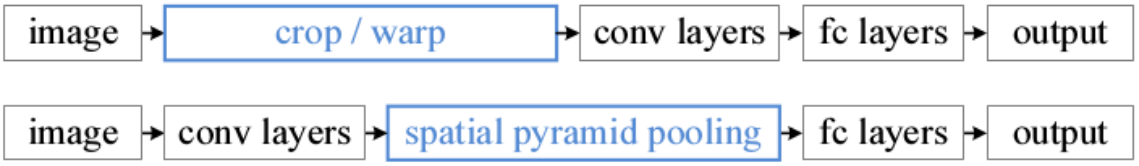
SPPnet论文阅读笔记

1.前言

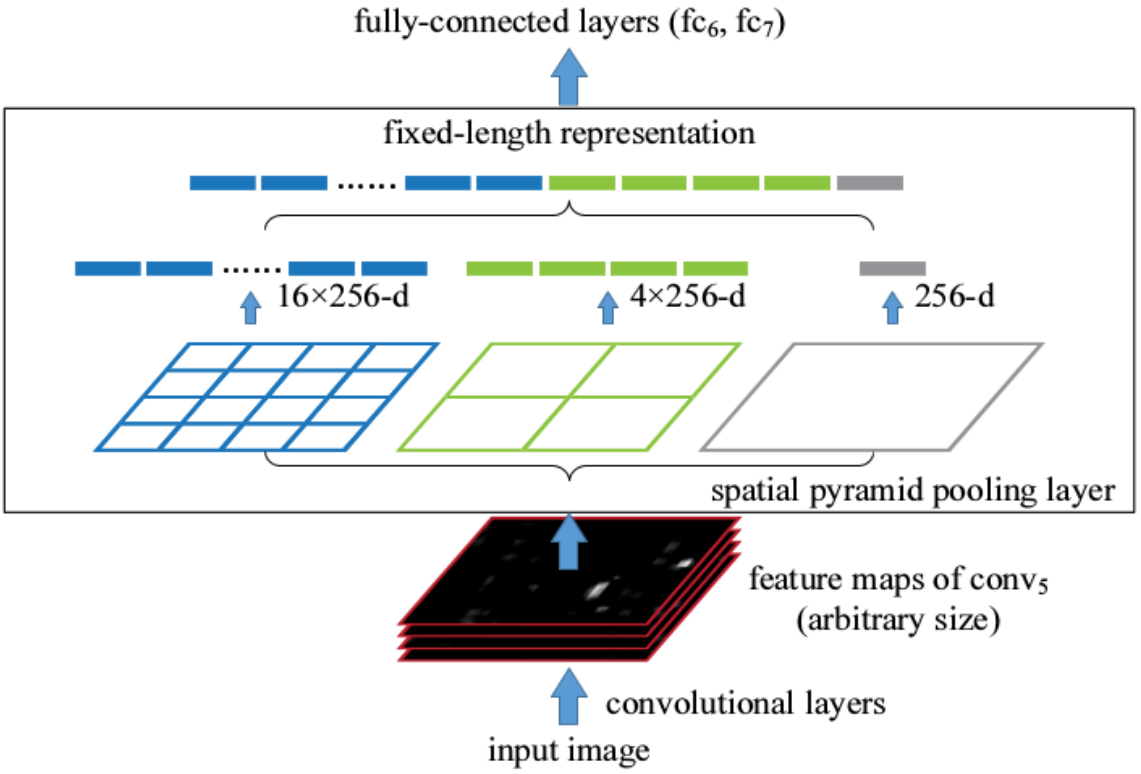
Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition这篇文章主要是提出SPPnet,应用不同的尺寸对特征图进行pooling.

2.SPPnet的改进

其主要对R-CNN有两大方面的改进:
a).采用共享卷积层特征.
RCNN是先将图片提取出2000个roi.然后对这2000个分别进行卷积计算,这样其实造成了极大地重复计算,也是RCNN计算速度慢的原因之一.而sppnet将整张图片进行卷积计算,接下的2000个roil共享这些特征图.可以极大地减小计算量和加上计算.而这其中就需要将ss选出的2000region映射到conv5的特征图(这个暂时没有看懂,大概的意思是将region的两个对角坐标映射到feature map上面,具体看参考资料3)
b).SPP
RCNN中为了后续的网络计算,采用crop/warp定了大小,使得输入的图片有了形变或者是不完整,对后续特征的提取有一定的影响.而SPPnet中提出其实conv层是不需要固定的size,只是FC层需要将固定的size.



