

# Indexing Images for Content Based Retrieval

蒋梦青

2014013443

jm14@mails.tsinghua.edu.cn

张子昭

2014013430

zzz\_14@126.com

叶佩

2014013456

yep14@mails.tsinghua.edu.cn

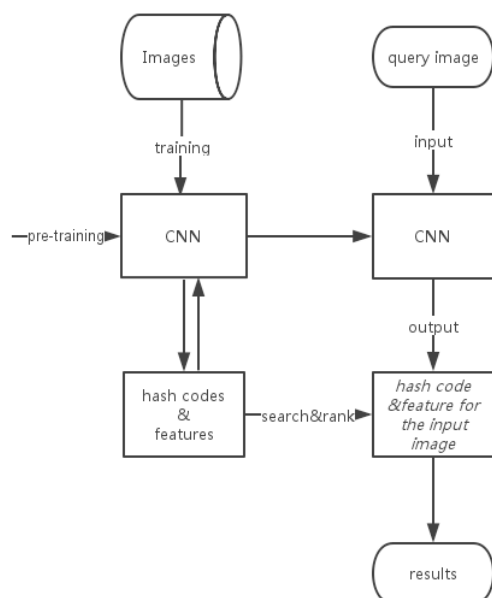
## 1. 引言

随着互联网时代数据爆炸式地增长, 哈希方法越来越多地被用于最近邻搜索。之前有监督的哈希方法使用了相似度矩阵来得到图片的语义信息。但由于本次实验较难得到相似度矩阵, 所以我们使用了基于前向卷积神经网络的哈希方法。

本文的架构如下: 第 2 部分介绍基于内容的图片检索系统的构架, 第 3 部分介绍使用深度神经网络提取特征, 第 4 部分介绍实验中的索引机制和搜索策略, 第 5 部分展示我们的实验过程和结果分析, 第 6 部分为结论。

## 2. 基于内容的图片检索系统

该系统的架构分为三个部分。第一部分: 使用大量的图像进行有监督的预训练, 第二部分: 使用隐层对网络进行微调, 以同时学习表示图片的特征向量和一系列哈希函数。第三部分: 通过粗略搜索和精细搜索得到最终检索结果。这次实验我们使用了开源的 Caffe CNN 库预训练了 CNN 模型, 在此基础上进行第二和第三部分的工作。



## 3. 提取特征: 深度神经网络

最近, 随着深度学习的发展, 卷积神经网络 (CNN, convolutional neural network) 在图像识别领域取得了重要进展, 如图像分类和物体检测。Razavian 等工作指出, 通过在大

规模图像库上预训练得到的 CNN 模型可以用来提取图像的视觉特征 (CNN visual features), 并应用于各类图像识别任务, 如场景分类和多标签分类等。本次实验提取图像特征的方法为:

- 1) 在大规模图像数据集 (ImageNet) 上利用开源的 Caffe CNN 库预训练了 CNN 模型。利用预训练的 CNN 模型提取目标图像库的 CNN 视觉特征, 用于图像检索。
- 2) 利用目标图像库对预训练的 CNN 模型参数进行微调, 并利用微调的 CNN 模型提取目标图像库 CNN 视觉特征, 用于图像检索。

## 4. 检索系统

### 4.1 建立索引: 深度哈希学习

最近的研究发现如果使用神经网络 F6-8 层生成的特征向量, 图像分类、检索以及其他一些操作的性能会大大优化。但是这些特征向量维度很高, 会降低检索图像的效率。为了解决这个问题, 我们可以将高维的特征向量转换为二进制码。这种二进制码可以通过计算海明距离快速地进行比较。

这次实验中, 我们假设最终输出层的分类结果是基于一个大小为  $h$  的属性集, 每个属性可以打开或关闭。换句话说, 同一类的图片哈希码应该相似。所以我们在 F7 和 F8 之间加了一个隐层 H, H 层是完全连接的。它不仅是 F7 层高维特征向量的抽象, 也在低级的特征和高级的语义之间建立了桥梁。我们将 H 层神经元的激活函数设为 sigmoid 函数, 这样神经元的值可以近似为  $\{0, 1\}$ 。

