

基于偏差抑制对比学习的无监督深度哈希图像检索^①



苏 海, 钟雨辰

(华南师范大学 软件学院, 广州 528225)
通信作者: 苏 海, E-mail: suhai@m.scnu.edu.cn

摘要: 在时下的无监督深度哈希研究领域中, 基于对比学习而提出的方法占主流地位。但是对比学习所采用的随机抽取负样本方式, 会带来采样偏差问题, 从而对图像检索精度造成负面影响。为解决以上问题, 本文提出了一种基于偏差抑制对比学习的无监督深度哈希 (unsupervised deep hashing based on bias suppressing contrastive learning, BSCDH)。本文在对比学习框架的基础上提出了偏差抑制方法 (bias suppression, BSS), 将错误负样本近似为极困难负样本, 并设计了偏差抑制系数来抑制极困难负样本从而缓解采样偏差的负面影响。本文根据当前负样本与查询样本的相似度来确定其对应的抑制系数取值, 并引入当前负样本与邻近的聚类中心间的距离关系对抑制系数进行取值修正, 降低正常负样本被过度抑制的可能性。最终 BSCDH 的 64 位哈希码 mAP@5000 指标在 CIFAR-10、FLICKR25K、NUS-WIDE 数据集上分别达到 0.696、0.833、0.819, 相较 baseline 具有显著的性能优势。本文开展的大量实验证明了 BSCDH 在无监督图像检索方法中拥有较高的检索精度, 且能有效应对采样偏差问题。

关键词: 无监督哈希; 对比学习; 图像检索; 采样偏差

引用格式: 苏海,钟雨辰.基于偏差抑制对比学习的无监督深度哈希图像检索.计算机系统应用,2025,34(2):165–173. <http://www.c-s-a.org.cn/1003-3254/9774.html>

Unsupervised Deep Hashing Image Retrieval Based on Bias Suppressing Contrastive Learning

SU Hai, ZHONG Yu-Chen

(School of Software, South China Normal University, Guangzhou 528225, China)

Abstract: In the contemporary field of unsupervised deep hashing research, methods predicated on contrastive learning are predominant. However, sampling bias brought about by the random extraction of negative samples in contrastive learning deteriorates image retrieval accuracy. To address the issue, this study proposes a novel unsupervised deep hashing based on bias suppressing contrastive learning (BSCDH). It proposes a bias suppression method (BSS) based on a contrastive learning framework. This method approximates incorrect negative samples as extremely hard negative samples and designs a bias suppression coefficient to suppress these extremely hard negative samples, thereby alleviating the negative impact of sampling bias. The corresponding suppression coefficient value is determined based on the similarity between the current negative sample and the query sample. Distance relationship between the current negative sample and adjacent hash centers is introduced to correct the suppression coefficient value, reducing the possibility of excessive suppression of normal negative samples. Ultimately, the mAP@5000 of the BSCDH method (64 bits) achieves 0.696, 0.833, and 0.819 respectively on the CIFAR-10, FLICKR25K, and NUS-WIDE datasets, demonstrating a significant performance advantage over the baseline. Extensive experiments conducted in this paper verify that BSCDH exhibits high retrieval accuracy in unsupervised image retrieval methods and can effectively address sampling bias.

① 基金项目: 广东省基础与应用基础研究基金 (2022A1515140110, 2021A1515110673, 2020B1515120089)

收稿时间: 2024-07-29; 修改时间: 2024-08-20; 采用时间: 2024-09-10; csa 在线出版时间: 2024-12-16

CNKI 网络首发时间: 2024-12-17

Key words: unsupervised hashing; contrastive learning; image retrieval; sampling bias

1 前言

在图像检索研究领域,无监督深度哈希方法^[1-3]是一个受到广泛关注的重要课题。因为不需要像有监督方法进行大规模的数据标注,无监督方法更加适应图像检索实际应用场景的需求,但目前无监督方法的图像检索精度并不让人满意。无监督深度哈希方法需要在标签缺失的情况下为模型提供其他监督信号,这些监督信号会帮助模型从新的角度学习到样本的特征。其中,使用正负样本对比产生监督信号的对比学习框架^[4-7]一经提出,便获得无监督深度哈希领域的关注。

对比学习框架对同一图像进行随机图像增强分别获得查询样本与正样本,并随机抽取数据集中不同图像作为负样本。该框架希望查询样本接近正样本,远离负样本,从而利用对比关系提供监督信号。近年来的一些工作^[8-12]通过将对比学习与无监督深度哈希结合,使图像检索精度相较先前的工作获得了一定的提升,这也为无监督深度哈希的研究提供了新的可行方向。

但随着进一步地研究,研究者们发现原本的对比学习框架存在着一些缺点。对比学习随机抽取图片作为当前图片负样本时,可能会将与当前图片同一标签的其他图片错误抽取为负样本。这部分负样本会被称为错误负样本,而这一现象则被称为采样偏差(sampling bias)^[13,14]。对比学习框架会受到采样偏差的负面影响,将错误负样本推远,这意味着同标签样本之间的距离将会扩大,导致同标签样本形成的聚类较为松散。

