



基于高阶统计信息的深度哈希学习模型

顾 岩, 赵崇宇, 黄 平

(太原理工大学 物理与光电工程学院, 山西 晋中 030600)

摘 要: 深度哈希因其检索效率和存储代价上的优势而被广泛应用于大规模图像检索领域。为增强哈希编码的区分能力并提高检索准确率和效率, 建立一种基于高阶统计信息的深度哈希学习模型 BCI-DHH。采用改进的 VGG-m 模型分别提取输入图像基于层内的自相关特征和基于层间的互相关特征, 并生成归一化的高阶统计向量。通过引入权重参数对训练样本中的正负样本数目进行平衡, 提出一种基于数据平衡性的对比损失函数。在此基础上, 对不相似图像对之间对应的多级索引哈希块进行差异化操作, 增大不相似图像与其查询图像之间的汉明距离, 优化多级哈希索引的兼容性。在基准数据集上的实验结果表明, 该模型在检索准确率和效率方面优于 BDH、DSH 等方法。

关键词: 深度哈希; 图像检索; 哈希学习; 高阶统计; 对比损失; 多级索引

开放科学(资源服务)标志码(OSID):



中文引用格式: 顾岩, 赵崇宇, 黄平. 基于高阶统计信息的深度哈希学习模型[J]. 计算机工程, 2020, 46(7): 260-267, 276.

英文引用格式: GU Yan, ZHAO Chongyu, HUANG Ping. Deep hash learning model based on high-order statistical information[J]. Computer Engineering, 2020, 46(7): 260-267, 276.

Deep Hash Learning Model Based on High-Order Statistical Information

GU Yan, ZHAO Chongyu, HUANG Ping

(College of Physics and Optoelectronics, Taiyuan University of Technology, Jinzhong, Shanxi 030600, China)

[Abstract] Deep hashing has been widely used in large-scale image retrieval for its advantage in retrieval efficiency and storage cost. To enhance the identification ability of hash code and improve the retrieval accuracy and efficiency, this paper proposes a deep hash learning model, BCI-DHH, based on high-order statistical information. The improved VGG-m model is employed to extract intra-layer auto-correlation features and inter-layer cross-correlation features from input images, and to generate a normalized high-order statistical vector. Then the weighting parameters are introduced to balance the number of positive and negative samples, and on this basis a contrastive loss function based on data balance is proposed. Then multi-level index hash blocks corresponding to dissimilar image pairs are differentiated to increase the hamming distance between the dissimilar image and the to-be-retrieved image, so as to optimize the compatibility of multi-level hash index. Experimental results on the benchmark datasets demonstrate that the proposed model outperforms BDH, DSH and other methods in terms of retrieval accuracy and efficiency.

[Key words] deep hash; image retrieval; hash learning; higher-order statistics; contrastive loss; multi-level index

DOI: 10.19678/j.issn.1000-3428.0055259

0 概述

面向大规模图像的近似最近邻(Approximate Nearest Neighbor, ANN)检索方法已经成为学术界和工业界的研究热点之一^[1-2], 哈希方法由于其特有的高查询速度和低存储代价而被广泛应用于 ANN 检

索领域。现有的哈希学习方法大致可以分为两类, 即数据独立哈希和数据相关哈希^[3-4]。其中, 数据独立哈希在训练过程中通常不依赖任何数据集, 其采用随机映射的方式进行哈希映射函数的学习, 而数据相关哈希通过训练数据集以学习哈希函数, 因此其又被称为学习哈希(Learning to Hash, L2H)^[5]。

基金项目: 山西省自然科学基金(201801D121020, 201801D221132)。

作者简介: 顾 岩(1992—), 男, 硕士研究生, 主研方向为深度学习、机器学习、数据分析与挖掘; 赵崇宇, 本科生; 黄 平(通信作者), 教授。

收稿日期: 2019-06-20

修回日期: 2019-07-22

E-mail: huangpinghuangka@163.com

根据是否利用训练样本的监督信息,学习哈希又可以进一步分为监督哈希、半监督哈希和无监督哈希^[6-8]。传统的学习哈希由于输入的手动特征在表达深层语义信息方面的局限性,导致其哈希检索的性能受限。随着深度神经网络技术的深入研究,许多研究人员采用卷积神经网络(Convolutional Neural Networks, CNN)等模型对特征进行自动提取,并提出了较多基于深度学习的哈希方法,大幅提高了哈希检索的性能^[9]。文献[10]利用 CNN 自动提取图像的深层语义特征,并在此基础上采用坐标下降方法对相似矩阵进行分解,但是该学习过程是基于两阶段的,难以进行端到端的训练。文献[11]采用三元组思想设计一种优化的排序损失函数,并将特征学习和哈希学习处于同一框架下,以对模型进行端到端的训练。文献[12]采用对比损失的方式来优化哈希学习过程,提高了哈希检索的性能。文献[13]提出了一种基于多标签的多层语义相似度排序方法,其采用代理损失的方式来优化哈希码学习过程。文献[14]通过 CNN 对图像进行特征提取,并设计损失函数对相似图像对和不相似图像对之间的汉明距离分别进行逼近和惩罚。

在现有基于深度哈希方法的训练数据集中,与目标图像相似的图像数量远小于不相似的图像数量,这导致了整个训练数据集的不平衡性^[15],同时造成在训练过程中模型对负样本学习的过拟合和对正样本学习的欠拟合现象。因此,对不平衡的样本数据进行哈希学习容易导致训练得到的模型泛化能力差,从而降低哈希检索的准确率。此外,大部分学习哈希注重损失优化的设计,忽略了对图像深层语义特征的表示,然而高阶细粒度特征不仅可以保持高层的语义相似度,而且可以提高所生成的哈希编码的区分能力^[16-17]。为提高哈希检索的时间效率,现有的多数深度哈希学

