

一种改进的二进制哈希编码方法

江泽涛^{1,2}, 简 雄^{1,2}, 刘小艳^{1,2}, 曾聪文^{1,2}

(1. 桂林电子科技大学广西高校图像图形智能处理重点实验室, 广西桂林 541004;
2. 桂林电子科技大学广西可信软件重点实验室, 广西桂林 541004)

摘 要: 为了应对手工视觉特征与哈希编码过程不能最佳地兼容以及现有哈希方法无法区分图像语义信息的问题, 提出一种基于深度卷积神经网络学习二进制哈希编码的方法. 该方法基本思想是在深度残差网络中增加一个哈希层, 同时学习图像特征和哈希函数; 以此同时提出一种更加紧凑的分级哈希结构, 用来提取更加接近图像语义的特征. 经 MNIST、CIFAR-10、NUS-WIDE 数据集的实验, 结果表明该方法优于现有的哈希方法. 该方法不仅统一了特征学习和哈希编码的过程, 同时深层残差网络也能得到更接近图像语义的特征, 进而提高了检索准确度.

关键词: 图像检索; 深度残差网络; 分级哈希; 语义信息

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 0372-2112(2019)02-0462-08

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2019.02.029

An Improved Binary Hash Coding Method

JIANG Ze-tao^{1,2}, JIAN Xiong^{1,2}, LIU Xiao-yan^{1,2}, ZENG Cong-wen^{1,2}

(1. Key Laboratory of Image and Graphic Intelligent Processing of Guangxi, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China;
2. Key Laboratory of Dependable Software of Guangxi, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China)

Abstract: To address the problems that hand-engineering visual features can't be optimally compatible with the Hash coding process and existing Hash methods can't differentiate images semantics information, a learning method of binary Hashing based on deep convolutional neural networks is proposed. The basic idea is to add a Hash layer into the deep residual network and to learn simultaneously image features and Hash functions. Meanwhile, we propose a more compact hierarchical Hashing structure to extract features closer to semantics information of images. Experimental results of MNIST, CIFAR-10 and NUS-WIDE datasets show that the method is superior to existing Hashing methods. This method not only unifies the process of feature learning and Hash coding, and at the same time, the deep residual network is able to get features closer to image semantics. Thus the retrieval accuracy is improved.

Key words: image retrieval; deep residual network; hierarchical Hashing; semantics information

1 引言

随着网络和社交媒体的快速发展, 网络多媒体与图像数据呈现爆炸式的增长. 如何从大规模的图像资源中进行快速而有效的检索以满足用户需求是亟待解决的问题. 在大规模数据集上基于内容的图像检索(Content-Based Image Retrieval, CBIR)已引起了越来越多的关注, CBIR 目的是从大型图像数据集中返回与视觉查询相匹配的相似图像. 最近邻搜索(Nearest Neighbor, NN)是 CBIR 最基本的方法, 但是由于 NN 搜索通常是将待查询图像与数据集中的每个样本进行详尽的比较, 所以当数据集变的很大的时候, 该方法变得不可行, 因此, 更多的关注集中在近似最近邻搜索(Approximate Nearest Neighbor, ANN)^[1]. 为了有效地进行 ANN 搜索, 许多方法被设计出来, 其中包括基于树的方法和基于哈希的方法. 基于树的方法, 如 KD 树^[2,3]通常在有效的数据结构中组织数据样本用来实现快速搜索, 但是这种方法在高维数据下搜索效率会极大地降低. 而以位

bor, NN)是 CBIR 最基本的方法, 但是由于 NN 搜索通常是将待查询图像与数据集中的每个样本进行详尽的比较, 所以当数据集变的很大的时候, 该方法变得不可行, 因此, 更多的关注集中在近似最近邻搜索(Approximate Nearest Neighbor, ANN)^[1]. 为了有效地进行 ANN 搜索, 许多方法被设计出来, 其中包括基于树的方法和基于哈希的方法. 基于树的方法, 如 KD 树^[2,3]通常在有效的数据结构中组织数据样本用来实现快速搜索, 但是这种方法在高维数据下搜索效率会极大地降低. 而以位

收稿日期: 2017-05-02; 修回日期: 2018-07-20; 责任编辑: 李勇锋

基金项目: 国家自然科学基金(No. 61572147, No. 61762066); 广西科技计划项目(No. AC16380108); 桂林电子科技大学图像图形智能处理重点实验室项目(No. GHP201405, No. GHP201701); 广西可信软件重点实验室项目(No. kx201502); 广西研究生教育创新计划(No. 2018YJCX46); 桂林电子科技大学校级教改重点项目(No. JGA201506); 广西云计算与大数据协同创新中心(No. YD16304)

