基于深度卷积神经网络和二进制哈希学习的图像检索方法

彭天强*① 栗 芳②
①(河南工程学院计算机学院 郑州 451191)
②(河南图像识别工程技术中心 郑州 450001)

摘 要：随着图像数据的迅猛增长，当前主流的图像检索方法采用的视觉特征编码步骤固定，缺少学习能力，导致其图像表达能力不强，而且视觉特征维数较高，严重制约了其图像检索性能。针对这些问题，该文提出一种基于深度卷积神径网络学习二进制哈希编码的方法，用于大规模的图像检索。该文的基本思想是在深度学习框架中增加一个哈希层，同时学习图像特征和哈希函数，且哈希函数满足独立性和量化误差最小的约束。首先，利用卷积神经网络强大的学习能力挖掘训练图像的内在隐含关系，提取图像深层特征，增强图像特征的区分性和表达能力。然后，将图像特征输入到哈希层，学习哈希函数使得哈希层输出的二进制哈希码分类误差和量化误差最小，且满足独立性约束。最后，给定输入图像通过该框架的哈希层得到相应的哈希码，从而可以在低维汉明空间中完成对大规模图像数据的有效检索。在 3 个常用数据集上的实验结果表明，利用所提方法得到哈希码，其图像检索性能优于当前主流方法。

关键词：图像检索；深度卷积神径网络；二进制哈希；量化误差；独立性
中图分类号：TP391.4 **文献标识码**：A **文章编号**：1009-5896(2016)08-2068-08
DOI: 10.11999/JEIT151346

Image Retrieval Based on Deep Convolutional Neural Networks and Binary Hashing Learning

PENG Tianqiang① LI Fang②
①(*Department of Computer Science and Engineering*, *Henan Institute of Engineering*, *Zhengzhou* 451191, *China*)
②(*Henan Image Recognition Engineering Center*, *Zhengzhou* 450001, *China*)

Abstract: With the increasing amount of image data, the image retrieval methods have several drawbacks, such as the low expression ability of visual feature, high dimension of feature, low precision of image retrieval and so on. To solve these problems, a learning method of binary hashing based on deep convolutional neural networks is proposed, which can be used for large-scale image retrieval. The basic idea is to add a hash layer into the deep learning framework and to learn simultaneously image features and hash functions should satisfy independence and quantization error minimized. First, convolutional neural network is employed to learn the intrinsic implications of training images so as to improve the distinguish ability and expression ability of visual feature. Second, the visual feature is putted into the hash layer, in which hash functions are learned. And the learned hash functions should satisfy the classification error and quantization error minimized and the independence constraint. Finally, an input image is given, hash codes are generated by the output of the hash layer of the proposed framework and large scale image retrieval can be accomplished in low-dimensional hamming space. Experimental results on the three benchmark datasets show that the binary hash codes generated by the proposed method has superior performance gains over other state-of-the-art hashing methods.

Key words: Image retrieval; Deep convolutional neural networks; Binary hashing; Quantization error; Independence

1 引言

随着大数据时代的到来，互联网图像资源迅猛增长，如何对大规模图像资源进行快速有效的检索以满足用户需求亟待解决。图像检索技术由早期的基于文本的图像检索(Text-Based Image Retrieval, TBIR) 逐渐发展为基于内容的图像检索(Content-Based Image Retrieval, CBIR), CBIR 通过提取图像视觉底层特征来实现图像内容表达。视觉底层特征包括基于梯度的图像局部特征描述子，

收稿日期：2015-12-01；改回日期：2016-04-29；网络出版：2016-06-24
*通信作者：彭天强 ptq_drumboy@163.com
基金项目：国家自然科学基金(61301232)
Foundation Item: The National Natural Science Foundation of China (61301232)

如 SIFT[1] (Scale-Invariant Feature Transform), HOG[2] (Histogram of Orientated Gradients)等。与人工设计的特征相比，深度卷积神经网络(Convolutional Neural Networks, CNNs)更能够获得图像的内在特征，且在目标检测、图像分类和图像分割等方面都表现出了良好的性能。利用深度 CNNs 学习图像特征，文献[3]首先提出了一个提取图像特征的框架，且在 ImageNet 数据集上取得了不错的效果。

针对大规模数据的检索问题，哈希技术被广泛用于计算机视觉、机器学习、信息检索等相关领域。为了在大规模图像集中进行快速有效的检索，哈希技术将图像的高维特征保持相似性地映射为紧致的二进制哈希码。由于二进制哈希码在汉明距离计算上的高效性和存储空间上的优势，哈希码在大规模图像检索中非常高效。