针对以上问题,本文提出了基于偏差抑制对比学习的无监督深度哈希方法(unsupervised deep hashing based on bias suppressing contrastive learning, BSCDH)。当一个负样本与查询样本越相似,那么模型就越难以区分该负样本与查询样本的区别,因此本文将与查询样本最相似的若干负样本称为极困难负样本。而错误负样本因为与查询样本标签相同,他们之间的相似度往往较高,这意味着错误负样本通常满足极困难负样本的标准。因此,本文认为可以将一定范围内的极困难负样本近似认为是错误负样本,并设计抑制系数来抑制查询样本与极困难负样本之间的余弦相似度。这会让模型在计算对比损失函数时减少对这部分负样本的关注,在一定程度上缓解采样偏差带来的对同标签样

本聚类的不利影响,从而提升该无监督图像检索模型的检索精度。

2 相关工作

2.1 无监督深度哈希

随着深度神经网络技术在人工智能的诸多领域取得成功,研究者们选择将深度神经网络作为骨干网络来进一步提升无监督哈希方法的特征提取能力。其中,生成式的无监督深度哈希^[15-19]会通过重建图片原始信息或者生成对抗等方式来进行训练模型。自编码双瓶颈哈希(auto-encoding twin-bottleneck hashing, TBH)^[18]提出一种自编码器变体,并引入双瓶颈协同交换关键信息。但是生成式的方法很可能在生成的过程中给哈希码引入冗余信息,损害模型性能。

除了生成式方法以外,也有研究者提出了利用相似度特征的无监督深度哈希方法^[20,21],通过构建具有可信语义相似关系的图像对来为无监督训练提供监督信号。深度二进制描述符哈希(learning compact binary descriptors with unsupervised deep neural network, DeepBit)^[20]认为可以将图像的不同旋转视图视为相似对,进而用于优化哈希码。

而随着对比学习框架的提出,利用相似度特征的方法得到了进一步发展^[8,11,12,22]。基于对比信息瓶颈的无监督哈希(unsupervised hashing with contrastive information bottleneck, CIBH)^[8]修改对比损失函数使其更加符合哈希特定要求,并引入概率二进制表示层实现端到端训练。基于相似度特征的无监督深度哈希方法目前是无监督图像检索的重要方向,但如果想要进一步提升检索精度,仍需要进行不断的完善。

2.2 对比学习

对比学习框架的提出为无监督表示学习领域取得了新的突破。而在对比学习框架中,实例判别任务^[5,23-30]是一个重要概念。具体来说,该任务的目标是将查询样本的正样本拉近,而将来自其他图像的负样本推远,利用相似性关系让模型学习到关键的判别语义信息。无监督视觉表征的动量对比学习(momentum contrast for unsupervised visual representation learning, MoCo)^[25]提出了基于动量编码器的架构,使用动量编码器来生成

正样本对应的特征，并在本轮次结束后用于负样本队列的更新。

随着对比学习研究的不断推进，研究者们发现实例判别任务中的一些问题有待解决。因为实例判别任务的负样本是通过从数据集随机抽取获得的，这过程中可能将图像的同类样本误选为负样本。这种现象被称为采样偏差^[12-14]，它将干扰嵌入空间中同类样本的聚类，导致模型并不能学到最优的表示(representation)。部分研究者通过设计相关实验，发现对比学习中最困难的0.1%负样本可能对模型训练有害^[31]，这也是采样偏差现象影响的体现。针对这一问题，去偏对比学习(debiased contrastive learning, Debiased CL)^[12,13]等工作认为可以设计新的去偏对比目标来对负样本的采样偏差加以纠正，从而模拟无偏采样的理想情况，缓解采样偏差带来的影响。

3 方法

本文所提出的方法，基于偏差抑制比学习的无监督深度哈希(即BSCDH)的框架如图1所示，其中主要包含两个部分：(1)对比学习框架，负责将数据增强得到的查询样本与正样本图片经过两个编码器转变为哈希码，并更新负样本队列；(2)偏差抑制模块，负责找出极困难负样本并计算对应的抑制系数用以抑制。原始图像 x_i 通过旋转裁剪等数据增强方式变换为查询样本 v_i 与正样本 v'_i ， v_i 与 v'_i 经过各自对应的编码器分别转变为哈希码 h_i 与 h'_i 。在偏差抑制(BSS)模块中， h_i 将与负样本队列进行相似度计算，再按照相似度对负样本队列进行排序，最后应用抑制系数 λ_s 对极困难负样本进行抑制。被抑制处理后的负样本哈希码，将与查询样本哈希码 h_i 以及正样本哈希码 h'_i ，进行SSNCE损失函数计算。

