

深度哈希在图像检索中的研究综述

陆正昀¹, 金露¹, 唐金辉^{2*}

(1. 南京理工大学计算机科学与工程学院, 江苏南京 210094;

2. 南京林业大学信息科学技术学院、人工智能学院, 江苏南京 210037)

摘要:随着图像数据规模的迅速增长,大规模图像检索对效率提出了更高要求。深度哈希能够将高维图像特征映射为紧凑二值码,兼顾深层语义学习与高效图像检索,是这个领域的重要研究方向。现有方法依据监督信息的利用程度,可以划分为无监督、弱监督 and 全监督三类:无监督方法通过建模数据内在结构从无标签数据中挖掘潜在语义信息;弱监督方法从带有噪声或不完整的用户标签中提取有效监督信号;而全监督方法依托完整的类别标签来精确建模语义关系。针对上述三类方法,系统梳理了其核心思想与代表性成果,并在多个主流数据集上对典型方法的检索性能进行综合比较。尽管深度哈希技术已取得显著进展,但在对动态新增数据的适应能力、跨模态场景下的协同建模等方面仍面临严峻挑战。未来的研究应聚焦于基于增量学习的可扩展哈希、基于预训练模型的跨模态哈希等方向,以推动深度哈希向更高效、可扩展和实用的方向发展。

关键词:图像检索;哈希;深度学习

中图分类号:TP391 **文献标志码:**A **文章编号:**1001-2486(2026)03-291-25

A survey on deep hashing for image retrieval

LU Zhengyun¹, JIN Lu¹, TANG Jinhui^{2*}

(1. School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China;

2. College of Information Science and Technology & Artificial Intelligence, Nanjing Forestry University, Nanjing 210037, China)

Abstract: With the rapid expansion of image data, large-scale image retrieval faces increasingly stringent efficiency requirements. Deep hashing is a key research direction in this field by mapping high-dimensional features into compact binary codes, thereby simultaneously enabling deep semantic learning and efficient image retrieval. Existing methods can be classified into three categories according to the extent of supervision utilized: unsupervised, weakly supervised, and fully supervised. Specifically, unsupervised methods mine latent semantic information from unlabeled data by modeling intrinsic data structures; weakly supervised methods extract effective supervisory signals from noisy or incomplete user-provided tags; and fully supervised methods rely on complete class labels to accurately model semantic relationships. The core ideas and representative achievements across these three categories were systematically reviewed, and comprehensive comparisons of retrieval performance for representative methods were conducted on multiple mainstream datasets. Moreover, despite significant progress, deep hashing still confronts substantial challenges in adapting to dynamically arriving data and achieving effective collaborative modeling in cross-modal scenarios. Future research should prioritize incrementally scalable hashing via continual learning, cross-modal hashing leveraging pre-trained models and so on, thereby promoting deep hashing toward greater efficiency, scalability, and real-world applicability.

Keywords: image retrieval; hashing; deep learning

随着互联网上图像数据规模的快速扩张,大规模图像检索问题引发了广泛关注。研究者希望在保证检索精度的前提下,设计高效的算法以降低存储和计算成本。其中,近似最近邻

(approximate nearest neighbor, ANN)搜索通过牺牲部分检索精度实现更低的计算开销,成为图像检索的常用策略。哈希^[1]作为ANN中的主流技术,通过构造哈希函数将高维图像特征映射为简

收稿日期:2026-01-03

基金项目:国家自然科学基金资助项目(62372233)

第一作者:陆正昀(1995—),男,江苏南京人,博士,E-mail:zhengyunlu@njust.edu.cn

*通信作者:唐金辉(1981—),男,江苏丹阳人,教授,博士,博士生导师,E-mail:tangjh@njfu.edu.cn

引用格式:陆正昀,金露,唐金辉.深度哈希在图像检索中的研究综述[J].国防科技大学学报,2026,48(3):291-315.

Citation: LU Z Y, JIN L, TANG J H. A survey on deep hashing for image retrieval [J]. Journal of National University of Defense Technology, 2026, 48(3): 291-315.

短的二进制编码,使得在汉明空间中能够通过简单的异或运算快速判断图像相似性,从而大幅提升了大规模图像检索系统的效率与可扩展性。

近年来,深度学习^[2-3]在图像分类^[4-6]、目标检测^[7-9]、语义分割^[10-13]等多个视觉任务中取得突破性的进展,同时也为图像检索任务提供了新的解决思路。凭借深度神经网络在表示学习方面的强大能力,深度模型能够从原始图像中提取更具有判别性、包含更多语义信息的图像特征,为高质量哈希编码奠定基础。在此背景下,深度图像哈希方法逐步发展,将图像表示学习与哈希函数学习整合到统一的网络架构中,从而生成兼具表达能力与压缩效率的二进

