



计算机科学

COMPUTER SCIENCE

有监督相似性保持的深度二阶哈希方法

张建新, 吴悦, 张强, 魏小鹏

引用本文

张建新, 吴悦, 张强, 魏小鹏. [有监督相似性保持的深度二阶哈希方法](#) [J]. 计算机科学, 2022, 49(11A): 210900021-8.

ZHANG Jian-xin, WU Yue, ZHANG Qiang, WEI Xiao-peng. [Supervised Similarity Preserving Deep Second-order Hashing](#) [J]. Computer Science, 2022, 49(11A): 210900021-8.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[融合多层次信息的海关同义词识别方法](#)

Customs Synonym Recognition Fusing Multi-level Information

计算机科学, 2022, 49(11A): 210800197-5. <https://doi.org/10.11896/jsjcx.210800197>

[一种鲁棒的双教师自监督蒸馏哈希学习方法](#)

Robust Hash Learning Method Based on Dual-teacher Self-supervised Distillation

计算机科学, 2022, 49(10): 159-168. <https://doi.org/10.11896/jsjcx.210800050>

[基于注意力机制的医学影像深度哈希检索算法](#)

Deep Hash Retrieval Algorithm for Medical Images Based on Attention Mechanism

计算机科学, 2022, 49(8): 113-119. <https://doi.org/10.11896/jsjcx.210700153>

[融合快速注意力机制的节点无特征网络链路预测算法](#)

Link Prediction for Node Featureless Networks Based on Faster Attention Mechanism

计算机科学, 2022, 49(4): 43-48. <https://doi.org/10.11896/jsjcx.210800276>

[基于预训练和深度哈希的大规模文本检索研究](#)

Study on Text Retrieval Based on Pre-training and Deep Hash

计算机科学, 2021, 48(11): 300-306. <https://doi.org/10.11896/jsjcx.210300266>

有监督相似性保持的深度二阶哈希方法

张建新^{1,2} 吴悦² 张强^{2,3} 魏小鹏^{2,3}

1 大连民族大学计算机科学与工程学院 辽宁 大连 116600

2 大连大学先进设计与智能计算省部共建教育部重点实验室 辽宁 大连 116622

3 大连理工大学计算机科学与技术学院 辽宁 大连 116024

摘要 近年来深度哈希方法因其存储效率高和查询速度快的优势在大规模图像检索领域受到了广泛关注。为改善深度成对有监督哈希方法在图像检索上的性能,从提高深度哈希获取图像特征的全局性和同类样本相似性角度出发,提出了一种有监督相似性保持的深度二阶哈希方法。该方法采用成对样本图像进行特征建模,并利用协方差估计来捕获样本图像的深度二阶信息,以获取具有良好全局表达能力的深度二阶哈希码;在此基础上,借鉴类哈希近似二值化来解决哈希映射过程中的非凸性问题,以更好地避免量化误差,同时基于多损失函数集成思想构建类别监督和相似性保持的联合约束,进而采用交替迭代的优化方式实现网络的端到端训练,最终确定样本图像的最优哈希码。在3个通用数据集上进行了广泛的实验,结果有效表明了所提出有监督相似性保持的深度二阶哈希方法的有效性。

关键词: 深度哈希;二阶统计建模;类别监督;相似性保持;图像检索

中图分类号 TP391.41

Supervised Similarity Preserving Deep Second-order Hashing

ZHANG Jian-xin^{1,2}, WU Yue², ZHANG Qiang^{2,3} and WEI Xiao-peng^{2,3}

1 School of Computer Science and Engineering, Dalian Minzu University, Dalian, Liaoning 116600, China

2 Key Lab of Advanced Designed and Intelligent Computing, Ministry of Education, Dalian University, Dalian, Liaoning 116622, China

3 School of Computer Science and Technology, Dalian University of Technology, Dalian, Liaoning 116024, China

Abstract Recently, deep hashing technology, with its advantages of high storage efficiency and quick query speed, has been widely investigated in the field of large-scale visual image retrieval. However, the fundamental image features obtained by current deep hashing methods mainly depend on the first-order statistics of deep convolutional features, and they seldom take the global structure into consideration, leading to the limitation of retrieval accuracy to a certain degree. Focusing on the global representation capability and the similarity of intra-class samples, this paper proposes a novel supervised similarity preserving deep second-order hashing (S2PDSoH) based on deep pairwise supervised hashing, gaining effective performance improvement in the image retrieval task. Based on the pair-wise deep hashing model, S2PDSoH first employs covariance estimation based on matrix power normalization method to capture the deep second-order information of sample images, so that hash codes can possess good global presentation ability. Then, to gain more robust hash codes, it further constructs a joint constraint of category supervision and similarity preservation motivated by the idea of multi-loss integration, followed by an alternate iteration optimization algorithm to realize the end-to-end training. Therefore, with the semantic information added to the dual channel deep second-order hashing framework, S2PDSoH establishes a mechanism for common constraints on category supervision and similarity maintenance. In addition, it also introduces a hash-like function to achieve the approximate binarization result of hash codes, which solves the problem of non-convexity and avoids the quantization error in the hash mapping process. Extensive experimental results on three commonly used data sets show the effectiveness of the proposed deep-order hashing method with supervised similarity preservation.

Keywords Deep hashing, Second-order statistic modeling, Category supervision, Similarity preserving, Image retrieval

1 引言

近年来哈希技术凭借其存储空间小、查询速度快等优势

在图像检索领域被广泛研究。随着深度卷积神经网络(Convolutional Neural Networks, CNNs)^[1]取得突破性进展,基于深度学习的深度哈希方法^[2-4]迅速引起了相关领域研究者的

基金项目:国家自然科学基金(61972062);辽宁省重点研发计划项目(2019JH2/10100030);辽宁省自然科学基金(2019-MS-011);国家民委中青年英才培养计划项目

This work was supported by the National Natural Science Foundation of China(61972062), Key R & D Program of Liaoning Province(2019JH2/10100030), Natural Science Foundation of Liaoning Province(2019-MS-011) and Young and Middle-aged Talents Program of the National Civil Affairs Commission.

