体三维边界的预测，这种方法不需要复杂的事先预训练和事后额外处理，训练收敛速度快，检测精度高，无论是在三维物体检测还是俯视图显示方面显著优于其它方法。
4. 主要贡献：（1）提出了一个单阶段的单目三维物体检测算法，可以精确高效的获得三维物体的立体空间信息。（之前的所有方法都需要多阶段完成，都需要先生成2维的边界框，进而预测三维的立体相关信息）（2）提出了一个多阶段分解的方法改善了三维参数的训练，参数收敛性较好，训练速度较快，训练结果精确度也较高。（3）所提到的方法在KITTI数据集上优于所有单目3D Object Detection算法。
5. 3D Object Detection 最新进展 **基于激光雷达的方法**：（1）将点云投影到二维鸟瞰图中，再在用于2D Object Detection的网络中进行分类和三维立体回归。（2）在体素网格中提取点云特征，通过2D/3D CNN神经网络计算出3D立体框。（3）激光雷达和图像融合的方法：根据点云和图像信息进行相关特征信息提取，然后输入到一个进行检测和分类联合神经网络。**基于立体图像的方法**：（1）伪激光雷达方法，将立体图像转化为一个假的点云，然后利用基于点云的三维物体检测网

络。（2）先利用立体的R-CNN网络创建2维平面的边界框。然后利用关键点预测的方法重建一个粗糙的3维立体边界框。最后再利用立体图像不断细化3维立体边界框，提高其准确度。基于单目图像的方法：

6. Figure2：按照1：4的比例提取特征得到特征图，进行两个分支的计算，关键点结合，三维变量回归，最终结合两者的运算结果得到3D物体框架。



7. Smoke方法 (h,w,l,x,y,z,seita) 物体高度，物体宽度，物体长度，相对于物体中心点的xyz坐标，seita为立方体框架偏移方向。H,W,C,B,分别为图片的高度，图片的宽度，物体的标签，物体的三维立体边框。与以往利用二维图像的边界框来预测三维边界框的技术相比，Smoke方法可以用一个简单的单个阶段来检测三维信息。三个部分：（1）backbone （2）3D detection （3）loss function 首先从主干中进行特征提取，然后进行两个分支的计算，关键点结合，三维变量回归的3D物体检测计算。最后利用多阶段分解的方式进行loss function的设计和计算，提高参数计算收敛速度。
8. 具体实现预处理：没有对数据集较复杂的预处理机制，只是消除那些在图像平面上的三维投影中心点超出图像范围的对象，这些对象的数量较少。数据增强：随机水平的翻转，随机的距离位移，只用于物体分类，不用于3D空间形象的获取。选择超参数、训练
9. 3D Object Detection 和Bird's Eye View测试任务最能检测3D Detection无论是在训练集还是测试集上的测试结果都很优异，2D Object Detection也远远高于3D Object Detection。探测速度明显提高，在中等和较难的训练集测试集上表现都有改善，在简单的训练集和测试集上也取得不错的结果。在大于40米的距离外，深度误差明显优于其它方法。