高级数据结构课程实验二 Proposal

洪方舟 2016013259 hongfz16@163.com

李帅 2016013270 lishuai16THU@163.com

周展平 2016013253 zhouzp16@163.com

摘要

本文对五篇图像检索相关论文进行综述，并且提出了我组最终项目的设想及架构。

关键词

图像检索；深度学习；二进制哈希；

1. 论文综述

本文作者选取了三篇图像检索系统论文，一篇 AlexNet 深度学习网络论文，一篇模糊近邻查询综述论文。

1.1 Deep learning of Binary Hash Codes for the Fast Image Retrieval

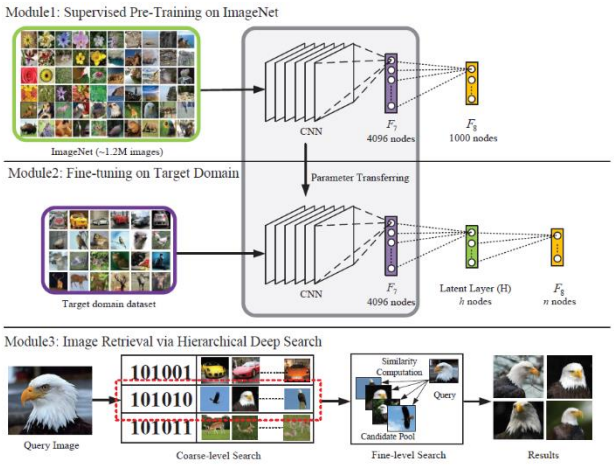
1.1.1 Motivation

在图像检索问题中，基于哈希的算法常常需要通过一个相似度矩阵来计算两幅图像之间的相似程度。然而，相似的矩阵的构建与计算会耗费大量的时间与空间资源。受到深度学习算法的启发，论文作者尝试利用深度学习提取哈希特征。

1.1.2 Framework

在已有的神经网络的基础上，论文在最后增加了与二进制哈希函数相关的网络层。训练时，利用训练好的网络的参数，利用数据集对新增加的网络层进行训练。

整体的思路与框架如下图：



整个算法由以下几部分组成：首先，训练 ImageNet 网络；然后在 CNN 的最后添加一层与产生哈希特征表示相关的神经网络层；最后，可以利用训练好的网络进行检索，检索的步骤为：

（1）将图像输入到训练好的网络中，得到一个输出的向量

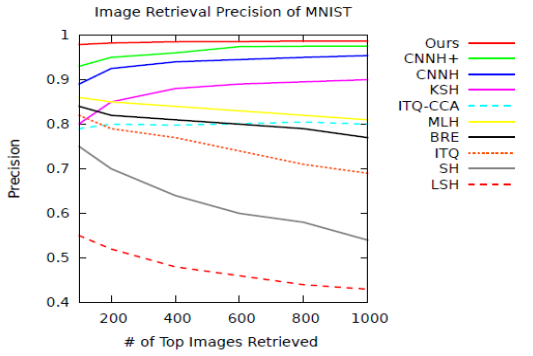
（2）将向量的每一维量化为 0 或 1

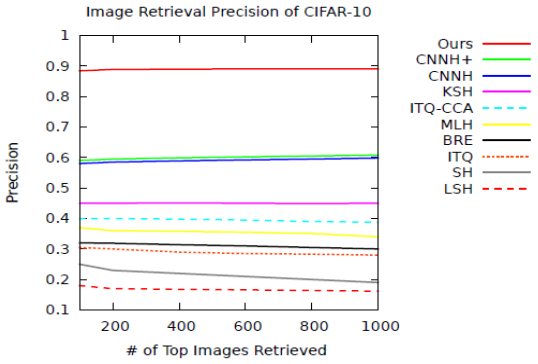
（3）用 Hamming 距离衡量图像之间的相似程度，设置合适的阈值，得到 candidate set