习方法在对不相似图像进行哈希学习时,通常采用多级索引的思想生成哈希编码,即将生成的哈希编码分解为多个连续且不相交的哈希块,并为每个哈希块创建一个单独的块索引^[18]。在检索过程中,将查询图像的哈希块与候选图像的哈希块进行匹配,并对块匹配不成功的候选图像进行过滤,在此基础上,生成候选图像集并根据汉明距离实现排序^[19]。然而,这种方式生成的哈希块内的哈希编码往往不均匀,可能会导致不相似图像的哈希索引块与相似图像的哈希索引块具有相同的哈希编码。假设将图像哈希编码划分为 m 个哈希块单元,每个块内有 n bit 的哈希码,传统的多级哈希检索由于编码的不均匀性,导致查询图像、相似图像以及不相似图像的哈希块单元内具有相同的哈希编码,这将大幅增加查询过程中候选图像的数量,降低多级索引检索的效率。

本文建立一种基于高阶统计信息的深度哈希学习模型(BCI-DHH)。采用改进的 VGG-m 模型提取输入图像基于层内的自相关特征以及基于层间的互相关特征,在此基础上生成归一化的高阶统计向量。为防止模型训练过程中由于数据的不平衡性导致对负样本学习的过拟合和对正样本学习的欠拟合现象,提出一种基于平衡权重的对比损失优化方法。为降低检索的时间复杂度,对不相似图像对之间的哈希块进行差异化学习,增大其与目标图像之间的汉明距离,从而减少候选图像的数量并提高检索效率。

1 网络架构

本文 BCI-DHH 模型将基于图像的高阶统计学习和哈希学习集成在统一框架下,并采用端到端的方式进行训练,整个模型的架构如图 1 所示。

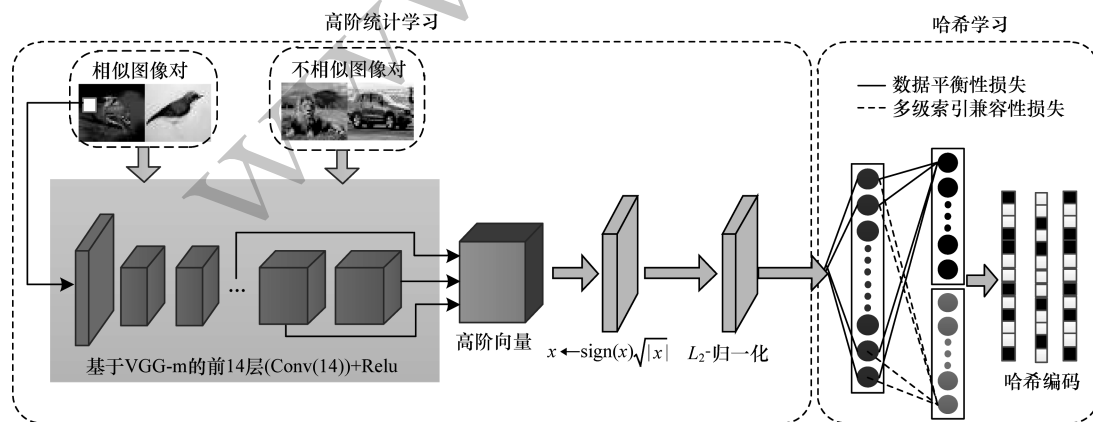


图1 BCI-DHH 模型总体框架

Fig. 1 Overall framework of BCI-DHH model

BCI-DHH 模型主要由高阶统计学习和哈希学习 2 个部分构成。在高阶统计学习部分,为了更高效地提取细粒度深层语义特征,模型同时获取了基于层内的自相关信息和基于层间的互相关信息,并在此基础上生成归一化的高阶统计向量。在哈希学习部分,首先对输入的训练图像对进行平衡优化,对不相似图像对的训练采用具有兼容性的多级索引进行优化,从而提高了哈希编码的检索效率和准确率。由于本文的目的在于研究基于高阶统计信息的深度哈希检索的有效性,因此仅采用基于 VGG-16 模型的前 14 层以及 Relu 激活函数作为特征提取的基本架构,其他 CNN 模型也可以用来验证本文方法的有效性。

1.1 问题定义

假设存在 N 张图像 $I = \{I_i\}_{i=1}^N$, 其中, I_i 表示数据集中的第 i 张图像, N 是数据集中图像的数目。在深度监督哈希学习中,一般采用 $s_{ij} \in S$ 表示图像对 I_i 和 I_j 之间的语义相似性, $s_{ij} = 1$ 表示输入的图像对相似, $s_{ij} = 0$ 表示输入的图像对不相似。哈希学习的目的是构建一个映射函数 $F: I \rightarrow B$, 从而实现图像 I 到哈希编码 $B = \{b_i\}_{i=1}^N$ 的基于语义保持的高效映射, 其中, $b_i \in \{-1, 1\}^k$, k 表示哈希编码的长度。本文在 CNN 的基础框架上增加哈希层, 从而产生有效的哈希编码。

与已有研究不同, 本文模型输入到哈希层的图像特征是基于归一化的高阶特征, 从而使得产生的哈希编码更加具有区分能力^[18-20]。在 BCI-DHH 模型中, 假设第 i 张图像的高阶统计信息 f_i 可以由式(1)得到:

$$f_i = \phi(I_i; \Theta) \quad (1)$$

其中, ϕ 表示可以进行归一化高阶统计学习的网络架构, Θ 表示网络参数。BCI-DHH 模型中哈希层输出的哈希编码可以由式(2)获得:

$$h_i = \text{sign}(\phi(W_h^T f_i + v_h)) = \text{sign}(\phi(W_h^T \phi(I_i; \Theta) + v_h)) \quad (2)$$

其中, $\phi(\cdot)$ 表示双曲正切函数, 即 $\phi(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$,