通信作者:张建新(jxzhang0411@163.com)

广泛关注。深度哈希方法能有效模拟图像深度特征的非线性变换,学习有效表达图像特征的哈希编码,能显著提升传统哈希方法的检索性能。深度哈希方法的最初探索是将 CNN 作为特征提取器来取代传统的手工特征提取模式,进而基于提取的深度特征来学习哈希编码。卷积神经网络哈希(Convolutional Neural Network Hashing, CNNH)^[5]是早期的代表工作之一, CNNH 将深度哈希学习分解为一个成对关系的目标哈希码学习阶段和一个同时学习图像特征、哈希函数的学习阶段,尽管其在第一阶段的编码能够得到更好的图像特征表示,但其在后一阶段中获得的特征表示却无法有效促进近似哈希码的训练。为此,研究者提出了若干可整体训练的改进哈希方法^[6-11],借助于整体可训练架构来使深度特征和哈希编码学习相辅相成,以确定从图像到哈希码的最优映射。深度哈希方法利用了深度卷积特征来增强哈希码,但大都依赖于一阶卷积特征信息,忽略了哈希建模中的全局特性交互,在一定程度上影响了哈希编码的可区分性。

近年来,鉴于高阶统计建模方法在图像深度特征全局信息表达上的优势,及其在多个视觉任务上较其一阶模型取得了明显的性能提高,深度高阶统计建模方法逐渐受到研究者的关注,代表性模型包括深度二阶池化、双线性池化和协方差池化等^[12-15]。其中,协方差池化模型可有效捕捉深度特征的二阶统计量,并利用特征通道相关性挖掘出了更具判别性的全局特征表示。Wu 等曾尝试探索在深度哈希方法中引入全局统计特征表达,证实了其良好的效果^[16-17],尽管有效利用协方差统计信息提高了编码的判别性,但在保持同类图像间相似性和哈希量化处理上仍有一定不足。为获取更加判别的哈希编码表达,这里进一步提出了一种有监督相似性保持的深度二阶哈希方法(Supervised Similarity Preserving Deep Second-order Hashing, S2PDSOH),其采用图像对方式作为网络输入,通过引入新的损失函数来建立类别监督的相似性以保持共同约束机制,并着重考虑离散哈希在训练过程中的不平滑问题,在哈希映射模块采用类哈希方法来获得近似二值哈希编码,在 3 个图像数据集上的广泛实验的结果表明了所提出有监督相似性保持的深度哈希方法的有效性。

2 相关工作

2.1 有监督深度哈希

鉴于 S2PDSOH 属于依托于图像数据及其标签信息进行学习的有监督哈希方法,因此在此仅对有监督深度哈希方法进行介绍,并按照监督方式将其大致分为 3 类:基于逐点标签的哈希方法、基于成对标签的哈希方法和基于三元组的哈希方法。基于逐点标签的哈希方法以单张图像为网络输入,借助单张图像及其类别信息来约束和指导哈希函数的学习过程,该方式对哈希函数的可伸缩性来说较好,但在约束指导上具有较大局限性。代表性方法如 Lin 等的二进制深度哈希码方法(Deep Learning of Binary Hash Codes, DLBHC)^[6],该方法使用一张图像作为输入来训练具有二值化隐含层的深度模型,同时使用精细化搜索策略来获得最终的编码。基于成对标签的哈希方法同时以两张图像为输入,采用成对图像信息来建立约束和指导哈希函数学习,并可通过构造相似性保持约束来保持同类样本哈希码的一致性,从而提升哈希码的表达能力。卷积神经网络哈希方法^[5]是其中一种代表性方法,其先使用成对相似关系矩阵来获取哈希编码,然后进一步

基于深度网络来逼近哈希编码。Li 等则提出深度成对有监督哈希方法(Deep Pairwise Supervised Hashing, DPSH)^[18],该方法主要利用成对图像标签来设计相似矩阵和损失优化函数,通过整体学习方式优化网络与获得更具判别性的哈希码表达结果。Li 等提出了深度有监督离散哈希(Deep Supervised Discrete Hashing, DSDH)^[19],其将对成对输入图像的标签信息和分类语义相整合,以此来构建更具区分能力的编码结果。Wan 等则提出了自监督的深度离散哈希方法(Self-Supervised Deep Discrete Hashing, SSDDH)^[10],引入自监督思想来实施深度哈希函数的学习,同时利用离散优化算法来降低二进制定量化误差。Fan 等在深度哈希网络中提出了可微极化损失概念,通过最大化类间哈希编码距离与最小化类内哈希编码距离来优化网络,构建出深度极化网络(Deep Polarized Network, DPN)^[9]。基于三元组的哈希方法则同时以 3 张图像为网络输入,其中两张图像来自同类别,另一张图像来自其他类别,通过建立约束来保证同类图像的相似性,同时尽量扩大异类图像间的区分性。典型方法如 Lai 等的深度神经网络哈希方法(Deep Neural Network Hashing, DNNH)^[7],其采用三元组形式下的约束函数端对端训练网络来提升哈希码的表达力;深度语义排序监督哈希(Deep Semantic Ranking Based Hashing, DSRH)^[8]则对最终评测指标进行优化,按照图像查询相关性返回检索结果;此外, Wang 等提出了三元组深度监督哈希(Deep Supervised Hashing with Triplet Labels, DTSH)^[20],利用三元组损失和单个语义损失来共同保证训练的约束力。但是,该类方法主要缺点是在构造网络输入时需要手动挑选三元组,工作量较大且易影响检索精度。

2.2 深度高阶统计建模

深度高阶统计建模方法是在图像建模过程中使用深度特征的高阶统计信息,具有代表性的深度高阶统计建模方法主要包括深度二阶池化、双线性池化和协方差池化等^[12-15]。其中, Ionescu 等^[12]提出了一种典型的深度二阶池化网络结构,该模型首先将全局化特征计算单元集成到卷积神经网络模型,并可进行端对端的整体优化训练,为之后相关工作奠定了较好的理论基础。另一类具有代表性的工作是 Lin 等^[13]提出的双线性池化网络模型,其通过计算双流网络末端卷积层输出特征的外积来获得二阶特征表达结果,在细粒度图像分类任务上获得了突出的性能结果。此后, Li 等和 Wang 等^[14-15]提出了幂归一化协方差估计模型(Matrix Power Normalized Covariance Pooling, MPN-COV),其主要计算网络末端卷积层深度特征的协方差矩阵,基于协方差特征的几何结构,对协方差矩阵作幂归一化操作来得到图像全局特征鲁棒表达,显著提高了模型在大规模图像分类任务上的性能,并在 CVPR 2018 国际大规模细粒度物种识别竞赛中获得了第一名的成绩;为了提升方法的计算效率, Li 等进一步提出了快速 MPN-COV 模型,该模型的核心在于采用 Newton-Schulz 迭代来求解矩阵开平方根归一化操作,实现了在 GPU 上的并行运算。另外,高斯建模和核化方法等也被研究者引入到 CNN 模型中,用于构建深度特征的高阶表达模型。