置敏感哈希 (Locality Sensitive Hashing, LSH)^[4] 为代表的哈希方法能够高效地实现 ANN 搜索,因而在近几年受到更多的关注. 哈希方法旨在将高维数据映射到低维二进制代码表示中. 一方面,哈希编码一般产生仅占用恒定查询时间的哈希表;另一方面,由于基于哈希的方法只需要存储二进制代码,所以能够显著减少存储. 除了 LSH 之外,常见的典型哈希方法还包括谱哈希 (SH)^[5]、PCA-ITQ^[6]、BRE^[7] 和 KSH^[8] 等.

对于大多数现有的哈希方法,输入首先被映射到低维度子空间,然后跟着一个单独的量化过程来得到紧凑的二进制向量. 基于随机投影的 LSH 及其扩展^[9-12] 是最广泛使用的哈希方法之一. 上述所提到的所有这些哈希方法都是基于手工提取的特征描述算子,如 GIST^[13]、BoF^[14,15] 来学习哈希函数的. 但是这些手工提取的特征只能提取图像的低层次的表示,并不能保留图像的语义相似性,而且这些方法中哈希函数的学习过程和特征提取的过程是分开的,所以哈希码学习过程不能给特征学习过程提供反馈,反过来也一样.

由于深度卷积神经网络在特征学习上的优越性,深度学习在许多视觉任务中已经取得了显著的成功,比如图像分类^[16]、图像检索^[17]、人脸识别^[18,19] 和姿态评估^[20]. 近几年也出现了深度卷积神经网络和哈希技术相结合的算法,Xia Rongkai^[21] 提出了一种 CNNs 与哈希方法结合的算法,该论文把图像特征提取和哈希编码的学习分为两个阶段,但效果不理想. Lai Hanjiang^[22] 提出了一种利用深度卷积神经网络同时学习图像特征和哈希函数的算法,该论文把三元组作为监督信息,三元组的挑选质量直接影响检索的准确度,同时三元组的挑选是一项特别繁杂的过程. Lin K^[23] 也提出了一种利用深度卷积神经网络框架同时学习特征和哈希函数的算法,并且使用标签信息作为监督信息,避免了挑选三元组的过程,但是它得到的哈希码过于冗长.

由于深度卷积神经网络提取的特征随着网络层数的增加而更加抽象、更接近于图像的真实语义,而深度残差网络又是目前最深、性能最好的卷积网络,所以本文提出在深度残差网络框架中引入哈希层,利用图像标签信息同时进行图像特征和哈希函数的学习,并且通过分级处理让哈希码更紧凑,相同长度的哈希码能够比其它哈希方法检索精度更高的基于深度卷积神经网络深度残差网络 (Deep Residual Network) 和哈希策略来学习二进制哈希函数的编码方法,然后将学习得到的哈希码用于大规模的图像检索.

该方法在 MNIST、CIFAR-10 和 NUS-WIDE 数据集上的实验结果表明,本文提出的哈希方法的检索性能优于现有的方法.

2 改进的二进制哈希编码方法

本文方法的框架如图 1 所示. 模型主要包括四个主要部分:1) 数据输入层,输入为图像为数据集里的图像或者是待查询图像;2) 深度残差网络层,该层是用目前分类效果最好的深度卷积网络来学习图像的特征表示;3) 哈希层,本文提出一种新的有效分级哈希结构;4) 损失层,该层同时考虑 softmax 分类器的分类误差和量化误差. 概括地说,输入图像数据首先经过深度残差网络层得到图像的特征表示,然后图像特征经过哈希层得到哈希码,最后哈希码进入损失层,计算损失函数,并且对损失函数学习到的模型参数进行优化进而提高检索的准确度.

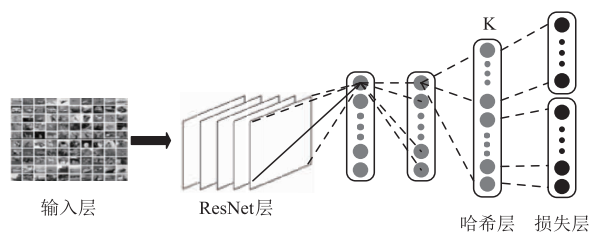


图1 本文方法框架

2.1 深度残差网络

本文使用深度残差网络来学习图像的特征表示,输入图像通过在拥有千万级别张图片的 ImageNet 数据集上预训练得到的深度残差网络可以得到图像的特征表示,该特征比手工制作的特征要更接近图像的语义,进而能够更有利于提高图像分类和图像检索的准确率,本文采用的是深度为 50 的 ResNet-50 网络模型作为基本架构.