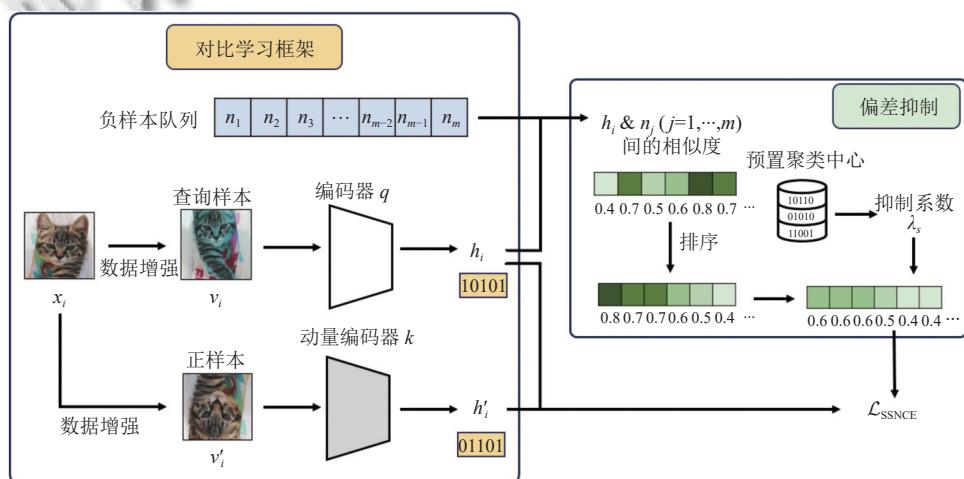


图1 BSCDH方法整体框架图

在本节内容中，我们将对BSCDH方法的训练框架进行更详细的叙述。首先会简要介绍使用对比学习方法训练哈希编码器的通用流程，然后介绍BSCDH在此基础上进行的改进。

3.1 基于对比学习的哈希编码器训练

给定一组具有 n 张图片的训练集 $X = \{x_1, x_2, \dots, x_n\}$ ，将每个图像随机变换为两个视图 v_i 与 v'_i (例如，旋转、裁剪、颜色失真等)。然后 v_i 与 v'_i 会被送入各自的哈希编码器 $f_\theta(\cdot)$ 转化为哈希码。在训练过程中，本文采用 $\tanh(\cdot)$ 作为最后的激活函数替代 $\text{sgn}(\cdot)$ 来产生二进制哈希码 h_i 与 h'_i 。

$$h_i = \tanh(f_\theta(v_i)) \quad (1)$$

对比学习框架依靠对比损失函数实现实例判别任务。以MoCo^[25]为例，他们会使用InfoNCE作为损失函数，定义如下：

$$L_{\text{InfoNCE}} = \frac{1}{N} \sum_{i=1}^N -\log \frac{e^{h_i \cdot h'_i / \tau}}{e^{h_i \cdot h'_i / \tau} + \sum_{j=0}^M e^{h_i \cdot n_j / \tau}} \quad (2)$$

其中， h_i 与 h'_i 都来自同一张实例图片 x_i ， h'_i 代表 h_i 的正样本， n_j 则是来自负样本队列的负样本。 M 代表负样本队列的长度。 τ 是温度超参数，起到控制分布平滑度的作用。MoCo中， h_i 与 h'_i 并非来自同一个编码器，实质上被训练的编码器只有 h_i 所对应的编码器 $f_{\theta q}(\cdot)$ ， h'_i 是由

动量编码器 $f_{\theta k}(\cdot)$ 产生的, 其更新遵循公式:

$$\theta_k = (1 - m)\theta_q + m\theta_k \quad (3)$$

每当一个 mini-batch 通过动量编码器转换为对应的哈希码, 它们将被用于更新负样本队列.

通过对比学习框架, 无监督深度哈希方法能更好地学习到图片之间的判别语义信息. 这能帮助模型生成更高质量的哈希码, 从而提升图像检索精度.

3.2 聚类中心的生成

为得出更加合适的极困难负样本抑制系数, 我们需要为哈希码确定所属的聚类中心. 通过计算哈希码与各个聚类中心的相似度, 选取相似度最高的聚类中心作为当前哈希码对应的聚类中心, 实现无监督条件下的伪标签分配.

受中心相似度量化 (central similarity quantization, CSQ)^[32] 启发, 本文使用哈达玛矩阵和伯努利分布两种方法为数据集生成预置聚类中心. 这两种方法不需要在训练过程中反复迭代更新也能保证预置聚类中心之间的合理分布, 相较于 K-means 等常规聚类算法大大降低了相关的时间成本.

作为一个 $K \times K$ 的矩阵, 哈达玛矩阵 H_K 具有任意两行或两列向量之间相互正交的性质, 则任意两行或两列向量之间的汉明距离 $D_H = K/2$, 距离分布均匀合理, 因此可以用于生成有效的预置聚类中心.

我们使用 Sylvester(·) 算法计算获得一个 $K \times K$ 的哈达玛矩阵 ($K = 2^n$), 该算法原理是使用克罗内克积将一个 2^{n-1} 阶的矩阵迭代为 2^n 阶的矩阵, 具体迭代公式如下所示:

$$H_K = \begin{bmatrix} H_{2^{n-1}} & H_{2^{n-1}} \\ H_{2^{n-1}} & -H_{2^{n-1}} \end{bmatrix} = H_2 \otimes H_{2^{n-1}} \quad (4)$$

其中, \otimes 代表克罗内克积, $H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$.

当所需的预置聚类中心数量 $m \leq K$, 我们可以直接选取矩阵的 m 行或 m 列作为预置聚类中心. 当所需的预置聚类中心数量 $K < m \leq 2K$, 则可以在构建两个哈达玛矩阵的组合 $H_{2K} = [H_K \quad -H_K]^T$ 后再选取预置聚类中心.