W_h 和 v_h 分别表示权重和偏置项。

1.2 高阶统计学习

为获取高层细粒度的语义特征, 多数学者采用双线性的卷积神经网络(Bilinear Convolutional Neural Networks, B-CNN)模型对图像进行深层语义特征提取, 以进行哈希学习并生成高质量的哈希编码。如图 2(a)所示, B-CNN 需要采用 2 个单独的 CNN 模型分别获取统计向量。然而, B-CNN 仅能获取不同层之间的互相关特征而忽略了层内的自相关特征,

相关研究也证明了自相关信息对生成具有区分能力的细粒度特征描述符具有重要作用^[16-17]。在本文 BCI-DHH 模型中, 对输入的图像分别提取层内的自相关特征和层间的互相关特征, 从而生成基于高阶统计信息的特征向量(如图 2(b)所示), 并在此基础上进行深度监督哈希学习。

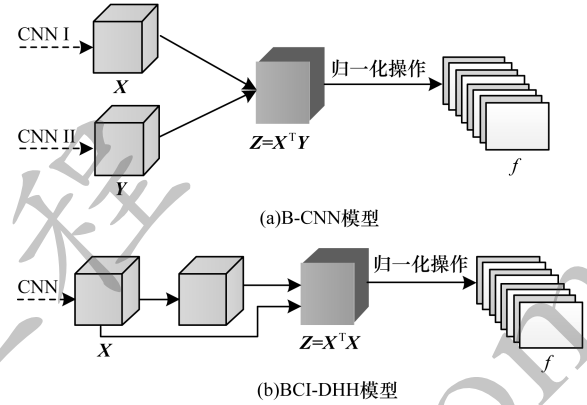


图 2 高阶统计学习示意图

Fig. 2 Schematic diagram of high-order statistical learning

在 B-CNN 模型架构中, 通过 2 个 CNN 对输入的图像分别进行局部特征提取, 得到特征矩阵 X 和 Y , 在此基础上进行双线性池化操作得到高阶特征 Z , 双线性池化操作具体如式(3)所示:

$$Z = X^T Y \quad (3)$$

其中, $X \in \mathbb{R}^{n \times d}$ 和 $Y \in \mathbb{R}^{n \times d}$ 分别表示通过 CNN I 和 CNN II 架构产生的特征矩阵, n 和 d 表示特征的数量和维度, 具体如图 2(a)所示。由于不同的卷积层和卷积操作能够获取不同级别和不同尺度的特征, 并且 CNN 网络中位置靠前的卷积层可以捕获更加精细的空间位置信息, 因此本文高阶统计学习不仅考虑层内的自相关特征, 而且获取了层间的互相关特征。此外, 本文采用单一的 CNN 模型进行特征学习, 从而大幅减少学习的代价, 具体如图 2(b)所示。采用单一的 CNN 模型可以提取到特征矩阵 X , 同样可以通过式(3)得到高阶统计特征, 在此基础上进行归一化操作从而产生用于哈希学习的图像特征描述符, 并将其输入到与损失函数相连接的哈希层, 以对网络进行端到端的训练。

在反向传播中, 假设存在损失函数 L 对高阶统计信息 Z 的梯度 $\frac{\partial L}{\partial Z}$, 则损失函数 L 对一阶统计特征 X 和 Y 的偏导数分别为 $\frac{\partial L}{\partial X}$ 和 $\frac{\partial L}{\partial Y}$, 两者可以通过如下的链式法则进行计算:

$$\frac{\partial L}{\partial X} = Y \left(\frac{\partial L}{\partial Z} \right)^T, \quad \frac{\partial L}{\partial Y} = X \left(\frac{\partial L}{\partial Z} \right)^T \quad (4)$$

在对本文模型进行反向传播的过程中, 基于式(3)和式(4), 损失函数 L 对单一 CNN 模型获取到的一阶统计特征 X 的偏导数 $\frac{\partial L}{\partial X}$ 可以通过式(5)计算:

$$\frac{\partial L}{\partial X} = X \left(\left(\frac{\partial L}{\partial Z} \right)^T + \left(\frac{\partial L}{\partial Z} \right) \right) \quad (5)$$

1.3 归一化操作

文献[21]研究表明, 对高阶统计特征进行归一化操作可以提高其哈希检索的性能。因此, 在进行哈希学习之前, 本文模型分别对学习得到的高阶特征描述符进行带符号的平方根操作 ($x \leftarrow \text{sign}(x) \sqrt{|x|}$) 和 L_2 -归一化操作。由于上述归一化操作都是可微的, 因此整个网络模型可以通过反向传播进行端到端的训练。

2 损失函数

为生成具有区分能力的哈希编码, 需要对模型的损失函数进行优化设计。模型通常采用损失函数来计算目标图像和查询图像之间的汉明距离, 从而生成最优的哈希编码。因此, 损失函数的设计对于整个哈希检索过程至关重要。本文在提取到的基于归一化高阶统计特征的基础上, 针对哈希学习中容易出现的数据不平衡性和多级索引的不兼容性, 提出一种更加有效的面向深度哈希学习的损失函数, 从而生成更具区分能力的哈希编码。

2.1 数据平衡性损失

假设存在图像对 $\{i_i, i_j\} \in I$ 及其对应输出的哈希编码 $\{b_i, b_j\} \in \{-1, +1\}^k$, 基于图像对的损失函数定义如下:

$$L(b_i, b_j, s_{ij}) = \frac{1}{2} \sum_{s_{ij} \in S} \{s_{ij} \cdot \max(D_h(b_i, b_j) - m_0, 0) + (1 - s_{ij}) \cdot \max(m_1 - D_h(b_i, b_j), 0)\} \\ \text{s.t. } b_i, b_j \in \{-1, +1\}^k, i, j \in \{1, 2, \dots, N\} \quad (6)$$

其中, $D_h(\cdot, \cdot)$ 表示 2 个哈希编码之间的汉明距离。 $m_0 > 0$ 和 $m_1 > 0$ 分别表示边缘阈值参数, $s_{ij} = 1$ 表示式(6)中第一项对汉明距离大于 m_0 的相似图像 i_i 和 i_j 进行惩罚, $s_{ij} = 0$ 表示式(6)中第二项对汉明距离小于 m_1 的不相似图像 i_i 和 i_j 进行惩罚, 由此可以保证所生成哈希编码的区分能力。

