最近更新：2019年9月25日

Light-Head R-CNN: In Defense of Two-Stage Object Detector

2017.11.7

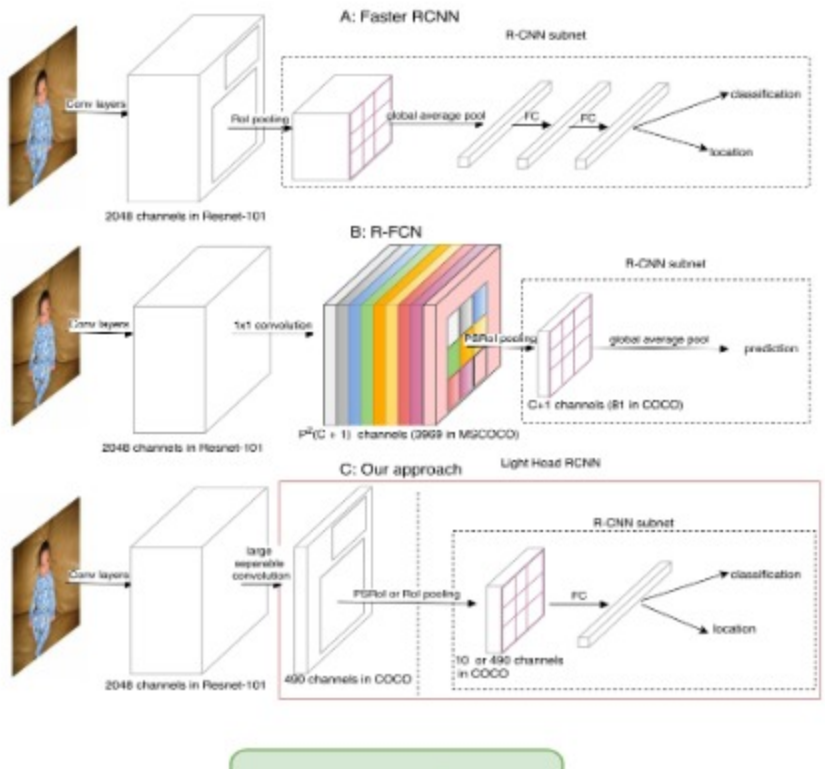
1. **Motivation**

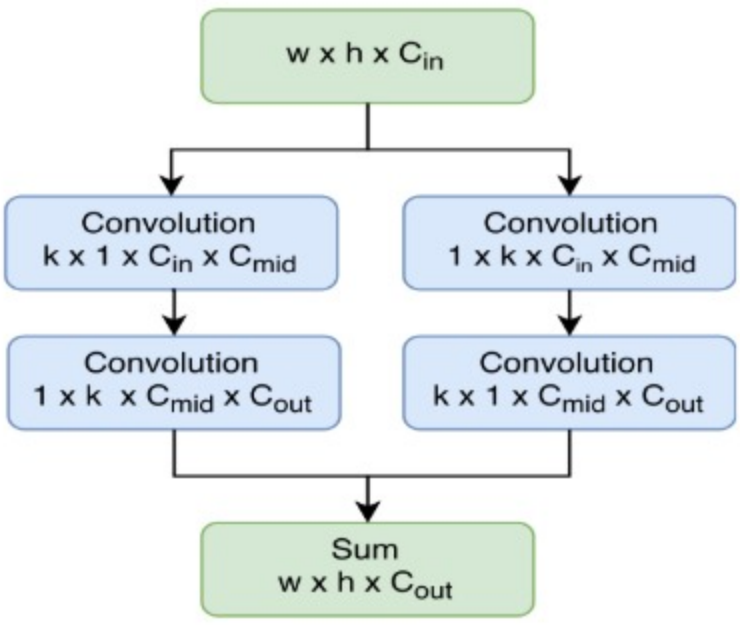
one-stage和two-stage分别侧重于速度和精度，那么能否进行权衡折衷？实验认为，two-stage检测器的时间消耗主要是heavy head即检测头部对anchors的分类和回归计算量过大导致，因此采取light head进行两者的权衡。

例如，对于Faster RCNN而言，time-consuming部分是对每个anchor都要进行大量的卷积和FC运算；而R-FCN则进行了改进，能够共享RoI的卷积特征，并且去掉了FC，代之以PSRoIPooling进行分类回归，但是这个score map通道数k^2(C+1)，仍然非常笨重。

2. **Light-Head RCNN**

网络结构图以及与另外两种的对比（可以看出实际上就是基于R-FCN进行的改动）





针对R-FCN中score map的channel数量较大的的问题，作者采用一个large separable convolution（见上右图，对于大的backbone如ResNet-101采用Cmid=256，小的如Xception选择Cmid=64，借鉴inception的1*n+n*1的卷积降低计算，这里作者用的k=15，大卷积确实能提升多一点的效果）生成thinner feature map，将原来P^2(C+1)的channel数量用10*（C+1）来代替，从7*7*81=3969降低到7*7*10=490，降低了后续Pooling和其他操作的计算开销。

得到的490维的feature map如果作为PSROI Pooling的输入之一（另一个输入是ROI，和R-FCN是一样的），就可以得到channel数为10的输出，再加上共享的2048维全连接层，然后分别接classification和location两个分支进行预测。

3. **Rethink**

（1）作者在R-FCN基础上直接压缩通道，将3969降为490发现在VOC上基本不掉点，COCO上掉零点几；即使进一步减少为5*5*10效果也是相差不多。说明两点：PSRoIPooling的冗余信息非常多；针对类别越多的任务这种heavy的feature才勉强有意义

（2）大卷积核的15换为7效果基本不差，直接用1*1才会有掉点，在1%以内

4. **Ablation Experiments**

- **Thin feature maps for RoI warping**

由于PSRoIPooling通道减少，计算大幅度减小，使得在该结构上引入FPN成为可能，并进一步提高性能（传统的R-FCN显存消耗太大，引入FPN后计算更加爆炸）。

可以看出，精度略有下降，而且集中在小目标上，大目标反而效果提升了

- **Large separable convolution**

引入inception类似的1*n+n*1的等效大卷积核，提升0.7个点

- **High Speed Larger**

ResNet-101改为类似Xception的小backbone；舍弃了计算消耗很大的膨胀卷积；RPN通道速改为256；大型可分离卷积核；PSRoIPooling采用Align。改动后的模型可达102fps，实际应用效果上佳。

流程图：