

# Large-Scale Image Annotation using Visual Synset

Image annotation/tagging    cluster

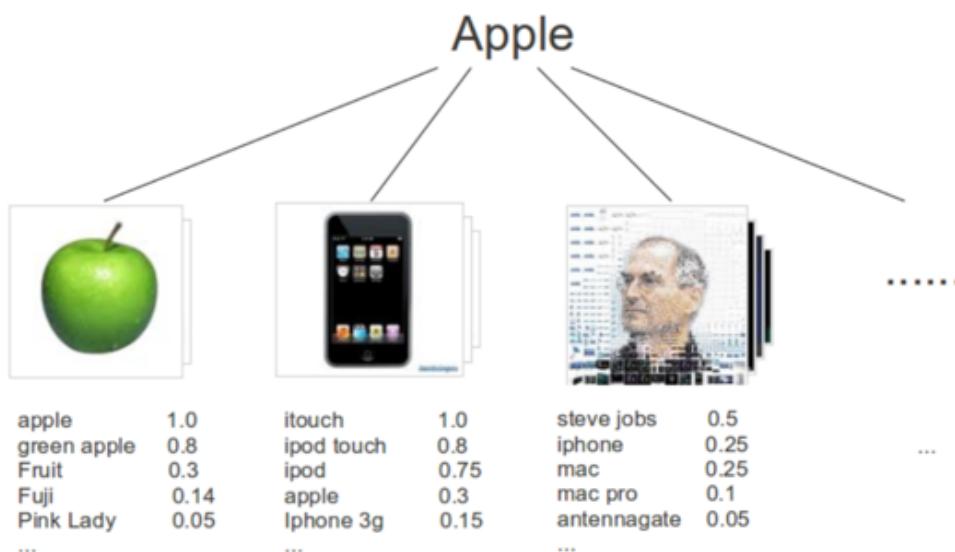
## Pipeline

### 1. 聚类

- 使用AP聚类算法，输入是两两图片之间的相似程度，该相似度由特征权重和的方式进行衡量，论文中使用的是SIFT和Geometric blur
- 聚类的迭代终止条件是：10次迭代中各类的exemplar没有变化
- 得到的类别就是一个visual synset，该synset中包含相似图片，同时 visual synset含有一组标签，这些标签是由google搜索图像时对应的 label集合组成

### 2. 赋权重

- 对于每个visual synset的中label set 中的每个label赋予权重，表示该 label对于visual synset的描述能力
- 权重的赋值方式就是TF-IDF



**Figure 1. Illustration of Visual Synsets.** An object class can be divided into various elementary partitions which share compact visual similarity, each associated with a set of key words to describe the semantic meaning.

### 3. 针对每个synset，训练一个svm。

- 对于一个输入图像，把其特征向量输入到每一个SVM中，如果得到的结果大于某个阈值，认为与该synset相关，最后的label vector就是把与之相关的各个synset中的label做一个element-wise的求和，然后label vector中值较大的标签，认为能描述该图像

# AP聚类

输入是：两两之间的相似度

输出是：自动聚类得到的簇中心

两个重要的描述性质

## 1. Responsibility(i,k)

描述的是数据对象k适合作为数据对象i的聚类中心的程度

## 2. Availability(i,k)

描述的是数据对象i选择数据对象k作为中心的合适程度

$R(i,k)$ 和 $A(i,k)$ 越大，说明数据点k作为聚类中心的可能性越大。

AP聚类就是不断的迭代求解R和A，直到迭代一定的次数，产生m个高质量的  
exemplar

- $R(i,k)$ 吸引度的迭代公式

$$R_{t+1}(i, k) = (1 - \lambda) \cdot R_{t+1}(i, k) + \lambda \cdot R_t(i, k)$$

$$R_{t+1}(i, k) = \begin{cases} S(i, k) - \max_{j \neq k} \{S(i, j)\}, & i = k \\ S(i, k) - \max_{j \neq k} \{A_t(i, j) + R_t(i, j)\}, & i \neq k \end{cases}$$

- $A(i,k)$ 的迭代公式

$$A_{t+1}(i, k) = (1 - \lambda) \cdot A_{t+1}(i, k) + \lambda \cdot A_t(i, k)$$

终止条件：当超过最大迭代次数或者聚类中心不再发生改变时。

优点是：不需要事前给出聚类中心个数

缺点是：需要保存两两之间的相似度，时间复杂度较高