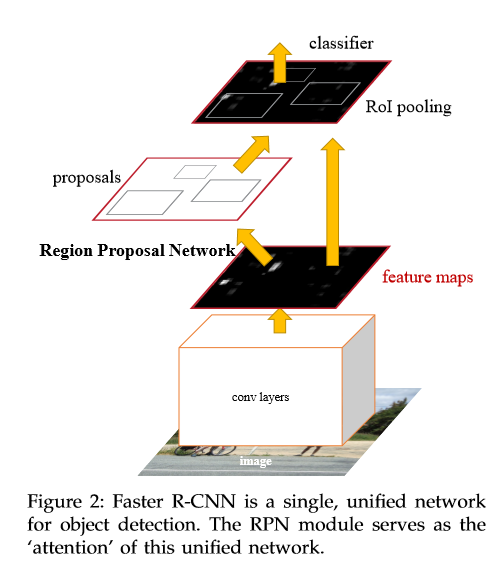
**Faster-rcnn解析**

**王忠 2017/4/23**

1. **网络结构**



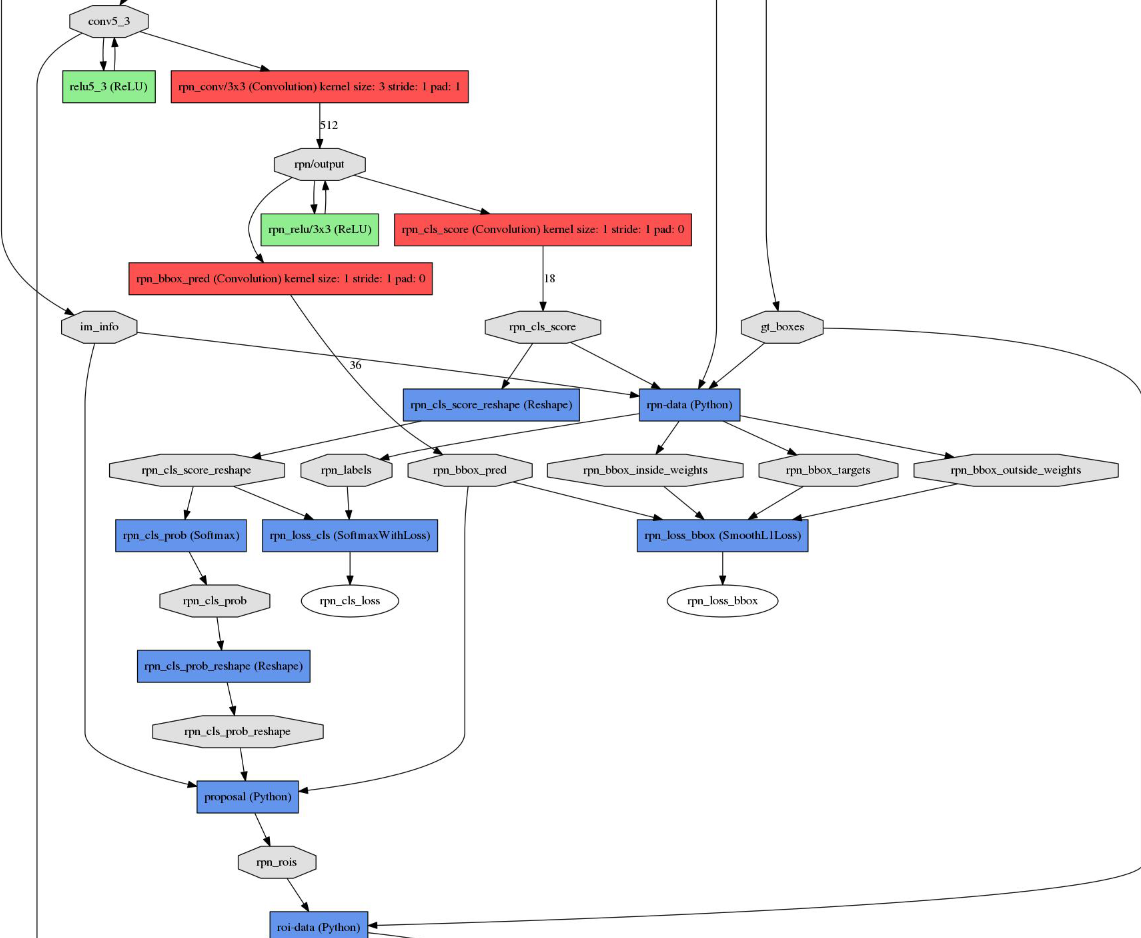
以上是原论文中po出的图，整个网络可以看作由以下三部分构成：

1. 基础卷积网络；
2. 区域生成网络（RPN）；
3. 全连接分类和回归精修网络；

介于第一部分就是一个简单的基础特征提取网络，此处不再赘述，下面我将对2，3部分进行详细的解读。

RPN网络

RPN结构图：



RPN结构说明：

1. 从基础网络提取的第五卷积层特征进入RPN后分为两个分支，其中一个分支进行针对feature map（上图conv-5-3共有512个feature-map）的每一个位置预测共（9\*4=36）个参数，其中9代表的是每一个位置预设的9种形状的anchor-box，4对应的是每一个anchor-box的预测值（该预测值表示的是预设anchor-box到ground-truth-box之间的变换参数），上图中指向rpn-bbox-pred层的箭头上面的数字36即是代表了上述的36个参数，所以rpn-bbox-pred层的feature-map数量是36，而每一张feature-map的形状（大小）实际上跟conv5-3一模一样的；
2. 另一分支预测该anchor-box所框定的区域属于前景和背景的概率（网上很对博客说的是，指代该点属于前景背景的概率，那样是不对的，不然怎么会有18个feature-map输出呢？否则2个就足够了），前景背景的真值给定是根据当前像素（anchor-box中心）是否在ground-truth-box内；
3. 上图RPN-data(python)运算框内所进行的操作是读取图像信息（原始宽高），groun-truth boxes的信息（bounding-box的位置，形状，类别）等，作好相应的转换，输入到下面的层当中。
