基于离散优化的哈希编码学习方法

刘昊淼 王瑞平 山世光 陈熙霖

(中国科学院计算技术研究所智能信息处理重点实验室 北京 100190)

(中国科学院大学计算机科学与技术学院 北京 100049)

摘 要 哈希作为近似近邻搜索的一种主流方法,通过将样本索引为紧致的二值编码,在计算效率和存储上都非常高效.由于二值码的离散特性,以往的哈希方法往往需要将二值码松弛为实数值才能高效地进行优化,因此在优化完成后重新将实数值的结果量化为二值时难免会由于二值的汉明空间与实数值的欧氏空间之间的差异而遇到性能上的损失问题.为了更好地解决量化损失的问题,本文提出了一种深度离散优化哈希(Deep Discrete Optimization Hashing,DDOH)方法.首先,设计了一种新的离散优化算法,通过直接在二值的汉明空间中对二值码进行优化,得到具有强判别性的二值编码.然后,训练卷积神经网络模型拟合上述二值码,得到用于编码的哈希函数.在 CIFAR-10 和 ImageNet-100 两个常用的评测数据集上的实验显示,本文提出的方法在 CIFAR-10 数据库上与目前最好的方法达到了同样的性能,在 ImageNet-100 数据库上的平均准确率指标与已有方法相比提升了约 2.2%,证明了该方法的有效性.

关键词 近似近邻搜索;高维特征索引;哈希学习;离散优化;卷积神经网络

中图法分类号 TP311 **DOI 号** 10.11897/SP.J.1016.2019.01149

Learning to Hash with Discrete Optimization

LIU Hao-Miao WANG Rui-Ping SHAN Shi-Guang Xilin Chen

(*Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing* 100190)

(*School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing* 100049)

Abstract In recent years, billions of images are uploaded to the Internet every day, making it extremely difficult to find an interested image according to a user's demand. This paper addresses the content-based image retrieval task, which aims at looking for database images that are similar to the given query image. However, due to the huge size of modern datasets, exact nearest neighbor search method cannot produce retrieval results in acceptable time. Therefore, approximate nearest neighbor search methods are proposed to sacrifice accuracy for acceptable retrieval time. As a mainstream approximate nearest neighbor (ANN) search method, hashing projects the original feature vectors of samples into very compact binary codes, and thus is very efficient in both computation and storage. As a result, hashing methods have received more and more research attention over the past twenty years. However, due to the discrete nature of binary codes, directly optimizing the binary codes is an NP-hard problem and the computation time required for obtaining the global optimum would be unacceptable. To deal with this problem, existing hashing methods can only perform optimization efficiently by relaxing the binary codes to real values, and

收稿日期:2019-01-29;在线出版日期:2019-03-27.本课题得到国家"九七三"重点基础研究发展规划项目基金(2015CB351802)、国家自然科学基金(61772500)、中国科学院前沿科学重点研究项目(QYZDJ-SSW-JSC009)、中国科学院青年创新促进会(2015085)资助.

刘昊淼,博士研究生,主要研究方向为图像检索、可解释的物体识别模型.E-mail:haomiao.liu@vipl.ict.ac.cn.**王瑞平**,博士,研究员,中国计算机学会(CCF)会员,主要研究领域为计算机视觉、模式识别、机器学习.**山世光**,博士,研究员,中国计算机学会(CCF)会员,主要研究领域为计算机视觉、模式识别、机器学习.**陈熙霖**(通信作者),博士,研究员,中国计算机学会(CCF)会士,主要研究领域为计算机视觉、多模式人机接口.E-mail:xlchen@ict.ac.cn.

optimize the real-valued counterpart of the objective function instead. After that, the optimum obtained in the relaxed real-valued space are again quantized to generate the real binary codes. However, there is no guarantee that the real-valued optimum would remain optimum after quantization, and thus existing methods inevitably suffer from performance drop when quantizing the real-valued optimization results into binary codes, due to the discrepancy between the binary Hamming space and the real-valued Euclidean space. To better deal with the problems of quantization, this paper proposes a novel hash learning method, named Deep Discrete Optimization Hashing (DDOH). First of all, the initial binary codes of all training image samples are obtained by one of the three binary code initialization methods proposed in this paper. After that, a discrete binary codes optimization algorithm is designed, which takes the initial binary codes of training images as well as their corresponding label information as inputs. The proposed optimization algorithm iteratively decides whether or not to flip certain binary bits in the binary codes with the Fisher's law, and it is theoretically proved in this paper that by doing so, the proposed method would improve or at least would not decrease the discriminability of the binary codes in terms of the Fisher's law. Next, to obtain the hash functions which would be used to encode new-coming images, a deep convolutional neural network (CNN) is trained to fit the aforementioned binary codes. Specifically, with the optimized binary codes, each bit can be seen as a binary classification problem, and all binary classifiers that share the same feature map of the CNN as training inputs are trained to perform as the hashing functions. Experiments on two widely studied datasets CIFAR-10 and ImageNet show that the proposed method achieves state of the art retrieval performance on CIFAR-10, and improves the performance of existing hashing methods by about 2.2% mean Average Precision (mAP) on ImageNet-100, validating the effectiveness of the proposed method.

Keywords approximate neighbor search; high dimensional feature indexing; hash learning; discrete optimization; convolutional neural network

1 引 言

随着便携式拍照设备的大量普及和社交网络的快速发展,互联网上的图片数量呈现爆炸式增长的趋势.面对海量的图像数据,在有限的时间和计算资源下,根据用户提供的查询图像搜索相似图像也变得极为困难.为了在可接受的时间内返回检索结果,近似近邻搜索算法在大规模以图搜图任务中得到了越来越多的关注.作为一种具有代表性的近似近邻搜索算法,哈希学习方法通过学习一组可以保持原始空间中相似性的哈希函数,将高维的图像样本编码为相对低维的二值编码.由于低维二值编码占用的存储空间非常小,同时可以通过 CPU 中集成的异或、比特计数等指令实现高效地距离计算,因此哈希学习方法逐渐成为了一种主流的近似近邻搜索方法,相应地也出现了越来越多新的哈希学习实现方法.

从本质上来说,哈希学习的目标是获得能够保持原始空间中相似性的二值编码及相应的哈希函数.由于二值码的离散性质,哈希学习算法中无法避免地要涉及到离散优化的问题.但是直接对二值码优化是一个 NP 困难的问题,无法进行高效地精确求解.为了解决这个问题,大多数已有的哈希学习方法[1-3]通常首先将离散取值的二值码松弛为实数值,并对定义在实数值上的近似问题进行优化.在对近似问题完成优化后,再重新将优化得到的实数值量化为二值的编码.但是由于实数值的欧氏空间与二值的汉明空间之间存在本质的差异,上述基于松弛-量化的方法即使可以在实数值空间中得到最优的实数值编码,也无法保证在量化之后得到的二值编码仍然是最优的.为了解决这个问题,本文提出一种新的哈希学习方法——深度离散优化哈希(Deep Discrete Optimization Hashing,DDOH).该方法可以通过直接在二值的汉明空间中进行离散优化来避免松弛-量化过程的缺点.具体来说,本文方法的框

架设计如图 1 所示. 首先通过随机初始化或主成分分析的方式获得样本初始的二值码，之后通过汉明空间中的离散迭代优化增强二值码的判别能力. 最后，通过训练卷积神经网络模型来拟合优化得到的二值码，并将训练得到的神经网络作为编码时使用的哈希函数.

