

Large-Scale Image Retrieval with Attentive Deep Local Features

Motivation

在图像检索领域，基于CNN建立图像全局表达的特征提取方法取得了重大的进步，该方法在中小规模的数据上取得了不错的效果，但是当数据集是大规模并且有背景复杂、视线阻挡、视角和光照变化等因素影响时，其性能受到了限制。由于全局表达不能很好地完成部分匹配的图像检索。在本论文中，作者提出了一种基于CNN的注意力局部特征表达，该CNN网络使用图像级类别标签进行训练，作者称之为DELF(DEep Local Feature)。

Framework

作者提出的大规模图像检索系统由四部分组成：

稠密局部化特征提取

首先通过全卷积网络（FCN）在带标注的地标图像数据集上进行训练，提取图像中的稠密特征，并通过构建图像金字塔来处理尺寸变化问题。训练结束后，可得与地标检索任务有关的局部特征表达。

关键点检测

训练得到注意力模型（Attention Model），用该模型对地标分类器来进行弱监督机制学习，以此来获得局部特征表达的相关得分。根据得分进行关键点挑选。

降维

先对选定的特征进行 l_2 正则化，然后运用PCA将维度降到40，最后再对特征使用一次 l_2 正则化。

图片检索系统

本论文中使用的图片检索系统基于最近邻搜索方法，在KD-tree和Product Quantization（PQ）基础上改进得来。该方法先将使用PQ将每个表达编码成50位code，并对没有编码的查询表达进行非对称的距离计算，以此提升最近邻检索的准确率。该方法还是用8K的码本为表达构建了倒排索引。

给定一张查询图像，先对从查询图像中提取的每个局部特征进行近似最近邻搜索，之后对于从索引中检索出的前K个局部特征，对数据库中的每张图像的所有匹配进行聚合，最后使用RANSAC进行集合验证，减少错误查询。

Faster R-CNN Features for Instance Search

在图像检索领域中，常常会涉及到最近邻查询的问题。为降低高维空间中最近邻查询的计算复杂度，人们提出了近似最近邻算法（Approximate Nearest Neighbor（ANN）），通过牺牲一定的精度来换取空间/时间复杂度的降低。本论文中综述了几种常用的ANN实现方法，包括基于局部敏感哈希算法（Locality Sensitive Hashing（LSH））、基于编码的方法、基于树的空间划分方法等。

这里主要介绍局部敏感哈希算法 (LSH)。LSH的主要思想是, 高维空间的两点若距离很近, 那么设计一种哈希函数对这两点进行哈希值计算, 使得他们哈希值有很大的概率是一样的。同时若两点之间的距离较远, 他们哈希值相同的概率会很小。本论文中, 作者讨论了两种改进算法: SRS和QALSH。

SRS通过将高维的数据集映射到的 m 维空间 (m 不超过10), 来进行查询。设原始空间中某点 o 到查询点 q 的距离为 $dist(o)$, 映射空间中该距离为 $\Delta(o)$, 则可观察到 $\frac{\Delta(o)^2}{dist(o)^2}$ 服从 $\chi^2(m)$ 分布, SRS便基于这一点来实现。实现方法分为以下两步:

(1) 通过在数据点的 m 维投影上发布具有 $k = T$ 的 k -NN查询来获得有序候选集合; (2) 如果满足提前终止测试 (如果存在一个 c -ANN点, 其概率至少达到给定的阈值), 则依次检查这些候选者的距离并返回迄今为止距离最小的点; 或者该算法已经耗尽了 T 点。通过设定 $m = 0(1)$, 该算法保证返回点不会远离 c 倍于最近邻距离且具有恒定概率; 空间和时间复杂度在 n 中是线性的并且与 d 无关。

QALSH (Query Aware Locality Sensitive Hashing), 感知查询LSH。该方法引入了感知哈希函数, 该函数是一个外加感知查询桶划分的随机投影, 它不需要传统LSH函数中的随机偏移。在预处理阶段, 所有数据通过感知哈希函数进行映射, 并用 B^+ -tree进行索引; 当一个查询到来时, 感知哈希函数计算该查询的映射, 并用 B^+ -tree来定位落到桶 $[h(q) - \frac{w}{2}, h(q) + \frac{w}{2}]$ 的点。