

# 高级数据结构课程实验二文档

洪方舟  
2016013259  
hongfz16@163.com

李帅  
2016013270  
lishuai16THU@163.com

周展平  
2016013253  
zhouzp16@163.com

## 摘要

本文介绍了我们实现图像检索所采用的深度学习得到的哈希值的方法。依次介绍了图像检索系统的结构、深度学习所使用的网络结构、查询方法、评价标准。最后给出实验结果。

## 关键词

图像检索；深度学习；二进制哈希；

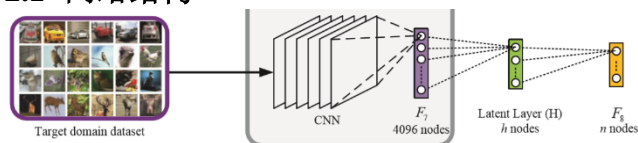
## 1. 图像检索系统综述

我们实现的图像检索系统主要包括两个模块：第一是基于深度神经网络的特征提取，第二是对提取出来的特征与哈希值进行近邻查找。最终使用 mAP 方法以及最近邻十个结果的准确率来度量该系统的有效性。

## 2. 深度神经网络模型

我们使用了 Pytorch 进行神经网络的构建及训练。

### 2.1 网络结构



我们采用了 Pytorch 预训练的 AlexNet 模型做迁移学习。我们将 AlexNet 最后一层输出层去掉，加上两层全连接层，倒数第二层采用 Sigmoid 函数作为激活函数。最后一层为 10 维向量，因为任务所给数据集一共有十个类别。倒数第二层的维数可以调整，我们在本次实验中选取 48 维与 64 维作比较。我们选取倒数第二层在进行 Sigmoid 函数激活之前的输出作为图片的特征，倒数第二层经过 Sigmoid 函数激活之后每一维的数字都映射到(0,1)区间，将所有小于等于0.5的位置为0，大于0.5的位置为1，从而得到一个二值编码作为对应输入图片的哈希值。

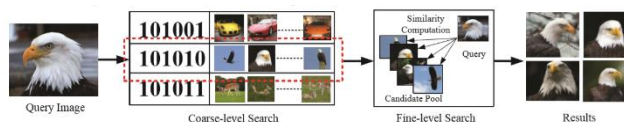
### 2.2 训练方法

由于任务给出的数据集并没有给出图片相关度信息，只给出了分类信息，因此我们训练网络采用的思路是按照分类任务训练。相关理论与实验都很好的证明了通过分类任务学习出的网络中，倒数第二层的输出可以很好的表示图像特征。

在给出最终的查询集之前，我们将所给 5613 张图片按照 4:1 的比例分为训练集与测试集。训练集每 100 张图片为一个 batch，一共迭代 128 次。

## 3. 哈希值与特征的近邻查询

为了加快查询时间，我们采用渐进求精的策略，先使用二值哈希快速的过滤出一个备选集，再用特征向量在备选集中精确查询前十个近邻。查询流程如下图所示。



### 3.1 利用哈希值的粗略查询

我们使用训练好的网络提取了图片数据库中所有图片的哈希值，同时也提取了查询图片对应的哈希值。计算查询图片哈希值与数据库中所有图片哈希值的 Hamming 距离，也即对应位置数字不一样的个数，依据此距离对整个数据库进行排序。取排序结果中 Hamming 距离小于阈值n的图片作为备选集，此处 n 的选取将在实验结果中给出最优值。

### 3.2 利用特征的精确查询

对粗略查询出的备选集中的图片，获取其特征向量，同时计算查询图片的特征向量。计算查询图片特征向量与数据库中所有图片特征向量的欧式距离，依据该距离对备选集进行排序，取出前十个图片作为查询结果。

## 4. 评价标准

### 4.1 mAP

mAP 全称 Mean Average Precision，计算公式为  $MAP = \frac{\sum_{q=1}^Q AveP(q)}{Q}$ ，其中  $AveP(q) = \int_0^1 p(r)dr$ ，p(r)为准确率关于召回率的函数。MAP的含义为准确率-召回率曲线下方面积的平均值。如果该值较高则说明在召回率较高的情况下，准确率仍然能够保持较高的水平。

我们主要对于利用哈希值进行粗略查询之后的排序结果进行MAP计算，计算过程如下。由于没有图片的相关度数据，下面做出假设，查询出与查询图片相同类别的图片，则认为查询正确。对排序后的图片从前到后依次取 1, 2, 3...张图片作为查询结果，计算对应准确率与召回率，得到准确率-召回率曲线。如果计算得到的MAP值较高，则说明排序后的图片中查询正确的图片大多集中在比较靠前的位置，从而证明了直接从排序结果前面取出一定大小备选集的做法的正确性。

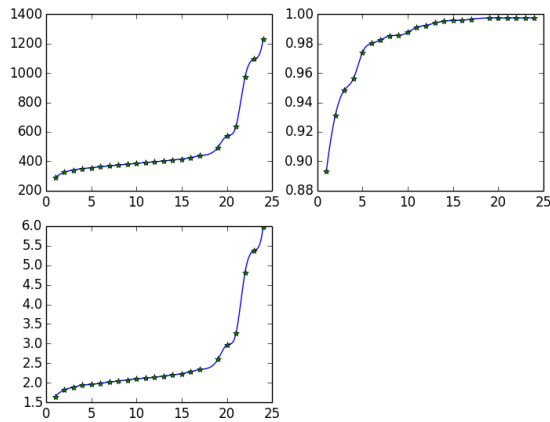
### 4.2 Top10 准确率

与上面相似，我们采用类别判断查询结果正确与否。计算 Top10 准确率的公式为  $Acc = \frac{Number\ of\ correct\ image}{10}$ 。