具体来说，在离散二值码优化部分，本文提出的方法基于 Fisher 准则，要求相似的图像具有相似的二值码，不相似图像对应的二值码也尽可能不同. 在优化的时候，通过在离散的汉明空间中计算损失函数的次梯度，并根据相应设计的规则，通过对满足一定条件的比特进行翻转的方式进行优化，以此提升二值编码的判别能力. 在拟合部分，以优化得到的二值编码作为监督，在预训练的深度卷积神经网络的基础上，使用交叉熵损失对卷积神经网络模型进行参数微调，得到相应的模型用来对图像进行编码. 当模型训练完成后，以图像作为卷积神经网络模型的输入，对模型的输出进行量化即可得到相应图像的二值编码. 在此基础上，可以通过汉明距离排序、哈希表查表等方式进行快速的近似近邻搜索. 为了验证本文方法的有效性，在 CIFAR-10 和 ImageNet-100 两个常用的图像检索评测数据库上进行了实验，在平均检索精度(mean Average Precision，mAP)指标上达到了与已有方法同样优秀或者超过已有方法的性能，特别是在 ImageNet-100 数据库上相比于已有方法提升了约 2.2 个百分点，证明了本文提出方法的有效性.

本文第 2 节对于已有的哈希学习方法进行综述，并讨论本文方法与已有方法的区别；第 3 节对本文方法中的各个环节进行详细的介绍；第 4 节在通用的大规模图像检索评测数据集上进行大量的实验，验证本文方法的有效性；第 5 节对本文工作进行总结和展望.

2　相关工作

近邻搜索任务的目标是在给定一个查询样本的条件下，对数据库进行搜索，并返回数据库中和查询样本相似的样本. 当数据库的规模非常大或者计算样本间的相似度非常耗时的情况下，精确的近邻搜索所需的计算代价也将增长到难以接受的程度. 因此，作为一种更实用的替代方案，近似近邻搜索凭借其更高的效率受到了越来越多的关注. 近似近邻搜索的关键在于使用一种高效的相似度计算方式替代原有的、低效的计算方式. 作为一种具有代表性的近似近邻搜索方法，哈希[1-5]方法由于占用的存储空间极小、计算极为高效，吸引了大量研究人员的关注.

在哈希方法中，早期的研究工作主要关注不依赖于数据的哈希算法，如局部敏感哈希(Locality Sensitive Hashing，LSH)[4]方法. 这类方法使用随机投影的方式对特征空间进行划分，生成二值码的不同比特. 理论上可以证明，随着二值码比特数量的增长，通过这类方法得到的二值码之间的汉明距离可以渐近地逼近相应样本在原始特征空间中的距离. 但是由于这类方法没有考虑数据的实际分布情况，往往需要较多的比特才能达到较好的检索结果，因此对于存储空间的需求较高.

为了获得更紧凑的二值码，基于数据进行学习的哈希学习算法逐渐成为主流. 这类哈希学习方法

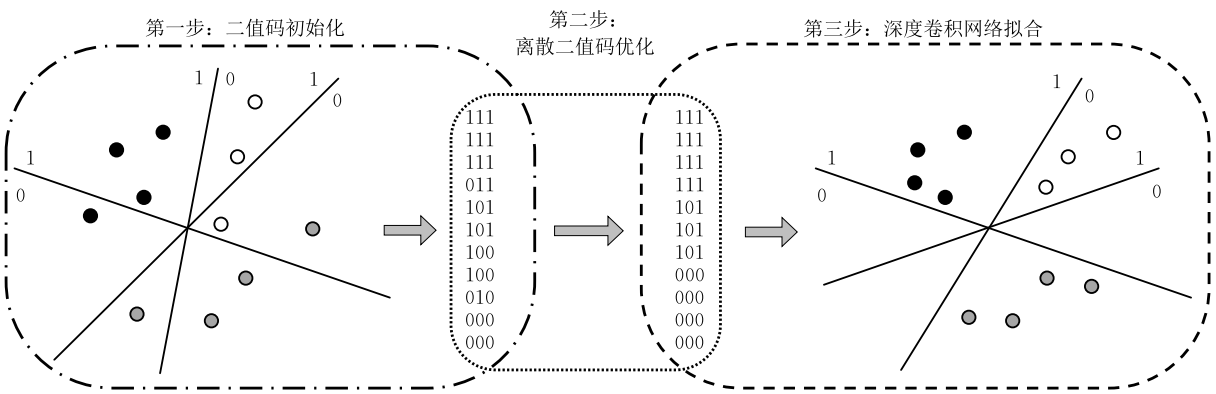


图 1　深度离散优化哈希(Deep Discrete Optimization Hashing，DDOH)方法流程示意图(本文方法主要包括三个步骤：第一步，通过使用任意的图像特征，并在该特征的基础上通过随机初始化、随机投影或主成分分析(Principle Component Analysis，PCA)的方式，获得初始的二值编码；第二步通过离散二值码优化对初始二值码进行迭代更新，获得判别力更强的二值码；第三步，通过训练深度卷积神经网络拟合优化得到的二值码，同时学习具有强判别能力的图像特征并获得相应的用于编码的哈希函数)

通过在一个训练集上进行学习，相比于不依赖数据的哈希算法，可以得到与数据更加适配的哈希函数，因此通常在使用同样数量的比特时，可以达到更好的检索精度.按照训练数据是否有人工标注的标签，哈希学习方法又可以进一步分为无监督方法和有监督方法两类.其中，无监督方法适用于数据没有标签的情况，可以直接在无标注的训练数据上学习哈希函数.在无监督哈希学习方法中，谱哈希(Spectral Hashing，SH)[6]以样本对在原始特征空间中的相似度作为权重，通过最小化投影后得到的二值码之间的加权汉明距离学习哈希函数；迭代量化(Iterative Quantization，ITQ)[1]通过主成分分析(Principle Component Analysis，PCA)将样本投影到低维空间后，寻找一个正交变换矩阵尽可能减小量化损失，以正交变换后的主成分作为哈希函数.

为了更好地应对面向复杂的语义相似度进行检索的任务，研究者进一步提出使用人工标注的标签信息辅助进行有监督的哈希函数学习.在有监督哈希学习方法中，典型相关分析迭代量化(Canonical Correlation Analysis Iterative Quantization，CCA-ITQ)[1]的思想与迭代量化大致相同，只是在投影到低维空间时，通过利用样本的标签，使用典型相关分析(CCA)对样本的特征进行投影；最小损失哈希(Minimal Loss Hashing，MLH)[7]为了缓解松弛-量化带来的负面影响，在松弛之后通过优化一个近似的损失上界来学习哈希函数.上述方法均使用线性投影作为哈希函数对数据的原始特征进行变换，因此无法很好地处理数据线性不可分的情况.为了解决这个问题，核监督哈希(Kernel Supervised Hashing，KSH)[8]和二值重建嵌入(Binary Reconstructive Embedding，BRE)[9]提出在核空间中学习非线性的哈希函数；深度哈希(Deep Hashing，DH)[10]使用高度非线性的深层感知机制学习判别能力更强的非线性哈希函数.

尽管上述方法在哈希学习任务中取得了一定的成功，但是由于在学习的过程中使用的样本特征表示并不能完全和数据匹配，因此在检索精度上有一定的局限.

为了解决这个问题，近期的一些深度哈希方法提出使用深度卷积神经网络，同时学习图像的特征表示和非线性的哈希函数.其中，卷积神经网络哈希(Convolutional Neural Network Hashing，CNNH)[11]首先对样本间的相似度矩阵进行二值矩阵分解来获得样本的目标二值码，再训练卷积神经网络拟合目标二值码，但是由于该方法中两个步骤之间是相互割裂的，无法保证学到的二值编码的质量；深度神经网络哈希(Deep Neural Network Hashing，DNNH)[12]、深度语义排序哈希(Deep Semantic Ranking Hashing，DSRH)[13]、深度正则化相似度比较哈希(Deep Regularized Similarity Comparison Hashing，DRSCH)[14]通过在汉明空间中设置锚点的方法，通过约束“与锚点相似的样本到锚点的距离要比与锚点不相似的样本到锚点的距离更近”的方式端到端地学习哈希函数，得到了判别能力更强的二值码；为了减小量化损失的影响，深度监督哈希(Deep Supervised Hashing，DSH)[2]、深度样本对监督哈希(Deep Pairwise Supervised Hashing，DPSH)[15]、哈希网络(HashNet)[3]通过最小化/最大化相似/不相似样本对之间的距离端到端地学习具有强判别力的哈希函数，同时通过显式地惩罚松弛后的实值特征与目标二值码之间的量化损失，进一步提升了二值码的检索精度；有监督语义保持深度哈希(Supervised Semantic-preserving Deep Hashing，SSDH)[16]首先通过 sigmoid 激活函数将卷积网络中某一层的输出限定在 0 到 1 的范围内，并端到端地训练模型在尽可能减小量化误差的条件下，使用这一层的输出预测样本的标签，以此将语义信息编码到这一层的输出中，之后通过量化的方式获得真正的二值编码.