4. 要注意的是RPN内部有两个loss层，一个是BBox的loss,该loss通过减小ground-truth-box与预测的anchor-box之间的差异来进行参数学习，从而使RPN网络中的权重能够学习到预测box的能力。实现细节是每一个位置的anchor-box与ground-truth里面的box进行比较，选择IOU最大的一个作为该anchor-box的真值，若没有，则将之class设为背景（概率值0，否则1），这样背景的anchor-box的损失函数中每个box乘以其class的概率后就不会对bbox的损失函数造成影响。另一个loss是class-loss,该处的loss是指代的前景背景并不是实际的框中物体类别，它的存在可以使得在最后生成roi时能快速过滤掉预测值是背景的box。也可实现bbox的预测函数不受影响，使得anchor-box能（专注于）正确的学习前景框的预测，正如前所述。所以，综合来讲，整个RPN的作用就是替代了以前的selective-search方法，因为网络内的运算都是可GPU加速的，所以一下子提升了ROI生成的速度。可以将RPN理解为一个预测前景背景，并将前景框定的一个网络，并进行单独的训练，实际上论文里面就有一个分阶段训练的训练策略，实际上就是这个原因。
5. 最后经过非极大值抑制，RPN层产生的输出是一系列的ROI-data，它通过ROI的相对映射关系，将conv5-3中的特征已经存入ROI-data中，以供后面的分类网使用。

另外两个loss层的说明：

也许你注意到了，最后还有两个loss层，这里的class-loss指代的不再是前景背景loss，而是真正的类别loss了，这个应该就很好理解了。而bbox-loss则是因为rpn提取的只是前景背景的预测，往往很粗糙，这里其实是通过ROI-pooling后加上两层全连接实现更精细的box修正（这里其实是我猜的）。

ROI-Pooing的作用是为了将不同大小的Roi映射（重采样）成统一的大小输入到全连接层去。

以上。