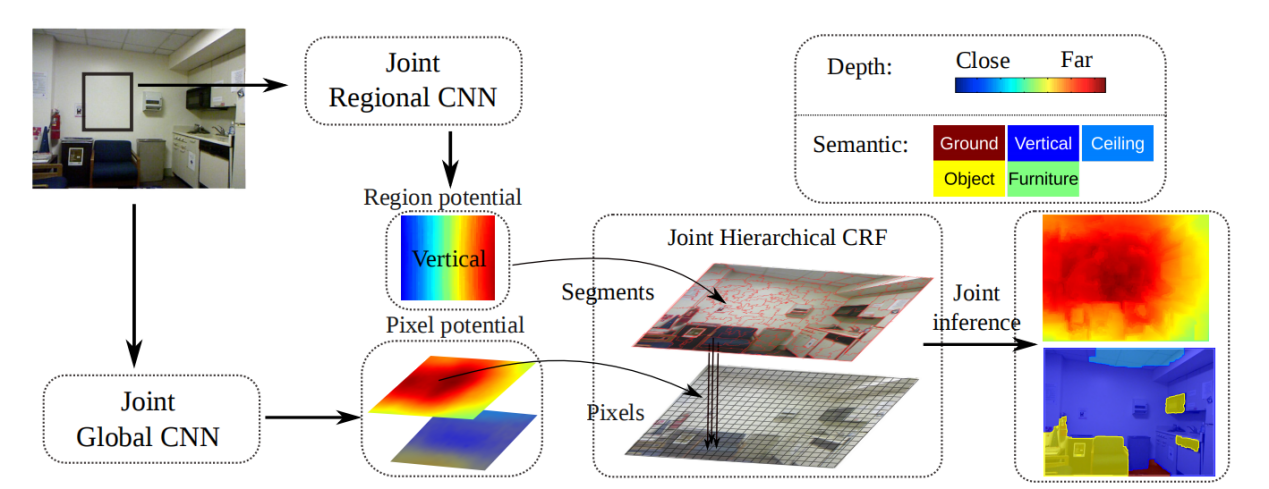
论文阅读笔记：

Towards Unified Depth and Semantic Prediction from a Single Image

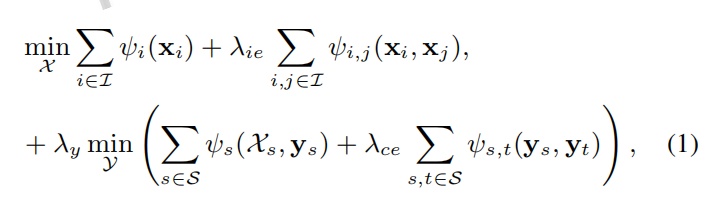
**内容**：提出了一个统一的联合深度和语义预测框架。给定图像，首先使用训练的卷积神经网络（CNN）来联合预测由像素深度值和语义标签组成的全局布局。为了进一步获得精细级别的细节，在全局布局的指导下将图像分解为局部片段以进行区域级深度和语义预测。利用逐像素全局预测和区域局部预测，在两层分层条件随机场（HCRF）中制定推理问题，以产生最终深度和语义映射。

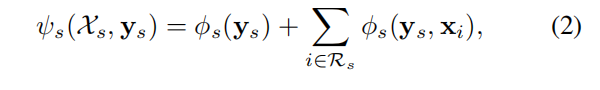


**理解**：在这篇论文中，我的理解就是主要分成三个部分。第一步主要是用CNN构造深度和语义预测的全局信息，之后在全局的基础上进行片段式细节优化（图像分解成块进行深度和语义预测）最后它是在两层分层条件随机场（HCRF）进行预测。

**细节方面的理解：**

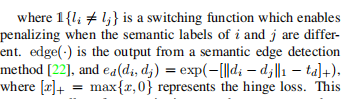
1. **关于HCRF**

由两层节点和边组成。在底层，节点是图像中的像素。是全局预测得到的；在上层，分解图像到段，并使用段来表示节点。HCRF存在三种edge的关系：一种是相邻的像素之间，一种是相邻的段之间，还有端和标签之间

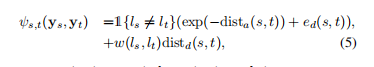
其中

分别是ψi（xi）是底层的像素级一元电位（全局CNN提供的，预测像素的深度值和语义标签的概率。φs（ys）由在区域的基于CNN的回归器生成），ψi,j（xi，xj）是像素之间的成对边缘电位，ψs,t（ys，yt）是上层中的段之间的边缘电位。（翻译叫做电位，网上也有说叫几元势）