尽管上述哈希方法通过对量化损失的显示约束，在近似近邻搜索任务上达到了很好的性能指标，但是由于方法中固有的松弛-量化步骤，仍然无法保证松弛后得到的实数值的最优解，在量化为二值后仍然是最优的.为了更好地解决这种量化带来的问题，一些近期的哈希学习方法提出直接在离散的汉明空间中进行优化.其中离散图哈希(Discrete Graph Hashing，DGH)[17]和有监督离散哈希(Supervised Discrete Hashing，SDH)[18]通过逐个比特的优化，在不需要松弛的条件下，直接得到了具有判别力的二值码；深度离散监督哈希(Deep Discrete Supervised Hashing，DDSH)[19]通过将离散优化与特征表示学习融合在一起，进一步提高了二值码的检索精度.但是 DDSH 方法只能在每个训练的小批量样本(mini-batch)中得到最优的二值码，因此无法保证得到的二值码相对于整个数据集是最优的.为了解决这个问题，本文提出了一种新的离散优化算法，通过在整个训练集上进行优化，得到最优的二值码；并在最优二值码的基础上，训练深度卷积网络同时学习图像

表示和相应的哈希函数.因此,与 DDSH 方法相比,本文方法优化得到的二值码判别力更强,在检索时可以达到更高的精度.

3　方　法

为了进一步缓解松弛-量化步骤对哈希学习方法的负面影响,提出了一种新的离散优化算法,直接在二值的汉明空间中进行优化,提升初始二值码的判别能力.为了在得到优化之后的二值码基础上得到相应的用于编码的哈希函数,训练了一个深度卷积神经网络同时学习更适配于数据的特征表示并对二值码进行拟合,整个方法的流程示意如图 1 所示.下面将分别详细介绍本文方法中的三个关键步骤,即二值码初始化、离散二值码优化、哈希函数学习.

3.1　形式化

令 Ω 表示 RGB 彩色图像空间,$S_{\mathrm{tr}}\subseteq\Omega$ 表示训练图像构成的集合,哈希学习方法的目标是在训练集 S_{tr} 上学习一个从图像空间 Ω 到 K 比特二值码的变换 $f:\Omega\rightarrow\{0,1\}^K$,使得相似的图像在变换之后的二值码也相似,不相似的图像的二值码也不相似.对于深度哈希方法,上述变换可以进一步分解为 $f(\boldsymbol{X})=h(\psi(\boldsymbol{X}))$,其中 $\boldsymbol{X}\in\Omega$ 表示一张 RGB 彩色图像,$\psi:\Omega\rightarrow R^d$ 表示使用深度卷积网络提取 d 维图像特征的过程,$h:R^d\rightarrow\{0,1\}^K$ 表示将深度特征变换为二值码的操作.具体来说,$h(\psi(\boldsymbol{X}))=I[\boldsymbol{W}^{\mathrm{T}}\psi(\boldsymbol{X})+\boldsymbol{b}>0]$,其中 $\boldsymbol{W}\in R^{d\times K}$ 表示哈希函数的权重矩阵,$\boldsymbol{b}\in R^K$ 表示相应的偏置项,$I[\text{条件}]$代表指示函数,当括号中的条件为真时取值为 1,否则取值为 0.为了便于表示,在下文中将使用 $\boldsymbol{B}$ 表示所有训练图像的二值码,$\boldsymbol{B}_i^{(t)}$ 表示第 i 个样本在第 t 次离散优化之后对应的二值码,$\boldsymbol{B}_i^{(t)}(k)$ 表示相应二值码的第 k 个比特.

在本文方法中,需要学习 $\boldsymbol{W}$ 和 $\boldsymbol{b}$ 中参数的值,同时也要对深度卷积网络中特征提取部分 ψ 的参数值进行学习(如卷积神经网络中卷积核的权重、全连接层的变换参数等).

3.2　二值码初始化

一种最简单直接的获得初始二值码 $\boldsymbol{B}_i^{(0)}$ 的方法是通过随机采样的方式初始化二值码,完全不考虑图像特征的分布情况:

$$\boldsymbol{B}_i^{(0)}=I[randn(K)>0] \tag{1}$$

其中 $randn(k)$ 表示从标准正态分布中随机采样 K 次.该方法通过直接对这 K 个采样结果分别进行量化得到初始的 K 比特二值码.

上述方法虽然可以快速地得到初始二值码,但是由于没有考虑到数据的真实分布情况,因此可能会给之后的优化步骤带来一些困难.为了解决这个问题,本文进一步考虑基于数据的真实分布的初始化方法.通常来说,在 ImageNet 的 1000 类物体识别任务[20]上预训练的深度卷积网络中,最后一层的输出对应的是图像属于相应的 1000 个类别的概率,而倒数第二层的输出对应的则是更加通用的图像表示.大量已有工作[21-22]表明,在 ImageNet 上预训练的深度卷积网络倒数第二层的特征可以应用于各种各样的计算机视觉任务.因此,一种可行的得到初始的二值码方法为:对于训练集中的第 i 张图像 $\boldsymbol{X}_i\in S_{\mathrm{tr}}$,本文方法首先将图片作为输入送入预训练的深度卷积神经网络模型,并使用网络倒数第二层的输出作为训练图像的特征 $\psi(\boldsymbol{X}_i)$.相应地,在该特征的基础上考虑两种初始方法:第一,通过对特征进行随机线性变换得到初始的二值码 $\boldsymbol{B}_i^{(0)}$:

$$\boldsymbol{m}=1/|S_{\mathrm{tr}}|\times\sum_{\boldsymbol{X}_i\in S_{\mathrm{tr}}}\psi(\boldsymbol{X}_i) \tag{2}$$

$$\boldsymbol{B}_i^{(0)}=I[\boldsymbol{W}_{\mathrm{rand}}^{\mathrm{T}}(\psi(\boldsymbol{X}_i)-\boldsymbol{m})>0] \tag{3}$$

其中 $\boldsymbol{m}$ 表示所有训练数据的平均特征,用于将样本的特征平移到特征空间的原点附近;$|\cdot|$ 表示集合的元素个数;$\boldsymbol{W}_{\mathrm{rand}}\in R^{d\times K}$ 表示一个 $d\times K$ 维的随机矩阵,该矩阵中的每一个元素都是从标准正态分布中随机采样得到的;第二,通过使用主成分分析的方法对特征进行降维,并以此获得初始的二值码:

$$\boldsymbol{B}_i^{(0)}=I[\boldsymbol{W}_{\mathrm{pca}}^{\mathrm{T}}(\psi(\boldsymbol{X}_i)-\boldsymbol{m})>0] \tag{4}$$

其中 $\boldsymbol{m}$ 为式(2)中定义的样本平均特征,$\boldsymbol{W}_{\mathrm{pca}}\in R^{d\times K}$ 表示训练数据的前 K 个主成分方向.