为防止模型在训练过程中由于数据的不平衡性而导致对负样本进行拟合学习的现象, 在式(6)的基础上引入权重 $w_{i,j}$, 对训练数据集中的正负样本数目进行平衡, 具体表示如下:

$$w_{i,j} = \begin{cases} \frac{|S_0|}{|S_0| + |S_1|}, s_{ij} = 1 \\ \frac{|S_1|}{|S_0| + |S_1|}, s_{ij} = 0 \end{cases} \quad (7)$$

其中, $S_1 = \{(x_i, x_j) \mid s_{ij} \in S \cap s_{ij} = 1\}$ 表示相似图像对, $S_0 = \{(x_i, x_j) \mid s_{ij} \in S \cap s_{ij} = 0\}$ 表示不相似图像对。基于此, 式(6)可以重写如下:

$$L(b_i, b_j, s_{ij}) = \frac{w_{i,j}}{2} \sum_{s_{ij} \in S} \{s_{ij} \cdot \max(D_h(b_i, b_j) - m_0, 0) + (1 - s_{ij}) \cdot \max(m_1 - D_h(b_i, b_j), 0)\} \\ \text{s.t. } b_i, b_j \in \{-1, +1\}^k, i, j \in \{1, 2, \dots, N\} \quad (8)$$

在式(8)中, 由于 $b_i, b_j \in \{-1, +1\}^k$ 为离散化值, 因此整个模型不能直接进行反向传播与训练。为提高整个模型的收敛速度, 本文采用具有连续属性的欧式距离来替代离散化的汉明距离。此外, 为了减少量化误差, 通过增加正则项使得模型的输出值达到离散化 $(-1/+1)^{[14]}$ 。因此, 基于正则化的数据平衡性损失函数 l_w 如下:

$$l_w = L(b_i, b_j, s_{ij}) = \frac{w_{i,j}}{2} \sum_{s_{ij} \in S} \{s_{ij} \cdot \max(\|b_i - b_j\|_2^2 - m_0, 0) + (1 - s_{ij}) \cdot \max(m_1 - \|b_i - b_j\|_2^2, 0)\} + \alpha (\|b_i\|_1 - 1 + \|b_j\|_1 - 1) \quad (9)$$

其中, α 为超参数, 其表示正则化的程度, $\|\cdot\|_2^2$ 表示 L_2 范数, $\|\cdot\|_1$ 表示 L_1 范数, $|\cdot|$ 表示按位的绝对值操作, $\mathbf{1}$ 表示所有元素为 1 的向量。

2.2 多级索引兼容性损失

除了训练数据集的不平衡性之外, 本文模型的主要目的还在于尽量减少对不相似图像进行匹配时的哈希块数目, 构建兼容性多级哈希索引, 从而提高多级哈希检索的效率。多级哈希索引的基本思想是将哈希编码划分为 b 个哈希索引块, 在此基础上, 本文模型对不相似图像对之间的对应哈希块进行差异化操作, 以增加其与查询图像之间的哈希距离。具体而言, 假设存在 b'_i 和 b'_j 分别表示不相似图像对 i_i 和 i_j 对应哈希编码的第 t 个哈希块, 则数据平衡性损失中不相似图像对 i_i 和 i_j 的多级哈希索引的兼容性损失 l_{index} 定义如下:

$$l_{\text{index}} = \frac{w_{i,j}}{2} \sum_{s_{ij} \in S} (1 - s_{ij}) \sum_{t=1}^b \max(m_1 - \|b'_i - b'_j\|_2^2, 0) \\ \text{s.t. } b_i, b_j \in \{-1, +1\}^{k/b} \quad (10)$$

从式(10)可以看出, 兼容性损失使得不相似图像对的哈希编码中每个哈希块的哈希距离增大为 m_1 , 从而减少了检索过程中候选图像的数量, 提高了检索效率。

综合考虑模型训练过程中数据平衡性和多级哈希索引的兼容性,本文 BCI-DHH 模型对应的目标函数 $l_{w-index}$ 如下:

$$l_{w-index} = \frac{w_{i,j}}{2} \sum_{s_{ij} \in S} \{s_{ij} \cdot \max(\|b_i - b_j\|_2^2 - m_0, 0) + \beta \sum_{s_{ij} \in S} (1 - s_{ij}) \sum_{t=1}^b \max(m_1 - \|b'_i - b'_j\|_2^2, 0)\} + \alpha(\|b_i\|_1 + \|b_j\|_1) \quad (11)$$

其中, $\beta > 0$ 表示平衡性损失 l_w 和兼容性损失 l_{index} 之间的权衡参数。

3 参数优化

本文采用基于小批量的梯度下降算法对 BCI-DHH 模型中的损失函数 $l_{w-index}$ 进行优化,故需要计算式(9)对 $b_{i,j}$ 的梯度。此外,由于 \max 操作和绝对值操作在个别情况下是不可导的,因此在上述不可导点本文采用次梯度(训练过程中设置次梯度为 1)来代替梯度。本文目标函数 $l_{w-index}$ 的梯度计算如下:

$$\begin{aligned} \frac{\partial \text{Term}_1}{\partial b_{i,j}} &= w_{i,j} (-1)^{j+1} (1 - s_{ij}) (b_i - b_j), \|b_i - b_j\|_2^2 > m_0 \\ \frac{\partial \text{Term}_2}{\partial b_{i,j}} &= w_{i,j} \beta \begin{cases} (-1)^j (b_i - b_j), s_{ij} = 1 \cap \|b'_i - b'_j\|_2^2 < m_1 \\ 0, \text{otherwise} \end{cases} \\ \frac{\partial \text{Term}_3}{\partial b_{i,j}} &= \alpha \delta(b_{i,j}), \delta(x) = \begin{cases} 1, -1 \leq x \leq 0 \text{ or } x \geq 1 \\ -1, \text{otherwise} \end{cases} \end{aligned} \quad (12)$$

