**Redmon\_You\_Only\_Look\_CVPR\_2016\_paper**

阅读笔记:

YOLO解决提图像检测问题中“ 慢”的问题

利用整张图作为网络的输入,每个网格预测B个boundingbox,每个boundingbox预测(x,y,w,h, confidence)，直接在输出层回归boundingbox的位置和bounding box所属的类别。

YOLO的实现方法：

将一幅图像分成S×S个网格，每一个网格负责自己的object的预测,每个网格要预测B个bounding box，以及C个类别概率Pr，C是网络分类总数，由训练时决定，在YOLO中每个格子只有一个C类别，即相当于忽略了B个boundingboxes，每个格子只判断一次类别，这样做非常简单粗暴。每个bounding box除了要回归自身的位置之外，还要附带预测一个confidence值，这个confidence值代表了所预测的box中含有object的置信度和这个box预测的多准两重信息，其值计算公式为：

其中若有object落在一个网格中，第一项取1，否则取0，第二项是预测的bounding box和实际的ground truth之间的IOU值。

每个bondingbox要预测(x,y,w,h)和confidence共5个值，每个网格还要预测一个类别信息，记为C类，则S×S个网格，每个网格要预测B个bounding box，还要预测C个categories，输出就是S×S×(5\*B+C)的一个tensor（class是针对每个网格的，confidence信息是针对每个bounding box的）。

在测试的时候，每个网格预测的class信息和bounding box预测的confidence信息相乘，就得到每个bounding box的class-specific confidence score。得到每个box的class-specific confidence score之后，设置阈值，滤掉得分低的boxes，对保留的boxes进行NMS处理，就得到最终的检测结果。

YOLO的缺点：

YOLO对相互靠的很近的物体，还有很小的群体检测效果不好，这是因为一个网格中只预测了两个框，并且只属于一类；对测试图像中，同一类物体中出现新的不常见的长宽比和其他情况时，泛化能力偏弱；由于损失函数的问题，定位误差是影响检测效果的主要原因，尤其是大小物体的处理上，还有待加强。

**Scalable Recognition with a Vocabulary Tree\_nister\_stewenius\_cvpr2006**

阅读笔记:

从图像中提取局部描述子，再对这些描述子进行K-means聚类(视觉中心)，聚类中心成为visual words。 visual words的集合使用TD-IDF（Term Frenquency Inverse Document  Frenquency）打分，来评价查询图像和数据库图像的相关性。作者用这些visual words构造了一个词汇树(Vocabulary Tree)，并提出了一种层次化的TF-IDF打分策略

构造词汇树: 使用大量具有代表性的描述子矢量来进行树的无监督训练，在这里使用K-means聚类方法，不过K不是代表最终的聚类中心的数量，而是代表每一层的分类数。首先，对原始的训练数据进行K-means聚类，定义K个聚类中心。然后把训练数据按照聚类中心分为K个组，每个组的数据都有同样的聚类中心。然后同样的聚类过程应用到每个组中，把每个组再划分为K个组，不停地迭代，直到词汇树到达了预设的最大深度L。

怎么打分:

在线搜索阶段，每个描述子只用按照跟训练相似的步骤逐层地寻找聚类中心，直到叶结点

这是通过比较查询图像和数据库图像的特征在词汇树中的分布相似程度衡量的。一个图像往往会有很多个特征。同时每个特征的重要程度不同,赋予不同的权值,每一个特征和图像之间建立类似于倒排索引,将特征聚类和图像的特征对比,逐层寻找,直到找到叶子结点,将所有匹配特征值结果累加,然后怎么打分最高的图片