讨论:上述三种初始化方法中,第一种完全没有考虑数据的分布情况,因此该方法得到的初始二值码无法保持样本间的相似度关系,但是计算十分高效;第二种做法与局部敏感哈希[4]相似,可以在一定程度上保持图像原始特征之间的相似度关系,而且计算上也比较高效;第三种做法利用数据分布的主方向,可以得到冗余度较小的初始二值码,但是由于需要计算主成分分析,因此在计算效率上低于前两种方法,并且当训练集规模增大时,计算量和内存需求也会随之增大.综上所述,上述三种初始化方法中,前两种即使在训练数据规模非常大的情况下,依然可以很高效地得到初始的二值码.

3.3　离散二值码优化

初始二值码 $\boldsymbol{B}_i^{(0)}$ 的判别力较弱,因此如果直接

将其用于近似近邻搜索任务将很难达到令人满意的效果.为了提高二值码的判别力,本文提出一种新的离散二值码优化方法,直接在汉明空间中学习最优的二值码.具体来说,对于拥有类别标签的样本,在检索任务中,同类的图像应该具有相似的二值码,而不同类的图像的二值码也应该不同.令 $y_i \in \{1,2,3,\cdots,C\}$ 表示第 i 张训练图像所属的类别标号,对于相应的训练图像 $\boldsymbol{X}_i$,优化的目标是减小第 i 张图像的二值码与同类图像的二值码之间的距离,同时增大第 i 张图像与其他类图像的二值码之间的距离.在本文方法中,各个比特相互独立,因此不失一般性的,对于第 k 个比特,相应的优化目标为

$$L_i = 1\Big/\Big(\sum_j I[y_j = y_i]\Big) \times \sum_{y_j = y_i} (\boldsymbol{B}_i^{(0)}(k) - \boldsymbol{B}_j^{(0)}(k))^2 - 1\Big/\Big(\sum_j I[y_j \neq y_i]\Big) \times \sum_{y_j \neq y_i} (\boldsymbol{B}_i^{(0)}(k) - \boldsymbol{B}_j^{(0)}(k))^2 \quad (5)$$

其中第一项为归一化的类内差异,第二项为归一化的类间差异.

值得注意的是,对于二值码来说,其取值只有两种可能的结果,因此式(5)可以改写为

$$\boldsymbol{B}_i^{(0)}(k) = 0 : L_i = \sum_{y_j = y_i} \boldsymbol{B}_j^{(0)}(k) - \lambda \sum_{y_j \neq y_i} \boldsymbol{B}_j^{(0)}(k),$$

$$\boldsymbol{B}_i^{(0)}(k) = 1 : L_i = \sum_{y_j = y_i} \neg \boldsymbol{B}_j^{(0)}(k) - \lambda \sum_{y_j \neq y_i} \neg \boldsymbol{B}_j^{(0)}(k) \quad (6)$$

其中 $\lambda = \sum_j I[y_j = y_i] \Big/ \sum_j I[y_j \neq y_i]$, $\neg$ 表示取反操作.

为了使式(6)取得最小值,分情况进行讨论,当 $\boldsymbol{B}_i^{(0)}(k) = 0$ 时,不难证明当 $\sum_{y_j = y_i} \boldsymbol{B}_j^{(0)}(k) - \lambda \sum_{y_j \neq y_i} \boldsymbol{B}_j^{(0)}(k) < 0$ 时,维持 $\boldsymbol{B}_i^{(0)}(k) = 0$ 可以使式(6)取得最小值;反之,当 $\sum_{y_j = y_i} \boldsymbol{B}_j^{(0)}(k) - \lambda \sum_{y_j \neq y_i} \boldsymbol{B}_j^{(0)}(k) > 0$ 时,翻转 $\boldsymbol{B}_i^{(0)}(k)$ 的值可以使式(6)取得最小值.具体的证明见附录 1.

同理可得,当 $\boldsymbol{B}_i^{(0)}(k) = 1$ 且 $\sum_{y_j = y_i} \neg \boldsymbol{B}_j^{(0)}(k) - \lambda \sum_{y_j \neq y_i} \neg \boldsymbol{B}_j^{(0)}(k) < 0$ 时,维持 $\boldsymbol{B}_i^{(0)}(k)$ 的值可以使式(6)取得最小值;反之,当 $\sum_{y_j = y_i} \neg \boldsymbol{B}_j^{(0)}(k) - \lambda \sum_{y_j \neq y_i} \neg \boldsymbol{B}_j^{(0)}(k) > 0$ 时,翻转 $\boldsymbol{B}_i^{(0)}(k)$ 的值可以取得最小值.因此,对所有比特按照上述方式进行更新,即可降低损失的值.通过迭代地更新,即可得到具有强判别力的二值码,该过程的总体更新算法详见算法 1.由于二值码优化是一个 NP 难的问题,目前已有的理论无法保证本算法可以收敛到全局最小值.因此,本文将在第 4 节中,通过大量实验验证该算法在多种初始条件下的收敛性及有效性.

算法 1. 离散二值码优化算法.

输入:训练数据的初始二值码矩阵 $\boldsymbol{B}^{(0)}$,训练数据标签 y,优化迭代次数 n

输出:优化后的二值码矩阵 $\boldsymbol{B}^{(n)}$

```
for iter=1:n
  for k=1:K
    C_0 = Σ_i ¬B_i^(iter-1)(k) //训练集所有样本在第 k 个比特取 0 的个数
    C_1 = Σ_i B_i^(iter-1)(k) //训练集所有样本在第 k 个比特取 1 的个数
    for c=1:C
      S_{c,0} = Σ_{y_i=c} ¬B_i^(iter-1)(k) //第 c 类样本在第 k 个比特上取 0 的个数
      S_{c,1} = Σ_{y_i=c} B_i^(iter-1)(k) //第 c 类样本在第 k 个比特上取 1 的个数
      D_0 = C_0 - S_{c,0} //除第 c 类外,其他所有类样本在第 k 个比特上取 0 的个数
      D_1 = C_1 - S_{c,1} //除第 c 类外,其他所有类样本在第 k 个比特上取 1 的个数
      g_0 = S_{c,1} - λD_1 //第 c 类在第 k 个比特上取值为 0 的样本的损失
      g_1 = S_{c,0} - λD_0 //第 c 类在第 k 个比特上取值为 1 的样本的损失
      for y_i == c
        if B_i^(iter-1)(k)
          B_i^(iter-1)(k) = ¬(g_1 > 0)
        else
          B_i^(iter-1)(k) = (g_0 > 0)
return B^(n)
```

由于本文方法在进行二值码优化的时候,只需要考虑二值码的判别性,而不需要像大多数已有的深度哈希方法一样考虑量化损失的问题,因此可以避免在判别性损失和量化损失之间进行权衡,从而相比于已有的哈希方法,本文方法可以更加专注于二值码的判别性,从而可以获得具有更强判别能力的二值码.