基于式(12)中梯度的计算方法,可以对本文 BCI-DHH 中的参数进行反向传播以训练网络模型,从而得到最优参数。

4 实验结果与分析

4.1 实验设置

在 CIFAR-10 和 NUS-WIDE 2 个被广泛使用的基准数据集上对本文 BCI-DHH 模型进行对比实验。CIFAR-10 是一个单标签图像数据集,其包括属于 10 个类别的共计 60 000 张 32×32 图像,本文随机选择每个类别中的 100 张图像(共计 1 000 张图像)作为测试数据集,剩余每个类别中的 5 000 张图像(共计 50 000 张图像)作为训练数据集。NUS-WIDE 是来自于 Flickr 的 269 648 张多标签图像数据集,其中手动标注了 81 个类别与图像之间的关联关系,与文献[14,22]类似,本文使用与其中 21 个最常见类别相关联的图像作为输入,每个类别至少包含 5 000 张图像,从而有共计 1 958 334 张图像。在一般情况下,如果 2 张图像共享至少一个类标签,则认为该图像对为相似图像对,否则为不相似图像对。在 NUS-WIDE 数据集中,本文随机选择 2 100 张图像(每类 100 张图像)作为测试数据集,剩余数据用于模型训练。在本文实验中,分别采用深度哈希方

法和传统哈希方法与 BCI-DHH 模型进行对比实验。其中,深度哈希方法包括 RICH^[23]、BDH^[15]、DSH^[14]、DHN^[22] 和 DPSH^[12],传统哈希方法包括 ITQ^[24] 和 SH^[25]。为提高传统哈希方法的检索性能,本文采用 CNN 对输入数据进行特征提取并将其作为哈希学习的输入特征。

本文 BCI-DHH 模型基于开源 Caffe 框架实现。在模型训练过程中,采用基于小批量的梯度下降算法进行参数优化,设置动量为 0.9,权重衰减为 0.004。在 $10^{-5} \sim 10^2$ 区间内采用基于步长为 10 的交叉验证的方式对超参数 α 和 β 进行选取^[23]。对于边缘阈值参数 m_0 ,本文根据经验值,在 CIFAR-10 中将其设置为 4,在 NUS-WIDE 中将其设置为 8。此外,本文实验中设置 $m_1 = 2k$ 。在多级索引的兼容性损失中,将哈希编码划分为 $k/2$ 个哈希块,这样可以保证不相似图像的每个哈希块中至少有一位不同的二进制编码。本文提出的多级索引的方式基于高性能全特征的文本搜索引擎库 Lucene-6.4.2 实现。

在实验过程中,对本文 BCI-DHH 模型和其他方法在检索的准确度和检索效率方面进行对比分析,采用均值平均精度(mean Average Precision, mAP)来评价模型的检索准确度,通过基于汉明距离的图像查询时间衡量检索效率。

4.2 检索准确度

在 CIFAR-10 和 NUS-WIDE 2 个基准数据集上,当设置哈希编码长度为 16 bit、32 bit、48 bit 和 64 bit 时,分别计算本文模型和其他方法的 mAP,结果如表 1 所示,其中最优结果加粗表示。从表 1 可以看出,多数深度哈希方法的检索性能远优于传统哈希方法,表明相比于传统手工设计的特征,深度神经网络模型可以提取深层语义特征。本文 BCI-DHH 模型不仅提取到不同层之间的互相关信息,而且获取到了层内的自相关信息,从而生成了更加细粒度的深层语义特征。因此,在 CIFAR-10 数据集上,相比于其他深度哈希方法,本文模型基于不同哈希长度下检索的 mAP 得到全面提升,而在 NUS-WIDE 数据集中,相比于其他深度哈希方法,本文模型哈希检索的 mAP 提升了 1.8% ~ 4.3%。DSH 方法在 CIFAR-10 数据集上的 mAP 优于 DPSH 方法,而在 NUS-WIDE 数据集上,DSH 方法的 mAP 低于 DPSH 方法,表明 DSH 在不同数据集上的健壮性较低。本文模型在进行训练的过程中考虑到正样本数量过少且负样本数量过多的问题,采用数据平衡性损失对模型进行优化,防止对正样本学习欠拟合而对负样本学习过拟合的现象,因此,BCI-DHH 模型在 CIFAR-10 和 NUS-WIDE 数据集上都表现出了最佳性能。

表 1 CIFAR-10 和 NUS-WIDE 数据集中不同位数哈希编码的 mAP 对比
Table 1 mAP comparison of different bits of hash codes in CIFAR-10 and NUS-WIDE datasets

方法	CIFAR-10 数据集				NUS-WIDE 数据集			
	16 bit	32 bit	48 bit	64 bit	16 bit	32 bit	48 bit	64 bit
BCI-DHH	0.891 2	0.903 6	0.907 7	0.910 8	0.820 3	0.837 9	0.841 6	0.848 3
RICH	0.880 6	0.900 3	0.905 1	0.908 4	0.781 2	0.796 3	0.812 3	0.830 2
BDH	0.782 1	0.798 6	0.801 3	0.813 1	0.695 6	0.701 5	0.730 6	0.733 7
DSH	0.849 6	0.882 0	0.899 6	0.907 5	0.749 5	0.751 3	0.765 2	0.768 1
DHN	0.865 1	0.869 3	0.877 6	0.888 5	0.804 6	0.812 3	0.823 1	0.822 5
DPSH	0.731 2	0.753 1	0.759 6	0.770 5	0.773 2	0.782 7	0.807 9	0.824 7
ITQ + CNN	0.245 3	0.247 8	0.253 8	0.266 0	0.423 1	0.436 5	0.461 1	0.431 2
SH + CNN	0.159 6	0.164 3	0.167 5	0.171 8	0.378 4	0.361 2	0.380 5	0.352 3