何凯明团队提出的残差网络 Residual Network 在 ImageNet 的图像分类、目标检测方面取得非常理想的效果,与 AlexNet、VGG 和 GoogLeNet^[24] 对比,ResNet^[25] 主要是解决了所谓的“退化”问题,即当模型的层次加深时,错误率却提高了. 该问题产生的原因归结于优化难题,当模型变复杂时,SGD 的优化变得更加困难,导致了模型达不到好的学习效果. 针对这个问题,何凯明团队提出了一个如图 2 的 Residual 的结构,即增加一个恒等映射 (identity mapping),将原始所需要学的函数 $H(x)$ 转换成 $F(x) + x$,两种表达的效果相同,但是 $F(x)$ 的优化比 $H(x)$ 要简单的多. 其中残差模块 (Residual block) 通过捷径连接 (short connection) 实现,通过 shortcut 将 block 的输入和输出进行叠加,这个简单的加法不会给网络增加额外的参数和计算量,同时却可以大大增加模型的训练速度、提高训练效果,并且当模型的层数增加时,这个简单的结构能够很好的解决退化问题,这也是残差网络 ResNet 能够做到比之前的网络深

得多的根本原因.

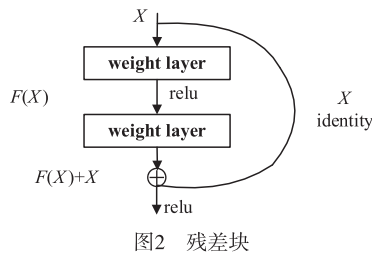


图2 残差块

2.2 哈希编码 (Hash coding)

根据是否独立于训练数据,哈希方法可以分为数据无关和数据相关哈希方法.典型的数据无关的哈希方法包括局部敏感哈希 LSH 和移位不变内核哈希 SIKH 等.与数据相关哈希方法相比较,达到相同预期的性能,数据无关哈希方法一般需要更长的哈希编码位数.数据相关哈希方法又可以进一步分为无监督和有监督的哈希方法,分类的依据在于是否提供监督信息.无监督哈希方法在学习哈希函数时只利用训练数据,没有使用任何监督信息,典型的无监督哈希方法包括 SH、BRE 和 ITQ 等;有监督哈希方法利用标签数据来学习哈希编码,通常监督信息以点标签、成对标签或排序标签这三种形式中的一种给出.典型的有监督哈希方法包括 CCA-ITQ、KSH 和 CNNH 等.

本文的分级哈希结构如图 3,分为 A 和 B 两个部分,A 部分的长度为 K_1 位,最多可用来表示 2^{K_1} 种不同的类别,如 CIFAR-10 有 10 个类别的图像,则 $K_1 = 4$ (最多能表示 16 种类别)就能完全表示其包含的类别,而 B 部分的长度为 K_2 位,最多可用来表示 2^{K_2} 种不同的子类别,还是以 CIFAR-10 为例,假如其中包含 100 种不同狗的类别,则 $K_2 = 7$ (最多能表示 128 种子类)就能完全表示所有不同狗的子类别.其中 $K_1 + K_2 = K$,这里的 K 为指定的哈希码的编码长度.本文提出的这种哈希结构有点类似于网络地址中的网络号 + 主机号的分级结构,A 确定图像属于哪一个大类,B 确定图像属于该类中具体哪一种,这样的哈希码不仅能够大致区分不同的类,还能区分更接近图像语义的信息,从而返回更准确的检索结果.

2.3 损失层

本文损失层函数包括 Softmax 分类器损失和量化误差损失. Softmax 分类误差损失,记为 L_c . 同时,将连续值哈希码二值化为离散哈希码时也会产生一个量化误差损失 L_q :

$$L_q = \frac{1}{2} \|h - s\|^2 \quad (1)$$

其中 $h = (g(s^{(1)}), g(s^{(2)}), \dots, g(s^{(q)}))^T$ 为阈值化输出的哈希码, s 为输出的连续值哈希码. 量化误差损失函数的优化目标是通过学习让激活层的输出值尽可能地

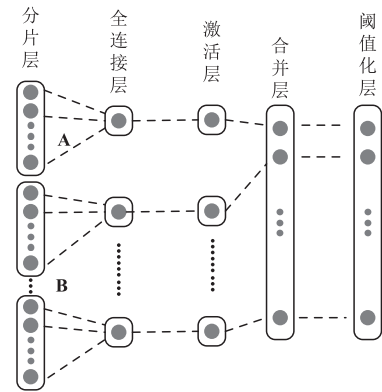


图3 哈希层

接近量化值 0 和 1,从而达到降低阈值化过程带来的误差.