（4）计算 candidate set 中的图像与查询图像的特征之间的距离，这里的特征是指神经网络输出的未经量化的向量。根据欧氏距离对图像进行排序输出。

1.1.3 Performance

在 MNIST 与 CIFAR-10 数据集上，检索的精度与之前最好的结果相比分别提高了 1%与 30%。





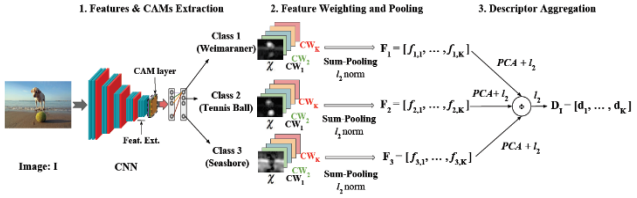
1.2 Class-Weighted Convolutional Features for Visual Instance Search

1.2.1 Motivation

之前许多利用 CNN 进行图像检索的方法都需要针对给定的数据集进行 fine-tune,一些研究也发现了网络中蕴含的有关空间信息的知识，然而对于用于 fine-tune 的数据集的处理花费大量的精力。论文作者希望仅仅使用网络中的知识完成 fine-tune 步骤。

1.2.2 *Framework*

在已有的神经网络的基础上，论文将最后的全连接层替换成了卷积层、GAP 层与池化层，以此提取出 CAMs 并对原来网络得到的特征进行加权求和、池化，得到新的特征表示。



整个算法由以下几部分组成：首先，训练 CNN 网络；然后输入用于 fine-tune 的图像集，针对每一个类别利用 CAM 加权计算出新的特征，并通过求和池化增强效果；之后，对于特征向量中的维度加权计算出新的向量，以此减少通道冗余；最后，依次进行 I2 normalization、PCA-whitening、I2 normalization，将各个类别得到的特征描述合并成一个向量。

1.2.3 *Performance*

在某些数据集上，论文的算法在查询精度上有所提升：

Method	Dim	Oxford5k	Paris6k	Oxford105k	Paris106k
SPoC [3]	256	0.531	-	0.501	-
uCroW [14]	256	0.666	0.767	0.629	0.695
CroW [14]	512	0.682	0.796	0.632	0.710
R-MAC [28]	512	0.669	0.830	0.616	**0.757**
BoW [15]	25k	0.738	0.820	0.593	0.648
Razavian [21]	32k	**0.843**	0.853	-	-
Ours(OnA)	512	0.736	**0.855**	-	-
Ours(OfA)	512	0.712	0.805	**0.672**	0.733

(a)

Method	Dim	R	QE	Oxford5k	Paris6k	Oxford105k	Paris106k
CroW [14]	512	-	10	0.722	0.855	0.678	0.797
Ours(OnA)	512	-	10	0.760	0.873	-	-
Ours(OfA)	512	-	10	0.730	0.836	0.712	0.791
BoW [15]	25k	100	10	0.788	0.848	0.651	0.641
Ours(OnA)	512	100	10	0.780	0.874	-	-
Ours(OfA)	512	100	10	0.773	0.838	0.750	0.780
RMAC [28]	512	1000	5	0.770	**0.877**	0.726	**0.817**
Ours(OnA)	512	1000	5	**0.811**	0.874	-	-
Ours(OfA)	512	1000	5	0.801	0.855	**0.769**	0.800