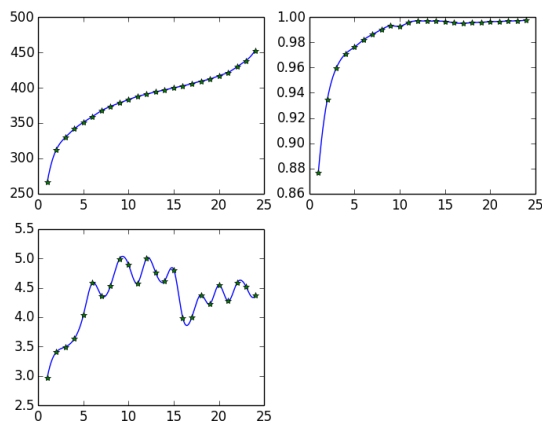
## 5. 实验结果

### 5.1 Hamming 距离的阈值的选取

在粗略查询中，需要确定 Hamming 距离的阈值来筛选出备选集，如果阈值设置过大，备选集过大，则会导致后面精确查询的效率降低；如果阈值设置较小，则可能将一部分正确的结果筛选掉，从而降低准确率。下面通过一组实验数据给出该任务中合适的阈值。



上图为 48 维版本的实验结果。上图中从左到右，从上到下依次为备选集大小、Top10 准确率、1000 次查询所耗时间与 Hamming 距离的阈值的曲线。可以明显看到当阈值为 7 的时候，准确率已经到达 98% 以上，阈值再提高也无法获得更多的准确率的提升。并且当阈值为 7 的时候查询耗时也在可以接受的范围内，平均每 1000 次查询耗时 2 秒左右。因此选取 7 为 48 维版本的 Hamming 距离的阈值。



上图为 64 维版本的实验结果，同 48 维的情况的分析，选取 7 为该版本的 Hamming 距离的阈值。

### 5.2 48 维与 64 维哈希值查询结果对比与分析

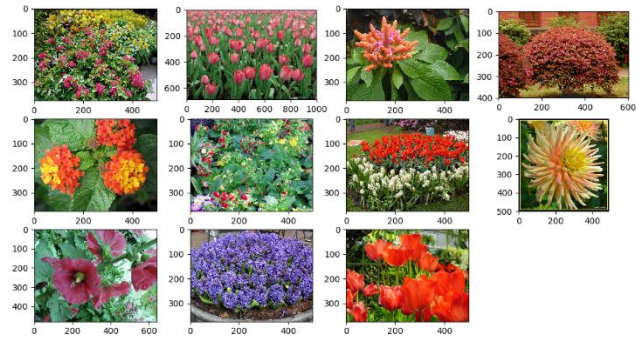
Bits of binary hash	48-bits	64-bits
mAP (%)	83.8	87.1

Top 10 Accuracy (%)	98.2	98.6
Query Time (s/1000 queries)	2.02	4.37

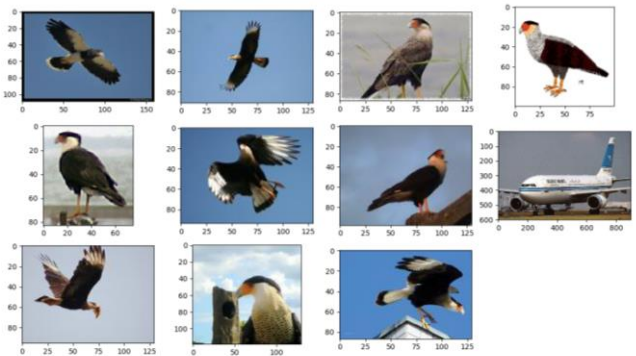
首先从 MAP 值可见使用哈希值的方法进行粗略筛选所得到的结果具有很好的效果，大多数正确结果都被排到数组开头，从而保证了精筛的结果准确率。从 Top10 准确率可见该方法效果非常好，当然如此之高的准确率仅仅是在小样本（4000 张图片作为训练集，1000 张图片作为测试集）下得到的结果，并不具有很高的代表性。横向比较发现，虽然 64 维的版本准确率要好于 48 维版本，但是查询时间却远慢于 48 维版本，因此特征维数不一定越高越好，综合考量准确率与效率才能得出较为合适的维数。

### 5.3 具体的查询样例与分析

下面给出两个查询样例，第一张图片为查询图片，后面的十张图片为查询结果图片。



从上图可以看出，如果仅仅从分类的角度来讲，查询结果非常好，找到的全部是花。但是如果仔细观察，会发现其实查询出来的花的图片与给定的查询图片并不很相似，因为如果考虑要查询出来的是“一丛花”，那么第二、四、七、八张图片均不符合要求。出现这种现象的原因可能在于我们训练集中所给信息只包含类别，而不包含更高级别的信息，例如“花的数量”、“花的颜色”，因而该方法具有较大的局限性。



第二个查询样例存在着同样的问题，如果考虑要查询的结果为“张开双翅的物体”，那么第七个查询结果将不会被认为是错误结果，所以在图像检索中效果的评价标准是多元的，而不能简单将分类作为评价的唯一指标。

## 6. 结论

我们组使用了深度学习学习出图片的特征以及哈希值，先使用哈希值快速筛选出备选集，再用特征距离排序查询出 Top10 的查询结果。通过实验数据可知，我们的方法在以分类为评价准则的评价框架下得到了很好的结果。但是该方法也存在一定的缺陷，由于训练数据集缺失了图片相关度信息或者更高级别的信息。因此对我们训练得到的网络所得出的结果人工分析，会发现虽然同一类但是并不一定相似的情况。

## 7. 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] 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).
- [3] 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.