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。

图像理解是一系列实际应用的重点，面向图像的重要步骤是全景标签，也被称为场景解析，这一任务要对图像的每个像素进行标记其所属的类别。在一个完美的场景解析中，每个区域每个物体都被描绘和标签。场景理解的挑战是将检测，分割，多标签识别集合在了一起。

在场景解析下，有两个最重要的问题：如何产生良好的视觉信息内部表示，如何结合上下文的信息

这篇论文解决这一场景理解系统依赖于深度学习方法，主要是用一个卷积网络操作在一个大的输入窗口中去产生每个像素的定位。卷积网络是用一个原始图像像素作输入，用一个全标记的的图像去产生每个像素定位的类。卷积网络由多阶段构成。 每一阶段包含滤波模块，一个非线性模块，一个空间池化模块，通过端到端的学习自动学习多层次的特征表示。

不幸的是，标记每个像素在一个小的区域是很困难的。一个像素的类别通常依赖于相对短范围的信息，但是也依赖于一个长范围的信息。举例，证明一个灰度像素属于道路，人行道，车，建筑物或是天空要求一个宽的上下文框和展示足够的周围区域去做一个合理的决定。为此我们提出了一种多范围卷积神经网络，他能够考虑大的输入窗口，同时将无效的参数降到最小。

通常的场景理解方法需要首先产生分割假设使用图的方法。参考假设是按照设计好的特征进行编码。最后（一个条件随机场或是其他类型的图模型）被训练去产生标签为每个参考的分割，来确保标记在全局上是连续的。

本文所采用的论文所采用的特征就是使用一个大的上下文信息的窗口来标注像素减少了对于后处理方法的要求以确保标注是全局一致的。

更准确的说这里所提出的场景解析由图1所表示，它依赖于两个组件。

（1）多范围，全卷积的表示：我们的多范围，紧密的特征提取器产生一系列特征向量对于在每个像素为中心的多范围的区域，覆盖了一个大的背景信息。这个多范围的卷积网络包含一个单独网络的拷贝（都分享相同的权重）被应用到输入图像的拉普拉斯金字塔中。对于一个像素，网络对于信息进行集体编码在一个大的窗口中显示像素（此处描述的是18\*118个像素。这个卷积网络使用原始像素进行端到端的训练，因此可以缓解对于人工设计特征的需要。当进行合适的训练的时候，特征可以产生一个纹理的表示。当在FSL中一个多层次的表示看起来是自然的，在特征学习的背景下这种方法很少被使用。学习到的多尺度的表示通常是足够完整的可以检测出场景下的物体和区域。然而，它并没有精确地指出边界在这些区域中，需要一些后期处理去产生清晰的预测。

2基于图的分类

从图像中进行一个过度分割，并用于对特征描述符进行分组。其中的一些过度分割被考虑，通过三种技术产生最终的图像标签。

a.超像素的方法，图像被分割成互斥的组件，一个像素级的分类器被训练在卷积特征向量上，这里对每个组件进行了简单的投票机制，假设每个组件是一个单独的分类。这种方法是简单的并且搞笑的，但是形成的是次优的固定层次的分割，

b.基于超像素的条件随机场。一个条件随机场在一系列超像素上被定义。相对于过去的，简单的方法，这种后处理模型在场景的层次上将概率联合起来，避免了局部的畸变。这种方法在计算机视觉领域被广泛使用，并且我们展示了我们学习的多尺度特征表示本质上使得全局随机场的使用更加有用：大多数场景级关系似乎已经被它捕获。