所以SPPnet将卷积层的最后一层换成了空间金字塔池化（Spatial Pyramid Pooling),将图像送入五层conv,之后spp,最后两层卷积.



1.把一个图像区域分成了16个bin，也就是每个bin的大小就是(w/4,h/4)；2.再把这个图像区域划分了4个bin，每个bin的大小就是(w/2,h/2)；3.再把这个图像区域作为了一个bin，也就是bin的大小为(w,h)；4：对这些bin都采用maxpooling，得到21值，然后再送入全连接层。所以不管输入图像的大小是多少，给全连接层的都是21个值。这样解决了不同的size的特征图统一size的问题.

```
[pool3x3]
type=pool
pool=max
inputs=conv5
sizeX=5
stride=4

[pool2x2]
type=pool
pool=max
inputs=conv5
sizeX=7
stride=6

[pool1x1]
type=pool
pool=max
inputs=conv5
sizeX=13
stride=13

[fc6]
type=fc
outputs=4096
inputs=pool3x3,pool2x2,pool1x1
```

3.SPnet的训练与结果

训练还是RCNN一样,要经过多个阶段，首先要提取特征微调ConvNet，再用线性SVM处理proposal，计算得到的ConvNet特征，然后进行用bounding box回归2. 训练时间和空间开销大。要从每一张图像上提取大量proposal，还要从每个proposal中提取特征，并存到磁盘中。
训练模型主要采用了:ZF-5,Convnet*-5,Overfeat-5/7.

		top-1 error (%)			
		ZF-5	Convnet*-5	Overfeat-5	Overfeat-7
(a)	no SPP	35.99	34.93	34.13	32.01
(b)	SPP single-size trained	34.98 (1.01)	34.38 (0.55)	32.87 (1.26)	30.36 (1.65)
(c)	SPP multi-size trained	34.60 (1.39)	33.94 (0.99)	32.26 (1.87)	29.68 (2.33)

		top-5 error (%)			
		ZF-5	Convnet*-5	Overfeat-5	Overfeat-7
(a)	no SPP	14.76	13.92	13.52	11.97
(b)	SPP single-size trained	14.14 (0.62)	13.54 (0.38)	12.80 (0.72)	11.12 (0.85)
(c)	SPP multi-size trained	13.64 (1.12)	13.33 (0.59)	12.33 (1.19)	10.95 (1.02)

结果可以看到,使用SPPnet错误率下降了.

	SPP (1-sc) (ZF-5)	SPP (5-sc) (ZF-5)	R-CNN (Alex-5)
$pool_5$	43.0	44.9	44.2
fc_6	42.5	44.8	46.2
$ftfc_6$	52.3	53.7	53.1
$ftfc_7$	54.5	55.2	54.2
$ftfc_7$ bb	58.0	**59.2**	58.5
conv time (GPU)	0.053s	0.293s	8.96s
fc time (GPU)	0.089s	0.089s	0.07s
total time (GPU)	0.142s	0.382s	9.03s
speedup (*vs.* RCNN)	**64**×	**24**×	-

Table 9: Detection results (mAP) on Pascal VOC 2007. "ft" and "bb" denote fine-tuning and bounding box regression.

	SPP (1-sc) (ZF-5)	SPP (5-sc) (ZF-5)	R-CNN (ZF-5)
$ftfc_7$	54.5	55.2	55.1
$ftfc_7$ bb	58.0	**59.2**	**59.2**
conv time (GPU)	0.053s	0.293s	14.37s
fc time (GPU)	0.089s	0.089s	0.089s
total time (GPU)	0.142s	0.382s	14.46s
speedup (*vs.* RCNN)	**102**×	**38**×	-

Table 10: Detection results (mAP) on Pascal VOC 2007, **using the same pre-trained model** of SPP (ZF-5).

4.总结

SPP-net 采用SPP技术解决了图像固定size的问题,对图片分类的准确度有了一定程度的提高,采用共享卷积层也一定程度上提高了模型的训练和测试的运行时间.但是在Spatial Pyramid Poolin时,由于他分为了不同的BIn,反向传播梯度的时候,无法将梯度传递到conv层,使得训练时 conv层的参数不能进行更新.训练速度并不是特别的快.提取region proposal部分依然用的是selective search,分类器也是用了SVM，后处理也是用了cls-specific regression.

其实SPPnet的思想来源与SPW(参考2)

目录————-KAM_FANG

参考文献:
1.SPPnet原文:https://arxiv.org/abs/1406.4729
2.spm和BOM:http://blog.csdn.net/v_JULY_v/article/details/6555899
3.ROI如何映射到到feature map:https://zhuanlan.zhihu.com/p/24780433