位置敏感哈希[4](Locality Sensitive Hashing, LSH)按照其应用可以分为两类[5]：一类是以一个有效的方式对原始数据进行排序，以加快搜索速度，这种类型的哈希算法称为“original LSH”；另一类是将高维数据嵌入到 Hamming 空间中，并进行按位操作以找到相似的对象，将这种类型的哈希算法称为二进制哈希(binary hashing)。二进制哈希方法可以分为无监督的哈希算法、半监督的哈希算法和监督的哈希算法。无监督的哈希方法不考虑数据的监督信息，包括 Isotropic hashing[6]、谱哈希[7] (SH)、PCA-ITQ[8]等；半监督的哈希方法考虑部分的相似性信息，包括 SSH[9]；监督的哈希方法利用数据集的标签信息或者相似性点对信息作为监督信息，包括 BRE[10]、监督的核哈希[11] (KSH)等。这些哈希算法的目标均是构造出能够保持数据在原空间中的相似性且能够生成紧致二进制哈希码的哈希函数。在谱哈希[7]中给出了度量哈希函数好坏的 3 个标准：(1)将原始数据空间中相似的对象映射为相似的二进制编码；(2)需要较少的位数来对整个数据集进行编码；(3)给定一个新的输入易求出相应的二进制编码。其中第(2)个标准的目标要求生成紧致的二进制码，即不同哈希函数之间应该是独立的。在 PCA-ITQ[8]中在哈希函数构造利用量化误差最小作为优化目标，最后生成了表达能力很强的二进制哈希码。

基于深度学习的方法[12–14]在图像分类、目标检测等方面都展现了其优越性。从 2012 年文献[13]提出的 AlexNet 模型到 2014 年文献[15]提出的 NIN(Network In Network)模型和文献[16]提出的深层 VGG 模型都成功地验证了基于深度卷积神经网络的方法在学习图像特征表示上的能力。

由于深度卷积神经网络在特征学习上的优越性以及哈希方法在检索中计算速度和存储空间上的优越性，近几年也出现了深度卷积神经网络与哈希技术相结合的方法。文献[17]提出了一种 CNNs 与哈希方法相结合的算法，该算法分为两个步骤，第 1 步首先利用数据的相似性信息构建相似性矩阵，然后得到训练样本的近似哈希编码；第 2 步将第 1 步学习得到的哈希码作为目标利用深度卷积网络框架学习哈希构造函数，该论文将哈希编码的学习和特征的提取分为两个阶段，效果不够好。文献[18]提出了一种利用深度卷积网络同时学习特征和哈希函数的算法，它利用图像三元组作为监督信息，优化目标函数是在最终的变换空间中相似的图像对之间距离比不相似图像对的距离近，且有一定的间隔；该论文将三元组作为监督信息，三元组对的挑选质量直接影响着检索的精度且三元组的挑选需要较大的工作量。文献[19]也提出了一种利用深度卷积网络框架同时学习特征和哈希函数的算法，该论文中采用标签信息作为监督信息，避免了需要挑选三元组的工作量，但是它没有考虑到将连续值阈值化为二进制码时产生的量化误差以及哈希函数之间的独立性。

结合深度卷积神经网络和哈希算法的优势，本文提出了一种基于深度卷积神经网络学习二进制哈希函数的编码方法，学习得到的二进制哈希码可用于大规模的图像检索。本文的基本思想是在 CNNs 框架中引入哈希层，利用图像标签信息同时学习图像特征和哈希函数，且哈希函数需要满足独立性和量化误差最小的约束。本文提出的二进制哈希函数学习算法，考虑哈希函数之间的独立性和阈值化产生的量化误差，与其他相关的方法相比，本文有以下特点:

(1)在原有的 CNNs 框架中，引入哈希层，将哈希层得到的编码输入分类器进行分类，将 Softmax 分类损失作为优化目标之一。

(2)哈希层中包括两部分，第 1 部分包括分片层(slice layer)、全连接层、激活层以及合并层(concat layer)，将特征映射为连续的编码，用于生成具有独立性的哈希函数；第 2 部分是阈值化层，将连续编码二值化，得到二值哈希码，用于计算量化误差。

(3)在整个框架模型中考虑量化误差的影响，将连续值阈值化为二进制哈希码时产生的误差加入到优化目标中，从而得到表达能力更强的哈希码。

实验结果表明，本文提出的二进制哈希学习方法的检索性能优于现有的方法。

2　本文方法

本文方法的框架图如图 1 所示。该模型接受的输入为图像及其相应的标签信息。该模型主要包括

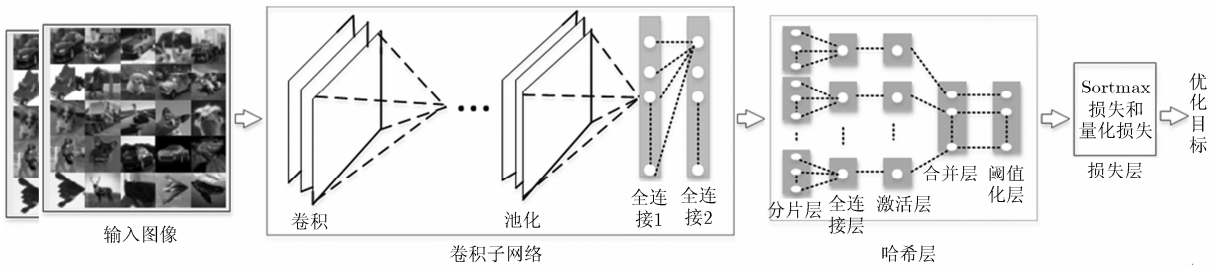


图 1 本文卷积神经网络的框架图

3 个部分：(1) 卷积子网络，用于学习图像的特征表示；(2) 哈希层，用于构建独立的哈希函数得到相应的哈希码；(3) 损失层，包括 Softmax 分类损失和量化误差损失。首先，输入图像通过卷积子网络层得到图像的特征表示；其次，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，并优化该损失函数学习得到模型参数。