此外,基于 CIFAR-10 和 NUS-WIDE 2 个数据集,分别设置哈希编码长度为 16 bit、32 bit、48 bit 和 64 bit,在返回样本数量为 Top500 和检索的汉明半径不超过 2 的情况下对 BCI-DHH 模型和其他方法的准确率进行对比分析,结果如图 3 ~ 图 6 所示。从中可以看出,由于 BCI-DHH 模型在进行哈希学习时输入的特征向量是基于高阶统计信息的,因此当返回的样本数目为 Top500 时,其在不同编码长度下的哈希检索中都表现出了最佳性能。同时,随着哈希编码长度的增加,大部分哈希学习方法的检索性能也不断提升,特别是在 NUS-WIDE 数据集中,当哈希编码长度由 32 bit 增加到 64 bit 时,BCI-DHH 模型检索性能提升了 5% 左右。当设定检索的汉明距离半径长度不超过 2 时,无论是在 CIFAR-10 数据集还是 NUS-WIDE 数据集中,本文 BCI-DHH 模型相比于其他方法都表现出较好的性能。随着用于检索的哈希编码的不断加,固定汉明半径,则本文模型检索性能会有一定程度的下降,而在 NUS-WIDE 数据集中,哈希编码由 48 bit 增加到 64 bit 时,BCI-DHH 模型检索性能没有发生显著变化,充分证明了本文模型的健壮性。

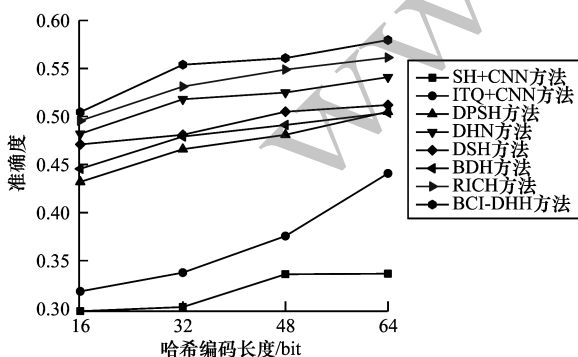


图 3 CIFAR-10 数据集中的 Top500 准确度对比
Fig. 3 Comparison of Top500 accuracy in CIFAR-10 dataset

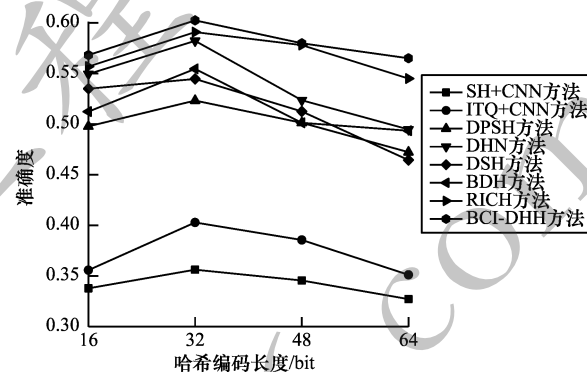


图 4 CIFAR-10 数据集中汉明距离小于 2 时的准确度对比
Fig. 4 Comparison of accuracy when Hamming distance is less than 2 in CIFAR-10 dataset

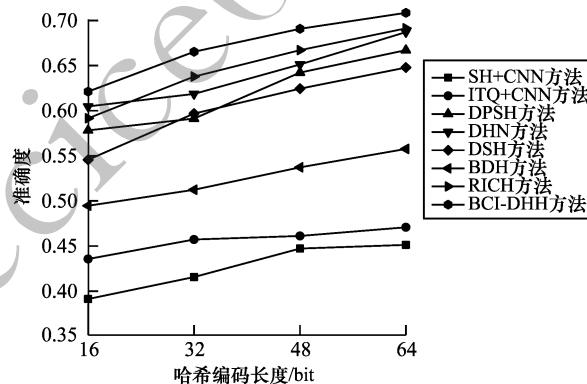


图 5 NUS-WIDE 数据集中的 Top500 准确度对比
Fig. 5 Comparison of Top500 accuracy in NUS-WIDE dataset

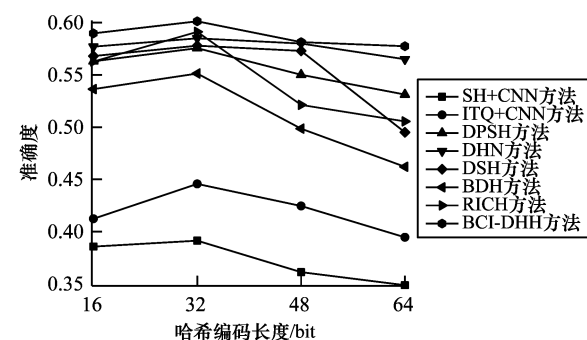


图 6 NUS-WIDE 数据集中汉明距离小于 2 时的准确度对比
Fig. 6 Comparison of accuracy when Hamming distance is less than 2 in NUS-WIDE dataset

4.3 检索时间

一般而言,多级索引的检索过程包括搜索和检查 2 个阶段。在搜索阶段,查询图像通过哈希函数生成哈希编码并划分为不同的哈希块,然后搜索基于多级索引的二进制代码,从而查找出至少与查询图像具有一个相同哈希块的哈希编码。在检查阶段,根据查询图像与候选图像之间的汉明距离对候选图像数据集进行排序。在本文检索时间实验中,主要考虑这 2 个阶段消耗的时间代价,实验过程中所有图像的哈希代码均存储在内存中。

