Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition(SPP-net)

基础框架:

CNN 网络需要固定尺寸的图像输入 ,SPPNet 将任意大小的图像池化生成固定长度的图像表示，提升 R-CNN 检测的速度 24-102 倍。

固定图像尺寸输入的问题，截取的区域未涵盖整个目标或者缩放带来图像的扭曲。

事实上，CNN 的卷积层不需要固定尺寸的图像，全连接层是需要固定大小输入的，因此提出了 SPP 层放到卷积层的后面，改进后的网络如下图所示：

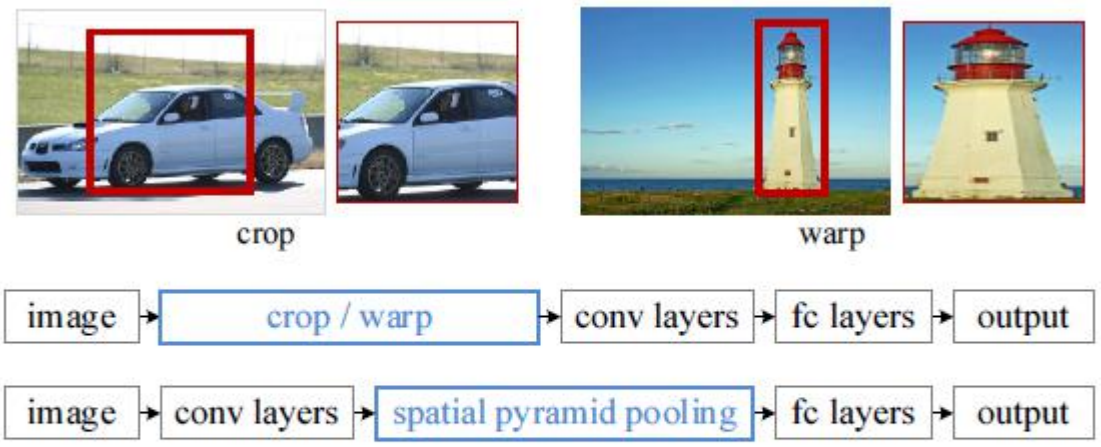

Figure 1: Top: cropping or warping to fit a fixed size. Middle: a conventional CNN. Bottom: our spatial pyramid pooling network structure.

SPP 的优点：

1）任意尺寸输入，固定大小输出，解决了 rcnn 由于图像扭曲问题造成的信息丢失。

2）层多

3）可对任意尺度提取的特征进行池化。

4）R-CNN 提取特征比较耗时，需要对每个 warp 的区域进行学习，而 SPPNet 只对图像进行一次卷积，之后使用 SPPNet 在特征图上提取特征。结合 ss 提取的 proposal，系统处理一幅图像需要 0.5s。

最后我对 sppnet 的最终感觉是：

通过对 feature map 进行相应尺度的 pooling，使得能 pooling 出 4×4, 2×2, 1×1 的 feature map，再将这些 feature map concat 成列向量与下一层全链接层相连。这样就消除了输入尺度不一致的影响。

但是关于如何映射 ROI 我还没搞太明白。