2.1 卷积子网络

卷积子网络层用于学习图像的特征表示，输入图像通过该卷积子网络层可以得到图像的特征表示。

本文采用具有深度为 16 的 VGG[16]模型结构作为基本架构。卷积子网络中包含 5 个大卷积层、5 个池化层和两个全连接层。前 2 个大卷积层中分别包含了两个核大小为 3×3，步幅为 1 的卷积层；后 3 个大卷积层分别包含了两个核大小为 3×3，步幅为 1 的卷积层和一个核大小为 1×1，步幅为 1 的卷积层。在使用该卷积子网络模型时需要根据图像大小，调整相应的卷积层的输出个数。对于图像大小为 32×32 的小图，本文的卷积子网络配置见表 1。

表1 小图的卷积子网络配置

类型	核大小/步长	输出大小
卷积 1-1	3×3/1	32×32×32
卷积 1-2	3×3/1	32×32×32
池化 1	2×2/1	32×16×16
卷积 2-1	3×3/1	64×16×16
卷积 2-1	3×3/1	64×16×16
池化 2	2×2/1	64×8×8
卷积 3-1	3×3/1	96×8×8
卷积 3-2	3×3/1	96×8×8
卷积 3-3	1×1/1	192×8×8
池化 3	2×2/1	192×4×4
卷积 4-1	3×3/1	128×4×4
卷积 4-2	3×3/1	128×4×4
卷积 4-3	1×1/1	256×4×4
池化 4	2×2/1	256×2×2
卷积 5-1	3×3/1	160×2×2
卷积 5-2	3×3/1	160×2×2
卷积 5-3	1×1/1	320×2×2
池化 5	2×2/1	320×1×1

2.2 哈希层和优化目标

LSH[20]中给出了保持内积相似性的哈希函数的定义：给定特征 $\boldsymbol{x}\in R^m$，构造 q 个 m 维随机向量构成矩阵 $\boldsymbol{W}\in R^{q\times m}$，则 q 个哈希函数产生的哈希码为 $(h_1,h_2,\cdots,h_q)^{\mathrm{T}}=(\mathrm{sign}(\boldsymbol{W}\boldsymbol{x}))^{\mathrm{T}}$。

在本文中，哈希层是由分片层、各子块的全连接层、各子块的激活层、合并层和阈值化层组成。其中，分片层、各子块的全连接层、各子块的激活层和合并层用于构造相互独立的哈希函数；阈值化层将连续值编码二值化，用于计算量化误差。

从卷积子网络的第 2 个全连接层得到图像特征 $\boldsymbol{x}$ 之后，将它传入哈希层。首先，进入哈希层的分片层，对图像特征 $\boldsymbol{x}$ 进行分片，假设图像特征 $\boldsymbol{x}$ 的维数为 m，需要生成哈希码的长度为 q，则需要将图像特征分为 q 片，记为 $\boldsymbol{x}^{(i)}(i=1,2,\cdots,q)$，每一片中包含的特征维数为 m/q(这里最好 m 为 q 的倍数，可以通过控制卷积子网络第 2 个全连接层的输出单元数确定图像特征的维数)。

从分片层得到的 q 个子特征 $\boldsymbol{x}^{(i)}(i=1,2,\cdots,q)$，分别进入全连接层，且每个全连接层的输出均为 1 维的，表示为

$$f_i(\boldsymbol{x}^{(i)})=\boldsymbol{W}_i\boldsymbol{x}^{(i)},\quad i=1,2,\cdots,q \tag{1}$$

其中，$\boldsymbol{W}_i\in R^{\dim(\boldsymbol{x}^{(i)})\times 1}$ 为第 i 个全连接层的权重矩阵。

每个子块分别进入激活层，激活层使用双正切激活函数将每个子块输出的 1 维数值映射为值域在 $[-1,1]$ 之间的数值，表示为

$$\tanh\left(v^{(i)}\right)=\frac{1-e^{\beta v^{(i)}}}{1+e^{\beta v^{(i)}}},\quad i=1,2,\cdots,q \tag{2}$$

其中 $v^{(i)}=f_i(\boldsymbol{x}^{(i)})$，参数 β 用于控制平滑度。本文方法用双正切激活函数近似代替符号函数，使用分片层且分别为每个子块分配随机权重矩阵 $\boldsymbol{W}_i$ 使得每位哈希码仅与特征的部分是相关的，从而达到哈希

函数构造的独立性。

然后进入合并层，合并层主要是将 q 个子块的 1 维输出合并为一个 q 维向量，表示为

$$\boldsymbol{s}=\left(v^{(1)},v^{(2)},\cdots,v^{(q)}\right)^{\mathrm{T}} \tag{3}$$

合并层的输出即为哈希函数输出值的近似值，为连续的编码值。

最后进入阈值化层，阈值化层主要是将合并层得到值域在[−1,1]之间的 q 维连续值编码进行量化，量化为 −1 和 1，表示为

$$g(s^{(i)})=\begin{cases}1, & s^{(i)}\geq 0\\ -1, & s^{(i)}<0\end{cases} \tag{4}$$

