@魏晋

https://www.zhihu.com/question/35887527

首先膜拜RBG（Ross B. Girshick）大神，不仅学术牛，工程也牛，代码健壮，文档详细，clone下来就能跑。断断续续接触detection几个月，将自己所知做个大致梳理，业余级新手，理解不对的地方还请指正。

传统的detection主流方法是DPM(Deformable parts models)， 在VOC2007上能到43%的mAP，虽然DPM和CNN看起来差别很大，但RBG大神说“Deformable Part Models are Convolutional Neural Networks”（http://arxiv.org/abs/1409.5403）。

CNN流行之后，Szegedy做过将detection问题作为回归问题的尝试（Deep Neural Networks for Object Detection），但是效果差强人意，在VOC2007上mAP只有30.5%。既然回归方法效果不好，而CNN在分类问题上效果很好，那么为什么不把detection问题转化为分类问题呢？

RBG的RCNN使用region proposal（具体用的是Selective Search Koen van de Sande: Segmentation as Selective Search for Object Recognition）来得到有可能得到是object的若干（大概10^3量级）图像局部区域，然后把这些区域分别输入到CNN中，得到区域的feature，再在feature上加上分类器，判断feature对应的区域是属于具体某类object还是背景。当然，RBG还用了区域对应的feature做了针对boundingbox的回归，用来修正预测的boundingbox的位置。

RCNN在VOC2007上的mAP是58%左右。RCNN存在着重复计算的问题（proposal的region有几千个，多数都是互相重叠，重叠部分会被多

次重复提取feature），于是RBG借鉴Kaiming He的SPP-net的思路单枪匹马搞出了Fast-RCNN，跟RCNN最大区别就是Fast-RCNN将proposal的region映射到CNN的最后一层conv layer的feature map上，这样一张图片只需要提取一次feature，大大提高了速度，也由于流程的整合以及其他原因，在VOC2007上的mAP也提高到了68%。

探索是无止境的。Fast-RCNN的速度瓶颈在Region proposal上，于是RBG和Kaiming He一帮人将Region proposal也交给CNN来做，提出了Faster-RCNN。Fater-RCNN中的region proposal netwrok实质是一个Fast-RCNN，这个Fast-RCNN输入的region proposal的是固定的（把一张图片划分成n*n个区域，每个区域给出9个不同ratio和scale的proposal），输出的是对输入的固定proposal是属于背景还是前景的判断和对齐位置的修正（regression）。Region proposal network的输出再输入第二个Fast-RCNN做更精细的分类和Boundingbox的位置修正。

Fater-RCNN速度更快了，而且用VGG net作为feature extractor时在VOC2007上mAP能到73%。个人觉得制约RCNN框架内的方法精度提升的瓶颈是将dectection问题转化成了对图片局部区域的分类问题后，不能充分利用图片局部object在整个图片中的context信息。

可能RBG也意识到了这一点，所以他最新的一篇文章YOLO（http://arxiv.org/abs/1506.02640）又回到了regression的方法下，这个方法效果很好，在VOC2007上mAP能到63.4%，而且速度非常快，能达到对视频的实时处理（油管视频：https://www.youtube.com/channel/UC7ev3hNVkx4DzZ3LO19oebg），虽然不如Fast-RCNN，但是比传统的实时方法精度提升了太多，而且我觉得还有提升空间。