To achieve this goal, we propose position-sensitive score maps to address a dilemma between translation-invariance in image classiﬁcation and translation-variance in object detection

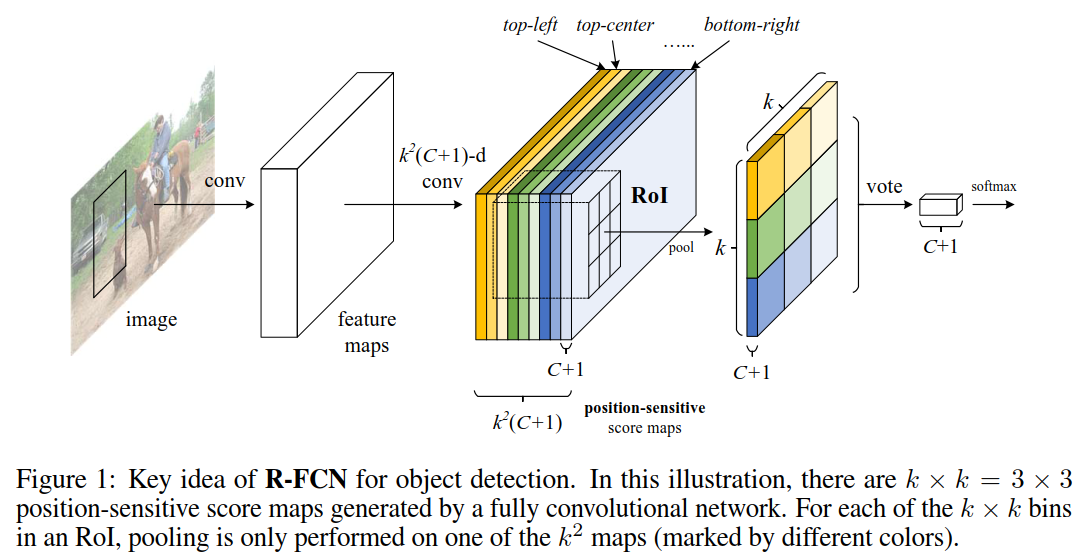
我们提出了，position-sensitive score maps，以阐述图像分类的平移不变性和目标检测的变性。

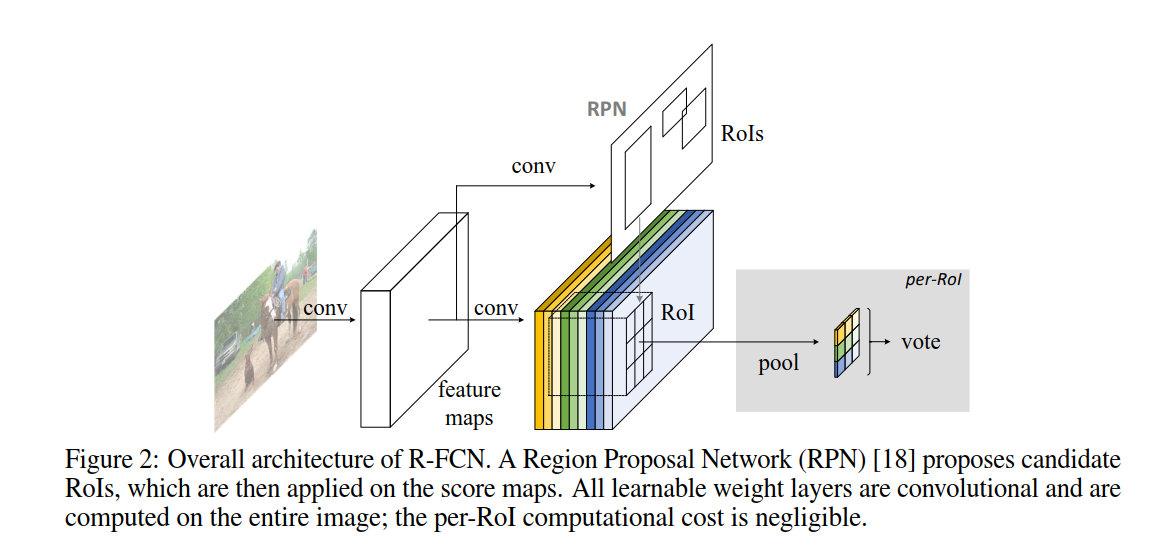
We present region-based, fully convolutional networks for accurate and efﬁcient object detection. In contrast to previous region-based detectors such as Fast/Faster R-CNN [6, 18] that apply a costly per-region subnetwork hundreds of times, our region-based detector is fully convolutional with almost all computation shared on the entire image. To achieve this goal, we propose position-sensitive score maps to address a dilemma between translation-invariance in image classiﬁcation and translation-variance in object detection. Our method can thus naturally adopt fully convolutional image classiﬁer backbones, such as the latest Residual Networks (ResNets) [9], for object detection. We show competitive results on the PASCAL VOC datasets (e.g., 83.6% mAP on the 2007 set) with the 101-layer ResNet.Meanwhile, our result is achieved at a test-time speed of 170ms per image, 2.5-20×faster than the Faster R-CNN counterpart.