3 有监督相似性保持的深度二阶哈希

有监督相似性保持的深度二阶哈希方法是在成对有监督的深度二阶哈希模型的基础上,进一步加强同类样本相似性保持性和改善离散哈希在训练过程中的平滑性,具有基础

特征、高阶池化、类哈希和约束优化 4 个典型模块,方法框架如图 1 所示。其中,基础特征模块用于提取图像的一阶深度卷积特征,这里以 VGG-F 网络为基础构建,也可采用更为先进的 ResNet 等网络来构建,为了与现有典型深度哈希网络相一致,选择使用 VGG-F 网络来更好地验证后续模块的效果;高阶池化模块由协方差(Covariance, COV)、特征值分解(Eigenvalue, EIG)和幂归一化(Power normalization, Power)组成,作用是将基础特征模块输出的所有通道的深度特征图谱进行二阶协方差统计建模,以此

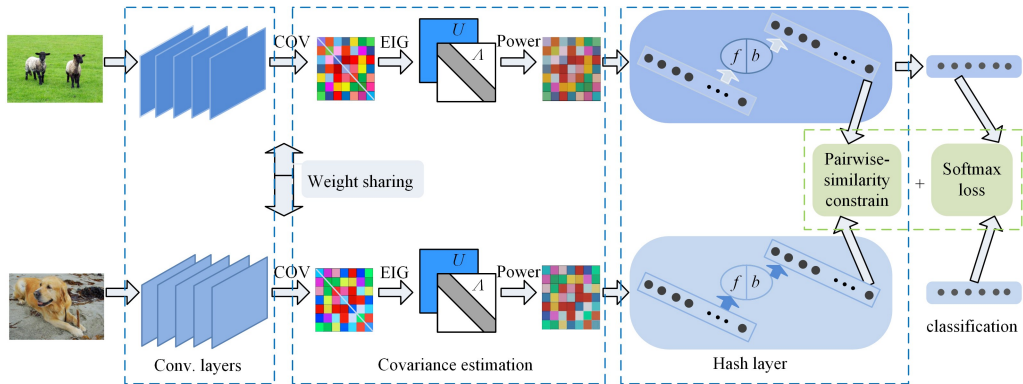


图 1 有监督相似性保持的深度二阶哈希方法框架图

Fig. 1 Architecture of supervised similarity preserving deep second-order hashing

3.1 深度二阶哈希建模

给定图像数据集 $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N)$, N 为图像样本数,单张图像记作 $\mathbf{x}_i, \{i = 1, \dots, N\}$, 以 $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_N)$ 代表图像数据集对应的标签信息, $\mathbf{y}_i \in \{0, 1\}^c$, c 为类别数。将输入信息定义为 $\{(x_i, x_j, s_{ij}), s_{ij} \in \mathbf{S}\}$, 其中 (x_i, x_j) 代表成对输入图像数据, \mathbf{S} 表示数据集的相似矩阵; s_{ij} 对应样本 x_i 与 x_j 的语义相似信息, $s_{ij} = 1$ 代表两个样本具有相同标签, 反之具有不同标签。该方法旨在构造哈希函数, 学习输入 x_i 的二值编码 b_i , 并在哈希空间尽可能保持样本在原始空间的语义分类信息和相似关系。

S2PDSOH 以成对输入的双流网络为依托, 采用两个权重共享的基础深度网络架构来执行特征提取, 此处基于 ImageNet 预训练的 CNN-F 网络^[21], 也可使用 ResNet^[22] 和 AlexNet^[23] 等。输入图像经过由卷积层、池化层及激活层等组合成的基础特征子网络完成逐层映射, 实现基础深度特征的提取。考虑到双流网络采用权重共享结构, 在此仅描述其中一支的特征表述, 另一支过程相同。借助上述结构执行特征学习, 基础特征模块的输出 \mathbf{z}_i 可表示为:

$$\mathbf{z}_i = \Phi(\mathbf{x}_i; \boldsymbol{\theta}) \quad (1)$$

其中, Φ 代表深度神经网络逐层传递的映射关系, $\boldsymbol{\theta}$ 为网络参数, \mathbf{z}_i 是 d 维基础深度特征。

为了充分利用特征信道的相关性和特征分布, 挖掘特征交互的全局统计信息, 在深度网络执行接受域的操作之后, 嵌入协方差池化操作来从二阶统计更好描述空间矩阵的几何特征。鉴于协方差矩阵具有较多非线性矩阵计算, 特别是对于特征值分解^[12] 和矩阵幂归一化操作^[15]。给定上述基础深度特征 \mathbf{Z} , 每个区域内局部描述符的初步全局特征统计可由样本的协方差矩阵 \mathbf{m}_i 建立, 具体表示为:

$$\mathbf{m}_i = \mathbf{z}_i \bar{\mathbf{I}}_i \mathbf{z}_i^T \quad (2)$$

其中, \mathbf{I} 表示 $n \times n$ 的单位矩阵, $\bar{\mathbf{I}} = \frac{1}{n} \left(\mathbf{I} - \frac{1}{n} \mathbf{J} \mathbf{J}^T \right)$, \mathbf{J} 表示 n

获得更为全局性与更具判别性的深度二阶特征; 类哈希模块类旨在将量化的高维二阶深度特征进行哈希映射和编码激活, 着重解决哈希映射过程中的非凸性问题来避免量化误差, 从而获得更为鲁棒的哈希表达; 最后, 约束优化模块是由类别监督和相似性约束集成构成, 其中相似性约束由两个输入图像哈希编码的汉明距离为基础, 结合类监督信息以保障同类样本具有更高的编码相似度, 在集成的类别监督和相似性约束指导下交替迭代优化, 可获得更优的哈希码结果。

维列向量, $\mathbf{J} = [1, \dots, 1]^T$, \mathbf{T} 表示矩阵转置操作。

尽管可使用最大似然估计方法来计算 \mathbf{m}_i , 但其在卷积特征的高维小样本情况下鲁棒性较弱。为了获得稳健协方差估计, 充分利用协方差矩阵的黎曼流形结构, 参照 MPN-COV^[14] 处理方式, 利用矩阵幂归一化方法缩小特征值。考虑到协方差矩阵 \mathbf{m}_i 的对称半正定性质, 其特征值分解可表示为:

$$\mathbf{m}_i = \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^T \quad (3)$$

其中, $\boldsymbol{\Lambda} = (\lambda_1, \dots, \lambda_d)$ 是由特征值构成的对角阵, \mathbf{U} 为特征向量构成的正交矩阵。

依托于协方差矩阵 \mathbf{M} 的流形结构, 利用对角矩阵原理和式(3)中的特征值分解, 全局特征统计可借助幂操作确定, 全局二阶统计特征 \mathbf{p}_i 经简化处理之后可表示为:

$$\mathbf{p}_i \cong \mathbf{m}_i^\alpha = \mathbf{U} \boldsymbol{\Lambda}^\alpha \mathbf{U}^T \quad (4)$$

其中, α 是取值范围为 $(0, 1)$ 的正实数。为增强二阶统计特征的描述能力, 采用元素幂归一化^[14-15] 处理特征交互的全局统计信息, 进而使用 L_2 或 F 范数来完成矩阵元素的幂归一化。