为了使网络适应本次实验的数据集, 我们还通过 BP 算法对神经网络的参数进行了微调。CNN 的权重初始值设为预训练后的权重值。隐层 H 和输出层 F8 的权重初始值随机设置。然后通过有监督的学习使得神经网络的参数更适应实验提供的数据集。

### 4.2 搜索策略: 渐进求精

本次实验我们采用了渐进求精的搜索策略来快速并准确地检索图片。我们首先给定一张图片  $I_q$ , 其哈希码为  $H_q$ 。假设我们用于检索的数据集为  $\Gamma = \{I_1, I_2, \dots, I_n\}$ , 每张图片对应的哈希码为  $H_H = \{H_1, H_2, \dots, H_n\}$ , 如果  $H_q$  和  $H_i \in \Gamma_H$  之间的海明距离小于一个阈值, 则将  $I_q$  加入候选池中。这样我们就得到了  $m$  个候选图片, 然后用高维特征向量对  $m$  张候选的图片进行相似度排序, 输出排序后的结果。

## 5. 实验

### 5.1 实验环境

- 操作系统: Mac OS X

- 编译环境: python 2.7.10
- 依赖库: caffe, os, tkinter
- 编程语言: python

## 5.2 实验过程

实验过程首先是基于神经网络的训练，其次是根据需要，提取网络输出的特征信息、哈希值和类别，并利用这个信息建立检索系统。

我们的具体实现如下：

- 使用 caffe 官方提供的基于 Image-net 大规模图像数据集训练好的 AlexNet<sup>1</sup>模型
- 处理数据集，从提供的 5613 张、10 个类别的图片中，没类抽 100 张图片、共 1000 张图片作为测试集，其余 4613 张图片作为训练数据集，对所有图片生成标签文件和用于 caffe 数据读入的 lmdb 文件
- 对网络结构进行修改，主要是在 AlexNet 的 fc7 和 fc8 之间插入一层全连接 inner product 层，用于产生哈希码，另外还有调整 fc8 层的输出神经元数量等
- 用现有的数据集对网络参数进行微调，迭代次数 1000 次，学习速率 learning rate 为  $10^{-5}$ ，momentum 为 0.9，weight\_decay 为 0.0005
- 用 python 包含 pycaffe 库进行测试和比较，包含 tkinter 库实现图形界面
- 检索过程中，先用 Hamming Distance 来计算哈希码的距离，并以此筛选出一部分图片，然后对它们的特征值向量于查询图片的特征值向量计算欧式距离，并以此作为返回的检索结果的排序

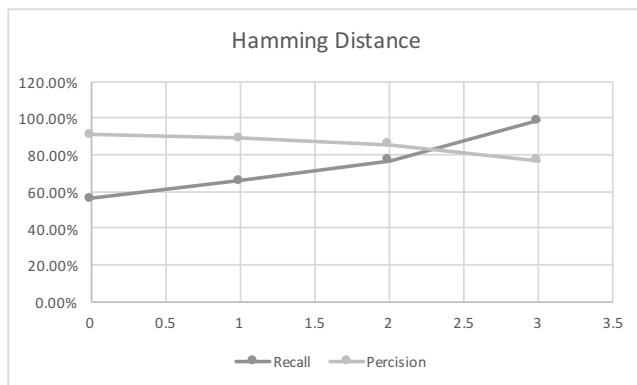
最后，为了得到更好的网络结构和系统参数，我们对测试结果进行了实验：

- 对 Hamming Distance 的不同阈值进行实验
- 对哈希值的不同长度进行实验

实验过程中，在配置 caffe 的环境和调整模型参数上遇到比较大的困难，因为我们没有 GPU，所以运行速度很慢。但数据集总体比较小，所以用有限的资源配合开源的 pretrained AlexNet 模型，我们依然得到了很好的结果。