制哈希码。

典型的深度图像哈希框架由两个主要组成部分构成:一个主干网络用于提取图像特征,一个哈希层用于将高维特征压缩为低维二值码,如图 1 所示。然而,深度哈希模型在优化过程中面临离散性约束问题,使得模型难以直接通过反向传播进行训练。为此,常用连续激活函数(如 tanh 函数)替代不可导的符号函数以实现近似优化。此外,另一个核心挑战在于相似性保持机制,即使哈希空间中的相似图像编码彼此接近、不同图像编码互相远离。为解决该问题,已有研究提出多种损失函数,用以缩小哈希空间与语义空间之间的差异。

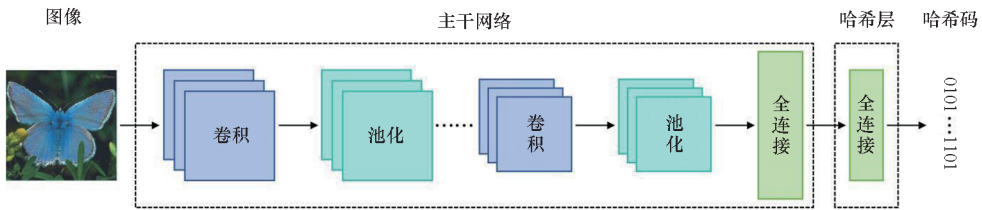


图 1 深度图像哈希框架示意图

Fig. 1 Framework of the deep image hashing

现有深度图像哈希方法可以根据监督信息的使用方式划分为无监督、弱监督和全监督三类,分类框架如图 2 所示。其中,无监督方法在无标签数据上学习图像结构关系,常见子类包括重构型方法、图结构方法与自监督学习方法;弱监督方法利用不完整或带噪声的用户标签进行哈希学习,通过标签重构、标签优化和多模态知识蒸馏提升哈希模型性能;全监督方法依赖类别标签引导哈希码学习,可分为相似度学习、类别标签学习、联合学习以及其他。

尽管已有综述工作^[14]对深度图像哈希方法进行了初步总结,但现有综述集中于无监督深度图像哈希与常规全监督深度图像哈希,对于近年来发展迅速的弱监督深度图像哈希,以及全监督深度图像哈希的相关前沿研究方向(如全监督增量图像哈希,全监督长尾图像哈希和全监督细粒度图像哈希)关注不足。为此,本文从监督信息的使用这一角度出发,构建统一的深度图像哈希方法分类体系,系统梳理各类哈希方法的发展脉络与代表性模型,并在多个主流数据集上对典型哈希方法的检索性能进行比较分析。同时,结合当前研究中的挑战与局限,提出未来可行的研究方向,为后续的研究工作提供参考。

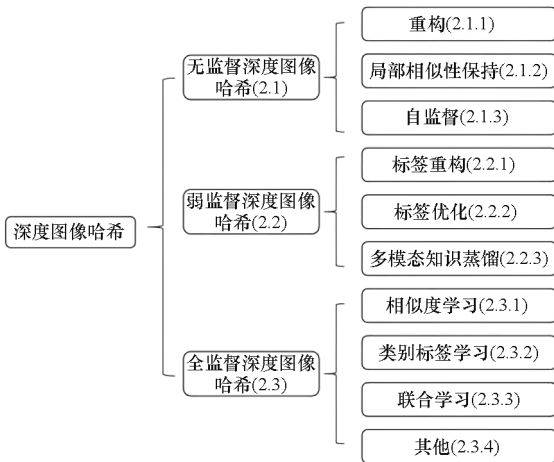


图 2 深度图像哈希分类框架示意图

Fig. 2 Framework of the classification of deep image hashing

1 基础知识

1.1 问题定义

深度图像哈希的目的是利用深度神经网络学习一个哈希映射,能够在保留数据相似性的同时,将高维图像数据转换为低维二值码。这种结合的理论优势在于:一方面利用深度学习强大的非线性表征能力,在特征提取过程中自适应地消除冗余并提取判别性语义;另一方面,通过将连续特征离散化,将高维欧氏空间中的浮点数距离计算转变为汉明空间中的位运算,从而降低存储开销,实

现高效的近似最近邻搜索。

假设训练集为 $D^r = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{N^r}$, 其中 \mathbf{x}_i 表示原始图像, \mathbf{y}_i 为其对应的标签。给定图像 \mathbf{x}_i , L 位哈希码可通过以下方式获得:

$$\mathbf{b}_i = \text{sign}(H(F(\mathbf{x}_i))) \in \{-1, 1\}^L \quad (1)$$

其中, F 表示提取图像特征的主干网络, H 表示投影高维特征向量为低维连续表示的哈希函数, sign 表示将连续表示离散化为二值码的符号函数。然而, 由于 sign 函数在原点处不可导, 在其余位置的导数值都是 0, 深度神经网络无法通过反向传播算法更新网络参数。为了解决这一离散优化难题, 现有方法形成了两种主要的技术路径: 基于连续松弛的一阶段方法和基于预生成码的两阶段方法。