在接下来的类哈希模块中, 主要是将高维的图像特征转化为有较强判别性的哈希特征表达结果, 采用的操作是图像高维特征的哈希映射与编码激活。该模块借助哈希潜层将全局特征统计特性 \mathbf{P} 映射到 $[0, 1]$ 范围 K 比特位编码, 并根据层节点数灵活地调整编码位数, 这里取 12, 24, 32 和 48 4 种编码位设置。根据哈希潜层建立二阶统计向量到可编码特征的基础哈希映射, 表示为:

$$\mathbf{h}_i = \mathbf{W}_h^T \mathbf{p}_i + \boldsymbol{\sigma}_i \quad (5)$$

其中, \mathbf{W}_h^T 为映射权重, $\boldsymbol{\sigma}_i$ 为偏置向量。

式(5)得到的哈希码是一系列连续值, 最终的哈希码是经过特定阈值函数二值化的结果, 但对于非凸的二值化结果在梯度传播中无法完成模型的整体训练。为能够完成网络的整体训练, 同时避免因哈希松弛操作带来的误差, 受平滑函数的启发^[24], 采用类哈希激活的方式在 \mathbf{h}_i 基础上来获得近似二值哈希码结果, 具体计算式如下:

$$\mathbf{b}_i = \text{sign}(\mathbf{h}_i) = \lim_{\beta \rightarrow +\infty} \tanh(\beta \cdot \mathbf{h}_i) \quad (6)$$

式(6)引入了一个超参数 β 作为辅助来构建激活函数, 函数 $\tanh(\cdot)$ 会受到 β 值的改变相对应出现坡度的调整。具体来说, 较大的 β 值会加速激活函数的改变, 使其更逼近于 $\text{sign}(\cdot)$ 函数, 同时能够较好保证函数的连续性。图 2 给出了函数 $\tanh(\cdot)$ 与辅助参数 β 值变化的关系情况。在实际网络模型训练时, 采用迭代方式来逐渐改变辅助参数 β 值以改变图像平滑度, 进而优化二值化的哈希码。实际上, 当 β 值设定为 10 时类哈希函数就能够非常好地逼近 $\text{sign}(\cdot)$, 能够有效实现阈值函数的替代效果。

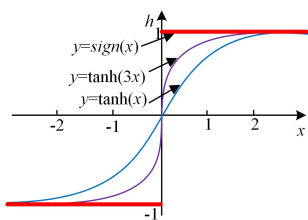


图 2 类哈希函数在不同 β 值下的变化情况

Fig. 2 Hash-like function curves under various β values

由于借助协方差池捕获二阶统计并极大利用特征信道的相关性, 获取的哈希编码能够有效体现图像全局特征分布信息。此外, 解决了离散哈希在端对端训练中的非凸性问题, 避免了量化带来的正则损失, 进一步保证了哈希码的性能。

3.2 目标函数

在建立深度二阶哈希特征表达后, 依据多损失函数集成的思路建立类别监督和相似性保持共同约束的机制, 从而以更优的语义表示来进一步提升哈希码性能, 目标函数可被定义为:

$$L = Q + \mu R \quad (7)$$

其中, Q 和 R 分别代表相似性保持和类别监督的约束, 权衡因子 μ 用来平衡两类约束。下文将具体阐述约束项构成形式及求解方式。

首先, 考虑到 S2PDSoH 方法采用类哈希码近似等效替代二值哈希编码, 在此省略量化误差的正则项, 相似性保持的约束 Q 可表示为:

$$Q = - \sum_{s_{ij} \in S} (s_{ij} \Omega_{ij} - \log(1 + e^{\Omega_{ij}})) \quad (8)$$

其中, $\Omega_{ij} \in [-K, K]$ 是编码 \mathbf{b}_i 和 \mathbf{b}_j 之间的汉明距离, 在具体计算上是使用两个编码向量的内积 $\langle \mathbf{b}_i, \mathbf{b}_j \rangle$, 能够有效表达两个图像之间的相似性程度。

其次, 在类别监督约束 R 的构造上, 主要采用语义分类和交叉熵损失, 利用深度哈希网络前端传递的哈希特征来完成相对应哈希码类别的确定, 具体如下:

$$\mathbf{c}_i = \mathbf{W}_f^T \mathbf{h}_i + \boldsymbol{\tau}_i \quad (9)$$

其中, \mathbf{W}_f^T 和 $\boldsymbol{\tau}_i$ 分别表示分类层的映射权重和偏置向量。需要注意的是, 式(9)是基础的类别预测方式。对于多分类和多标签分类任务, 还需要对 $\mathbf{C} = \{c_i\}_{i=1}^N$ 做 Softmax 软分配和 Sigmoid 二分配, 使用的具体函数如下:

$$\hat{y}_i = \begin{cases} e^{c_i} / \sum_{j=1}^M e^{c_j}, & \text{多分类任务} \\ 1 / (1 + e^{-c_i}), & \text{多标签任务} \end{cases} \quad (10)$$

在得到特征语义信息的有效预测的基础上, 再将其与图像的原始语义信息相融合, 可基于交叉熵监督真实标签 y_i 与预测标签 \hat{y}_i 的差异来保障图像类别约束 R , 具体如下:

$$R = - \frac{1}{N} \sum_{i=1}^N y_i \log \hat{y}_i \quad (11)$$

将式(8)与式(11)进行集成, 可获得约束目标的完整形式, 具体如下:

$$L = \sum_{s_{ij} \in S} (\log(1 + e^{\Omega_{ij}}) - s_{ij} \Omega_{ij}) - \frac{\mu}{N} \sum_{i=1}^N y_i \log \hat{y}_i \quad (12)$$

3.3 优化算法

考虑到集成的目标函数中包括相似性保持和语义标签两类约束条件, 将式(12)中完整的约束目标分解来得到 Q 和 μR 两个子约束, 通过交替优化迭代两个子约束的方式来完成整体约束目标的优化求解。该过程在保证整体目标优化的前提下, 能够保障相似关系和类别语义的联合效果, 下文将详细介绍该优化求解过程。

首先, 固定相似性保持约束 Q , 使用类别监督约束 μR 对类哈希码 \mathbf{h}_i 进行梯度求解, 具体表达式如下:

$$\begin{aligned} \frac{\partial(\mu R)}{\partial \mathbf{h}_i} &= \frac{\partial(\mu R)}{\partial \hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial c_i} \cdot \frac{\partial c_i}{\partial \mathbf{h}_i} \\ &= \begin{cases} \sum_{i=1}^N \mu \cdot (\hat{y}_i - y_i) \cdot \frac{\partial c_i}{\partial \mathbf{h}_i}, & \text{多分类任务} \\ \sum_{i=1}^N \mu \cdot y_i \cdot (1 - \hat{y}_i) \cdot \frac{\partial c_i}{\partial \mathbf{h}_i}, & \text{多标签任务} \end{cases} \end{aligned} \quad (13)$$