3.4 哈希函数学习

上述离散二值码优化过程虽然可以在训练集上得到具有强判别力的二值码,但是无法对训练集之外的样本进行编码.为了得到用于编码的哈希函数,本文方法通过训练深度卷积神经网络,在训练集上拟合优化得到的二值码.因此,对于第 k 个比特,优化的目标是使深度卷积网络输出的二值码尽可能地接近优化得到的二值码:

$$\min \sum_i E(h(\psi(\boldsymbol{X}_i)), \boldsymbol{B}_i^{(n)}) \quad (7)$$

其中 $E(\cdot,\cdot)$ 是对两个向量之间差异的度量,可以

采用欧氏损失、Hinge 损失、Sigmoid 交叉熵损失等方式定义. 本文中采用 Sigmoid 交叉熵的形式，对于第 i 个训练样本，损失函数的具体定义为

$$E(h(\psi(\boldsymbol{X}_i)),\boldsymbol{B}_i^{(n)})=$$
$$-\sum_k[\boldsymbol{B}_i^{(n)}(k)\times\log(\sigma(\boldsymbol{W}_k^{\mathrm{T}}\psi(\boldsymbol{X}_i)+\boldsymbol{b}_k))+$$
$$(1-\boldsymbol{B}_i^{(n)}(k))\times\log(1-\sigma(\boldsymbol{W}_k^{\mathrm{T}}\psi(\boldsymbol{X}_i)+\boldsymbol{b}_k))] \quad (8)$$

其中 $\sigma(a)=1/(1+\exp(-a))$ 是 Sigmoid 函数，$\boldsymbol{W}_k$ 表示矩阵 $\boldsymbol{W}$ 的第 k 列，$\boldsymbol{b}_k$ 表示向量 $\boldsymbol{b}$ 的第 k 个元素. 为了便于表示，令 $h_i(k)=\sigma(\boldsymbol{W}_k^{\mathrm{T}}\psi(\boldsymbol{X}_i)+\boldsymbol{b}_k)$，上述损失函数对 $h_i(k)$ 是可导的，其导数定义为

$$\partial E/\partial h_i(k)=\sigma(\boldsymbol{W}_k^{\mathrm{T}}\psi(\boldsymbol{X}_i)+\boldsymbol{b}_k)-1,\ \boldsymbol{B}_i^{(n)}(k)=1,$$
$$\partial E/\partial h_i(k)=\sigma(\boldsymbol{W}_k^{\mathrm{T}}\psi(\boldsymbol{X}_i)+\boldsymbol{b}_k),\ \boldsymbol{B}_i^{(n)}(k)=0 \quad (9)$$

因此，利用链式法则，该模型可以通过标准的反向传播算法进行优化.

模型训练结束后，对于样本 $\boldsymbol{X}$，可以通过对卷积网络输出进行量化的方式，得到样本的二值码：

$$\boldsymbol{B}(k)=0.5\times\mathrm{sign}(\boldsymbol{W}_k^{\mathrm{T}}\psi(\boldsymbol{X})+\boldsymbol{b}_k)+0.5 \quad (10)$$

其中 sign(·)表示符号函数，当自变量的值大于 0 时，函数值为 1，否则函数值为−1.

4　实验验证

在本节中，通过在两个常用的图像检索数据集上进行对比实验，验证本文方法相对于已有方法的优越性. 另外，通过大量的消融实验和模块测试，验证本文方法中采用的各个模块的有效性，并对参数的敏感性进行分析.

4.1　实验环境和数据集

实验环境. 本次实验在一台 GPU 服务器上进行，离散二值码优化部分使用 MATLAB，二值码拟合部分使用的深度学习平台为 Caffe，基于 CUDA 和 cudnn 进行 GPU 加速，实际实验中只使用一块 GPU 进行实验，机器的配置见表 1.

表 1　实验机器配置信息

操作系统	Ubuntu 16.04.5 LTS
CPU	Intel(R) Xeon(R) CPU E5-2620 v3@2.40 GHz
内存	32 GB
GPU	4×Titan X
显卡驱动版本	396.54
CUDA 版本	9.2
cudnn 版本	7.0

数据集. 为了和已有哈希学习方法进行公平的对比，实验过程中沿用 CIFAR-10[23] 和 ImageNet-100[3] 两个已有方法中常用的图像检索评测数据集来测试本文方法的有效性. CIFAR-10：该数据库包含 60 000 张分辨率为 32×32 的彩色图像. 这些图像属于 10 个互斥的类别，其中每类 6000 张图像. 在本次实验中，使用 CIFAR-10 数据库标准的训练、测试数据划分，使用 50 000 张图像（每类 5000 张）训练模型并作为检索的数据库，10 000 张图像（每类 1000 张）作为查询图像. ImageNet-100：该数据库是 ImageNet 物体识别任务的一个子集，包含 100 类物体. 其中来自 ImageNet 训练集的 128 503 张图像作为检索的数据库，其中每类 130 张图像（共 13 000 张）用于训练模型，来自 ImageNet 校验集的每类 50 张图像（共 5000 张）作为测试时使用的查询图像. 由于目前只有极少数方法在更大规模的数据集上进行了评测，为了保证本文汇报结果的准确性，减小复现对比方法中可能的错误对对比方法性能的不利影响，本文仅在上述两个数据库上进行评测.

码长. 由于 CIFAR-10 数据库相对比较简单，本文中使用已有工作中常用的评测协议[2,16]，在 12、24、32、48 比特四个不同的码长上进行模型的评测. 对于 ImageNet-100，本文沿用文献[3]的评测方式，在 16、32、48、64 四个码长上进行测试.

评测方法及指标. 在上述两个数据集上，使用查询数据对数据库进行检索，当返回结果与查询图像来自同一个类别时，认定为一个正确的检索结果. 通过使用检索的查准率（precision）、召回率（recall）、平均准确率（mAP）、汉明距离小于等于 2 的样本的准确率作为指标评价各个方法的性能并进行对比. 特别的，对于 ImageNet-100 数据库，由于在数据库中一个类别只有约 1300 张图像，因此本文沿用文献[3]的评测方式，使用前 1000 个返回结果的平均准确率（mAP）作为评测的指标.

4.2　方法实现细节

网络结构. 对于大规模图像检索任务来说，编码速度是一个重要的指标. 考虑到编码速度与检索性能的权衡，本文采用 AlexNet[24] 的网络结构进行哈希学习. 为了减少模型的可训练参数，降低过拟合的风险，本文方法在 ImageNet 物体识别任务上预训练的网络参数的基础上进行参数微调. 具体来说，基于深度卷积网络的哈希函数中的参数 $\boldsymbol{W}$ 和 $\boldsymbol{b}$ 是预训练模型中所没有的，因此需要从随机初始化开始从头学习；另一方面，ψ 表示卷积网络中的卷积、池化等操作，其中的参数已经在 ImageNet 物体识别任务上进行了训练，因此能够从图像中提取表示能力较强的特征，在学习哈希函数的时候只需根据

相应的数据进行微调.

参数设置. 对于对比方法,本文使用原作者在各自论文中建议的结果设置参数. 对于本文提出的 DDOH 方法,离散二值码优化时的迭代次数 n 设置为 10 次. 训练卷积网络时,新加入的哈希层的学习率设置为 0.003,前面在 ImageNet 识别任务上预训练过的层学习率设置为 0.0003. 在 CIFAR-10 数据库上,因为训练数据量比较大,且训练数据与 ImageNet 预训练的数据差异较大,因此需要训练的时间也较长,共训练 50000 次,其中在 25000 次的时候将学习率降低到原来的 1/10. 在 ImageNet-100 数据库上,由于训练数据较少,为了避免过拟合,总共训练 1000 次,其中在 500 次的时候将学习率降低到初始学习率的 1/10. 此外,在两个数据库上,模型的权重衰减系数均设置为 0.0005,训练时使用的小批量大小为 256. 在训练的过程中,所有输入图像首先缩放为 256×256 的尺寸,之后从中随机裁剪出 227×227 的图像块作为模型的输入. 在测试阶段,同样首先将输入图像缩放为 256×256,然后选取位于图像中心的大小为 227×227 的图像块作为输入,通过对网络前向计算后,对输出进行量化得到最终的二值码.

4.3 与已有方法对比

为了验证本文方法的有效性,每次测试均与已有的哈希方法进行了对比. 具体来说,对比的方法包括局部敏感哈希(LSH)[4]、迭代量化(ITQ)[1]、基于典型相关分析的迭代量化(CCA-ITQ)[1]、有监督离散哈希(SDH)[18]、卷积神经网络哈希(CNNH)[11]、深度卷积网络哈希(DNNH)[12]、深度监督哈希(DSH)[2]、哈希网络(HashNet)[3]、有监督语义保持深度哈希(SSDH)[16].