最后,结合 Softmax 分类损失 L_c 和量化误差损失 L_q ,将两者进行加权求和,得到整体的损失函数:

$$\text{Loss} = \lambda_1 L_c + \lambda_2 L_q \quad (2)$$

其中 λ_1 和 λ_2 为权重因子,它们的值取决于分类误差和量化损失在整体损失中的重要性.在模型训练中, λ_1 和 λ_2 参数值的选取既可以使用固定的经验值,也可以和其它网络参数一样采用相同的选取方式,即先设定一个初始值,然后随着模型的迭代进行优化,本文采用的是后者.

2.4 通过深度残差网络学习哈希码

通过在目前分类效果最好的深层卷积神经网络,即深度残差网络 ResNet 上添加哈希层,同时进行特征表示的学习和哈希编码的学习,整个网络通过不断迭代和调参的过程,对模型进行优化,最终生成一个用于大规模图像检索的最优网络模型.具体过程如图 1 所示,作为输入的图像,首先经过残差网络得到特征表示,然后经过哈希层,得到二进制哈希码,再通过损失层,计算误差,然后反向传播计算梯度,并根据梯度调整哈希层的权重参数,随着网络的迭代而不断优化哈希层的参数,最终随着网络的收敛得到最优的参数,进而得到最优的哈希函数.

3 实验结果

为了验证本文方法,在数据集 MNIST、CIFAR-10 和 NUS-WIDE^[26] 上把本文方法和现有的其它哈希方法,包括 LSH、BRE、CNNH 和 CNNH+ 进行对比实验,其中 LSH 是一种无监督的哈希方法,BRE 是有监督的哈希方法,两者的共同点就是 LSH 和 BRE 使用的都是手工提取的特征;而 CNNH 和 CNNH+ 和本文提出的方法一样,都是利用卷积神经网络提取的图像特征并且都是使用有监督的标签信息,其中 CNNH+ 是 CNNH 的改进版. CNNH、CNNH+ 和本文都是直接使用图像作为输

入,LSH 和 BRE 在数据集 MNIST 和 CIFAR-10 上采用 Gist 特征表示图像,在 NUS-WIDE 上图像使用 bag-of-words 向量表示图像.为了使最后的对比结果更加清楚和更具说服力,本文使用 mAP、top-k 近邻域检索准确率和汉明距离小于 2 的准确率 3 个性能指标来展示实验结果.下面对本文用到的数据集和性能指标以及最后的实验结果做具体的说明和分析.

3.1 数据集

MNIST 该数据集包含 70000 张 28X28 的灰度图像,包括手写数字 0 到 9 共 10 个类别,其中 60000 张是训练数据集,10000 张是测试数据集.

CIFAR-10 该数据集包含 60000 张 32X32 的彩色图像,包括飞机、卡车等 10 个类别,其中 50000 张是训练数据集,10000 张是测试数据集.

NUS-WIDE 该数据集包含 270000 张图像,包括 81 个大类,每张图片具有一个或多个标签.本文使用 21 个常用类,其中每一个类至少包括 5000 张图像.

3.2 性能指标

为了评估本文方法的图像检索性能,并与已有的方法作比较,本文采用 mAP、top-k 近邻域检索准确率和汉明距离小于 2 的准确率 3 个性能指标对实验结果进行评估.

Mean average precision (mAP) 一般我们都是根据与查询图像的汉明距离和计算得到的 mAP 值来对图像进行排序.mAP 的值就是 Precision-Recall 曲线所包围的面积,mAP 的值是哈希函数整体性能的指标.

top-k 近邻域检索准确率 该指标表示的是与查询图像汉明距离最小的 k 张图像中正确的结果所占的比例,换言之,就是查询出来排名前 k 的图像中有正确结果的比例.

汉明距离小于 2 的准确率 该指标表示的是与查询图像的汉明距离小于 2 的图像中正确结果所占的比例.

3.3 实验结果与分析

3.3.1 MNIST 数据集上的检索结果

表 1 给出的是在 MNIST 数据集上本文方法与 LSH、BRE、CNNH 和 CNNH + 四种哈希方法在分别采用 12 位、32 位和 48 位哈希码的 mAP 值对比结果,表 2 给出的是对应的汉明距离小于 2 的准确率对比结果,图 4,5,6 给出的是在 MNIST 数据集上的实验对比曲线.