1.3 Large-Scale Image Retrieval with Attentive Deep Local Features

1.3.1 *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)。

1.3.2 *Framework*

作者提出的大规模图像检索系统由四部分组成：

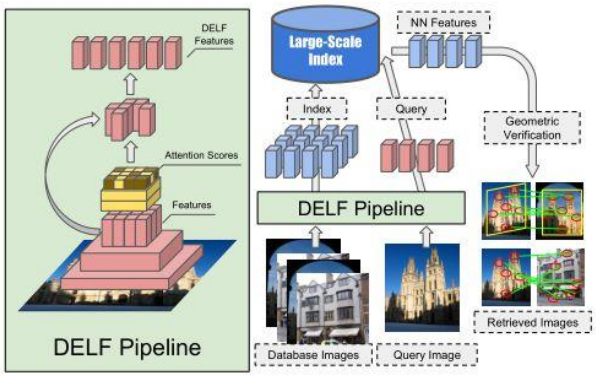


Figure 1: Overall architecture of our image retrieval system, using DEep Local Features (DELF) and attention-based keypoint selection. On the left, we illustrate the pipeline for extraction and selection of DELF. The portion highlighted in yellow represents an attention mechanism that is trained to assign high scores to relevant features and select the features with the highest scores. Feature extraction and selection can be performed with a single forward pass using our model. On the right, we illustrate our large-scale feature-based retrieval pipeline. DELF for database images are indexed offline. The index supports querying by retrieving nearest neighbor (NN) features, which can be used to rank database images based on geometrically verified matches.

1.3.2.1 *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。

1.3.2.2 *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。

1.3.2.3 *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。

1.3.2.4 *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。

1.4 ImageNet Classification with Deep Convolutional Neural Networks

该论文详细讲解了 2012 年 ImageNet 竞赛第一名的网络 AlexNet。相较于之前较为成功的深度学习网络 LeNet-5，该模型中提出了很多新的思想，使得深度学习重新受到人们的关注。

1.4.1 Network Structure

该论文提出了如下创新结构，提高了学习的效率与效果：1. ReLUs 线性整流函数，提高了梯度下降的效率，一定程度上解决了梯度爆炸与梯度消失的问题；2. 局部归一化，避免了对输入数据归一化的需求，使得模型能够更好的泛化；3. 重叠池化，使得模型更加难过拟合。

1.4.2 Training Methods

同时该论文还提出了多种提高训练效率的模型训练方法：1. 数据集增强，减少过拟合，更好的捕捉原始图像的重要特征；2. Dropout 方法，减少参数个数，防止过拟合。最终该网络应用于 LSVRC-2010 数据集，将 Top-1 错误和 Top-5 错误分别降到了 37.5%和 17.0%。

1.5 Approximate Nearest Neighbor Search on High Dimensional Data---Experiments, Analyses, and Improvement

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。

2. 图像检索架构的提出

综合阅读的论文中的方法，我们提出如下图像检索系统的架构:

（1）在已有的神经网络框架的基础上，将全连接层改为卷积层、GAP 层与池化层，以此得到输入图像的特征以及 CAMs，在此基础上利用 CAMs 对特征进行加权求和，计算出新的特征表示。

（2）得到特征表示之后，在网络最后增加一层网络用于表示二进制哈希函数。对于前面得到的新的图像特征，输出一串二进制哈希值，用于快速检索到图像类别。

与前人的方法比较而言，我们的方法同时吸收了图像特征表示的空间体现以及二进制哈希函数检索速度快这两个优点，可以同时提高准确度与检索速度。

3. REFERENCES

[1] Lin, K., Yang, H. F., Hsiao, J. H., & Chen, C. S. (2015). Deep learning of binary hash codes for fast image retrieval. 27-35.

[2] Jimenez, A., Alvarez, J. M., & Giro-I-Nieto, X. (2017). Class-weighted convolutional features for visual instance search.

[3] Noh, H., Araujo, A., Sim, J., Weyand, T., & Han, B. (2017). Large-Scale Image Retrieval with Attentive Deep Local Features. IEEE International Conference on Computer Vision (pp.3476-3485). IEEE.

[4] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[5] Li, W., Zhang, Y., Sun, Y., Wang, W., Zhang, W., & Lin, X. (2016). Approximate Nearest Neighbor Search on High Dimensional Data---Experiments, Analyses, and Improvement (v1. 0). arXiv preprint arXiv:1610.02455.