对比结果见表 2、图 2 和图 3. 在 CIFAR-10 数据库上(表 2 左侧和图 2),本文提出的方法在平均准确率、精度-召回率曲线、汉明距离小于等于 2 的样本的精度三项指标上都达到了和现有最好的方法一样好的性能指标. 特别是在精度-召回率曲线上,即使在召回率接近 1 的情况下,本文的方法仍然可以保持非常高的精度,这说明本文的方法中,同类的样本之间非常紧凑,而且每一类的编码都有很强的差异. 在 ImageNet-100 数据库上,本文提出的方法在所有指标上都显著超越了已有的方法,这证明了该方法的优越性.

表 2 与现有哈希学习方法在平均准确率指标(mAP)上的对比(本文方法(DDOH)的结果见表中最后一行,对比方法的最好结果用下划线标出)

方法	CIFAR-10(mAP)				ImageNet-100(mAP@1000)			
	12-bit	24-bit	32-bit	48-bit	16-bit	32-bit	48-bit	64-bit
LSH[4]	0.125	0.150	0.169	0.186	0.080	0.160	0.224	0.274
ITQ[1]	0.230	0.241	0.253	0.259	0.307	0.457	0.516	0.553
CCA-ITQ[1]	0.573	0.614	0.625	0.634	0.321	0.494	0.589	0.650
SDH[18]	0.205	0.637	0.632	0.660	0.481	0.567	0.600	0.617
CNNH[11]	0.856	0.860	0.863	0.864	0.281	0.450	0.525	0.554
DNNH[12]	0.694	0.821	0.825	0.835	0.290	0.461	0.530	0.565
DSH[2]	0.920	0.929	0.933	0.935	0.558	0.632	0.650	0.663
HashNet[3]	0.943	0.950	0.952	0.953	0.506	0.631	0.663	0.684
SSDH[16]	0.927	0.942	0.945	0.947	0.621	0.680	0.688	0.700
DDOH	**0.949**	**0.948**	**0.949**	**0.950**	**0.647**	**0.697**	**0.712**	**0.720**

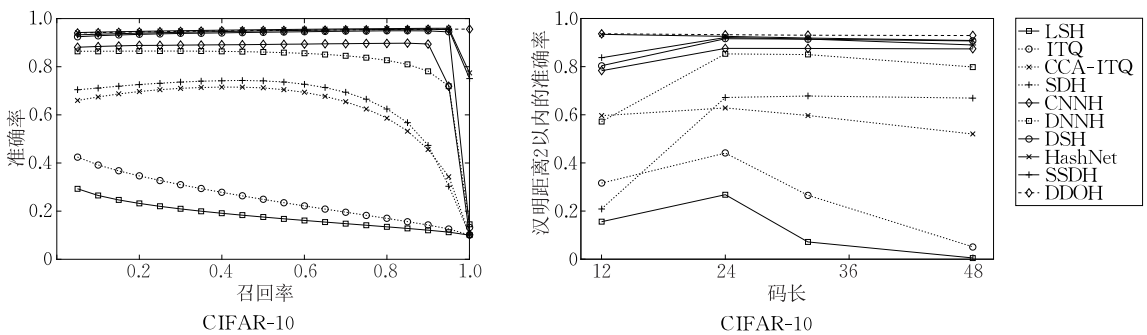


图 2 在 CIFAR-10 数据库上进行测试时得到的 48 比特模型的精度-召回率(Precision-Recall,PR)曲线(左图)和不同码长模型的汉明距离小于等于 2 的样本的精度曲线(右图)

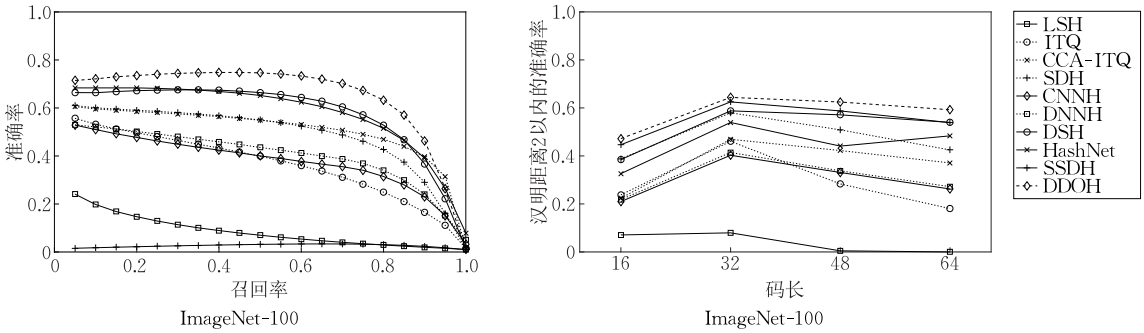


图 3　在 ImageNet-100 数据库上进行测试时得到的 64 比特模型的精度-召回率(PR)曲线(左图)和不同码长模型的汉明距离小于等于 2 的样本的精度曲线(右图)

4.4　离散二值优化有效性验证

为了验证本文提出的离散二值码优化方法能够提升初始二值码的判别力，训练卷积神经网络模型对初始使用随机投影进行初始化的二值码进行拟合，并与使用完整算法训练的模型进行对比. 具体来说，在 CIFAR-10 和 ImageNet-100 两个数据库上，分别使用 48 比特和 64 比特的二值码进行实验.

表 3 展示了对比的实验结果，从表中的结果可以看出，使用本文提出的离散优化方法可以显著提升检索的性能，证明了本文提出的离散二值码优化方法的有效性.

表 3　使用深度卷积网络拟合离散优化前/后的二值码得到的模型，在两个数据库上的检索平均准确率对比(在两个数据库上使用的分别是 48 比特(CIFAR-10)和 64 比特(ImageNet-100)的二值码)

目标二值码	CIFAR-10	ImageNet-100
优化前	0.192	0.447
优化后	0.950	0.720

4.5　不同二值码初始化方法对比

前文的 3.2 节中描述了三种不同的二值码初始化方式. 本节通过实验对比这三种初始化方式在两个数据库上的检索性能. 具体来说，在 CIFAR-10 和 ImageNet-100 两个数据库上，分别使用 48 比特和 64 比特的二值码进行实验，通过三种不同的初始化方法得到初始二值码后，经过离散二值码优化、深度卷积神经网络拟合两个步骤，得到相应的编码模型并进行测试. 特别地，为了更好地验证本文方法对不同初始化方式的鲁棒性，对于随机投影初始化方法，汇报 10 次随机试验的平均结果.

表 4 展示了不同初始化方式得到的模型在检索时的平均准确率，三种初始化方式在性能上的差异很小，说明本文的离散二值码优化方法对于二值码的初始化不敏感，体现出了该方法的鲁棒性. 出于计算时间和效率上的考虑，在实际应用中，前两种初始化方法要优于采用主成分分析的初始化方法.

表 4　使用不同初始化方式获得的初始二值码在经过离散优化之后，使用卷积神经网络拟合的模型的检索性能(在两个数据库上使用的分别是 48 比特(CIFAR-10)和 64 比特(ImageNet-100)的二值码. *：10 次实验的平均结果)

方法	CIFAR-10	ImageNet-100
随机采样	0.946	0.719
随机投影*	0.951	0.720
主成分分析	0.947	0.725

为了进一步验证本文提出的离散二值码优化方法的有效性，使用 HashNet[3] 方法在训练集上得到的 48 比特(CIFAR-10)和 64 比特(ImageNet-100)二值码作为本文方法的初始二值码. 在 CIFAR-10 和 ImageNet-100 上，模型的检索平均准确率分别为 0.950 和 0.700. 其中，在 CIFAR-10 上，该结果与 HashNet 的结果基本相同，在 ImageNet-100 上则显著超越了 HashNet 的检索性能，这证明了本文方法并不局限于 3.2 节中提出的三种二值码初始化方法. 并且当使用已有的哈希方法作为初始化的时候，本文方法有能力进一步提升检索的性能. 对这个现象一种可能的解释是，本文方法在进行离散二值码优化的时候只需要关注二值码的判别能力，而不需要对量化损失进行权衡，因此相比于已有方法，可以找到判别力更强的解. 此外，上述结果也表明，使用更好的二值码初始化方法并不能保证本文方法可以得到更好的检索性能(相比于本文提出的三种初始化方法)，再次证明了本文方法对初始化并不敏感.