其中, $\frac{\partial c_i}{\partial \mathbf{h}_i}$ 代表网络分类层到哈希映射层的梯度回传结果。在链式法则和随机梯度下降法下, 按照依次计算 $\mathbf{h}_i \rightarrow \mathbf{p}_i \rightarrow \mathbf{z}_i \rightarrow \mathbf{x}_i$ 的顺序, 就能够得到对网络参数的一次更新, 进行多次的参数迭代更新后就可基于 $\frac{\partial(\mu R)}{\partial \mathbf{h}_i}$ 来实施后续操作。

此后, 固定类别监督约束 R , 使用相似性保持约束继续通过计算梯度的方式更新参数, 计算式如下:

$$\frac{\partial Q}{\partial \mathbf{h}_i} = \frac{1}{2} \left(\sum_{j: s_{ij} \in S} (a_{ij} - s_{ij}) \mathbf{h}_j + \sum_{j: s_{ji} \in S} (a_{ji} - s_{ji}) \mathbf{h}_j \right) \quad (14)$$

最后, 根据式(12)和式(13)计算的类哈希码 \mathbf{h}_i 的梯度表示, 回到整体优化层面来完整表示相似性保持和类别监督下的梯度求解, 具体形式如下:

$$\frac{\partial L}{\partial \mathbf{h}_i} = \frac{\partial Q}{\partial \mathbf{h}_i} + \frac{\partial(\mu R)}{\partial \mathbf{h}_i} \quad (15)$$

将式(12)和式(13)分别代入, 可具体表示为:

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{h}_i} &= \frac{1}{2} \left(\sum_{j: s_{ij} \in S} (a_{ij} - s_{ij}) \mathbf{h}_j + \sum_{j: s_{ji} \in S} (a_{ji} - s_{ji}) \mathbf{h}_j \right) \\ &\quad \begin{cases} \sum_{i=1}^N \mu \cdot (\hat{y}_i - y_i) \cdot \frac{\partial c_i}{\partial \mathbf{h}_i}, & \text{多分类任务} \\ \sum_{i=1}^N \mu \cdot y_i \cdot (1 - \hat{y}_i) \cdot \frac{\partial c_i}{\partial \mathbf{h}_i}, & \text{多标签任务} \end{cases} \end{aligned} \quad (16)$$

这样, 可根据联合偏导求解多分类与多标签任务, 并同样按照 $\mathbf{h}_i \rightarrow \mathbf{p}_i \rightarrow \mathbf{z}_i \rightarrow \mathbf{x}_i$ 的方式来优化参数, 从而完成一次联合更新过程。重复上述联合更新过程, 当目标函数收敛时即认为完成了整体的目标函数优化过程。该优化过程在保持图像间相似性的同时, 有效结合类标签指导信息, 能够获得更鲁棒的深度高阶哈希编码结果。

4 实验

4.1 基准数据集

为了评估方法的性能, 采用平均准确率均值 (Mean Average Precision, mAP)、top- k 精度及准确率-召回率 (Precision-recall, PR) 评价指标, 在 MNIST, CIFAR-10 及 NUS-WIDE 这 3 个基准数据集上开展测试。实验使用 MatConvNet 工具

箱^[25],PC机配备为3.30GHz CPU,64GB RAM及NVIDIA GTX 1080 GPU,程序软件为MATLAB R2014b。

MNIST^[26]数据集包含70000张手写数字灰度图像,图像大小为 28×28 像素,包括数字0至9共10个类别且每类图像数相等,其是一个单标签数据集。CIFAR-10^[27]数据集包含平均来自10个类别的60000张彩色图像,图像大小为 32×32 像素,同样属于单标签数据集。不同于MNIST和CIFAR10,NUS-WIDE^[28]数据集属于多标签数据集,其每个图像都与81个属性中的一个属性或多个属性相符合。此外,数据集规模也更庞大,总体数量接近270000张并且全部来源于公开的网络,各类中的图像数目也存在差异。实验选择NUS-WIDE中的195834幅图像,均来自21个最常见的属性类别,同时保证每类图像数不少于5000张^[18]。

为公平起见,选择使用深度哈希研究中较为常用的实验标准与设置方式^[18-20]。对于MNIST数据集,从数字0到9的每个类别中随机选择100张图片标记为查询图像,将600张图片作为训练使用。在CIFAR-10中,按照DLBHC^[6]进行数据划分,其中训练集包含50000张图片,测试集包含10000张图片。对于NUS-WIDE^[29]数据集,则分别使用每个属性类别中的500张(一共10500张)和100张(一共2100张)用于模型训练和测试。另外,在所有3个数据集上,哈希编码的训练仅仅使用划分后选择的训练集数据样本,不会使用其他图像数据。

4.2 评价指标

(1) top-k 精度:前 k 张检索到的图像中相关图像所占的百分比可表示为:

$$P@K = I_r / K \quad (17)$$

其中, I_r 和 K 分别代表检索到的相关图片数和返回的图片总数。

(2) 准确率-召回率(PR):用于评测真阳样本数目在召回样本和阳性样本中所占的比例,具体如下:

$$P = TP / (TP + FP) \quad (18)$$

$$R = TP / (TP + FN) \quad (19)$$

其中, P 和 R 分别是准确率和召回率, TP 表示召回图片中真阳性图片数, FP 表示召回图片中假阳性图片数, FN 表示未召回真阳性样本数。

(3) 平均准确率均值(mAP):平均准确率均值 mAP 能够有效衡量哈希函数的整体检索性能,是基于平均精度(AP)的计算结果,表示准确率-召回率曲线下的面积情况,计算式如下:

$$AP(x_q) = \frac{1}{I_r} \sum_{k=1}^N P@K \cdot \psi(k) \quad (20)$$

$$mAP = \frac{1}{M} \sum_{q=1}^M AP(x_q) \quad (21)$$

其中, x_q 表示检索图像, T 是检索数据集的图像数目, $\psi(k)$ 为指标函数;当返回图像与检索图像 x_q 为相同类别时其结果为1,否则为0。

4.3 实验相关设置

考虑到哈希模块中的类哈希码的获取与端到端整体训练的过程一致性,为保证类哈希码能在优化完成后有效地逼近二值编码,在实验初始时将辅助参数 β 设置为1,然后在参数优化的迭代过程中按一定步长动态对辅助参数的值进行调整,当目标函数整体收敛且辅助参数值上升到10时,就能够达到逼近二值化编码的哈希码。另外,将优化相似性保持和