其中，$s^{(i)}$ 表示合并层的输出 q 维向量 $\boldsymbol{s}$ 的第 i 个分量。阈值化层的输出为二进制哈希码。

本文中深度卷积神经网络框架的优化目标结构图如图 2 所示。损失层函数包括 Softmax 分类器损失和量化误差损失。激活层得到的编码进入 Softmax 分类器进行分类，在这个过程中产生 Softmax 分类误差损失，记为 L_{sl}。另一方面，考虑到哈希码为离散值，需要加入将连续值二值化为离散值时带来的误差，在目标损失函数中，加入合并层输出的连续值编码与阈值化层输出的哈希码之间的误差损失，即量化误差损失，表示为

$$L_{\mathrm{q}}=\frac{1}{2}\|\boldsymbol{h}-\boldsymbol{s}\|_2^2 \tag{5}$$

其中 $\boldsymbol{h}=(g(s^{(1)}),g(s^{(2)}),\cdots,g(s^{(q)}))^{\mathrm{T}}$ 为阈值化层输出的哈希码，$\boldsymbol{s}$ 为激活层输出的连续值编码。该损失函数的目标是通过学习使得激活层的输出值尽可能地接近量化值 −1 和 1，从而降低阈值化带来的误差。

结合 Softmax 分类器的损失函数和量化误差损失，得到该框架的整体损失函数：

$$L_{\mathrm{T}}=L_{\mathrm{sl}}+\lambda L_{\mathrm{q}} \tag{6}$$

其中，λ 为权重因子，决定着量化损失所占的重要性。

2.3 哈希码的生成

在利用本文卷积神经网络的框架训练之后，给定一张图像作为输入，通过该网络框架可以得到 q 位二进制哈希码。生成流程如图 3 所示，给定输入图像，首先经过卷积子网络层，然后经过哈希层，哈希层中的最后一层为阈值层，直接输出了二进制哈希码。

3　实验设置与性能评价

3.1 实验设置

为验证本文方法的有效性，在以下 3 个图像集上对本文方法进行了评估。MNIST 数据集[21]，该数据集是包括了 70000 张 28×28 的灰度图像，手写数字从 0 到 9 共 10 个类别。CIFAR-10 数据集[22]，包括了 60000 张 32×32 的彩色图像，其类别包括飞机、卡车等 10 类。NUS-WIDE 数据集[23]，包括了将近 270000 张图像，每张图像具有一个或者多个标签。借鉴文献[24]的使用方式，仅使用 21 个常用类，常用类中每一类中至少包括 5000 张图像。另外在训练时我们统一将图像大小重设置为 256×256。

将本文的方法的检索性能与其它的哈希方法做比较，包括非监督的哈希方法 ITQ[8]，监督的哈希方法 KSH[11]，以及深度学习与哈希技术相结合的哈希方法 CNNH[17]，改进 CNNH[18], DCNNH[19]。

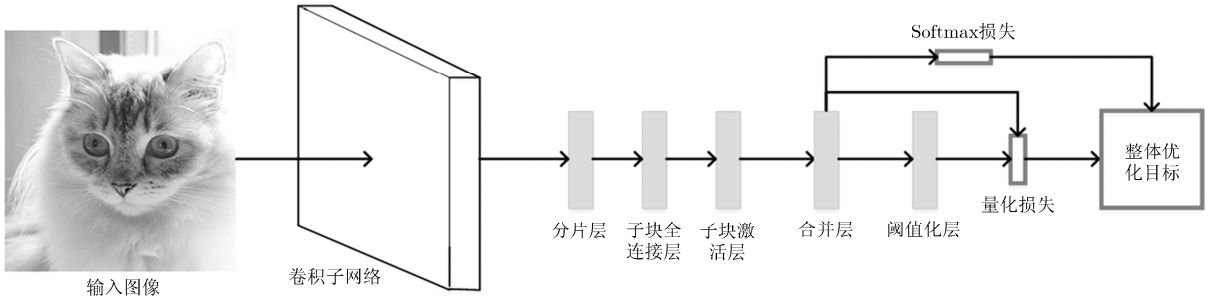


图 2 本文卷积神经网络的优化目标结构图

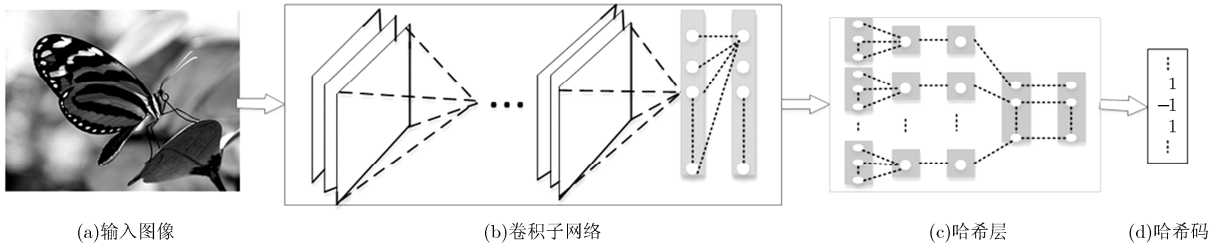


图 3 哈希码生成流程图

在 MNIST 和 CIFAR-10 数据集中，每一类选择 1000 张图像构成包含 10000 图像的测试集。对于无监督的哈希方法，其余的数据作为训练集。对于监督的哈希方法，每类选择出 500 张，组成包括 5000 张图像的训练集。在 NUS-WIDE 数据集中，随机从每一类中选择出 100 张图像组成 2100 张图像的测试图像集。对于无监督的哈希方法，其余的数据作为训练集。对于有监督的哈希方法，每类选择出 500 张，组成包括 10500 张图像的训练集。

