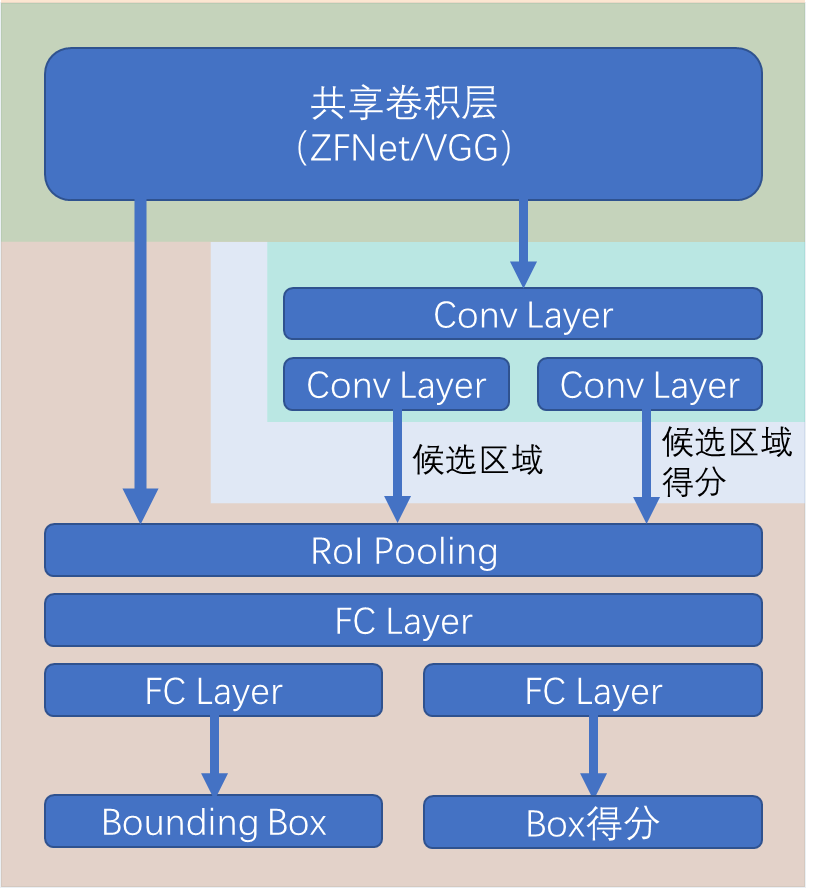
自从Ross Girshick等人通过R-CNN将CNN网络引入到目标识别任务中后，第一次成功将深度学习应用到目标检测上。后续的SPP-Net和Fast R-CNN模型分别从不同角度对R-CNN网络进行了加速。相比R-CNN，两者在精度与速度上均有提升。但对于生成候选区域这一阶段，两个模型还是使用的是传统的计算机视觉方法，而这也成为了让模型进一步提速的瓶颈。人们开始思考能否使用深度学习的方法来替代当前的算法来生成region proposal。2016年，SPP-Net的作者Kaiming He与R-CNN的作者Ross Girshick等人联合提出了Faster R-CNN模型。通过使用与检测网络共享参数的区域生成网络，整个模型的速度与精度都有了质的提升。

与R-CNN和Fast R-CNN网络一样，Faster R-CNN同样是一个two step的模型。其由两部分组成：RPN+Fast R-CNN。RPN负责生成候选区域，Fast R-CNN负责对生成的候选区域进行检测分类与bounding box回归。两个网络通过共享卷积层参数来减少模型参数。Faster R-CNN的网络结构如下（橙色为Fast R-CNN，蓝色为RPN，绿色为两者共有部分）。

1. 共享卷积层：

共享卷积层对

1. 区域生成网络（RPN）

RPN的任务是通过经卷积层处理得到的feature map来预测可能存在物体的候选区域的坐标以及这个候选区域内有物体的概率。坐标包括4个值，分别是候选区域的中心点的坐标以及候选区域的宽度与高度。但是RPN预测的这个4个坐标不是绝对坐标，而是相对一些先验框的相对坐标。

在卷积层输出的feature map的每个位置上，会有个先验框。这个框的大小与长宽比需要人为来设定，在文章中，作者使用的是大小为，长宽比为共9个先验框。RPN会基于这9个先验框来进行bounding box回归，共预测出9\*4个输出。具体回归的参数为：