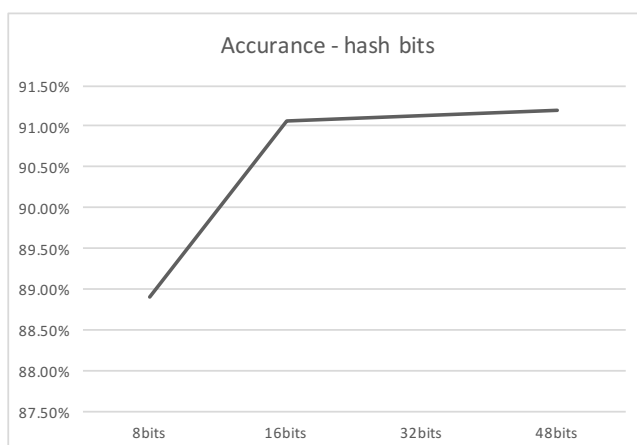
## 5.3 实验结果

在测定图片检索正确率随阈值海明距离变化的关系的时候，我们发现，设定的阈值海明距离越大，搜索得到的图片越多，图片检索正确率越低，这从侧面说明了海明距离越小的图片之间相似度越高，从而验证了算法的正确性。



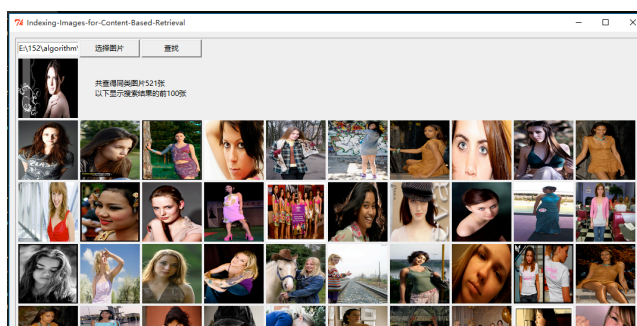
但同时我们也注意到，海明距离越大，recall rate 也越多，即搜索到的结果在相应的类别当中越全面。

我们还对哈希值的维度对正确率的影响进行了探究。



可以看到，本质上来讲，哈希值越长，准确率就越高，但是我们也发现，在 16 维之后，准确率的变化不再那么大，反而带来了很大的计算和空间成本，违背了我们选择哈希的初衷。所以我们牺牲掉一小部分准确度，最终选择 16 维的哈希值，构成了我们的检索系统。

为了更直观地看到我们的检索结果，我们用 python 实现了图形界面。



从图形界面展示的结果可以看到，图片搜索的准确率十分高，通过最后给的测试样例计算得，搜索的准确率高达 91.06%，用数据验证了图形界面带给我们的直观感受。

<sup>1</sup> [http://dl.caffe.berkeleyvision.org/bvlc\\_alexnet.caffemodel](http://dl.caffe.berkeleyvision.org/bvlc_alexnet.caffemodel)

## 6. 结论

通过让已经过监督学习的 CNN 进行给定数据集上的再训练，并调整其网络结构，令其生成哈希码，用 Hamming Distance 比较哈希码相似度来生成检索结果，经过实验选定合适的网络结果和系统参数，从实验结果我们可以看到，我们的系统既能实现搜索的高准确率，又能节约储存空间，实现快速搜索。

## 7. REFERENCES

- [1] A. Babenko, A. Slesarev, A. Chigorin, and V. Lempitsky. Neural codes for image retrieval. In *Computer Vision–ECCV 2014*, pages 584–599. Springer, 2014.
- [2] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys (CSUR)*, 40(2):5, 2008.
- [3] J. Guo and J. Li. Cnn based hashing for image retrieval. *arXiv preprint arXiv:1509.01354*, 2015.
- [4] K. Lin, H.-F. Yang, J.-H. Hsiao, and C.-S. Chen. Deep learning of binary hash codes for fast image retrieval. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 27–35, 2015.
- [5] H. Wang, Y. Cai, Y. Zhang, H. Pan, W. Lv, and H. Han. Deep learning for image retrieval: What works and what doesn’t. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pages 1576–1583. IEEE, 2015.
- [6] R. Xia, Y. Pan, H. Lai, C. Liu, and S. Yan. Supervised hashing for image retrieval via image representation learning. In *AAAI*, volume 1, page 2, 2014.