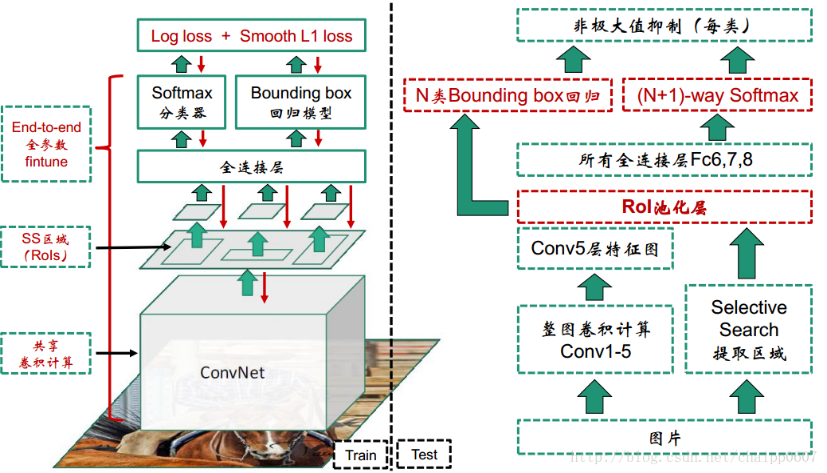
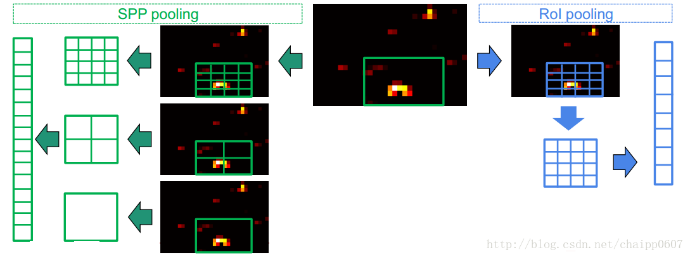
**[Fast R-CNN] ICCV 2015 Ross Girshick**

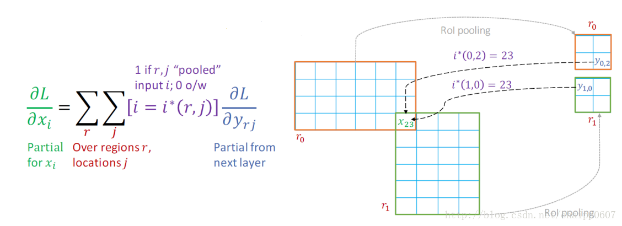
**主要思想：**。Fast RCNN测试过程是1)和SPP-Net类似，输入任意size图像，经过**VGG-16**卷积池化后得到feature map；2)ss生成2000个region映射到feature map上，对每个region采用**RoI池化**到HxW的size，经过fc层得到**fixed-length feature vector**； 3) fixed-length feature vector分别输入到两个fc层支路，得到**两个输出**，分别是K+1类上的**softmax分数**和K类上**bounding-box offset修正**；4)在每类上NMS得到最终输出。

**贡献/缺点：**Fast RCNN的**主要改进**1)把**SPP简化至RoI池化**，进一步缩减计算量；2)RCNN的训练是multi-stage，训练也是互相独立的，Fast RCNN中通过**multi-task loss**将网络的训练整成**one-stage**，实现了**end-to-end**的训练；3)所有的network layer在反向传播时都可以更新；4)mAP略有提升，速度大大加快。局限：Fast R-CNN依然没有脱离ss算法，达不到实时检测的要求的，好在ss算法在Faster R-CNN中被换成RPN（区域建议网络）。

**论文细节：**







**网络结构：**整体结构与SPP-Net类似，区域建议依然使用ss算法生成，并在卷积后的特征图上提取（充分共享卷积计算），初始模型从AlexNet换成了VGG16，区别在于：1) SVM分类器被换成了**SoftMax**；2) SPP-Net中的SPP换成了RoI pooling；3) 多任务损失函数的引入整合了分类网络的损失函数与bounding box回归模型的损失函数，使任务实现**one-stage训练**。

**感兴趣区域池化（RoI pooling layer）:**在conv层输出的feature map上，对每个region proposal，分割为HxW(H,W是超参)的网格，对每个网格区域每个通道都max pooling，然后拉伸得到fixed-length feature vector，输入到后面的全连接层，相当于只有**一个level的SPP**。看似计算量比SPP减不了多少，但若考虑到通道数，计算量的减少还是非常可观的，本质上是简化的SPP。

**多任务损失函数（Multi-task loss）:** Multi-task loss是Fast R-CNN**最重要的改进**，它将分类模型与bounding box模型的损失函数加到了一起，这样一来就不再需要分阶段的训练了，而是实现了end-to-end。分类损失是一个交叉熵，BB回归损失是GT和ss推荐框的L1距离。细节见原文和链接。

https://img-blog.csdn.net/20171108114257583?watermark/2/text/aHR0cDovL2Jsb2cuY3Nkbi5uZXQvY2hhaXBwMDYwNw==/font/5a6L5L2T/fontsize/400/fill/I0JBQkFCMA==/dissolve/70/gravity/SouthEasthttps://img-blog.csdn.net/20171107215629482?watermark/2/text/aHR0cDovL2Jsb2cuY3Nkbi5uZXQvY2hhaXBwMDYwNw==/font/5a6L5L2T/fontsize/400/fill/I0JBQkFCMA==/dissolve/70/gravity/SouthEasthttps://img-blog.csdn.net/20171107223203464?watermark/2/text/aHR0cDovL2Jsb2cuY3Nkbi5uZXQvY2hhaXBwMDYwNw==/font/5a6L5L2T/fontsize/400/fill/I0JBQkFCMA==/dissolve/70/gravity/SouthEast

**其他：梯度回传**。文章阐述RoI池化中**梯度回传**问题。当两个RoI没有重叠时，就是正常的max pooling梯度回传，当它们有交集时，重叠的像素区域采用多个区域的**偏导之和**，如右下图，细节见文下链接。

**分级抽样**。为了在训练时得到更好的效果，作者提出了一种**分级抽样**法，如果batch-size为128的话，那么这128个RoI由2张图片，各生产64个区域。细节见原文，这是很深层次的训练技巧了。

**全连接层加速：**图像分类中，卷积层计算时间比全连接层多，而在目标检测任务中，ss算法提取的框比较多,大半前向计算时间用于全连接层， 本文采用**SVD分解加速全连接层**计算。实现时，相当于把一个全连接层拆分为两个全连接层，第一个不含偏置，第二个含偏置；实验表明，SVD分解全连接层能使mAP只下降0.3%的情况下提升30%的速度，同时该方法也不必再执行额外的微调操作。

<https://blog.csdn.net/WoPawn/article/details/52463853> <https://blog.csdn.net/chaipp0607/article/details/78458883>

**对end-to-end的理解：1)端到端的训练**。RCNN是典型的非端到端的训练，其特征提取、SVM分类、BB回归都是分开训练最后组合在一起实现测试的。Fast RCNN对RCNN的改进之一就是采用Muti-task loss将特征提取、分类模型、BB回归的梯度反向传播放在一起更新，这样就不需要分阶段训练了，此为端到端的训练。**2)端到端模型。**系统中不再有独立的特征提取模型、分类模型、BB回归模型等模块，而是从输入到输出直接用一个神经网络承担所有功能，如YOLO。