语义标签两类约束的权衡因子 μ 设为0.01,深度高阶特征表达计算的幂次 α 参数设定为0.5。

4.4 实验结果与分析

实验评测 S2PDSOH 方法在上述 3 个基准数据集上的平均准确率均值(mAP)、top-k 精度和准确率-召回率结果,并与 DSDH^[19],DPSH^[18],DTSH^[20],HashGAN^[29],CNNH^[5],DLBHC^[6],DCEH^[16],DPN^[9]和 SSDDH^[10]等典型深度哈希方法进行比较。注意,结果中*表示重新运行作者发布代码获得的实验结果。

4.4.1 平均准确率均值(mAP)的性能比较

第一组实验在 MNIST 数据集上实施,测试 S2PDSOH 方法在使用 12 位、24 位、32 位和 48 位哈希码的 mAP 情况,并与 DLBHC^[6],CNNH^[5]和 DPSH^[18]等方法进行比较,表 1 列出了相应的结果。由表 1 可知,相比其他方法,所提出的 S2PDSOH 在 4 个编码位上均获得最优 mAP 结果,这表明了所提方法的有效性。此外,与同样取得较好结果的 DCEH 方法相比,S2PDSOH 仍然能够在 12 位、24 位和 48 位 3 个结果上获得 0.3%,0.1%和 0.2%的精度提高,主要原因可归结为 S2PDSOH 有效地将分类语义信息和相似关系相融合,在优化策略上也提升了哈希特征的可区分性。

表 1 MNIST 数据集上的平均准确率均值比较

Table 1 Compared mAP results on MNIST

Method	12 bits	24 bits	32 bits	48 bits
DLBHC ^[6]	0.932	0.931	0.937	0.935
CNNH ^[5]	0.957	0.963	0.956	0.960
DPSH ^[18]	0.965	0.963	0.967	0.973
DCEH ^[16]	0.983	0.990	0.989	0.988
S2PDSOH	0.986	0.991	0.989	0.991

第二组实验评测 S2PDSOH 在 CIFAR-10 数据集上的 mAP 结果,同样使用 12 位、24 位、32 位和 48 位 4 个编码位并与 7 种典型方法相比较,结果如表 2 所列。其中,“—”表示没有数值结果,原因是 HashGAN 方法原论文中没提供相应结果,在表 3 中也采用相同的表示方式。观察表 2 可知,S2PDSOH 在所有编码位上同样都优于其他几种典型的深度哈希方法。值得注意的是,S2PDSOH 较 DCEH 仍然有较好的提升,在 4 个结果上分别提升了 1.2%,0.8%,1.3%和 2.0%的精度。与 DSDH 方法相比,S2PDSOH 在 32 位和 48 位结果上分别取得了 1.4%和 0.6%的性能提升。尽管该方法也使用集成约束,但 S2PDSOH 方法在使用成对图像输入下提取二阶深度特征,并有机融合类别监督和相似性保持共同约束,在整体上综合考虑了鲁棒的特征表示与特征优化,使得其能够获得更优的哈希编码结果。

表 2 CIFAR-10 数据集上的平均准确率均值比较

Table 2 Compared mAP results on CIFAR-10

Method	12 bits	24 bits	32 bits	48 bits
DPSH ^[18]	0.713	0.727	0.744	0.757
DSDH ^[19]	0.740	0.786	0.801	0.820
DTSH ^[29]	0.710	0.750	0.765	0.774
HashGAN ^[30]	—	—	0.731	0.735
DCEH ^[16]	0.745	0.788	0.802	0.806
DPN ^[9]	0.755	0.759	0.789	0.769
SSDDH ^[10]	0.749	0.762	0.774	0.785
S2PDSOH	0.757	0.796	0.815	0.826

此后,在多标签 NUS-WIDE 数据集上评测 mAP 情况,并与 7 种典型方法进行比较,结果如表 3 所列。

表 3 NUS-WIDE 数据集上的平均准确率均值比较

Method	12 bits	24 bits	32 bits	48 bits
DPSH ^[18]	0.752	0.790	0.794	0.812
DS DH ^[19]	0.776	0.808	0.820	0.829
DTSH ^[20]	0.773	0.808	0.812	0.824
HashGAN ^[29]	—	—	0.737	0.744
DCEH ^[16]	0.781	0.816	0.827	0.839
DPN ^[9]	0.762	0.793	0.809	0.827
SSDDH ^[10]	0.787	0.811	0.816	0.823
S2PDSoH	0.786	0.826	0.834	0.846