对于深度学习与哈希计算相结合的算法，直接使用图像作为输入。而其余的方法，数据集 MNIST 和 CIFAR-10 采用 512 维的 Gist 特征表示图像；NUS-WIDE 图像使用 500 维的 bag-of-words 向量表示图像。

为了评估图像检索性能并与已有方法作比较，本文采用 MAP、查准-查全率(Precision-Recall, P-R)曲线、汉明距离小于 2 的准确率曲线以及检索返回 top-k 近邻域的准确率曲线这 4 个参数进行评估。其中，查准率是指查询结果中正确结果所占的比例，查全率是指查询结果中正确结果占全部正确结果的比例；P-R 曲线是指按照汉明距离从小到大的排序，所有测试图像的平均查全率和平均查准率的曲线图。MAP 是指 P-R 曲线所包围的面积。汉明距离小于 2 的准确率是指在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。top-k 近邻域的准确率是指与查询图像距离最小的 k 张图像中正确结果所占的比例。

本文的训练过程基于开源 Caffe 实现的。在所有实验中，量化损失的权重因子 λ 取值为 0.2。

3.2 实验性能分析

表 2 中给出在 MNIST 数据集上，本文方法和已有方法 MAP 值的比较结果。从表 2 中可以看出，本文算法的 MAP 值远远高于传统特征与哈希方法相结合的算法(KSH, ITQ)，因为本文利用深度卷积网络同时学习特征表示和哈希函数，大大提高了图像的表示能力。与其它的深度卷积网络与哈希技术相结合的方法相比，本文算法的 MAP 值最高，与 CNNH 算法和改进 CNNH 算法相比，本文算法采用了标签信息作为监督信息且考虑了量化误差，得到了表示能力更强的哈希码；与 DCNNH 算法相比，本文算法架构中考虑了量化误差和哈希函数之间的独立性，得到了更具有图像表示能力的哈希码，使得它在检索中 MAP 值较高。

表 3 给出在 CIFAR-10 数据集上，本文方法和已有方法 MAP 值的比较结果。从表 3 中可以看出，本文算法的 MAP 值远远高于传统特征与哈希方法相结合的算法(如 KSH)，提高了 50%；与现有的深度卷积网络与哈希技术相结合的方法相比，由于本文算法同时考虑了量化误差和哈希函数之间的独立性，且采用了标签信息作为监督信息，本文算法的 MAP 值最高。特别地，比改进 CNNH 算法的 MAP 值提高了 27%左右。

表2 在数据集MNIST上按汉明距离排序的MAP值对比

方法	12 位	24 位	32 位	48 位
本文算法	0.9941	0.9956	0.9958	0.9963
DCNNH	0.9899	0.9939	0.9938	0.9953
改进 CNNH	0.9840	0.9930	0.9920	0.9940
CNNH	0.9690	0.9750	0.9710	0.9750
KSH	0.8720	0.8910	0.8970	0.9000
ITQ	0.3380	0.4360	0.4220	0.4290

表3 在数据集CIFAR-10上按汉明距离排序的MAP值对比

方法	12 位	24 位	32 位	48 位
本文算法	0.838	0.845	0.851	0.855
DCNNH	0.816	0.826	0.829	0.832
改进 CNNH	0.552	0.566	0.558	0.581
CNNH	0.465	0.521	0.521	0.532
KSH	0.303	0.337	0.346	0.356
ITQ	0.162	0.169	0.172	0.175

表 4 给出在 NUS-WIDE 数据集上，本文方法和已有方法 MAP 值的比较结果。从表 4 中可以看出。本文算法的 MAP 值比传统特征+KSH 的 MAP 值提高了将近 20%；与现有的深度卷积网络与哈希技术相结合的方法相比，本文算法比改进 CNNH 算法的 MAP 值高了 6%左右，比 DCNNH 算法的 MAP 值提高了 2%左右，主要是因为本文算法同时加入了量化误差和哈希函数的独立性的约束，得到了表示能力更强的哈希码。