改进点：1.与fast/faster R-CNN相比，基于区域检测是完全卷积的，几乎在整个图像上共享所有计算，2.使用了最新的ResNet（101层）残差网络，进行区域图像分类，在2007的PASCAL VOL获得83.6%的mAP，并且检测速度，打到170ms每张图，比faster R-CNN要快2.5-20

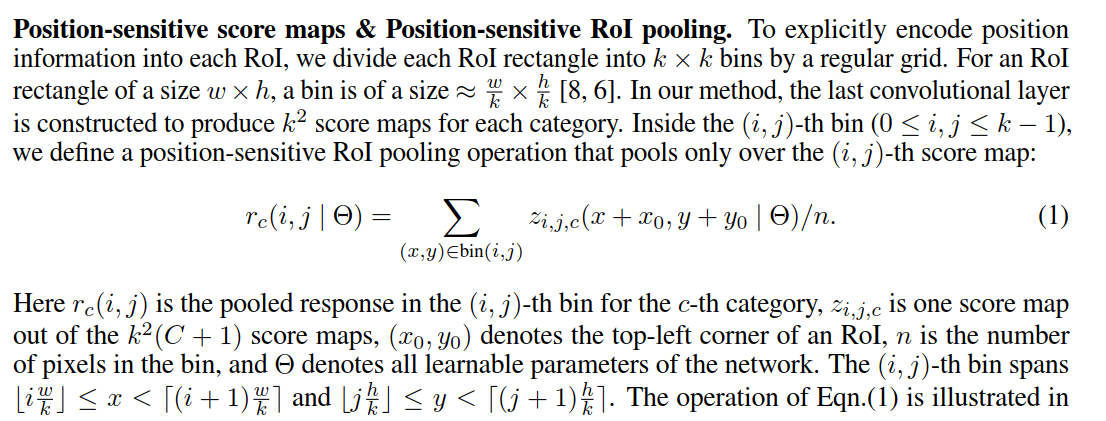
To remedy this issue, in the ResNet paper [9] the RoI pooling layer of the Faster R-CNN detector [18] is unnaturally inserted between two sets of convolutional layers this creates a deeper RoI-wise subnetwork that improves accuracy, at the cost of lower speed due to the unshared per-RoI computation

在ResNet的论文[9]中，Faster R-CNN检测器[18]把RoI池层，故意的插入到两组卷积层之间，这就创建了一个更深层的RoI-wise子网，它可以提高准确性，但代价是由于每个RoI的计算不共享，速度会降低。

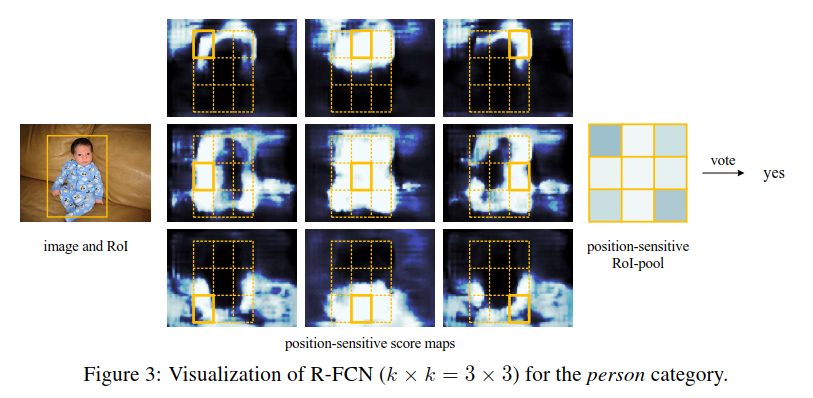




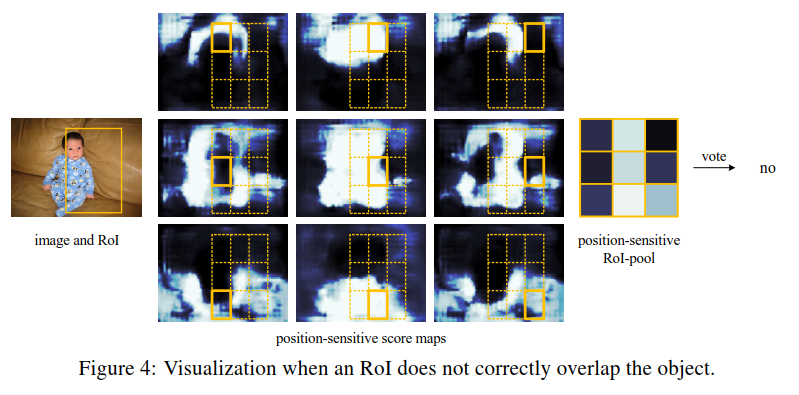
R-FCN ends with a position-sensitive RoI pooling layer. This layer aggregates the outputs of the last convolutional layer and generates scores for each RoI. Unlike [8, 6], our position-sensitive RoI layer conducts selective pooling, and each of the k × k bin aggregates responses from only one score map out of the bank of k × k score maps. With end-to-end training, this RoI layer shepherds the last convolutional layer to learn specialized position-sensitive score maps. Figure 1 illustrates this idea.



如何对Score Maps进行position-sensitive RoI-pool（见上）。



检测到了目标



没有检测到目标

In Figure 3 and 4 we visualize the position-sensitive score maps learned by R-FCN when k × k = 3 × 3. These specialized maps are expected to be strongly activated at a speciﬁc relative position of an object. For example, the “top-center-sensitive” score map exhibits high scores roughly near the top-center position of an object. If a candidate box precisely overlaps with a true object (Figure 3), most of the k2 bins in the RoI are strongly activated, and their voting leads to a high score. On the contrary, if a candidate box does not correctly overlaps with a true object (Figure 4), some of the k2 bins in the RoI are not activated, and the voting score is low

在图3和图4中，我们可视化了k = 3时R-FCN学习的位置敏感分数图。这些专门化的映射将在对象的特定相对位置被强烈激活。例如，顶部中心敏感的分数图显示的高分大致位于对象的顶部中心位置附近。如果一个候选框恰好与一个真实的对象重叠(图3)，RoI中的k2个箱子中的大多数都会被强烈激活，它们的投票结果会得到高分。相反，如果一个候选框没有正确地与一个真实的对象重叠(图4)，RoI中的一些k2箱子就不会被激活，并且投票分数很低。