Note:Fully Convolutional Networks for Semantic Segmentation

Sinkoo

March 26, 2022

来源：CVPR2015

1 综述

使用 fully convolutional networks （FCN）,端到端，像素到像素训练的FCN在语义分割方向超过的当时的最先进技术。而且不引入预处理等额外技术，

2 主体

1,传统recognition nets（Alexnet等）使用固定大小输入产生 non-spatial输出。把这些网络最后的全连接层换成卷积层就可以使用各种大小的输入并产生空间化的输出（对于语义分割也就是产生像素级输出的前提）。（全连接层某种意义以上也是一种卷积核size等于input size 的卷积）
2，卷积层对输入进行了 down sampling 如果要进行 像素级的 dense prediction 需要恢复输出大小。文中提到的方法主要有：
(1)shift and stitch： Dense predictions can be obtained from coarse outputs by stitching together output from shifted versions of the input. 基于下采样因子f（下采样的程度） 右方向下平移产f^2个输出，再把输出stitch together(?)。
(2)Interpolation and upsampling： 上采样，文中采用了这种方法。
3,直接最后进行上采样丢失了细节信息： 可以使用前面pooling层的结果进行上采样，结合这些结果优化输出已获得更细节的结果。

3 结论

语义分割开山之作，先看了解一下，现在的实用性不得而知。用卷积替代全连接层以获得spatial的输出，再加上upsampling来做dense prediction的想法值得借鉴。

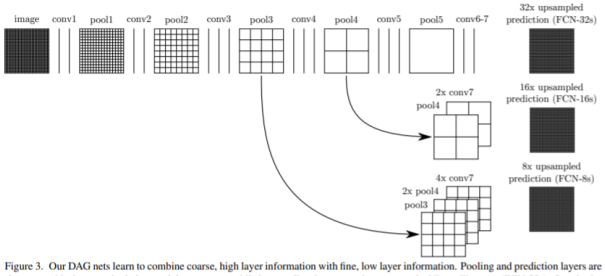

Figure 3. Our DAG nets learn to combine coarse, high layer information with fine, low layer information. Pooling and prediction layers are shown as grids that reveal relative spatial coarseness, while intermediate layers are shown as vertical lines. First row (FCN-32s): Our single-stream net, described in Section 4.1, upsamples stride 32 predictions back to pixels in a single step. Second row (FCN-16s): Combining predictions from both the final layer and the pool4 layer, at stride 16, lets our net predict finer details, while retaining high-level semantic information. Third row (FCN-8s): Additional predictions from pool3, at stride 8, provide further precision.