表 1 在数据集 MNIST 上 mAP 值对比

方法	12 位	32 位	48 位
本文	0.9892	0.9921	0.9986
CNNH +	0.9695	0.9751	0.9798
CNNH	0.9685	0.9702	0.9762
BRE	0.5142	0.5823	0.6115
LSH	0.1934	0.2122	0.2315

表 2 在数据集 MNIST 上汉明距离小于 2 的准确率对比

方法	12 位	32 位	48 位
本文	0.9824	0.9872	0.9884
CNNH +	0.9252	0.9265	0.9271
CNNH	0.9212	0.9225	0.9264
BRE	0.5931	0.7824	0.5847
LSH	0.1924	0.4322	0.3114

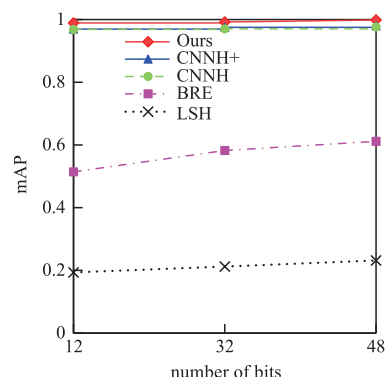


图 4 不同哈希码长度下的 mAP 曲线

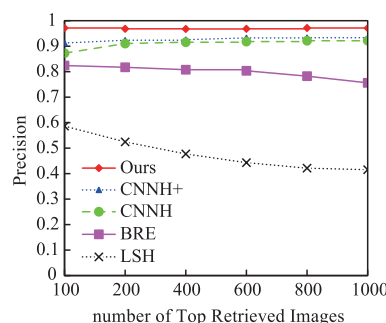


图 5 48 位哈希码长度下 top-k 近邻域检索准确率曲线

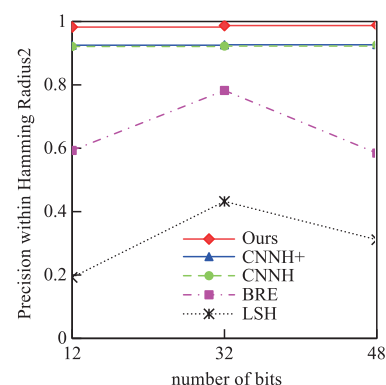


图 6 不同哈希码长度汉明距离小于 2 的准确率曲线

3.3.2 CIFAR-10 数据集上的检索结果

表 3 给出的是在 CIFAR-10 数据集上本文方法与 LSH、BRE、CNNH 和 CNNH + 四种哈希方法在分别采用 12 位、32 位和 48 位哈希码的 mAP 值对比结果,表 4 给出的是对应的汉明距离小于 2 的准确率对比结果,图 7,8,9 给出的是在 CIFAR-10 数据集上的实验对比

曲线.

表 3 在数据集 CIFAR-10 上 mAP 值对比

方法	12 位	32 位	48 位
本文	0.9141	0.9087	0.9316
CNNH +	0.4276	0.4434	0.4522
CNNH	0.4174	0.4425	0.4516
BRE	0.1841	0.1920	0.1937
LSH	0.1210	0.1241	0.1307

表 4 在数据集 CIFAR-10 上汉明距离小于 2 的准确率对比

方法	12 位	32 位	48 位
本文	0.8274	0.8869	0.8498
CNNH +	0.4878	0.5261	0.4613
CNNH	0.4876	0.5247	0.4764
BRE	0.2034	0.2218	0.0246
LSH	0.1224	0.1730	0.1814

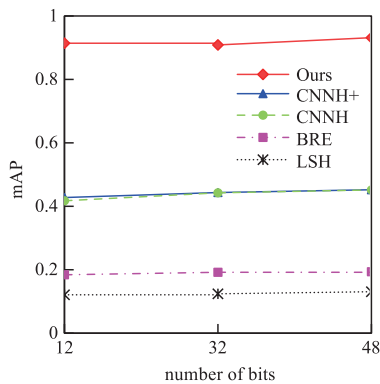


图7 不同哈希码长度下 mAP 曲线

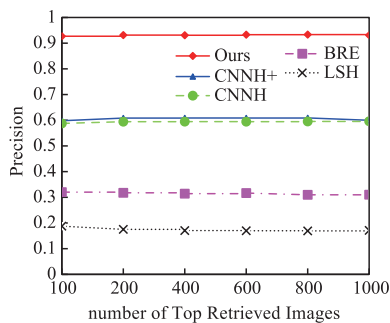


图8 48位哈希码长度下 top-k 近邻域检索准确率曲线

3.3.3 NUS-WIDE 数据集上的检索结果

表 5 给出的是在 NUS-WIDE 数据集上本文方法与 LSH、BRE、CNNH 和 CNNH + 四种哈希方法在分别采用 12 位、32 位和 48 位哈希码的 mAP 值对比结果,表 6 给出的是对应的汉明距离小于 2 的准确率对比结果,图 10,11,12 给出的是在 NUS-WIDE 数据集上的实验对比曲线.