然而当 $m > 2K$ 或者 $K \neq 2^n$ 时, 我们就需要利用伯努利分布来生成预置聚类中心. 具体来说, 通过让预置聚类中心的每一位都从伯努利分布 $Bern(0.5)$ 中随机采样, 则每一位元素 x 都满足 $p(x=1)=p(x=0)=0.5$. 通过

将 0 置换为 -1, 即可获得每一位取值均为 1 或 -1 的预置聚类中心. 因为每一位元素均由独立随机采样获得, 所以由伯努利分布生成的聚类中心之间, 对应位置的元素不同的概率为 0.5. 此时任意两个预置聚类中心 c_i 与 c_j 之间的汉明距离的期望则为 $E(D_H(c_i, c_j)) = K \times 0.5 = K/2$, 这表明由伯努利分布采样获得的预置聚类中心同样分布均匀合理.

3.3 偏差抑制

在不断研究的过程中, 研究者们发现 InfoNCE 损失函数也存在着一些缺点. 在无监督的情况下, 负样本队列使用随机抽取的样本进行更新. 这个过程中难以避免将与查询样本同标签的其他样本抽取作为负样本, 这样的现象被称为采样偏差. 同标签的其他样本与查询样本有着相似的语义特征, 但却在嵌入空间中被错误地推远, 这将干扰同标签样本的聚类. 采样偏差的发生对对比学习产生了负面影响, 不利于模型学习到同类样本间的重要语义特征.

如果用余弦相似度的高低来衡量区分负样本时的困难程度, 则与查询样本最相似的负样本则会被称为极困难负样本. 被错误抽取为负样本的样本一般被称为错误负样本 (false negative samples). 这些错误负样本因为与查询样本的标签相同, 所以彼此之间拥有着较高的语义相似度, 也就满足了极困难负样本的定义.

在无监督的条件下, 通过标签找到错误负样本的做法不能成立, 但是通过相似度计算找到极困难负样本则是可行的方案. 于是本文做出假设, 认为在一定范围内的极困难负样本可以被近似认为是错误负样本. 通过对对象概念的近似转换, 达成在无监督条件下找到错误负样本的这一任务可能性得以提升. 基于这一假设, 本文以语义相似度排序的方式, 从负样本队列中找到极困难负样本, 并通过设计抑制系数 λ_s 来抑制这部分样本与查询样本计算得到的余弦相似度. 这样能降低损失函数对这部分负样本的关注度, 来起到缓解采样偏差负面影响的作用.

本文将抑制系数 λ_s 设计为缓慢单调递增的函数, 以实现对抑制范围内的较困难的负样本进行更多抑制. 除此以外, 本文还计算当前负样本 n_j 与它邻近的聚类中心 c_{nj} 的相似度, 以及当前负样本 n_j 与查询样本所属的聚类中心 c 的相似度, 通过计算二者的比值来衡量当前极困难负样本作为错误负样本的可能性. 只有负样本与查询样本都靠近同一聚类中心时才认为这个负样

本是完全可信的错误负样本,其余情况则需要适当减少对当前负样本的抑制。 λ_s 值越小,表示抑制作用越大。按降序对 sim_{neg} 进行排序,并将其与 λ_s 一一匹配。 λ_s 的计算公式如下:

$$x = j/(M \times r) \quad (5)$$

$$\lambda_b = 0.1 \times \min(0, 1 - sim(n_j, c_{nj}) / sim(n_j, c)) \quad (6)$$

$$\lambda_{sj} = \lambda_b + (1 - \exp(-2 \times x)) / 10 + hp_s \quad (7)$$

其中, x 是当前负样本下标 j 与受抑制的负样本总数的比率。 M 是负样本总数, r 是抑制范围。 λ_b 是根据负样本与聚类中心的相似度来修正抑制系数大小的修正项,其中 c_{nj} 表示负样本邻近的聚类中心, c 表示查询样本邻近的聚类中心。 λ_{sj} 表示与负样本 n_j 对应的抑制系数。 λ_s 超参数 hp_s 负责控制 λ_s 的抑制强度。

负样本越远离查询样本所属中心,则认为此负样本越不可能为错误负样本,对此负样本施加的抑制应降低。如式(6)所示,当负样本 n_j 与其他聚类中心相似度较高时,则 λ_b 取值较高,用以修正对此负样本的抑制效果。而当负样本 n_j 所邻近的聚类中心 c_{nj} 就是查询样本所属的中心 c 时,此时应认为该负样本高概率属于错误负样本,应该被施加抑制,则 λ_b 为0。如式(7)所示,修正项 λ_b 用以修正抑制系数 λ_{sj} , λ_b 取值越高,则抑制力度越小。除去修正项,抑制系数整体呈单调递增,因负样本队列已根据与查询样本的相似度进行降序排序,则排序靠前的负样本所对应的抑制系数取值较小,受到的抑制力度则较大。

抑制系数有效降低了错误负样本在对比损失函数中产生的影响,能缓解采样偏差问题,同时修正项也尽可能减少了抑制系数对正常负样本的影响。在抑制系数的设计中,超参数 hp_s 主要决定了抑制系数的取值下限,用以整体控制抑制系数的抑制力度,本文进行了对应的超参数取值分析实验以探究其取值影响,具体结论详见第4.5节。