在 CIFAR-10 和 NUS-WIDE 数据集上,分别测试 BCI-DHH 和 RICH、BDH、DSH 哈希方法在哈希编码长度分别为 16 bit、32 bit、48 bit 和 64 bit 时检索目标图像所消耗的时间。由于本文 BCI-DHH 模型在进行训练优化时充分平衡了正负样本的数目,避免了参数过拟合或欠拟合的发生,因此产生了具有较强区分能力的哈希编码,保证了图像与哈希编码之间的语义相似度,从而在一定程度上减少了候选图像的数量。此外,在不相似图像的哈希编码学习过程中,通过对不相似图像的多级哈希索引块进行差异化操作,增加不相似图像与目标图像之间的汉明距离,从而大幅减少候选图像数目,降低检索代价,提高检索的效率。具体实验结果如图 7、图 8 所示,从中可以看出,DSH 模型的检索时间代价小于 BDH 模型,在 NUS-WIDE 数据集中,当哈希编码长度为 32 bit 时,DSH 的检索时间最长,从而说明上述模型的健壮性较低。而本文 BCI-DHH 模型在 CIFAR-10 和 NUS-WIDE 数据集中消耗的时间均最短,充分证明了模型的高效性和健壮性。

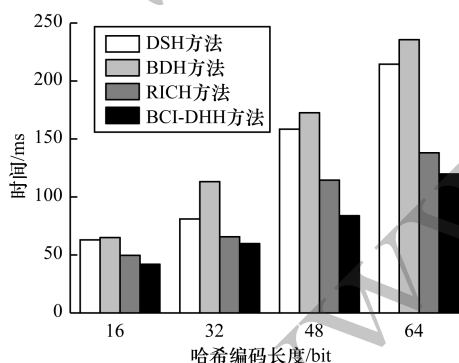


图 7 CIFAR-10 数据集上不同哈希编码的检索时间对比
Fig. 7 Comparison of retrieval time of different hash codes on CIFAR-10 dataset

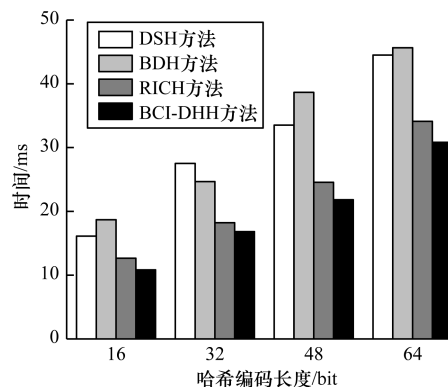


图 8 NUS-WIDE 数据集上不同哈希编码的检索时间对比
Fig. 8 Comparison of retrieval time of different hash codes on NUS-WIDE dataset

当哈希编码的长度为 32 bit 时,在 CIFAR-10 数据集和 NUS-WIDE 数据集上分别测试 BCI-DHH 和 RICH、BDH 以及 DSH 方法的候选图像数量,结果如表 2 所示。由于 BCI-DHH 模型不仅考虑数据的不平衡性,而且在对不相似图像进行多级索引哈希学习时,对哈希块进行了差异化操作,从而减少了候选图像的数量,因此,本文模型在检索过程中候选图像数量最小。从表 2 可以看出,相比其他方法,在 CIFAR-10 数据集上本文模型的候选图像数量减少了 3.5% ~ 24.5%,在 NUS-WIDE 数据集上减少了 7.6% ~ 47.7%。此外,当哈希编码长度设定为 32 bit 时,在 CIFAR-10 和 NUS-WIDE 数据集上分别对模型的召回率进行了比较分析,结果也展示在表 2 中。由于 CIFAR-10 数据集为单标签数据集,因此图像对之间共享标签数目最多为 1(表中用“—”表示)。可以看出,当哈希编码为 32 bit 时所有的哈希方法都在该数据集上实现了将近 100% 的召回率,表明在搜索阶段没有相似图像被过滤掉,进一步说明所有方法在 CIFAR-10 数据集中能表现出良好的性能。在 NUS-WIDE 数据集中,“ $\geq n$ ”表示与目标图像之间共享不少于 n 个相同标签的相似图像的召回率。通过实验结果可知,随着共享标签数的增加,召回率增加,表明与目标图像具有更高相似度的相似候选图像更有可能被检索到。本文 BCI-DHH 模型在 CIFAR-10 和 NUS-WIDE 数据集上的召回率均为最优,从而进一步验证了本文模型的有效性。

表 2 CIFAR-10 和 NUS-WIDE 数据集中的召回率和候选图像数量对比

Table 2 Comparison of recall rate and number of candidate images in CIFAR-10 and NUS-WIDE datasets

方法	CIFAR-10			NUS-WIDE			
	—	候选集数目	≥ 1	≥ 2	≥ 3	≥ 4	候选集数目
BCI-DHH	0.999 9	37 562	0.972 3	0.989 9	0.999 9	0.999 9	98 762
RICH	0.999 9	38 925	0.965 3	0.981 7	0.990 6	0.999 9	106 941
BDH	0.999 9	46 135	0.924 6	0.954 6	0.984 6	0.999 9	179 563
DSH	0.999 9	49 762	0.998 2	0.999 0	0.999 8	0.999 9	188 765

5 结束语

本文在高阶统计学习方法的基础上,充分考虑数据的不平衡性和多级哈希索引的不兼容性,建立一种深度哈希学习模型 BCI-DHH。在特征学习的过程中,采用改进的 VGG-m 模型提取具有高层语义特征的高阶统计向量。分别设计基于图像对的数据平衡性损失和多级哈希索引的兼容性损失,从而提高哈希检索过程的准确率和效率。在基准数据集 CIFAR-10 和 NUS-WIDE 上进行实验,结果验证了 BCI-DHH 模型的有效性。下一步将采用真实场景下的数据对本文模型进行优化与验证。