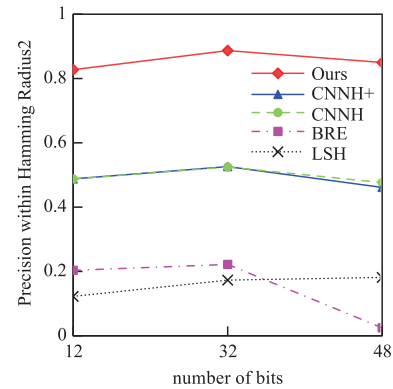


图9 不同哈希码长度汉明距离小于2的准确率曲线

表 5 在数据集 NUS-WIDE 上 mAP 值对比

方法	12 位	32 位	48 位
本文	0.8965	0.9147	0.9195
CNNH +	0.6186	0.6304	0.6157
CNNH	0.6147	0.6294	0.6098
BRE	0.4612	0.5144	0.5276
LSH	0.4076	0.4215	0.4196

表 6 在数据集 NUS-WIDE 上汉明距离小于 2 的准确率对比

方法	12 位	32 位	48 位
本文	0.7850	0.9046	0.8976
CNNH +	0.5998	0.5897	0.5664
CNNH	0.5864	0.5893	0.5644
BRE	0.4475	0.4524	0.2664
LSH	0.4075	0.1864	0.5466

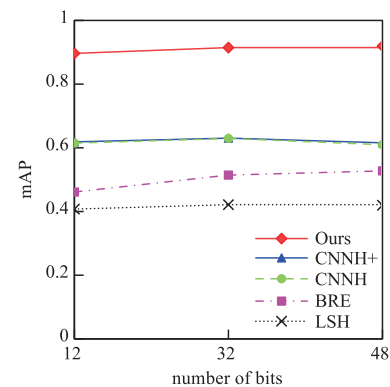


图10 不同哈希码长度下的 mAP 曲线

3.3.4 加入约束条件和未加约束条件的性能对比

为了验证和说明本文加入哈希码独立性和考虑阈值化产生的量化误差这两个约束条件的有效性,本文又在分别去掉哈希层中的分片层、去掉损失函数层中对量化误差的优化模块以及把两者都去掉,即未做任何优化后做了 3 组对比实验,其中去掉分片层就是破坏哈希码的独立性,去掉量化误差的优化模块就是忽略阈值化过程带来的量化误差,对比实验结果如表 7,8,9 所示.

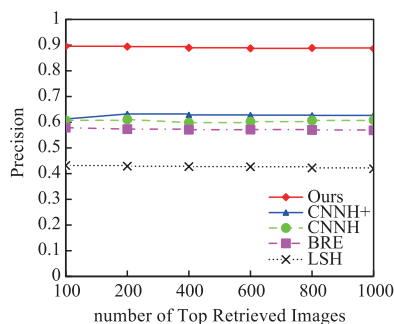


图11 48位哈希码长度下top-k近邻检索准确率曲线

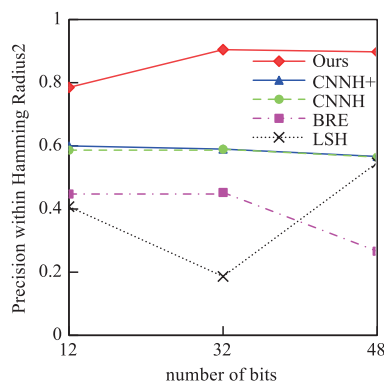


图12 不同哈希码长度汉明距离小于2准确率曲线

表 7 在数据集 MNIST 上 mAP 值对比

方法	12 位	32 位	48 位
本文	0.9892	0.9921	0.9986
不满足独立性	0.9860	0.9905	0.9977
忽略量化误差	0.9722	0.9764	0.9814
未做任何优化	0.9690	0.9712	0.9773

表 8 在数据集 CIFAR-10 上 mAP 值对比

方法	12 位	32 位	48 位
本文	0.9141	0.9087	0.9316
不满足独立性	0.9011	0.8956	0.9205
忽略量化误差	0.8734	0.8872	0.8995
未做任何优化	0.8605	0.8644	0.8795