偏差抑制算法的完整过程如算法1所示,偏差抑制算法的作用如图2所示,其中偏差抑制算法通过筛选极困难负样本并对其加以抑制,使其不被过度推离查询样本,更好地形成聚类。图2(a)为使用偏差抑制之前,图2(b)为使用偏差抑制之后。通过将偏差抑制算法应用在对比学习框架之中,本文提出了能更好地应对采样偏差问题的偏差抑制对比学习方法。

算法1. 偏差抑制算法

输入: 查询样本与负样本队列间相似度 sim_{neg} , 负样本的数量 M , 抑制范围 r , 抑制超参数 hp_s , 负样本邻近的聚类中心 c_{nj} , 查询样本邻近的聚类中心 c 。

输出: 抑制之后的查询样本与负样本队列间相似度 sim_{neg} 。

(1) 将 λ_s 初始化为长度为 M 的全1向量;

(2) repeat

$x \leftarrow j/(M \times r)$;

$\lambda_b \leftarrow 0.1 \times \min(0, 1 - sim(n_j, c_{nj}) / sim(n_j, c))$;

$\lambda_s[j] \leftarrow \lambda_b + (1 - \exp(-2 \times x)) / 10 + hp_s$;

until $i > M \times r$

(3) 对 sim_{neg} 进行降序排序;

(4) $sim_{neg}^* = \lambda_s$;

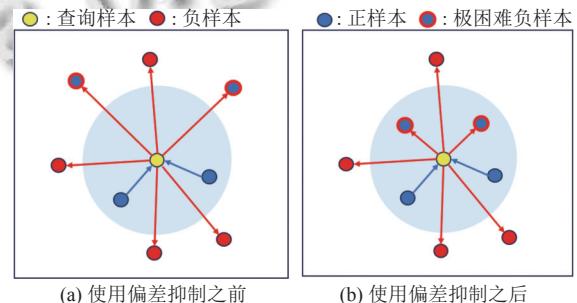


图2 使用偏差抑制算法的效果示意图

3.4 损失函数设计

在InfoNCE损失函数的基础上,为实现为对比学习框架进行偏差抑制这一目的,本文提出了SSNCE损失函数。

$$L_{SSNCE} = \frac{1}{N} \sum_{i=1}^N -\log \frac{e^{sim_{pos}}}{e^{sim_{pos}} + \sum_{j=0}^M e^{\frac{\lambda_{sj} h_i \cdot h_j}{\tau}}} \quad (8)$$

其中, h_i 代表查询样本的哈希码, h_i' 代表正样本的哈希码, n_j 代表负样本的哈希码, c_i 代表查询样本所属的聚类中心, λ_{sj} 代表负样本 n_j 对应的抑制系数, τ 代表温度超参数。

本文在训练过程中使用 $\tanh(\cdot)$ 替代 $\text{sgn}(\cdot)$ 对特征向量进行二值化产生哈希码。这一过程中将会产生一定的量化误差,因此还需要引入量化损失函数 L_Q 来约束哈希码 h 。

$$L_Q = \frac{1}{N} \sum_{i=1}^N \frac{1}{K} \sum_{j=1}^K (|ele_{i,j}| - 1)^2 \quad (9)$$

其中, N 为参与损失函数计算的哈希码数量, K 为哈希码的长度, $ele_{i,j}$ 是哈希码 h_i 的第 j 位元素。最终总体损失函数 L 如式(10)所示。

$$L = L_{SSNCE} + L_Q \quad (10)$$

基于总体损失函数 L , 本文使用动量编码器与负样本队列的对比学习框架对哈希码编码器进行训练。在此过程中, 本文所提出的 BSCDH 方法所采用的偏差抑制对比学习方法能缓解对比学习框架因负样本随机采样所带来的采样偏差问题。

4 实验

4.1 数据集

(1) CIFAR-10 由来自 10 个类别的 60 000 张图像构成。本文从每个类中随机选取 1 000 张图像作为测试查询集, 而检索集和训练集会从剩下的 50 000 张图像中选取。

(2) NUS-WIDE 是一个由来自 81 个类别的将近 270 000 张图像构成的多标签图像数据集。本文使用 21 个最受欢迎的类别进行评估, 从每个类中随机选取 100 张图像作为测试查询集, 而将剩下的图像作为检索集。除此之外, 本文会在检索集中的每个类中随机选取 500 张图像组成训练集。

(3) FLICKR25K 由来自 24 个类别的 25 000 张图像构成。本文随机抽取 1 000 张图像作为测试查询集, 将剩余的图像作为检索集, 并从检索集中随机抽取 10 000 张图像作为训练集。

4.2 实验设置

(1) 评价指标: 在本文的实验中, 我们主要采用 Top N 的平均检索精度 (mAP@ N) 衡量所获得的哈希码的质量。我们将以 mAP@5000 作为评价标准, 对 CIFAR-10、NUS-WIDE 和 FLICKR25K 这 3 个数据集的检索结果进行评价。

