This is a note for the paper: Semantic Object Parsing With Graph LSTM(2016)

问题

将图像进行语义分割，每个像素输出一个标签，代表其属于的类别。可以通过全卷积FCN进行，输出向量$C * W * H$ (被称为confidence maps)，其中C为类别数量。比如对人的分割，可以分为头、左臂、右臂、左腿、右腿、躯干(torso)，当然还包括不属于这些类别的像素，$C = 7$ 。

方法

在1*1卷积生成confidence maps之前的张量代表了每个像素的特征向量。传统上，LSTM作用于像素或规整的块儿，对像素或块儿信息进行融合。这篇文章首先生成了过分割的不规则图像块儿(SLIC算法，可以生成不同大小的过分割块，具体不了解)，然后将这些块儿看作节点，相邻关系看作边构建图。块儿的特征为其中包含像素特征的平均。一个块儿对应一个LSTM单元。虽然作者生成为LSTM，但是前向传播只在邻域进行，而且只前向传播两次，也没用到Long Term的信息，可能得到的是门控的好处。前向传播为异步更新，也就是说LSTM单元是逐个更新的，更新的顺序取决于confidence maps的置信度，每个块儿的置信度为其中包含元素的置信度的平均。注意：这里作者先训练了传统的FCN，得到confidence maps，然后再训练graph LSTM，之前的FCN只是微调(学习率设置的小一些)。

具体的门控包括一个自适应门控：对某个邻居的门控值与这个邻居节点的特征有关。

$$g_{ij} = \sigma(W f_i + U h_j + b)$$

其中h_j为邻居节点j的状态向量,f_i为i节点的输入向量。

作者使用graph LSTM的目的是希望学习到相邻区域间的依赖关系，从而增强特征，使打标签更加鲁棒。

实验

作者做了特别充分的实验，值得学习。

作者选取了5个数据集进行测试。一方面对比了基于LSTM的一系列方法，一方面对比了其他种类的方法。还通过去掉实验设置，验证模块的好处，比如去掉residual connection，

不同的更新LSTM单元顺序的算法，去掉自适应门控，实验不同的过分割块数，根据confidence maps选取块更新顺序时选用不同的特征。