表 9 在数据集 NUS-WIDE 上 mAP 值对比

方法	12 位	32 位	48 位
本文	0.8965	0.9147	0.9195
不满足独立性	0.8823	0.9005	0.9068
忽略量化误差	0.8675	0.8722	0.8764
未做任何优化	0.8501	0.8574	0.8642

3.3.5 结果分析

由表 1、表 3 和表 5 可知,在 MNIST、CIFAR-10 和 NUS-WIDE 三个数据集上,本文哈希方法在 mAP 值上分别高出 LSH 方法近 80%、78% 和 50%,高出 BRE 方法近 40%、70% 和 40%;与同样使用 CNN 特征的 CNNH 和 CNNH+ 方法对比,该方法的 mAP 也是最高的,特别

地,该方法在使用长度为 12 位哈希码的 mAP 值要远高于 CNNH 和 CNNH+ 使用 48 位哈希码的 mAP 值,这说明在达到相同检索准确率的要求下,该哈希方法可以用长度更短的哈希码来实现.对于除了 mAP 之外的另外两个性能分析指标 top-k 近邻检索准确率和汉明距离小于 2 的准确率的实验结果分析同 mAP,所以不再赘述.与现有哈希方法对比实验的结果中可以看出,本文方法由于采用的是残差网络提取的高级语义特征,所以与采用手工特征的 LSH、BRE 方法相比在性能上有非常大的提升,而与 CNNH、CNNH+ 这些也使用卷积网络特征的方法相比,由于残差网络具有比 CNNH 和 CNNH+ 网络更深的网络结构以及本文采用分级哈希结构,所以提取的特征也比这两种方法更接近图像的真实语义,进而检索性能也得到提升.

4 结束语

本文提出了一种基于深度残差网络的用于大规模图像检索的哈希方法.该方法与现有其它相关的方法对比,具有三个特点,首先,在现有分类效果最好的深度残差网络的基础上添加哈希层,并做相应的处理让预训练好的残差网络的参数不变,仅对新加入的哈希层的参数进行迭代优化,这样既充分利用了现有深度模型的优势,又避免了自己重新训练模型的各种困难;其次,与现有哈希方法不同的是,本文提出了一种全新的哈希方法来生成更紧凑的哈希码,在与其他哈希码相同长度的情况下,本文的哈希码能够利用多出来的哈希位来对图像做更细致的区分;最后,该哈希方法得到的特征更加接近图像的真实语义.

该方法由于运用深度残差网络以及分级哈希结构能得到更加接近图像语义的特征表示,从而提高了检索的准确率;也正因这个优势,本文方法和其它利用 CNN 特征的哈希方法对比,如本文中的 CNNH 和 CNNH+,达到相同的检索准确率所需要的哈希码的长度更短,这使得本文的方法更加适用于大规模的图像检索.此外本文方法还在整个框架中加入了哈希函数独立性的限制,同时考虑到量化误差对哈希函数的影响,把连续哈希码阈值化为离散哈希码的过程中产生的量化误差也加到损失函数中,因而本文构造出的哈希方法与现有的哈希方法相比,得到的哈希函数更好,检索准确率更高.

参考文献

- [1] Indyk P, Motwani R. Approximate nearest neighbors: towards removing the curse of dimensionality[A]. Proceedings of the ACM Symposium on Theory of Computing [C]. TX, USA: ACM, 1998. 604-613.