参考文献

- [1] LIU Yu, SONG Jingkuan, ZHOU Ke, et al. Deep self-taught hashing for image retrieval[J]. IEEE Transactions on Cybernetics, 2019, 49(6): 1-13.
- [2] LIU Haomiao, WANG Ruiping, SHAN Shiguang, et al. Learning to hash with discrete optimization[J]. Chinese Journal of Computers, 2019, 42(5): 1149-1160. (in Chinese)
刘昊淼,王瑞平,山世光,等.基于离散优化的哈希编码学习方法[J].计算机学报,2019,42(5):1149-1160.
- [3] YAN Shuangyong, LIU Changhong, JIANG Aiwen, et al. Discriminative cross-modal hashing with coupled semantic correlation[J]. Chinese Journal of Computers, 2019, 42(1): 164-175. (in Chinese)
严双咏,刘长红,江爱文,等.语义耦合相关的判别式跨模态哈希学习算法[J].计算机学报,2019,42(1):164-175.
- [4] GUO Yuchen, DING Guiguang, LIU Li, et al. Learning to hash with optimized anchor embedding for scalable retrieval[J]. IEEE Transactions on Image Processing, 2017, 26(3): 1344-1354.
- [5] LU Jiwen, DUAN Yueqi, CHEN Zhixiang, et al. Binary representation learning and its applications[J]. Pattern Recognition and Artificial Intelligence, 2018, 31(1): 12-22. (in Chinese)
鲁继文,段岳圻,陈志祥,等.二值表示学习及其应用[J].模式识别与人工智能,2018,31(1):12-22.
- [6] SHOU Zhenyu, QIAN Jiangbo, DONG Yihong, et al. EFH: an online unsupervised hash learning algorithm[J]. Telecommunications Science, 2020, 36(3): 71-82. (in Chinese)
寿震宇,钱江波,董一鸿,等.演化森林哈希:一种无监督的在线哈希学习算法[J].电信科学,2020,36(3):71-82.
- [7] YAO Tao, KONG Xiangwei, FU Haiyan, et al. Projective dictionary learning hashing for cross-modal retrieval[J]. Acta Automatica Sinica, 2018, 44(8): 1475-1485. (in Chinese)
姚涛,孔祥维,付海燕,等.基于映射字典学习的跨模态哈希检索[J].自动化学报,2018,44(8):1475-1485.
- [8] ZHANG Lili, WANG Li, ZHAO Qi, et al. Traffic state identification of intersection based on semi-supervised hash algorithm[J]. Journal of Transportation Systems Engineering and Information Technology, 2020, 20(1): 75-82. (in Chinese)
张立立,王力,赵琦,等.基于半监督哈希算法的交叉口交通状态识别[J].交通运输系统工程与信息,2020,20(1):75-82.
- [9] HUANG Wenming, WEI Peng, LIANG Jinhua. Application of CNN-based hashing in image retrieval[J]. Computer Engineering and Design, 2017, 38(2): 517-521. (in Chinese)
黄文明,魏鹏,梁金华.基于卷积神经网络的哈希在图像检索中的应用[J].计算机工程与设计,2017,38(2):517-521.
- [10] XIA Rongkai, PAN Yan, LAI Hanjiang, et al. Supervised hashing for image retrieval via image representation learning[C]//Proceedings of the 28th AAAI Conference on Artificial Intelligence. [S. l.]: AAAI Press, 2014: 2156-2162.
- [11] LAI Hanjiang, PAN Yan, LIU Ye, et al. Simultaneous feature learning and hash coding with deep neural networks[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2015: 3270-3278.
- [12] LI W J, WANG S, KANG W C. Feature learning based deep supervised hashing with pairwise labels[EB/OL]. [2019-05-28]. <https://www.ijcai.org/Proceedings/16/Papers/245.pdf>.
- [13] ZHAO Fang, HUANG Yongzhen, WANG Liang, et al. Deep semantic ranking based hashing for multi-label image retrieval[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2015: 1556-1564.
- [14] LIU H, WANG R, SHAN S, et al. Deep supervised hashing for fast image retrieval[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 2064-2072.
- [15] ZHOU Quan, QI Shuhan, WANG Xuan, et al. Balance the loss: improving deep hash via loss weighting and semantic preserving[C]//Proceedings of 2018 IEEE International Conference on Multimedia and Expo. Washington D. C., USA: IEEE Press, 2018: 1-6.
- [16] SUN Q, WANG Q, ZHANG J, et al. Hyperlayer bilinear pooling with application to fine-grained categorization and image retrieval[J]. Neurocomputing, 2018, 282: 174-183.
- [17] CHENG Jingdong, SUN Qiule, ZHANG Jianxin, et al. Deep high-order supervised hashing for image retrieval[C]//Proceedings of the 24th International Conference on Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 2693-2698.
- [18] GOG S, VENTURINI R. Fast and compact Hamming distance index[C]//Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, USA: ACM Press, 2016: 285-294.
- [19] WU Dayan, LIU Jing, LI Bo, et al. Deep index-compatible hashing for fast image retrieval[C]//Proceedings of 2018 IEEE International Conference on Multimedia and Expo. Washington D. C., USA: IEEE Press, 2018: 1-6.
- [20] GAO Y, BEIJBO O, ZHANG N, et al. Compact bilinear pooling[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 317-326.

(上接第 267 页)

- [21] LIN T Y, MAJI S. Improved bilinear pooling with CNNs [EB/OL]. [2019-05-28]. <https://ui.adsabs.harvard.edu/abs/2017arXiv170706772L/abstract>.
- [22] ZHU H, LONG M, WANG J, et al. Deep hashing network for efficient similarity retrieval [EB/OL]. [2019-05-28]. <http://ise.thss.tsinghua.edu.cn/ml/doc/2016/deep-hashing-network-aaai16.pdf>.
- [23] LIU J, WU D, ZHANG W, et al. Robust and index-compatible deep hashing for accurate and fast image retrieval [C]//Proceedings of Pacific Rim Conference on Multimedia. Berlin, Germany: Springer, 2018: 67-77.
- [24] 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.
- [25] WEISS Y, TORRALBA A, FERGUS R. Spectral hashing [EB/OL]. [2019-05-28]. <http://papers.nips.cc/paper/3383-spectral-hashing.pdf>.

编辑 吴云芳