这篇论文论述的的是场景标记问题，也就是对于图片中的每个像素进行标记其所属的类别。

这里使用原始像素训练的多尺度卷积网络来提取密集的特征向量，来编码以每个像素为中心的多个尺寸的区域，并产生捕捉纹理，形状和上下文信息的强大表示。最后使用多种后处理方法形成最终的标签，这里作者提出了一种技术自动检索最佳解释场景的最佳组件的集合，这里作者的方法在Sift Flow数据集，巴塞罗那数据集中创造了准确性记录，并在斯坦福数据集中获得了记录，并且快速。