表 3 中,除了在 12 位编码上与 SSDDH 方法获得相当性

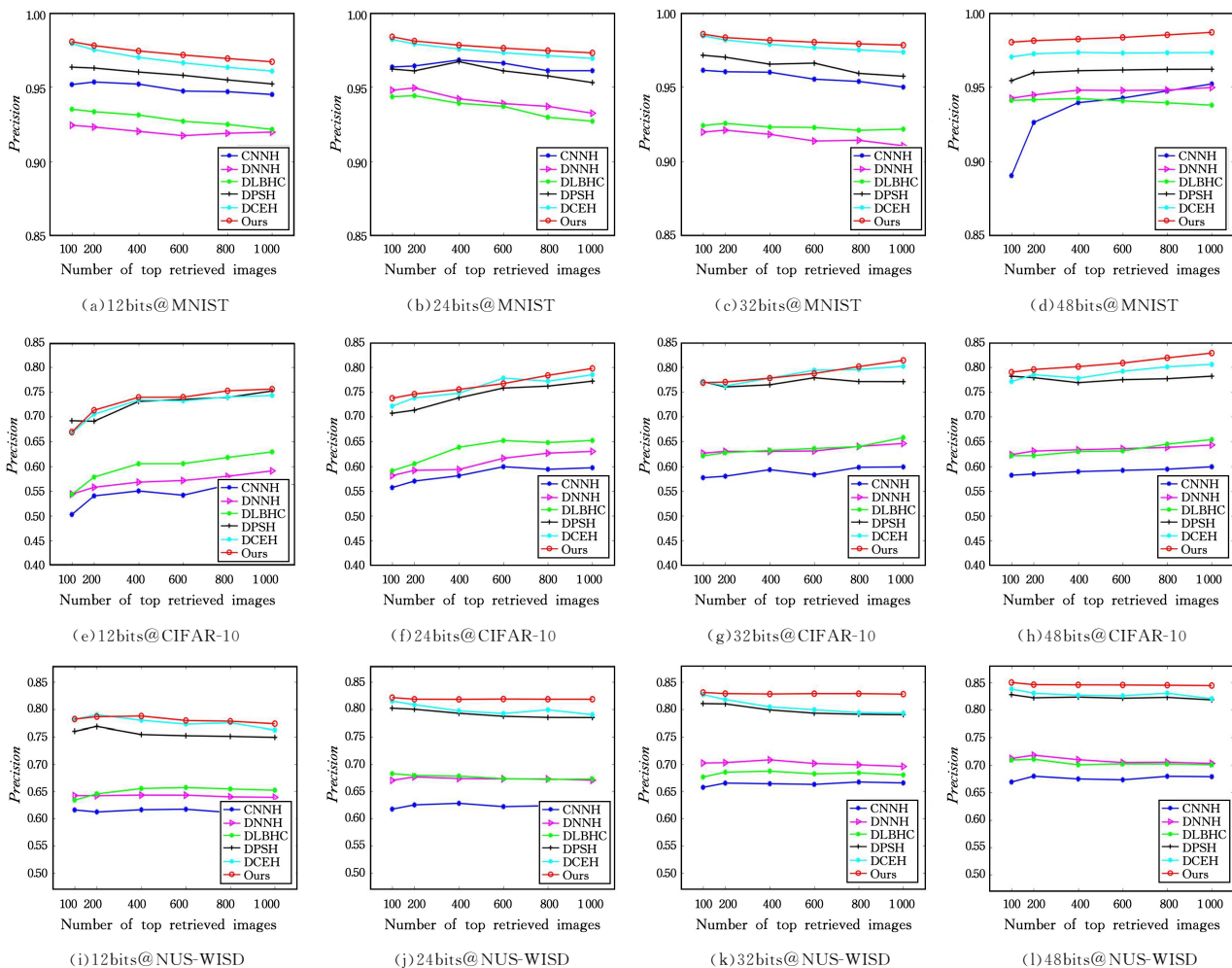


图 3 不同深度哈希方法的 top-k 精度对比结果图

Fig. 3 Compared top-k results of typical deep hashing methods with various hash bits on three datasets

由图 3 可知,除在 CIFAR10 数据库上使用 24 位和 32 位编码时的部分结果略低于 DCEH 方法外,S2PDSoH 在剩余的 10 种情况下均取得了最优的精度结果。同时可知,S2PDSoH 的 top-k 精度情况在高的编码位上具有更加明显的优势。因此,通过同时引入类别监督约束和相似性保持约束,并在集成的约束机制下优化深度二阶哈希深度网络框架,能够获得更具判别性的哈希码。

4.4.3 准确率-召回率性能分析

最后,测试了 S2PDSoH 与典型深度哈希方法^[5-7,16,18]在

能外,S2PDSoH 在其他情况上均获得最好的 mAP 结果,表明了本文方法引入多损失函数集成与优化学习的有效性。特别地,与未使用损失函数集成的 DCEH 方法相比,S2PDSoH 可在 4 个编码位上分别可获得 0.5%,1.0%,0.7% 和 0.7% 的精度收益,更有效地表明了使用多损失函数的创新是有效的,而采用的类哈希码也更好地保证了哈希码的二值离散特性,能够减小量化误差来获得更优的哈希码结果。

4.4.2 top-k 精度性能分析

为了进一步评测 S2PDSoH 的性能,利用 top-k 检索结果来测试哈希查找任务的精度。图 3 给出了在 3 个基准数据集上 S2PDSoH 与几种典型的深度哈希算法^[5-7,16,18]关于 4 种编码位的 top-k 精度情况。

12,24,32 和 48 比特位的准确率-召回率性能结果,如图 4 所示。从图中可以看出,该指标结果上 S2PDSoH 在 4 个编码位上同样获得了最优的评测结果,进一步验证了在深度哈希中获取图像特征全局性和同类样本相似性的可行性。以上比较结果表明,S2PDSoH 在成对图像输入下,利用协方差估计来捕获样本图像的深度二阶信息并集成类别监督和相似性保持约束,并在类哈希码映射下迭代优化,能够保证获得稳定的、具有高区分性的深度二阶哈希码,从而提升检索表现。

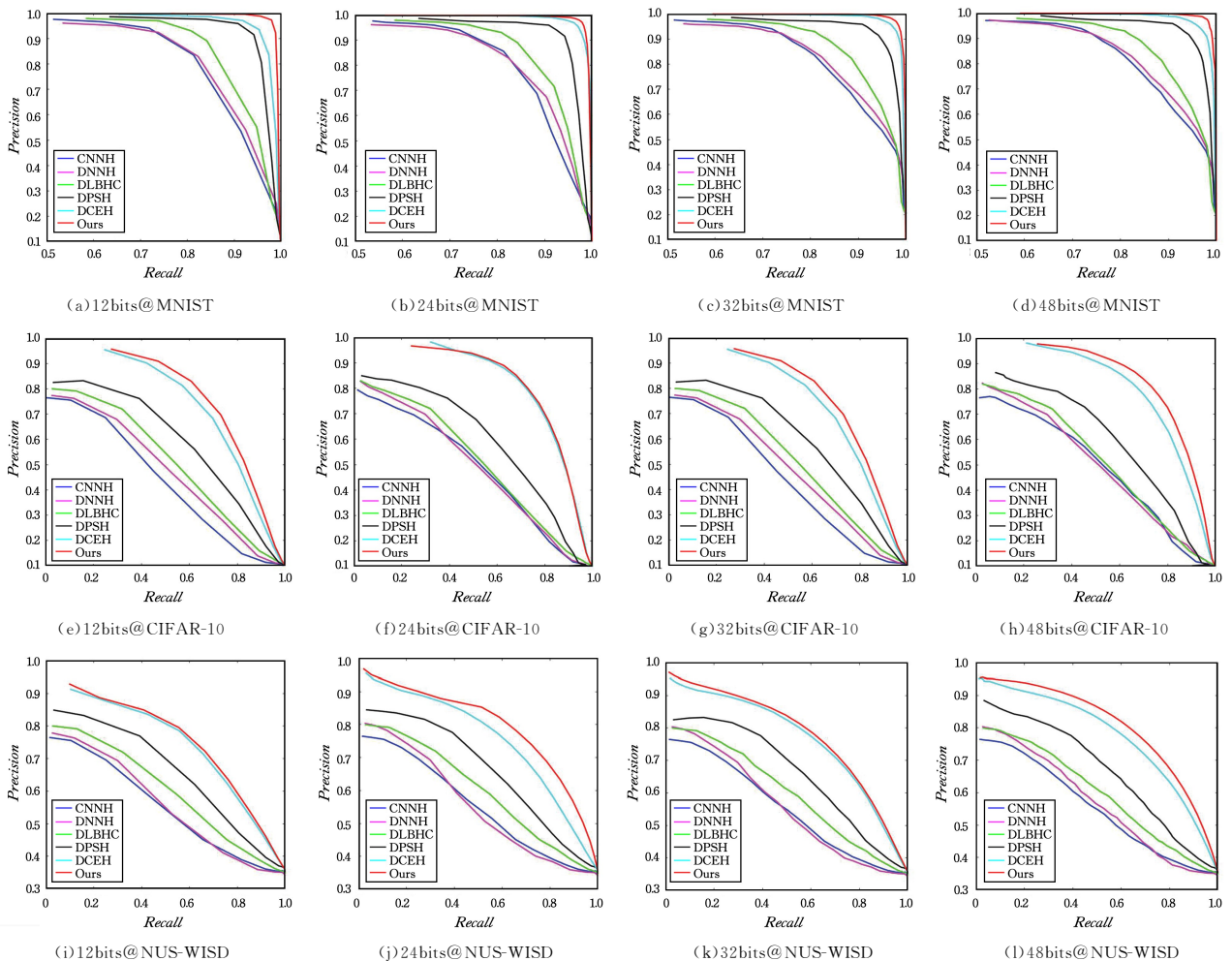


图4 不同深度哈希方法的准确率-召回率性能对比结果图

Fig. 4 Compared precision-recall of typical deep hashing methods with various hash bits on three datasets

