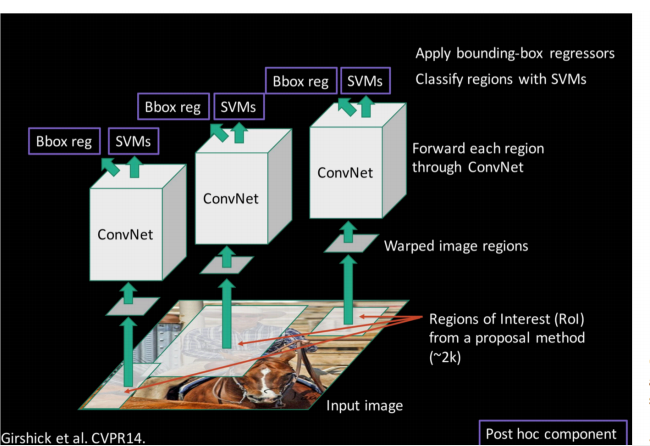
RCNN

　　1. 在图像中确定约1000-2000个候选框 (使用选择性搜索)

　　2. 每个候选框内图像块缩放至相同大小，并输入到CNN内进行特征提取

　　3. 对候选框中提取出的特征，使用分类器判别是否属于一个特定类

　　4. 对于属于某一特征的候选框，用回归器进一步调整其位置



Fast RCNN

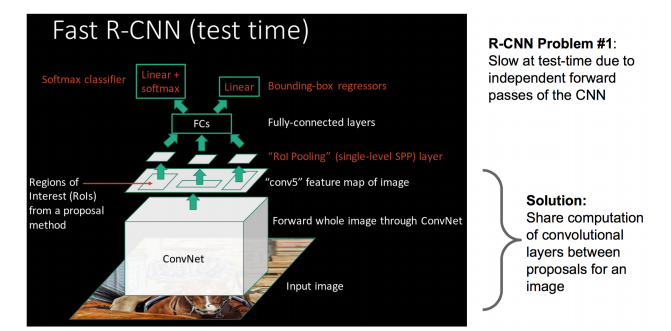
　　1. 在图像中确定约1000-2000个候选框 (使用选择性搜索)