图 4~图 6 给出了在 3 个数据集上，在其它检索性能(不同位数下汉明距离小于 2 的正确率、P-R 曲线、不同位数下 top-k 的检索正确率)上的比较结

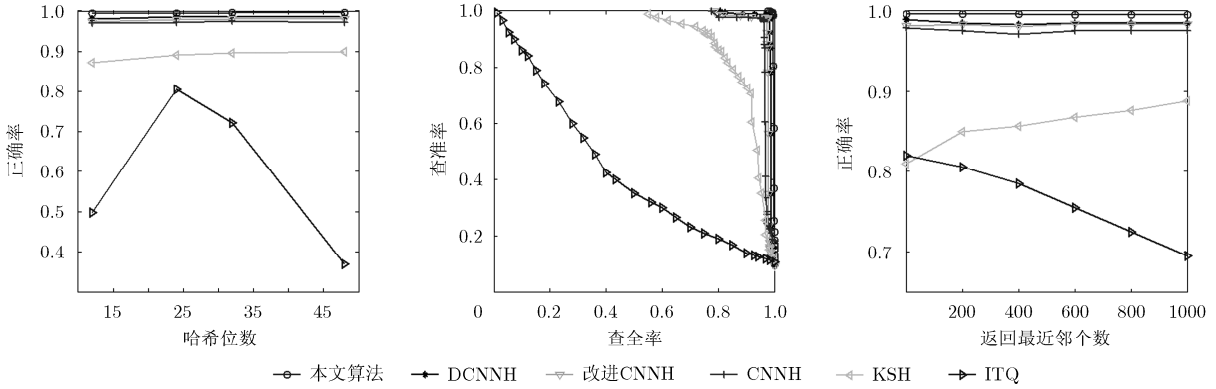
表4 在数据集NUS-WIDE上top-5000近邻域的MAP值对比

方法	12 位	24 位	32 位	48 位
本文算法	0.761	0.767	0.771	0.773
DCNNH	0.741	0.748	0.752	0.756
改进 CNNH	0.674	0.697	0.713	0.715
CNNH	0.623	0.630	0.629	0.625
KSH	0.556	0.572	0.581	0.588
ITQ	0.452	0.468	0.472	0.477

果。从这 3 个图中可以看出，本文的方法的检索性能均优于现有的其它方法。

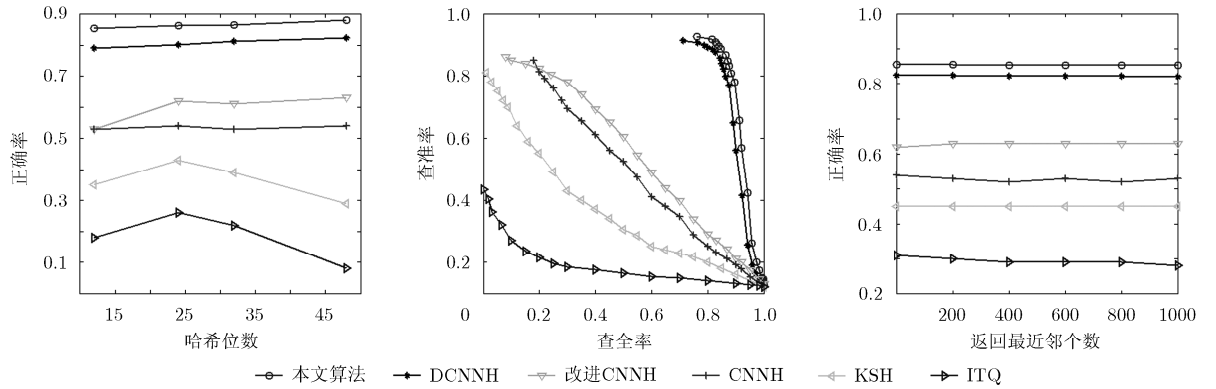
3.3 加入独立性和量化损失的性能对比

为了验证本文提出的框架的有效性，将本文算法与未加入分片层、阈值层和量化损失的算法(即未做任何约束，不考虑哈希函数间的独立性和量化误差)、以及在本算法的基础上未加入阈值层和量化损失的算法(即仅考虑哈希函数间的独立性，不考虑量化误差)分别做比较。不考虑哈希函数间的独立性和量化误差的算法框架见图 7 所示，在给定图像生成哈希码时，图像经过该框架仅得到了值域在[–1,1]之间的编码，需要对得到的编码进行二值化生成二进制哈希码，该算法框架类似于文献[19]提出的算法。仅包含独立性不包含量化损失的算法框架见图 8 所示，在该框架中也需要对该框架的输出编码也需要进行二值化生成二进制哈希码，该算法框架类似于改进 CNNH[18]的框架，但在改进 CNNH[18]中采用图像三元组损失函数，且不考虑量化损失。



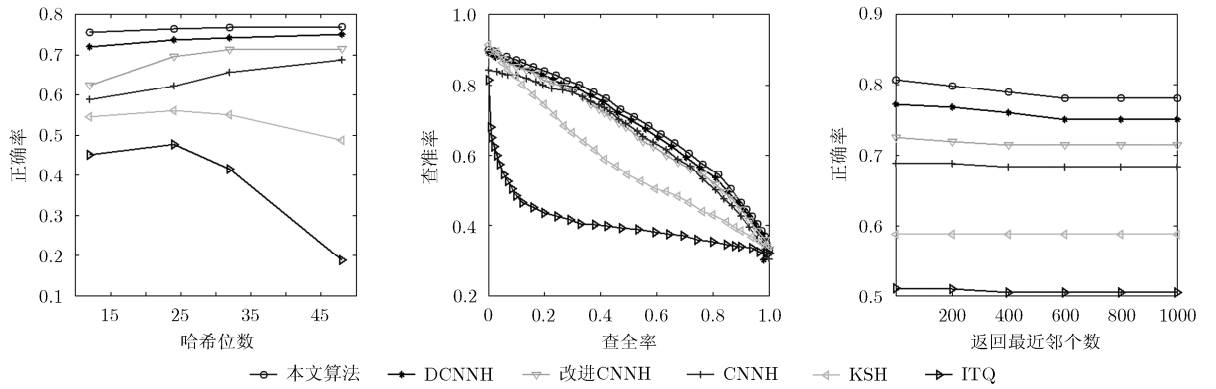
(a)汉明距离小于2的准确率曲线　(b) 48位按照汉明距离排序的P-R曲线　(c) 48位检索返回top-k近邻域的准确率曲线