捕获

像素边缘电位（只考虑相邻像素）。

对于每个片段s，我们计算片段内像素RGB值的均值和方差，以获得其局部外观特征fs。同时，对于每对相邻的片段s和t，我们基于由[生成的]语义边缘图计算它们之间的测地距离dist（s，t） 捕获

两个段之间边缘定位

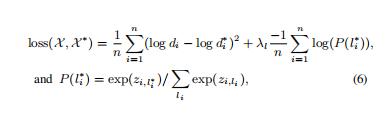
如果两个段之间的深度变化很大，则允许不同的语义标签。w（ls，lt）是从数据中学习的平滑权重矩阵，其中w（ls，lt）的较高值需要段s之间的更高深度平滑度，当他们的语义标签ls，lt 是一致的，反之亦然。（和上面的很相似）

1. **关于全局预测**

CNN结构参考论文的结构

除了最后一层中的深度节点之外，还引入了语义节点来预测语义标签。

在形式上，在网络train期间的损失功能由两部分组成



我们首先训练网络仅使用RGB-D训练数据预测深度值（RGB-D图不是已经有深度值吗。。。），然后使用带有可用语义标签的RGB-D数据，使用添加的语义节点微调网络。一旦训练，给定输入图像，网络将预测属于语义标签的每个像素的深度图和概率图。

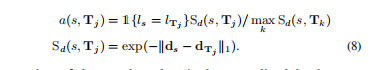
1. **关于分段预测**

在分段中心dc减去绝对深度值并将其重新缩放到范围[0,1]，将分段的绝对深度图变换为归一化的相对深度图。 给定归一化深度图，中心dc和尺度变化sc，我们可以精确地恢复段中每个像素的绝对深度值，其中dn是di =dn\*sc+dc，归一化深度中的相对深度值地图。

**局部和整体的联合**

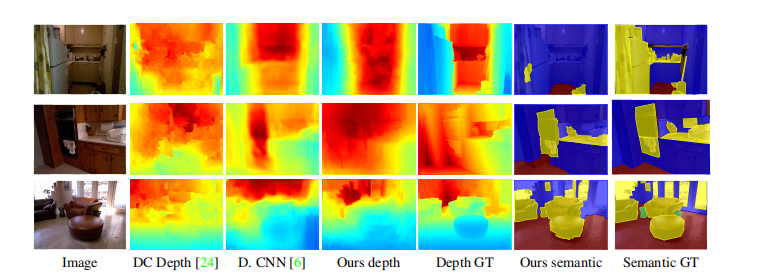
将局部深度估计问题表示为来自一组标准化模板的组合的预测。

首先根据它们的语义标签对片段进行聚类。对于具有相同语义标签的片段，使用L1 距离度量对其归一化深度图进行聚类以生成一组模板。给定一个片段s和一组模板Tj，该片段对模板的亲和力。

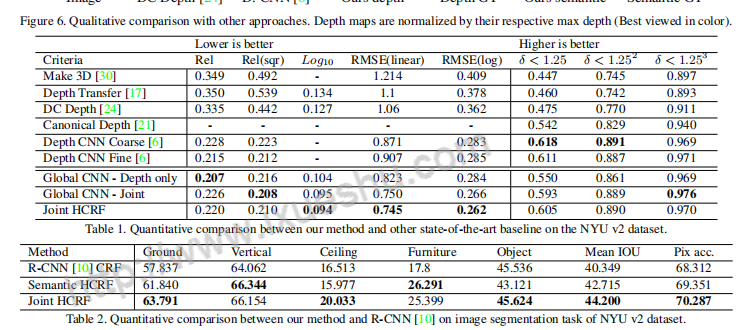
d表示归一化的深度图中的值，当语义标签不同时，片段和模板的亲和度为零。CNN作为局部训练模型把片段的边界框作为输入，损失函数表示为捕获

局部CNN通过微调全局CNN的部分来学习，回归之后选择具有最高亲和度的模板（不太懂这个亲和度这个东西，是不是可以这样理解就是局部深度和语义分割的一个关系啊）

**实验与结果**：



作者通过和其他情况的对比可以看出，他的深度图效果比D.CNN以及DC Depth都要好更加贴近深度图的情况，但是与此同时它本身深度图在物体的边缘部分并没有表现地很好。



根据数据的对比很明显可以看出该论文的方法在Rel RMSE都明显低于其他的方法，在分割方面他的GROUND 以及MEAN IOU都高于其他的方法，

**问题**：

文章在整体框架介绍之后，详细地介绍各个部分包括他们之间联合预测以及会遇到的问题，提出许多的辅助优化的方法，对于我自己的理解方面。一个在于那个亲和力的问题，第二个是HCRF（条件随机场）查阅许多资料，它主要是根据条件讨论的是概率问题，也就是像素的一个分布问题。