表 1 在 CIFAR-10, FLICKR25K, NUS-WIDE 这 3 个数据集上的 mAP@5000 对比结果

方法	年份	CIFAR-10			FLICKR25K			NUS-WIDE		
		16位	32位	64位	16位	32位	64位	16位	32位	64位
SSDH ^[21]	2018	0.322	0.339	0.348	0.751	0.763	0.779	0.746	0.762	0.774
BGAN ^[33]	2018	0.535	0.575	0.587	0.766	0.770	0.795	0.719	0.745	0.761
DistillHash ^[34]	2019	0.348	0.361	0.370	0.768	0.789	0.794	0.761	0.774	0.781
TBH ^[18]	2020	0.457	0.479	0.495	0.779	0.794	0.797	0.717	0.725	0.735
CIBH ^[8]	2021	0.564	0.603	0.622	0.773	0.781	0.798	0.756	0.777	0.781
DSCH ^[22]	2022	0.628	0.657	0.682	0.817	0.827	0.828	0.770	0.792	0.801
DDCH ^[12]	2023	0.583	0.629	0.639	—	—	—	0.781	0.798	0.808
S2H ^[11]	2023	0.604	0.648	0.663	—	—	—	0.795	0.805	0.806
UDTrHash ^[35]	2024	—	—	—	0.806	0.811	0.815	0.718	0.736	0.738
BSCDH (ours)	—	0.639	0.689	0.696	0.821	0.826	0.833	0.786	0.803	0.819

在 CIBH、DDCH、S2H 等基于对比学习的方法中, 我们发现同样关注采样偏差问题的 DDCH, 相比

(2) 对比方法: 在本文中, 我们将与下列的无监督哈希方法进行对比: SSDH^[21]、BGAN^[33]、DistillHash^[34]、TBH^[18]、CIBH^[8]、DSCH^[22]、DDCH^[12]、S2H^[11]、UDTrHash^[35]。

(3) 训练细节: 本文基于 PyTorch 框架实现 BSCDH 方法。对于 3 个数据集的图像, 我们都将其大小处理为 $224 \times 224 \times 3$ 。在训练过程中, 我们使用了 CIBH 中的数据增强方案将图像转换为不同的视图并输入到编码器网络中, 该方案结合了随机裁剪、随机水平翻转、随机施加的图像灰度化和颜色抖动以及高斯模糊。在我们的实验中, 编码器网络由预训练的 VGG-16 模型与新添加的前馈神经网络构成。在超参数设置方面, 我们使用优化器 SGD 进行优化, batch size 设置为 64, 最大 epoch 数设置为 200, 学习率设置为 0.01, 温度超参设置为 0.3。

4.3 对比实验结果

在本节中, 本文将与其他无监督深度哈希图像检索方法进行 mAP@5000 指标的对比, 以证明 BSCDH 具有更好的检索性能。在表 1 中, 本文展示了不同方法在 CIFAR-10、FLICKR25K 和 NUS-WIDE 上的 16 位、32 位、64 位哈希码的 mAP@5000 结果。在 3 个不同数据集上, BSCDH 的检索性能表现优秀。在 CIFAR-10 数据集上, 我们的方法相较于次优的方法 DSCH 提高了 1.75% (16 位)、4.87% (32 位)、2.05% (64 位)。在 FLICKR25K 数据集上, 我们的方法相较 DSCH 分别提升了 0.49% (16 位)、0.6% (64 位)。而在 NUS-WIDE 数据集上, 我们的方法相较 DSCH 则提升了 2.07% (16 位)、1.38% (32 位)、1.99% (64 位)。

图像检索方法的检索性能的。而本文提出的方法 BSCDH, 与以上基于对比学习的方法相比, 都有着明显的性能优势, 我们认为这是因为我们所提出的偏差抑制对比学习方法可以更好地应对采样偏差问题。

为进一步证明这一点, 本文与 DDCH、DCL 等讨论采样偏差问题的对比学习方法在 CIFAR-10 和 NUS-WIDE 两个数据集上开展了 64 位哈希码实验, 以对比图像检索性能。其中, 因为 DCL 方法本身并未应用于图像检索任务, 于是本文在对比实验采取了和本文方法 BSCDH 一致的模型设置, 仅将本文采用的偏差抑制对比学习方法替换为 DCL 方法进行训练。

由表 2 可知, 我们的方法 BSCDH 相较于另外两种针对采样偏差提出的方法在图像检索方面表现更佳。首先, 这可能是因为偏差抑制将极困难负样本近似认为是错误负样本并加以抑制的做法, 相比起通过修改损失函数来间接降低错误负样本影响的方法更为直接有效。除此以外, 我们在抑制系数的计算公式中考虑了负样本相较于前置聚类中心的位置关系, 从而能为当前负样本给出更合适的抑制系数。我们推测这可能是使我们的方法表现更佳的原因之一。

表 2 讨论采样偏差问题的不同方法在 CIFAR-10, NUS-WIDE 两个数据集上的 mAP@5000 对比结果

方法	CIFAR-10	NUS-WIDE
BSCDH (ours)	0.696	0.819
DCL	0.685	0.815
DDCH	0.639	0.808

4.4 对于第 3.3 节中所提出的假设的实验验证

在第 3.3 节中, 本文提出了极困难负样本可以被近似认为是错误负样本的假设, 并基于这一假设提出了偏差抑制对比学习方法。为验证这一假设的可信度, 我们统计了在训练过程的不同 epoch 中, 被偏差抑制算法所抑制的极困难负样本中真正的错误负样本所占的比例, 以证明二者的近似关系。