图 4 在数据集 MNIST 上结果对比



(a)汉明距离小于2的准确率曲线　(b) 48位按照汉明距离排序的P-R曲线　(c) 48位检索返回top-k近邻域的准确率曲线

图 5 在数据集 CIFAR10 上结果对比



(a)汉明距离小于2的准确率曲线　(b) 48位按照汉明距离排序的P-R曲线　(c) 48位检索返回top-k近邻域的准确率曲线

图 6 在数据集 NUS-WIDE 上结果对比

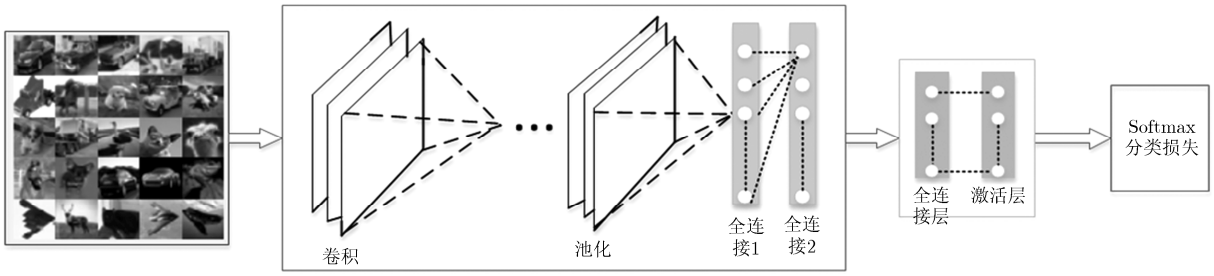


图 7 未加入分片层、阈值层和量化损失的算法框架

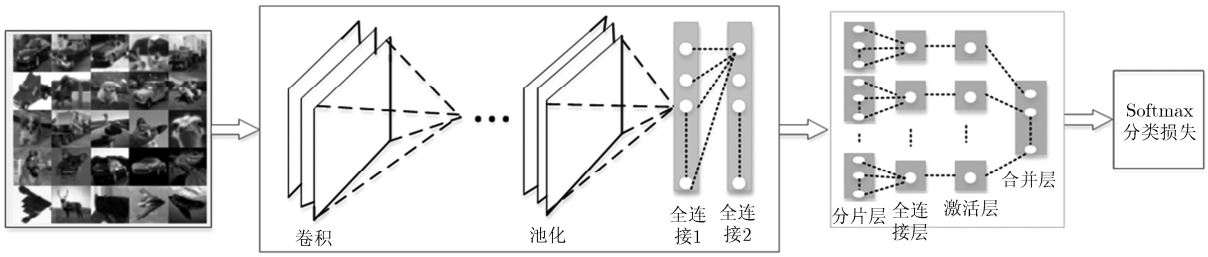


图 8 仅增加哈希函数间的独立性，未加入量化损失的算法框架

表 5~表 7 给出了 3 种算法在 3 个数据集上的 MAP 对比结果。从这 3 个表可以看出：仅加入独立性的算法比未做任何约束的算法的检索 MAP 值提高了 1%~2%；而本文算法包括独立性和量化损失约束，比仅有独立性约束的算法的 MAP 值又提高了 1%左右。在数据集 MNIST 上，虽然本文算法仅比未阈值化算法的 MAP 提高 0.5%左右，但本文算法 24 位的哈希码的检索 MAP 值高于未做任何约束算法的 48 位的哈希码检索 MAP 值，从而在大规模图像检索中可以用更短的哈希码来表示图像但能达到与较长哈希码相当的检索精度。在数据集 CIFAR-10 上，本文算法 24 位的哈希码的检索精度已经超过了未做任何约束算法的 48 位哈希码；在数据集 NUS-WIDE 上，本文算法 12 位的哈希码的检索精度也高于了未做任何约束算法的 48 位哈希码。从以上对比中可以看出，利用本文算法可以用较短的哈希码表示图像，且达到其他算法用较长的哈希码的检索精度。用较短的哈希码表示图像，使得在大规模图像检索中图像集占用的存储空间更少，距离计算速度更快，提高了图像检索在时间、空间上的性能，但同时保持了相应的检索精度。

表5 数据集MNIST上按汉明距离排序的MAP值对比

方法	12 位	24 位	32 位	48 位
未做任何约束	0.9901	0.9938	0.9940	0.9954
独立性	0.9936	0.9945	0.9954	0.9959
独立性+量化损失	0.9941	0.9956	0.9958	0.9963

表6 数据集CIFAR-10上按汉明距离排序的MAP值对比

方法	12 位	24 位	32 位	48 位
未做任何约束	0.8168	0.8266	0.8291	0.8328
独立性	0.8251	0.8308	0.8454	0.8488
独立性+量化损失	0.8385	0.8450	0.8507	0.8556

表7 数据集NUS-WIDE上top-5000近邻域的MAP值对比

方法	12 位	24 位	32 位	48 位
未做任何约束	0.7413	0.7482	0.7528	0.7560
独立性	0.7540	0.7603	0.7644	0.7689
独立性+量化损失	0.7608	0.7668	0.7710	0.7727

4 结束语

本文提出了一种基于深度卷积神经网络学习二进制哈希的方法，适用于大规模图像检索。在本文提出的框架中，采用类别信息作为监督信息，而不使用三元图像组作为监督信息，大大降低了人工标记量。另外，在整个框架模型中加入哈希函数独立性的限制，并考虑量化误差的影响，将连续值阈值化为哈希码时产生的误差加入到损失函数中，从而构造出了更好的哈希函数，得到了更具有图像表达能力的哈希码。与其他现有方法相比，本文算法的检索精度最优。

参 考 文 献

[1] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.