结束语 为提高深度哈希在图像检索中的全局表达和同类样本相似性保持能力,提出了一种有监督相似性保持的深度二阶哈希方法。该方法采用成对图像输入,将协方差估计嵌入深度哈希网络中来获取图像高阶全局特征表达,并构建类别监督与相似性保持结合的联合约束条件来指导学习过程,此外借鉴类哈希近似二值化模式来减小哈希映射过程中的量化误差,最终以交替迭代优化方式实现模型端到端训练与最优哈希码的获取,大量对比实验结果充分表明了所提方法在图像检索任务上的有效性。在下一步工作中,将考虑在深度哈希框架下挖掘更丰富的全局化表达信息,此外将继续引入更精简且高效的集成约束形式,以获取更具判别性的哈希表达结果。

参考文献

- [1] RAWAT W, WANG Z H. Deep convolutional neural networks for image classification: a comprehensive review [J]. *Neural Computation*, 2017, 29(9): 2352-2449.
- [2] WANG J D, ZHANG T, SONG J K, et al. A survey on learning to Hash [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 769-790.
- [3] LIU Y, CHENG M, WANG F P, et al. Deep hashing image retrieval methods [J]. *Journal of Image and Graphics*, 2020, 25(7): 1296-1317.
- [4] KE S C, ZHAO Y W, LI B C, et al. Image retrieval based on con-

- volutional neural network and kernel-based supervised hashing [J]. *Acta Electronica Sinica*, 2017, 45(1): 157-163.
- [5] XIA R K, PAN Y, LAI H J, et al. Supervised hashing for image retrieval via image representation learning [C] // *Proceedings of the 28th AAAI Conference on Artificial Intelligence*. Palo Alto: AAAI Press, 2014: 2156-2162.
- [6] LIN K, YANG H F, HSIAO J, et al. Deep learning of binary hash codes for fast image retrieval [C] // *Proceedings of the 28th IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2015: 27-35.
- [7] LAI H J, PAN Y, LIU Y, et al. Simultaneous feature learning and hash coding with deep neural networks [C] // *Proceedings of the 28th IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2015: 3270-3278.
- [8] ZHAO F, HUANG Y Z, WANG L, et al. Deep semantic ranking based hashing for multi-label image retrieval [C] // *Proceedings of the 28th IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2015: 1556-1564.
- [9] FAN L X, NG K W, JU C, et al. Deep polarized network for supervised learning of accurate binary hashing codes [C] // *Proceedings of the 2020 International Joint Conference on Artificial Intelligence*. Japan: IJCAI, 2020: 825-831.
- [10] WAN F, QIANG H P, LEI G B. Self-supervised deep discrete

- hashing for image retrieval [J]. *Journal of Image and Graphics*, 2021, 26(11): 2659-2669.
- [11] JIANG Q Y, LI W J. Discrete latent factor model for cross-modal hashing [J]. *IEEE Transactions on Image Processing*, 2019, 28(7): 3490-3501.
- [12] IONESCU C, VANTZOS O, SMINCHISESCU C. Matrix Back-propagation for deep networks with structured layers [C] // Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2015: 2965-2973.
- [13] LIN T Y, ROYCHOWDHURY A, MAJI S. Bilinear convolutional neural networks for fine-grained visual recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(6): 1309-1322.
- [14] LI P H, XIE J T, WANG Q L, et al. Is second-order information helpful for large-scale visual recognition? [C] // Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 2089-2097.
- [15] WANG Q L, XIE J T, ZUO W M, et al. Deep CNNs meet global covariance pooling: better representation and generalization [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(8): 2582-2597.
- [16] WU Y, SUN Q L, ZHANG J X, et al. Deep covariance estimation hashing for image retrieval [C] // Proceedings of the 26th IEEE International Conference on Image Processing. Los Alamitos: IEEE Computer Society Press, 2019: 2234-2238.
- [17] WU Y, SUN Q L, HOU Y Q, et al. Deep covariance estimation hashing [J]. *IEEE Access*, 2019, 7: 113223-113234.
- [18] LI W J, WANG S, KANG W C. Feature learning based deep supervised hashing with pairwise labels [C] // Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2016: 1711-1717.
- [19] LI Q, SUN Z N, HE R, et al. Deep supervised discrete hashing [C] // Proceedings of the 31st Annual Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2017: 2482-2491.
- [20] WANG X F, SHI Y, KITANI K M, et al. Deep supervised hashing with triplet labels [C] // Computer Vision – ACCV 2016. Lecture Notes in Computer Science. Cham: Springer, 2017, 10111: 70-84.
- [21] CHATFIELD K, SIMONYAN K, VEDALDI A, et al. Return of the devil in the details: Delving deep into convolutional nets [C] // Proceedings of the British Machine Vision Conference. Durham: BMVA Press, 2014: 1-11.
- [22] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C] // Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 770-778.
- [23] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C] // Proceedings of the 26th Annual Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2012: 1097-1105.
- [24] CAO Z, LONG M, WANG J, et al. HashNet: deep learning to hash by continuation [C] // Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 5609-5618.
- [25] VEDALDI A, LENC K. MatConvNet: convolutional neural networks for MATLAB [C] // Proceedings of the 23rd ACM International Conference on Multimedia. New York: Association for Computing Machinery, 2015: 689-692.
- [26] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [J]. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324.
- [27] KRIZHEVSKY A, HINTON G. Learning multiple layers of features from tiny images [D]. Toronto: University of Toronto, 2009.
- [28] CHUA T S, TANG J, HONG R, 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. New York: Association for Computing Machinery, 2009: 1-9.
- [29] CAO Y, LIU B, LONG M S, et al. HashGAN: deep learning to hash with pair conditional wasserstein GAN [C] // Proceedings of the 31st IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 1287-1296.



ZHANG Jian-xin, born in 1981, Ph.D. professor, MS supervisor, is a senior member of CCF. His main research interests include computer vision and intelligent medical data processing.