4.6　离散优化迭代次数的影响评测

相比于已有的深度哈希学习算法，本文的方法中需要设置的额外参数主要是离散二值码优化方法中的迭代优化次数 n. 在这一节中，进一步测试离散二值码优化迭代次数对模型的影响. 在 CIFAR-10 和 ImageNet-100 两个数据集上，分别使用 48 比特和 64 比特的二值码进行实验，根据式(5)画出所有训练样本的损失值 $\sum_i L_i$ 随迭代次数变化的情况，结果见

图 4. 从图 4 中可以看出，当迭代次数大于 4 次的时候，损失值不再有明显的变化，说明迭代次数设置为 4 次以上时就基本可以保证离散二值码优化的过程完全收敛.

4.7 离散优化必要性验证

本文提出的离散二值码优化方法基于 Fisher 准则，因此有必要验证直接使用 Fisher 准则训练的深度卷积网络模型在图像检索任务上的性能，以此验证离散优化的必要性. 因此，本文尝试在保持网络结构不变的条件下，使用如下基于 Fisher 准则的损失函数对模型进行训练：

$$\sum_{k}\Big[\sum_{y_j=y_i}(\sigma(\boldsymbol{W}_k^{\mathrm{T}}\psi(\boldsymbol{X}_i)+\boldsymbol{b}_k)-\sigma(\boldsymbol{W}_k^{\mathrm{T}}\psi(\boldsymbol{X}_j)+\boldsymbol{b}_k))^2-\alpha\sum_{y_j\neq y_i}(\sigma(\boldsymbol{W}_k^{\mathrm{T}}\psi(\boldsymbol{X}_i)+\boldsymbol{b}_k)-\sigma(\boldsymbol{W}_k^{\mathrm{T}}\psi(\boldsymbol{X}_j)+\boldsymbol{b}_k))^2\Big] \quad (11)$$

其中 α 是一个需要调节的参数，用于平衡相似样本对和不相似样本对之间的权重. 实验结果见表 5. 在 CIFAR-10 数据库上，通过调节参数 α 的值，该模型可以达到和本文方法相似的检索性能. 但是在更加复杂的 ImageNet-100 数据库上，该模型的检索性能远低于本文方法，这证明本文中提出的离散二值码优化方法在数据复杂的情况下，明显优于直接使用 Fisher 准则对模型的输出进行优化. 出现这种现象的原因可能是因为卷积神经网络只能基于小批量数据进行训练，当直接使用 Fisher 准则对模型进行训练的时候，模型只能利用每个小批量内的图像对进行学习，当类别数多的时候（如 ImageNet-100），该方法无法采样到足够丰富的相似样本对，因此无法学到判别力强的哈希函数. 具体来说，即使使用一个小批量数据中的所有可能的图像对，平均一个小批量中也只有约 190 个相似样本对，整个训练（1000 次迭代）过程中只利用了约 19 万个相似样本对. 而本文提出的方法可以直接利用训练数据集上的所有图像对（包含超过 80 个相似样本对）进行学习，因此可以利用更丰富的监督信息达到更好的检索性能. 此外，本文方法的另一个优势是不需要手动调节正负样本之间的平衡参数（α），因此相比于直接使用 Fisher 准则进行训练，本文方法的训练更加简单.

表 5 不使用离散二值码优化而只使用 Fisher 准则训练的模型在两个数据库上的检索性能（在两个数据库上使用的分别是 48 比特（CIFAR-10）和 64 比特（ImageNet-100）的二值码）

α	CIFAR-10	α	ImageNet-100
0.05	0.100	0.001	0.066
0.1	0.295	0.005	0.230
0.5	0.944	0.01	0.303
1	0.947	0.05	0.042
5	0.950	1	0.037

4.8 离散二值码优化速度评测

由于离散二值码优化步骤可以通过 MATLAB 中内置的矩阵计算实现高效地并行，因此本文方法在不同码长下的训练时间几乎完全相同.

离散二值码优化的速度也是本文方法应用中的一个关键因素. 为了测试本文方法中二值码优化的速度，我们重复进行了 10 次实验，每次实验中迭代优化的轮数设置为 10 次，在 CIFAR-10 和 ImageNet-100 两个数据集上，分别使用 48 比特和 64 比特的二值码进行实验，结果见表 6. 从表 6 中展示的结果可以看出，本文提出的离散二值码优化步骤速度非常快，与深度卷积神经网络拟合二值码的步骤相比，基本可以忽略不计. 其中，由于 CIFAR-10 的训练集明显大于 ImageNet-100 的训练集（50 000 vs. 13 000），因此 CIFAR-10 训练集的优化时间略长，但是即使如此，优化的时间也非常短.

表 6 离散二值码优化用时（两个数据库上分别使用 48 比特（CIFAR-10）和 64 比特（ImageNet-100）的二值码进行实验）

数据库	时间/s
CIFAR-10	2.120
ImageNet-100	0.860

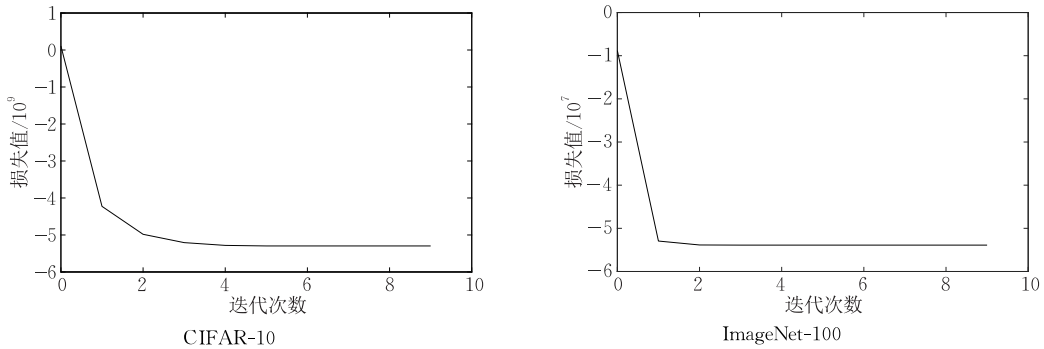


图 4 在 CIFAR-10 和 ImageNet-100 两个数据集上，经过不同迭代次数的离散二值码优化之后，二值码的判别损失函数值的变化情况（两个数据库上分别使用 48 比特（CIFAR-10）和 64 比特（ImageNet-100）的二值码进行实验）

5 结　论

本文提出了一种基于离散二值码优化的新的哈希学习算法 DDOH. 在通用的图像检索评测数据集上的大量实验表明，本文提出的方法达到了目前最好的水平，并且本文所提出的离散二值码优化算法可以在多种初始化下，高效地得到具有强判别力的二值码，证明了在汉明空间中直接对二值码进行优化的可行性和优越性. 由于离散优化可以避免量化损失的问题，本文方法可以避免目前大多数方法中常见的量化损失带来的问题，因此可以达到更好的检索性能. 后续工作中，为了更好地与真实应用场景匹配，验证本文方法在实际应用中的可用性，计划在难度更大的、更接近真实应用场景的数据库（如完整的 ImageNet 数据库）上进行测试，检验本文方法的通用性.