- [2] Bentley JL. K-d trees for semi-dynamic pointsets [A]. Proceedings of the Sixth Annual Symposium on Computational Geometry [C]. Berkeley, California, USA; ACM, 1990. 187 – 197.
- [3] Silpaanan C, Hartley R. Optimised KD-trees for fast image descriptor matching [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. Anchorage, Alaska, USA; IEEE, 2008. 1 – 8.
- [4] Gionis A, Indyk P, Motwani R. Similarity search in high dimensions via hashing [A]. Proceedings of 25th International Conference on Very Large Data Bases [C]. Edinburgh, Scotland, UK, 1999. 518 – 529.
- [5] Weiss Y, Torralba A, Fergus R. Spectral hashing [A]. Proceedings of the Conference on Neural Information Processing Systems [C]. Vancouver, Canada, 2008. 1753 – 1760.
- [6] Gong Y, Lazebnik S, Gordo A, et al. Iterative quantization: a Procrustean approach to learning binary codes for large-scale image retrieval [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35 (12): 2916 – 2929.
- [7] Kulis B, Darrell T. Learning to Hash with Binary Reconstructive Embeddings [A]. Advances in Neural Information Processing Systems [C]. Vancouver, Canada, 2009. 1042 – 1052.
- [8] Liu W, Wang J, Ji R, et al. Supervised hashing with kernels [A]. Proceedings of the Computer Vision and Pattern Recognition [C]. Providence, RI, USA; IEEE, 2012. 2074 – 2081.
- [9] Datar M, Immorlica N, Indyk P, et al. Locality-sensitive hashing scheme based on p-stable distributions [A]. Proceedings of the Twentieth Annual Symposium on Computational Geometry [C]. Brooklyn, New York, USA; ACM, 2004. 253 – 262.
- [10] Kulis B, Grauman K. Kernelized locality sensitive hashing for scalable image search [A]. Proceedings of the 12th International Conference on Computer Vision [C]. Kyoto, Japan; IEEE, 2009. 2130 – 2137.
- [11] Kulis B, Jain P, Grauman K. Fast similarity search for learned metrics [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31 (12): 2143 – 2157.
- [12] Mu Y, Yan S. Non-metric locality-sensitive Hashing [A]. Proceedings of the 24th AAAI Conference on Artificial Intelligence [C]. Atlanta, Georgia, USA, 2010. 2235 – 2246.
- [13] Oliva A, Torralba A. Building the gist of a scene: the role of global image features in recognition [J]. Progress in Brain Research, 2006, 155 (2): 23 – 36.
- [14] Csurka G, Dance C R, Fan L, et al. Visual categorization with bags of keypoints [J]. Workshop on Statistical Learning in Computer Vision ECCV, 2004, 44 (247): 1 – 22.
- [15] Sivic J, Zisserman A. Video google: A text retrieval approach to object matching in videos [A]. Proceedings of the 9th IEEE International Conference on Computer Vision [C]. Nice, France; IEEE, 2003. 1470 – 1477.
- [16] Krizhevsky A, Sutskever I, Hinton G E. Image-Net classification with deep convolutional neural networks [A]. Proceedings of the International Conference on Neural Information Processing Systems [C]. Lake Tahoe, USA, 2012. 1097 – 1105.
- [17] Babenko A, Slesarev A, Chigorin A, et al. Neural Codes for Image Retrieval [A]. Proceedings of the European Conference Computer Vision [C]. Zürich, Switzerland; ECCV, 2014. 584 – 599.
- [18] Taigman Y, Yang M, Ranzato M, et al. DeepFace: Closing the gap to human-level performance in face verification [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. Columbus, OH, USA; IEEE, 2014. 1701 – 1708.
- [19] Sun Y, Wang X, Tang X. Deep learning face representation from predicting 10,000 classes [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. Columbus, OH, USA; IEEE, 2014. 1891 – 1898.
- [20] Toshev A, Szegedy C. DeepPose: Human pose estimation via deep neural networks [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. Columbus, OH, USA; IEEE, 2014. 1653 – 1660.
- [21] Xia R, Pan Y, Lai H, et al. Supervised hashing for image retrieval via image representation learning [A]. Proceedings of AAAI Conference on Artificial Intelligence [C]. Quebec, Canada, 2014. 2156 – 2162.
- [22] Lai H, Pan Y, Liu Y, et al. Simultaneous feature learning and hash coding with deep neural networks [A]. Proceedings of the Computer Vision and Pattern Recognition [C]. Boston, MA, USA; IEEE, 2015. 3270 – 3278.
- [23] Lin K, Yang H F, Hsiao J H, et al. Deep learning of binary hash codes for fast image retrieval [A]. Proceedings of the Computer Vision and Pattern Recognition Workshops [C]. Boston, MA, USA; IEEE, 2015. 27 – 35.
- [24] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [A]. Proceedings of the Computer Vision and Pattern Recognition [C]. Boston, MA, USA; IEEE, 2015. 1 – 9.
- [25] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [A]. Proceedings of the Computer Vision and Pattern Recognition [C]. Las Vegas, Nevada, USA; IEEE, 2016. 770 – 778.
- [26] Chua T S, Tang J, Hong R, et al. NUS-WIDE: a real-world web image database from National University of Singapore

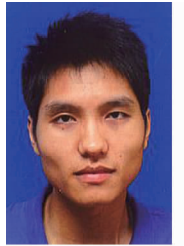
[A]. ACM International Conference on Image and Video

Retrieval[C]. SantorinIsland, Greece; ACM, 2009. 48.

作者简介



江泽涛 男,1961 年生于江西九江,教授,博士,研究方向为计算机视觉与信息安全.
E-mail: zetaojiang@126.com



简 雄(通信作者) 男,1990 年生于湖南邵阳县,硕士研究生,研究方向为计算机视觉.
E-mail: 799586040@qq.com