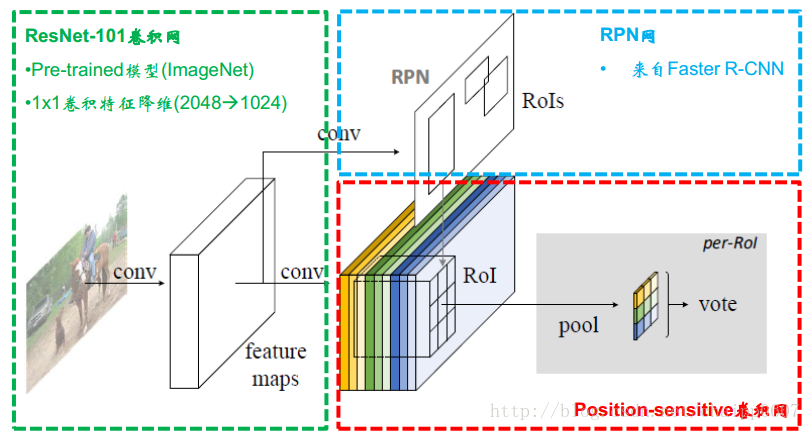
**[R-FCN: Object Detection via Region-based Fully Convolutional Networks]**

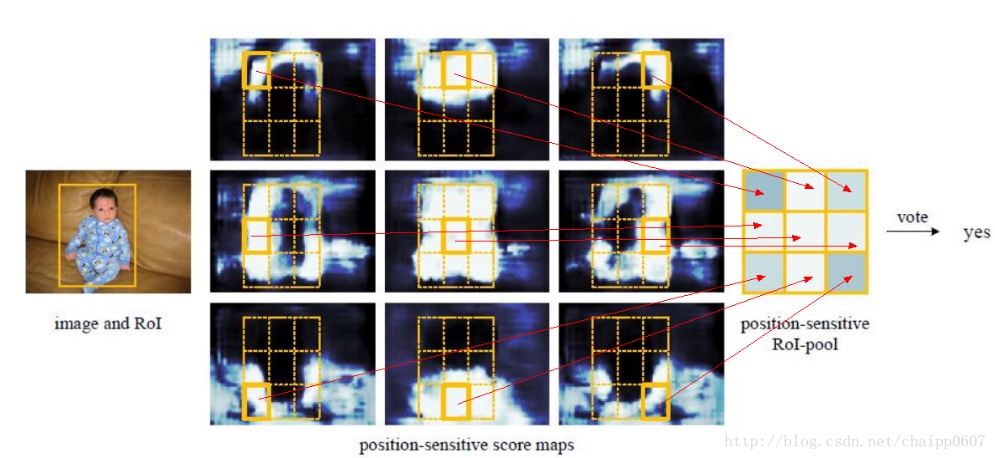
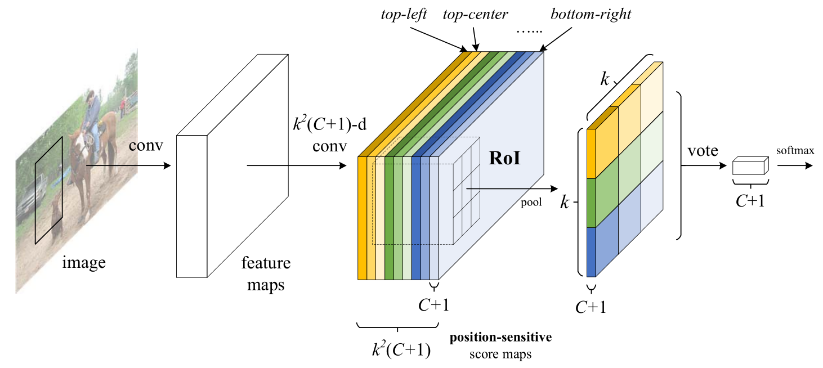
**NIPS 2016 | Jifeng Dai, Yi Li, Kaiming He, Jian Sun**

**主要思想：**首先，**分类**任务想要的是**变换不变性(Translation invariance)**，即目标在图像中的位置对分类结果不产生影响；相反地，**检测**任务需要**变换敏感性(Translation variance)**，需要对目标位置的变化特别敏感。其次，CNN越来越深，特征图的厚度越大，为了适应第一层的全连接的维度，会在最后一层特征图上做全尺寸的卷积，这层卷积的参数量是非常巨大的，所以新形态的CNN呈现全卷积化的趋势，比如ResNet，DenseNet，且最后一层一般采用全局池化而不是全尺寸卷积。但是**卷积的层数越深，变换不变性就越强，敏感性就会变弱**，因此Faster R-CNN的结构并不适合新形态的CNN来完成检测任务。R-FCN为了适应全卷积化的CNN结构，提出新概念**位置敏感卷积**和**位置敏感池化**。

**贡献/缺点：**之前的Faster R-CNN可以分为三个部分(1)CNN特征提取;(2)RPN(3)RoI Pooling + fc + regression等。当CNN提取模块采用较新的非常深的全卷积模型如ResNet，DenseNet，会失去位置敏感性(**卷积的层数太深**，感受野很大)，因此Faster R-CNN前面采用的CNN不会太深。R-FCN提出，CNN可以很深，只需要保持RPN不变，把第三部分的RoI Pooling+分类换成位置敏感卷积和位置敏感池化就可以了。R-FCN整个网络中**只有卷积，没有全连接**。

**论文细节：**结构如左上图，将Faster R-CNN红色部分换成下面要讲的位置敏感卷积和位置敏感池化。





**位置敏感卷积：**如右上图，feature map之后的红框就是位置敏感卷积。它的**卷积核个数是k^2(C+1)**，其中k是超参数，文中k=3，k^2是Grid的个数，与位置敏感池化中bin的数量一致。C是类别数，加1是因为背景类。这层卷积输出是**位置敏感分值图**，其宽高尺寸是与feature maps一致，分值图的通道数就是k^2(C+1)，即每一个类别都有k^2个通道。虽然这一层叫位置敏感卷积，并输出了位置敏感分值图，但它本质上是一个常规卷积，不同在于配合它一起使用的位置敏感池化。

**位置敏感池化：**如右上图蓝框内，是在分值图上的一种池化操作，它是RoI池化的变种，分值图的通道是K^2(C+1)，每一个类别有k^2个通道，即分值图上有k^2个C+1个通道的组合。假设k=3，就如图中所示，9个颜色都有C+1个通道，**RPN的区域建议扣在分值图上**后，位置敏感池化会把这个区域在每一个通道上平均分为K^2份，然后在每一个bin内做Max pooling，但是关键在**k\*k格子的位置是和通道对应的**，如上图中深黄色的通道数有C+1个，那么位置敏感池化操作时，只要深黄色通道为左上角的bin内的值，作为k\*k格子的左上角位置的值，所以k\*k格子的左上角位置也是深黄色的，得到一个k\*k的格子，厚度是C+1，也就是每一个通道代表一个类别。下面这张图可以更直观的说明位置敏感池化。红色的箭头就说明了这种对应关系。得到k\*k格子之后，再做一步全局平均池化，就得到了1\*1\*（C+1）的特征，刚好是C+1个，维度固定了，同样实现了将不同的输入整理成相同维度的输出，同时维度刚好与分类数相等。总之，在RPN输出的RoI内，分成k\*k个bin，第i个bin是敏感分值图中对应区域中第i个(C+1)个通道的池化结果，也就是实际**只池化了1/k\*k的通道数**。