　　2. 对整张图片输进CNN，得到feature map

　　3. 找到每个候选框在feature map上的映射patch，将此patch作为每个候选框的卷积特征输入到SPP layer和之后的层

　　4. 对候选框中提取出的特征，使用分类器判别是否属于一个特定类

　　5. 对于属于某一特征的候选框，用回归器进一步调整其位置



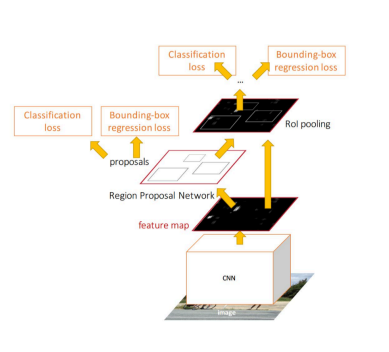
Faster RCNN

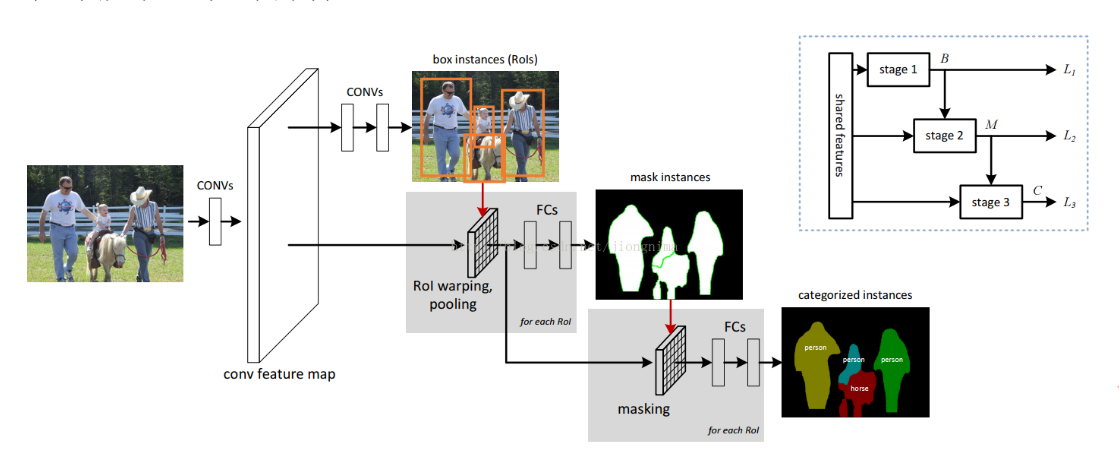
　　1. 对整张图片输进CNN，得到feature map

　　2. 卷积特征输入到RPN，得到候选框的特征信息

　　3. 对候选框中提取出的特征，使用分类器判别是否属于一个特定类

4. 对于属于某一特征的候选框，用回归器进一步调整其位置





问题1：由于全连接层的输入要求，所有ROI区域都要被转化为相同的尺度，这对于不同尺度的目标(尤其是面积比较大的目标)来说，细节信息损失巨大。

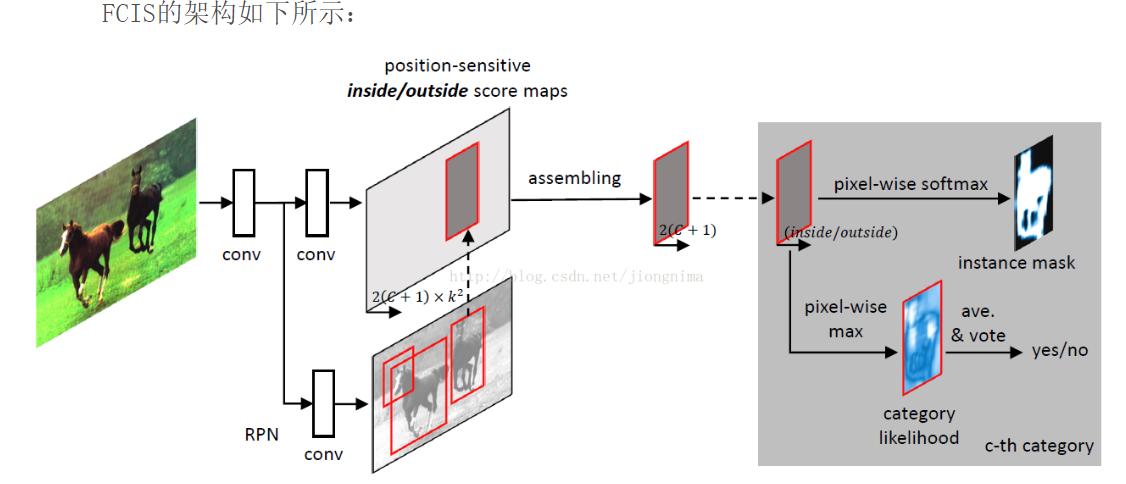
问题2：全连接层参数规模庞大，这种尾大不掉的架构很有可能发生过拟合。

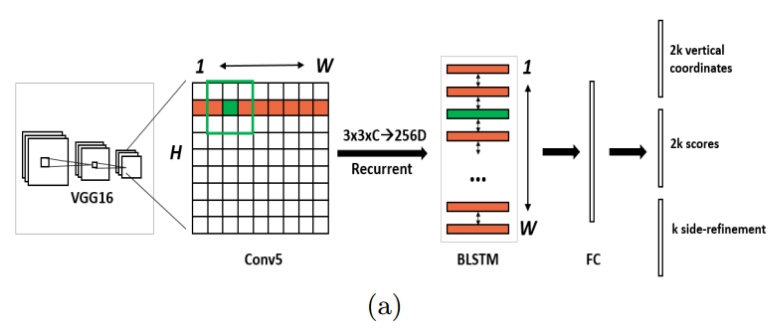
问题3：在ROI区域提出后，图像分割的子任务与图像分类的子任务之间没有共享参数。

首先针对问题1，****ROI-Pooling被取消了****，取而代之的是对ROI区域的****聚合****，实质就是复制粘贴。

然后针对问题2，****全连接层(FC layer)被取消了****，取而代之的是****分类器****(softmax)。

最后，针对问题3，图像分割与图像分类使用的是****相同的特征图****。





RPN很难准确预测单词的水平方向，因为单词中每个字符都是分离的

通过固定每个图片的水平位置，可以更准确的预测每个图片的竖直位置

CTPN增加了RNN，上下文信息对于减少文本错误非常有帮助

第一，用VGG16的前5个Conv stage（到conv5）得到feature map(W\*H\*C)

第二，在Conv5的feature map的每个位置上取3\*3\*C的窗口的特征，这些特征将用于预测该位置k个anchor（anchor的定义和Faster RCNN类似）对应的类别信息，位置信息。

第三，将每一行的所有窗口对应的3\*3\*C的特征（W\*3\*3\*C）输入到RNN（BLSTM）中，得到W\*256的输出

第四，将RNN的W\*256输入到512维的fc层

第五，fc层特征输入到三个分类或者回归层中。第二个2k scores 表示的是k个anchor的类别信息（是字符或不是字符）。第一个2k vertical coordinate和第三个k side-refinement是用来回归k个anchor的位置信息。2k vertical coordinate表示的是bounding box的高度和中心的y轴坐标（可以决定上下边界），k个side-refinement表示的bounding box的水平平移量。这边注意，只用了3个参数表示回归的bounding box，因为这里默认了每个anchor的width是16，且不再变化（VGG16的conv5的stride是16）。回归出来的box如Fig.1中那些红色的细长矩形，它们的宽度是一定的。

第六，用简单的文本线构造算法，把分类得到的文字的proposal（图Fig.1（b）中的细长的矩形）合并成文本线

****在inception结构中通过不同尺寸的卷积核设计达到检测不同大小和宽高比的文字，同时引入deformable卷积层操作和deformable PSROI pooling层提升任意方向文字的检测效果。****

