# Enriching Visual Knowledge Bases via Object Discovery and Segmentation阅读笔记

## 简介

目标识别问题有两种流行方法。其一是自底向上，从颜色和对比度出发，将像素分组，然后识别对象。另一种是使用滑动窗口，然后解决一个二分类问题。分割方法更加直观更符合心理学原理，第二种方法经验上优于第一种。

第二种方法的优势在于数据，其只需要标记有正负的样本即可，而基于分隔的方法需要更加准确的分割边界，手动标记边界框是非常消耗劳动力的，这使得其难以利用到大数据带来的优势。

本文着眼于从海量根据搜索引擎文字输入检索到的图片中自动找到实例和它们的分割边界。

## 方法

本文的方法考虑先建立视觉子类，也就是视觉上同质（类内变化更小）。比如飞机，视觉子类可以是前视图商务机、侧视图客机、前视图战斗机（即子分类还需要考虑类内视觉的同质）。但是如何在还未学习到子类分隔先验知识的情况下做分隔呢？

本文先不直接分类成视觉子类，而是先将视觉数据分类成存在重叠和冗余的聚类（即一个实例可能存在于多个聚类中）。这个过程使用滑动窗口检测器完成，因为滑动窗口，所以检测的实例可以很好的中心对齐。据此，作者引入一个新的约束：同一位置的像素应该具有相同的前景背景标签，结合据类，就可以得到良好的图像分隔。

得到的聚类是比较严格的（低召回，高测准），这样其中一部分聚类就是冗余的，比如5辆属于车分类的摩托车组成了一个聚类。通过合并以消除这些冗余的聚类，就可以获得视觉子类了。汇集之前的图像分割结果，就可以创建出至上而下的图像分隔先验知识了。

#### 寻找对齐和同质的分类

作者首先将每张图片都作为聚类种子，然后检测图片中类似的patch。作者基于Color-HOG特征进行训练ELDA检测器，并在所有图像中检测类似的部分并选择分数最高的前k个。因为CHOG关注于形状和轮廓，所以检测结果是良好对齐的，这些会作为随后联合分隔和探寻子分类的基础。因为最开始每个图片都是一个簇，那么对于一些噪声图像（不包含任何对象），最后会引入一些噪声簇。

#### 生成分隔种子

获取良好对其且聚类的图像后，利用同一聚类中相同位置的像素往往具有相同的前景背景标签构造约束，运用graph-cut算法进行前景分割。因为目标所处的环境可能千差万别，对所有图片同时分割的好处是一些背景纯净的图片的分割可以帮助分割那些难以分割的图片（背景复杂、被覆盖、分辨率低）。

数学角度来说，考虑这个问题为经典的graph cut问题，每个像素需要被标记为前景或者背景。我们具有某一聚类的张图片块。像素表示在图像中的像素位置，像素特征为，标记结果。一个标记代表了一个分割。

现在构造能量方程，其具有四个项

表示单点势能，由每张图片的外观模型决定。

表示单点势能，但由当前聚类的各图像所共享的外观模型决定。

表示对势项，这个项用来惩罚那些具有不同的标签但是却没有intervening contour（两像素间沿着一条直线存在很大的梯度）。

约束生成的分割掩码是对齐且连续的。

这其中，是每一个图像的外观模型，对于每一个，都由两个高斯混合模型（前景和背景）构成，高斯混合模型的每一个组成都是RGB色域上的全方差高斯，是所有图像共享的外观模型。

为了让生成的分割掩码在聚类内对齐且连续，作者构造了一个先验项，是聚类内分割掩码的均值，这表示了像素基于先验知识属于前景还是背景的概率。

然而在完成分割之前，我们是不知道和外观模型的。因此在做全局graph-cut优化的时候，每迭代一张图片就要重新估计模型参数和先验知识（通过求均值的方式），直到算法收敛。

#### 从聚类簇到视觉子类

因为从搜索引擎得到的图片存在不少噪声图片，用聚类簇中所有的结果进行整合是不合适的。初始的聚类处于高测准低召回的状态，分类非常严格。下一步就应该构建更大的聚类簇，提升边框盒的召回率，这需要利用到上一步的分割先验知识。

这一步可以保持视觉同质（先验的分割信息类似），并避免过拟合。这一步还可以摆脱噪音图像的干扰，因为这类图像数量很少且难以和其他簇合并（视觉差异较大），最终无法构建视觉子类。

作者把对一个簇都让一系列检测器在随机取样的图像块中做检测，利用检测的分数构造了一个矩阵，其中表示检测器个数，表示取样块数量也就是检测次数。表示第个检测器对第张图片的检测分数。每一行都是检测器对应的signature，这个矩阵则对应着当前的簇。计算矩阵的特征向量，取对应特征值最大的特征向量用来表示当前簇。应用k-means算法计算簇的聚类下标（这样就把零散的分类严格的簇再聚类成更大的聚类簇）。聚类完成后，对合并后的每一个聚类簇训练LSVM检测器。

#### 从视觉子类中生成分割

最后一步，作者利用视觉子类，训练一系列不包含判别训练的LSVM检测器。这些训练好的检测器就可以在数据集中进行检测并分类，得到它所属的视觉子类后，再利用先验的分割掩码（由视觉子类中各图片的分割掩码总结而来），对数据集图片进行graph-cut。