在表 3 中, 我们发现在 3 个数据集的实验中, 错误负样本在偏差抑制算法所抑制的极困难负样本中所占到的比例普遍在 75%–90% 范围之间。较高的比例证明了极困难负样本可以被近似认为是错误负样本, 而基于这一假设所提出的偏差抑制算法在普遍情况下可以成立。

除此之外, 我们发现错误负样本在受抑制的极困难负样本中的比例会随着训练的进行而逐渐增加。这

是因为模型的性能在训练过程中不断提升, 使得哈希码更能反映原本图像间的相似性关系。因此, 与查询样本标签相同的错误负样本, 现在会更容易因为较高的相似度被判定为极困难负样本。这也就意味着随着训练进行, 将极困难负样本近似认为是错误负样本这一假设会更加可信。

表 3 在不同 epoch 中错误负样本在受抑制的极困难负样本中所占的百分比 (%)

epoch	CIFAR-10	FLICKR25K	NUS-WIDE
50	75.97	82.31	85.64
100	76.65	85.63	87.42
150	77.48	87.39	90.11

4.5 超参数设置分析

在本文所提出的偏差抑制对比学习中, 使用了超参数 hp_s 来控制偏差抑制算法所得出的抑制系数对极困难负样本的抑制力度。在本节中, 将在 CIFAR-10、FLICKR25K 和 NUS-WIDE 上开展 64 位哈希码实验, 并分析不同超参数 hp_s 取值对 BSCDH 所带来的影响。

由表 4 可知, 在 3 个不同数据集上, BSCDH 方法的 mAP@5000 均会随着 hp_s 的减小而先上升再下降, 并在 $hp_s = 0.7$ 处取得最优值。 hp_s 减小代表着偏差抑制算法对极困难负样本的抑制力度加大。因此我们推测过小的抑制力度会使该算法无法有效发挥缓解采样偏差现象的作用, 而过大的抑制力度会干扰模型从困难负样本中学习相似性关系的过程, 导致训练效果表现不佳。

表 4 在不同 hp_s 取值下, BSCDH 的 mAP@5000 结果

epoch	CIFAR-10	FLICKR25K	NUS-WIDE
$hp_s = 0.9$	0.675	0.816	0.808
$hp_s = 0.8$	0.679	0.827	0.811
$hp_s = 0.7$	0.696	0.833	0.819
$hp_s = 0.6$	0.685	0.823	0.815
$hp_s = 0.5$	0.682	0.821	0.816

5 总结

本文提出了基于偏差抑制对比学习的无监督深度哈希图像检索方法 BSCDH。我们考虑到极困难负样本与错误负样本之间的关联, 提出了偏差抑制对比学习缓解采样偏差问题为对比学习带来的负面影响。我们通过在 CIFAR-10、FLICKR25K 和 NUS-WIDE 这 3 个数据集上开展大量实验, 证明了 BSCDH 方法在无