参 考 文 献

[1] Gong Y, Lazebnik S, Gordo A, Perronnin F. Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(12): 2916-2929

[2] Liu H, Wang R, Shan S, Chen X. Deep supervised hashing for fast image retrieval//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 2064-2072

[3] Cao Z, Long M, Wang J, Yu P S. HashNet: Deep learning to hash by continuation//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 5608-5617

[4] Gionis A, Indyk P, Motwani R. Similarity search in high dimensions via hashing//Proceedings of the International Conference on Very Large Data Bases. Edinburgh, UK, 1999: 518-529

[5] Wen Qing-Fu, Wang Jian-Min, Zhu Han, et al. Distributed learning to Hash for approximate nearest neighbor search. Chinese Journal of Computers, 2017, 40(1): 194-208 (in Chinese)
(文庆福，王建民，朱晗等. 面向近似近邻查询的分布式哈希学习方法. 计算机学报，2017，40(1)：194-208)

[6] Weiss Y, Torralba A, Fergus R. Spectral hashing//Proceedings of the Advances in Neural Information Processing Systems. Vancouver, Canada, 2009: 1753-1760

[7] Norouzi M, Blei D M. Minimal loss hashing for compact binary codes//Proceedings of the 28th International Conference on Machine Learning. Bellevue, USA, 2011: 353-360

[8] Liu W, Wang J, Ji R, et al. Supervised hashing with kernels//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Rhode Island, USA, 2012: 2074-2081

[9] Kulis B, Darrell T. Learning to hash with binary reconstructive embeddings//Proceedings of the Advances in Neural Information Processing Systems. Vancouver, Canada, 2009: 1042-1050

[10] Erin Liong V, Lu J, Wang G, et al. Deep hashing for compact binary codes learning//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 2475-2483

[11] Xia R, Pan Y, Lai H, et al. Supervised hashing for image retrieval via image representation learning//Proceedings of the 23rd International Joint Conference on Artificial Intelligence. Québec City, Canada, 2014: 2156-2162

[12] Lai H, Pan Y, Liu Y, Yan S. Simultaneous feature learning and hash coding with deep neural networks//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 3270-3278

[13] Zhao F, Huang Y, Wang L, Tan T. Deep semantic ranking based hashing for multi-label image retrieval//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 1556-1564

[14] Zhang R, Lin L, Zhang R, et al. Bit-scalable deep hashing with regularized similarity learning for image retrieval and person re-identification. IEEE Transactions on Image Processing, 2015, 24(12): 4766-4779

[15] Li W J, Wang S, Kang W C. Feature learning based deep supervised hashing with pairwise labels//Proceedings of the 25th International Joint Conference on Artificial Intelligence. New York, USA, 2016: 1711-1717

[16] Yang H F, Lin K, Chen C S. Supervised learning of semantics-preserving hash via deep convolutional neural networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(2): 437-451

[17] Liu W, Mu C, Kumar S, Chang S F. Discrete graph hashing//Proceedings of the Advances in Neural Information Processing Systems. Montreal, Canada, 2014: 3419-3427

[18] Shen F, Shen C, Liu W, Shen H T. Supervised discrete hashing//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 37-45

[19] Jiang Q Y, Cui X, Li W J. Deep discrete supervised hashing. IEEE Transactions on Image Processing, 2018, 27(12): 5996-6009

[20] Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge. International Journal of Computer Vision, 2015, 115(3): 211-252

[21] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA, 2014: 580-587

[22] Babenko A, Slesarev A, Chigorin A, Lempitsky V. Neural codes for image retrieval//Proceedings of the European Conference on Computer Vision. Zurich, Switzerland, 2014: 584-599

[23] Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images. Technical Report, University of Toronto, 2009, 1(4): 7

[24] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks//Proceedings of the Advances in Neural Information Processing Systems. Lake Tahoe, USA, 2012: 1097-1105

附录 1.

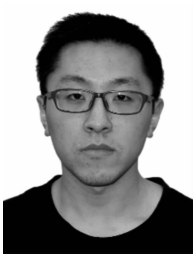
证明. 当 $\boldsymbol{B}_i^{(0)}(k)=0$ 且 $\sum_{y_j=y_i}\boldsymbol{B}_j^{(0)}(k)-\lambda\sum_{y_j\neq y_i}\boldsymbol{B}_j^{(0)}(k)<0$ 时,存在下列不等式:

$$\begin{aligned}0&>\sum_{y_j=y_i}\boldsymbol{B}_j^{(0)}(k)-\lambda\sum_{y_j\neq y_i}\boldsymbol{B}_j^{(0)}(k)\\&=\sum_{y_j=y_i}(1-\neg\boldsymbol{B}_j^{(0)}(k))-\lambda\sum_{y_j\neq y_i}(1-\neg\boldsymbol{B}_j^{(0)}(k)\\&=\sum_{y_j=y_i}1-\sum_{y_j=y_i}\neg\boldsymbol{B}_j^{(0)}(k)-\lambda\sum_{y_j\neq y_i}1+\lambda\sum_{y_j\neq y_i}\neg\boldsymbol{B}_j^{(0)}(k)\\&=-\Big(\sum_{y_j=y_i}\neg\boldsymbol{B}_j^{(0)}(k)-\lambda\sum_{y_j\neq y_i}\neg\boldsymbol{B}_j^{(0)}(k)\Big),\end{aligned}$$

因此有 $\sum_{y_j=y_i}\neg\boldsymbol{B}_j^{(0)}(k)-\lambda\sum_{y_j\neq y_i}\neg\boldsymbol{B}_j^{(0)}(k)>0$,

即 $\sum_{y_j=y_i}\boldsymbol{B}_j^{(0)}(k)-\lambda\sum_{y_j\neq y_i}\boldsymbol{B}_j^{(0)}(k)<\sum_{y_j=y_i}\neg\boldsymbol{B}_j^{(0)}(k)-\lambda\sum_{y_j\neq y_i}\neg\boldsymbol{B}_j^{(0)}(k)$,

即保持相应比特的值可以保证 L_i 的值不增加. 证毕.



LIU Hao-Miao, Ph. D. candidate. His research interests include image retrieval and interpretable classification model.

WANG Rui-Ping, Ph. D., professor. His research interests include computer vision, pattern recognition, and machine learning.

SHAN Shi-Guang, Ph. D., professor. His research interests include computer vision, pattern recognition, and machine learning.

Xilin Chen, Ph. D., professor. His research interests include computer vision, pattern recognition, image processing, and multimodal interfaces.

Background

Approximate nearest neighbor (ANN) search is the key technique in large-scale image retrieval tasks. As a representative ANN method, hashing methods index samples as compact binary codes, and thus is very efficient in both computation and storage. However, due to the discrete nature of binary codes, directly optimizing for optimal binary codes is an NP-hard problem. Therefore, existing hashing methods can only perform optimization efficiently by relaxing the binary codes to real values, and thus inevitably suffer from performance drop when quantizing the real-valued optimization results into binary codes, due to the discrepancy between binary Hamming space and real-valued Euclidean space. Although some recent works have explicitly take quantization error into consideration in the model training stage to relieve the negative effects of the relaxation and quantization, they might bring up some new problems. For example, more hyperparameters, difficulty in convergence, etc. To better deal with the problems of quantization, in this paper we propose a novel hash learning method, named Deep Discrete Optimization Hashing (DDOH). The proposed method obtains discriminative binary codes by directly using subgradient method in the binary Hamming space, and then fits the obtained binary codes with deep convolutional neural networks. Experiments on two widely studied datasets validate the effectiveness of the proposed method.

Before this work, our group has worked on the topic of hash learning for more than five years. Some methods proposed by our group have been published in top-tier computer vision conferences and journals such as CVPR, ICCV, TIP, etc..

This research was supported by the 973 Program under Grant No. 2015CB351802, the National Natural Science Foundation of China under Grant No. 61772500, the Frontier Science Key Research Project CAS No. QYZDJ-SSW-JSC009, and the Youth Innovation Promotion Association CAS No. 2015085.