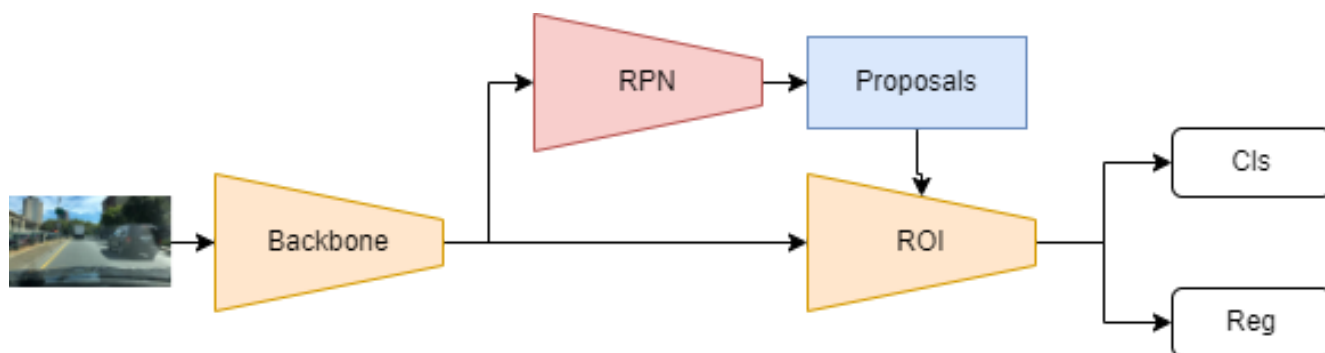


# Faster R-CNN



## 1.Method

Faster R-CNN的主要流程是：给定一张图片，先进入Backbone提取图像特征，基于输出的特征图利用RPN网络(全卷积网络)获取候选框，并筛选得分高的候选框输入ROI进行回归和分类。

### 1.1 Region Proposal Network(RPN)

Faster R-CNN的主要贡献是提出了RPN网络，将提取候选框这一步操作融入到了前向传播的过程中，实现了端到端的训练。RPN接受来自Backbone的特征图  $\mathcal{F}_b^{C \times H \times W}$ ，先经过一层3x3的滑动卷积层，然后分别经过两个1x1卷积分支，第一个分支用于分类，分为前景和背景，第二个分支用于回归坐标参数。由于输出的特征图有  $H \times W$  个锚点，根据给定的anchor个数  $\mathcal{N}$ ，每个锚点在分类分支上输出  $\mathcal{N} \times 2$  个通道特征图，在回归头上输出  $\mathcal{N} \times 4$  个通道特征图，2表示二分类，4表示中心点  $x, y, w, h$  的调整参数。

### 1.2 Forward

给定一批batch图像，批数为  $\mathcal{N}$ ，经过Backbone特征提取之后，输出  $\mathcal{N} \times \mathcal{C} \times \mathcal{H} \times \mathcal{W}$  的特征图，特征图分别进入两个网络，首先进入RPN提取候选框，假设预设的 *Anchor* 尺寸为[4, 8, 16]，比例为[0.5, 1.0, 2.0]，每个 *Anchor* 的长宽计算公式如下：

$$h = scale \times \sqrt{r}$$
$$w = scale \times \sqrt{\frac{1}{r}}$$

上述式子里scale表示尺寸，r表示比例，根据这两个公式计算预设 *Anchor* 的尺寸，每个锚点有9个 *Anchor*。然后将特征图送入RPN，先经过  $3 \times 3$  大小卷积，不改变特征图的通道数和尺寸，生成新的特征图，然后这个新的特征图进入两个分支，不分先后，我们先说分类分支，特征图首先经过  $1 \times 1$  卷积，特征图通道数变为18，尺寸不变，这里我的理解是特征图上有  $H \times W$  个锚点，每个锚点对应9个 *Anchor*，每个 *Anchor* 对应前景分数和背景分数两个概率值，所以输出18个通道，然后对特征图进行reshape，reshape到  $(\mathcal{N}, \mathcal{H} \times \mathcal{W} \times 9, 2)$ ，然后是回归头，和分类头类似，先经过  $1 \times 1$  卷积，不改变尺寸，但通道数变为36，这里同样，9个 *Anchor* 分别对应一个  $x, y, w, h$ ，所

以输出  $9 \times 4$  个通道，然后进行reshape，reshape到  $(\mathcal{N}, \mathcal{H} \times \mathcal{W} \times 9, 4)$ 。然后生成对应的步长特征图，将anchor加到步长特征图上，anchor在特征图上的所有坐标。然后将所有对应的anchors与调整参数结合获取真正的bbox，调整公式如下：

$$t_x = p_x * anchor_w + anchor_x$$

$$t_y = p_y * anchor_h + anchor_y$$

$$t_w = anchor_w * e^{p_w}$$

$$t_h = anchor_h * e^{p_h}$$

计算出bbox之后，开始进行过滤，过滤掉不符合条件的bbox，比如边界值超出限制范围(0-image\_size)的，长和宽的最大值超出限制值的，保留前  $\mathcal{T}$  个前景概率最高的proposal，然后进行nms过滤，过滤后的结果即为RPN得出的候选框，注意RPN输出的候选框的坐标值是对应原图的坐标值，而非特征图的。然后输入ROI进行判别分类。

### 1.3 笔记

☐ <https://zhuanlan.zhihu.com/p/617976226>

☐ <https://blog.csdn.net/wqwqqwqw1231/article/details/100823263>