[2] DALAL N and TRIGGS B. Histograms of oriented gradients for human detection[C]. Computer Vision and Pattern Recognition, San Diego, CA, USA, 2005: 886–893.

[3] KRIZHEVSKY A, SUTSKEVER I, and HINTON G E. ImageNet classification with deep convolutional neural networks[C]. Advances in Neural Information Processing Systems, South Lake Tahoe, Nevada, US, 2012: 1097–1105.

[4] DATAR M, IMMORLICA N, INDYK P, *et al*. Locality sensitive hashing scheme based on p-stable distributions[C]. Proceedings of the ACM Symposium on Computational Geometry, New York, USA, 2004: 253–262.

[5] ZHANG Lei, ZHANG Yongdong, ZHANG Dongming, *et al*. Distribution-aware locality sensitive hashing[C]. 19th International Conference on Multimedia Modeling, Huangshan, China, 2013: 395–406.

[6] KONG Weihao and LI Wujun. Isotropic hashing[C]. Advances in Neural Information Processing Systems, South Lake Tahoe, Nevada, US, 2012: 1646–1654.

[7] WEISS Y, TORRALBA A, and FERGUS R. Spectral hashing[C]. Advances in Neural Information Processing Systems, Vancouver, Canada, 2009: 1753–1760.

[8] GONG Yunchao, 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*, 2012, 35(12): 2916–2929.

[9] WANG Jun, KUMAR S, and CHANG Shihfu. Semi-Supervised hashing for large scale search[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(12): 2393–2406.

[10] KULIS B and DARRELL T. Learning to hash with binary reconstructive embeddings[C]. Advances in Neural Information Processing Systems, Vancouver, Canada, 2009: 1042–1052.

[11] LIU Wei, WANG Jun, JI Rongrong, *et al*. Supervised hashing with kernels[C]. Computer Vision and Pattern Recognition, Providence, RI, 2012: 2074–2081.

[12] GIRSHICK R, DONAHUE J, DARRELL T, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Computer Vision and Pattern Recognition, Ohio, Columbus, 2014: 580–587.

[13] OQUAB M, BOTTOU L, LAPTEV I, *et al*. Learning and transferring mid-level image representations using convolutional neural networks[C]. Computer Vision and Pattern Recognition, Ohio, Columbus, 2014: 1717–1724.

[14] RAZAVIAN A, AZIZPOUR H, SULLIVAN J, *et al*. CNN features off-the-shelf: an astounding baseline for recognition[C]. Computer Vision and Pattern Recognition, Ohio, Columbus, 2014: 806–813.

[15] LIN Min, CHEN Qiang, and YAN Shuicheng. Network in network[OL]. http://arxiv.org/abs/1312.4400, 2013.

[16] SIMONYAN K and ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [OL]. http://arxiv.org/abs/1409.1556, 2014.

[17] XIA Rongkai, PAN Yan, LAI Hanjiang, *et al*. Supervised hashing for image retrieval via image representation learning[C]. Proceedings of the AAAI Conference on Artificial Intelligence, Québec, Canada, 2014: 2156–2162.

[18] LAI Hanjiang, PAN Yan, LIU Ye, *et al*. Simultaneous feature learning and hash coding with deep neural networks[C]. Computer Vision and Pattern Recognition, Boston, MA, USA, 2015: 3270–3278.

[19] LIN K, YANG H F, HSIAO J H, *et al*. Deep learning of binary hash codes for fast image retrieval[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 2015: 27–35.

[20] GIONIS A, INDYK P, and MOTWANI R. Similarity search in high dimensions via hashing[C]. Proceedings of the International Conference on Very Large Data Bases, Edinburgh, Scotland, UK, 1999: 518–529.

[21] LECUN Y, CORTES C, and BURGES CJC. The MNIST database of handwritten digits[OL]. http://yann.lecun.com/exdb/mnist, 2012.

[22] KRIZHEVSKY A and HINTON G. Learning multiple layers of features from tiny images[R]. Technical Report, University of Toronto, 2009.

[23] CHUA TatSeng, TANG Jinhui, HONG Richang, *et al*. NUS-WIDE: A real-world Web image database from national university of singapore[C]. Proceedings of the ACM International Conference on Image and Video Retrieval, Greece, 2009: 48.

[24] LIU Wei, WANG Jun, Kumar Sanjiv, *et al*. Hashing with graphs[C]. Proceedings of the 28th International Conference on Machine Learning, Bellevue, Washington, USA, 2011: 1–8.

彭天强：　男，1978 年生，博士，副教授，主要研究方向为多媒体信息处理及模式识别.

栗　芳：　女，1986 年生，硕士，研究方向为图像检索与分类.