监督图像检索任务的有效性。

接下来，我们的工作是进一步优化抑制系数公式的设计，以实现在无监督条件下更准确地评估每个负样本潜在的采样偏差负面影响。

参考文献

- 1 Luo X, Wang HX, Wu DQ, *et al.* A survey on deep hashing methods. *ACM Transactions on Knowledge Discovery from Data*, 2023, 17(1): 15.
- 2 Weiss Y, Torralba A, Fergus R. Spectral hashing. *Proceedings of the 21st International Conference on Neural Information Processing Systems*. Vancouver: Curran Associates Inc., 2008. 1753–1760.
- 3 Gong YC, Lazebnik S, Gordo A, *et al.* 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. [doi: [10.1109/TPAMI.2012.193](https://doi.org/10.1109/TPAMI.2012.193)]
- 4 Kaya M, Bilge H\$. Deep metric learning: A survey. *Symmetry*, 2019, 11(9): 1066. [doi: [10.3390/sym11091066](https://doi.org/10.3390/sym11091066)]
- 5 Chen T, Kornblith S, Norouzi M, *et al.* A simple framework for contrastive learning of visual representations. *Proceedings of the 37th International Conference on Machine Learning*. PMLR, 2020. 149.
- 6 Albelwi S. Survey on self-supervised learning: Auxiliary pretext tasks and contrastive learning methods in imaging. *Entropy*, 2022, 24(4): 551. [doi: [10.3390/e24040551](https://doi.org/10.3390/e24040551)]
- 7 Jaiswal A, Babu AR, Zadeh MZ, *et al.* A survey on contrastive self-supervised learning. *Technologies*, 2020, 9(1): 2. [doi: [10.3390/technologies9010002](https://doi.org/10.3390/technologies9010002)]
- 8 Qiu ZX, Su QL, Ou ZJ, *et al.* Unsupervised hashing with contrastive information bottleneck. *Proceedings of the 30th International Joint Conference on Artificial Intelligence*. Montreal: ijcai.org, 2021. 959–965.
- 9 Luo X, Ma ZY, Cheng W, *et al.* Improve deep unsupervised hashing via structural and intrinsic similarity learning. *IEEE Signal Processing Letters*, 2022, 29: 602–606. [doi: [10.1109/LSP.2022.3148674](https://doi.org/10.1109/LSP.2022.3148674)]
- 10 Mikriukov G, Ravanbakhsh M, Demir B. Unsupervised contrastive hashing for cross-modal retrieval in remote sensing. *Proceedings of the 2022 IEEE International Conference on Acoustics, Speech and Signal Processing*. Singapore: IEEE, 2022. 4463–4467.
- 11 Lee W, Byun SY, Park M. Unsupervised soft-to-hard hashing with contrastive learning. *Computer Vision and Image Understanding*, 2023, 233: 103713. [doi: [10.1016/j.cviu.2023.103713](https://doi.org/10.1016/j.cviu.2023.103713)]
- 12 Wei RK, Liu Y, Song JK, *et al.* Deep debiased contrastive hashing. *Pattern Recognition*, 2023, 139: 109483. [doi: [10.1016/j.patcog.2023.109483](https://doi.org/10.1016/j.patcog.2023.109483)]
- 13 Chuang CY, Robinson J, Lin YC, *et al.* Debiased contrastive learning. *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Vancouver: Curran Associates Inc., 2020. 735.
- 14 Wang JP, Zeng ZY, Chen B, *et al.* Contrastive quantization with code memory for unsupervised image retrieval. *Proceedings of the 36th AAAI Conference on Artificial Intelligence*. AAAI Press, 2022. 2468–2476.
- 15 Dai B, Guo RQ, Kumar S, *et al.* Stochastic generative hashing. *Proceedings of the 34th International Conference on Machine Learning*. Sydney: PMLR, 2017. 913–922.
- 16 Shen YM, Liu L, Shao L. Unsupervised binary representation learning with deep variational networks. *International Journal of Computer Vision*, 2019, 127(11–12): 1614–1628. [doi: [10.1007/s11263-019-01166-4](https://doi.org/10.1007/s11263-019-01166-4)]
- 17 Do TT, Doan AD, Cheung NM. Learning to hash with binary deep neural network. *Proceedings of the 14th European Conference on Computer Vision*. Amsterdam: Springer, 2016. 219–234.
- 18 Shen YM, Qin J, Chen JX, *et al.* Auto-encoding twin-bottleneck hashing. *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020. 2818–2827.
- 19 Dizaji KG, Zheng F, Sadoughi Nourabadi N, *et al.* Unsupervised deep generative adversarial hashing network. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 3664–3673.
- 20 Lin K, Lu JW, Chen CS, *et al.* Learning compact binary descriptors with unsupervised deep neural networks. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016. 1183–1192.
- 21 Yang EK, Deng C, Liu TL, *et al.* Semantic structure-based unsupervised deep hashing. *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. Stockholm: AAAI Press, 2018. 1064–1070.
- 22 Lin QH, Chen XJ, Zhang Q, *et al.* Deep unsupervised hashing with latent semantic components. *Proceedings of the 36th AAAI Conference on Artificial Intelligence*. AAAI Press, 2022. 7488–7496.
- 23 Wu ZR, Xiong YJ, Yu SX, *et al.* Unsupervised feature

- learning via non-parametric instance discrimination. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 3733–3742.
- 24 Ye M, Zhang X, Yuen PC, et al. Unsupervised embedding learning via invariant and spreading instance feature. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 6210–6219.
- 25 He KM, Fan HQ, Wu YX, et al. Momentum contrast for unsupervised visual representation learning. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 9729–9738.
- 26 Misra I, van der Maaten L. Self-supervised learning of pretext-invariant representations. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 6707–6717.
- 27 Zhuang CX, Zhai A, Yamins D. Local aggregation for unsupervised learning of visual embeddings. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 6002–6012.
- 28 Hjelm RD, Fedorov A, Lavoie-Marchildon S, et al. Learning deep representations by mutual information estimation and maximization. Proceedings of the 7th International Conference on Learning Representations. New Orleans: OpenReview.net, 2019.
- 29 van den Oord A, Li YZ, Vinyals O. Representation learning with contrastive predictive coding. arXiv:1807.03748, 2018.
- 30 Tian YL, Krishnan D, Isola P. Contrastive multiview coding. Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer, 2020. 776–794.
- 31 Cai TT, Frankle J, Schwab DJ, et al. Are all negatives created equal in contrastive instance discrimination? arXiv:2010.06682, 2020.
- 32 Yuan L, Wang T, Zhang XP, et al. Central similarity quantization for efficient image and video retrieval. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 3083–3092.
- 33 Song JK, He T, Gao LL, et al. Binary generative adversarial networks for image retrieval. Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans: AAAI Press, 2018. 394–401.
- 34 Yang EK, Liu TL, Deng C, et al. DistillHash: Unsupervised deep hashing by distilling data pairs. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 2946–2955.
- 35 Meng LT, Zhang QY, Yang R, et al. Unsupervised deep triplet hashing for image retrieval. IEEE Signal Processing Letters, 2024, 31: 1489–1493. [doi: [10.1109/LSP.2024.3404350](https://doi.org/10.1109/LSP.2024.3404350)]

(校对责编: 张重毅)