1.1.1 基于连续松弛的一阶段方法

该类方法主张将特征提取与哈希映射整合在统一的优化框架下, 通过引入连续可微的激活函数 \tanh 代替 sign 来松弛离散约束。其理论核心在于语义结构对齐, 即通过设计损失函数, 约束哈希码在汉明空间中的相对分布, 使其尽可能还原数据在语义空间或标签空间中的拓扑结构。为了保持哈希码之间的语义关联, 设 $\langle \mathbf{b}_i, \mathbf{b}_j \rangle$ 为连续哈希表示 \mathbf{b}_i 和 \mathbf{b}_j 的内积, s_{ij} 表示数据间的相似度, 在一个批量 D^M 中, 整体损失函数可以表示为:

$$L = \sum_{i,j=1}^M L_{\text{sim}}(s_{ij}, \langle \mathbf{b}_i, \mathbf{b}_j \rangle) + \lambda \sum_{i=1}^M \|\mathbf{b}_i\|_2^2 \quad (2)$$

其中 $L_{\text{sim}}(s_{ij}, \langle \mathbf{b}_i, \mathbf{b}_j \rangle)$ 是设计的相似性损失, 用数据间的相似度来约束哈希码间的相似度, 而 $\sum_{i=1}^M \|\mathbf{b}_i\|_2^2$ 是量化损失函数, 用于降低松弛离散约束带来的量化误差, 具体训练流程见算法 1。

1.1.2 基于预生成码的两阶段方法

与一阶段方法的连续逼近不同, 两阶段方法采取“先求解离散码, 后拟合映射关系”的策略。在第一阶段, 直接基于标签矩阵或全局相似度矩阵求解二进制二次规划 (BQP) 问题, 预先为训练集生成最优的离散目标码 $\mathbf{K} = \{\mathbf{k}_i\}_{i=1}^{N^r}$, 该过程可表述为最小化数据相似度 s_{ij} 与离散目标码内积 $\langle \mathbf{k}_i, \mathbf{k}_j \rangle$ 之间的重构误差:

$$\min_{\mathbf{K}} \sum_{i,j=1}^{N^r} (\langle \mathbf{k}_i, \mathbf{k}_j \rangle - s_{ij})^2 \quad (3)$$

在第二阶段, 将预生成的离散码视为固定标签, 深度网络的目标是学习一个复杂的非线性映射, 使生成的连续值逼近对应的离散码。此时, 在一个批量 D^M 中, 损失函数为连续哈希码到离散

码的回归:

$$L = \sum_{i=1}^M \|\mathbf{b}_i - \mathbf{k}_i\|_2^2 + \lambda \sum_{i=1}^M \|\mathbf{b}_i\|_2^2 \quad (4)$$

相关训练流程见算法 2。

算法 1 基于连续松弛的一阶段深度哈希

Alg. 1 One-stage depth hashing based on continuous relaxation

输入: 训练集 $D^r = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{N^r}$, 码长 L , 平衡参数 λ , 迭代次数 T , 批量大小 M

输出: 优化后的网络 F, H

For 训练周期 $t = 1$ to T do

For 每个批量 $D^M \subset D^r$ do

图像特征 $\{\mathbf{f}_i\}_{i=1}^M = F(\{\mathbf{x}_i\}_{i=1}^M)$;

连续哈希 $\{\mathbf{b}_i\}_{i=1}^M = \tanh(H(\{\mathbf{f}_i\}_{i=1}^M))$;

基于数据和标签计算相似度 $\{s_{ij}\}_{i,j \in M}$;

基于式(2)计算相似性损失;

反向传播;

优化 F, H 的网络参数;

End for

End for

算法 2 基于预生成码的两阶段深度哈希

Alg. 2 Two-stage deep hashing based on pre-generated code

输入: 训练集 $D^r = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{N^r}$, 码长 L , 平衡参数 λ , 迭代次数 T , 批量大小 M

输出: 优化后的网络 F, H

Stage 1: 计算离散哈希码

基于数据和标签计算相似度 $\{s_{ij}\}_{i,j \in M}$;

基于式(3)生成离散码 $\mathbf{K} = \{\mathbf{k}_i\}_{i=1}^{N^r}$;

Stage 2: 非线性映射拟合

For 训练周期 $t = 1$ to T do

For 每个批量 $D^M \subset D^r$ do

图像特征 $\{\mathbf{f}_i\}_{i=1}^M = F(\{\mathbf{x}_i\}_{i=1}^M)$;

连续哈希 $\{\mathbf{b}_i\}_{i=1}^M = \tanh(H(\{\mathbf{f}_i\}_{i=1}^M))$;

基于式(4)计算和离散码的回归损失;

反向传播;

优化 F, H 的网络参数;

End for

End for

1.2 数据集介绍

近年来, 各类图像检索数据集相继发布, 表 1 总结了在深度图像哈希研究中广泛使用的 12 个检索数据集。

表 1 深度图像哈希中常用检索数据集的总结

Tab. 1 Summary of popular image retrieval datasets used in deep hashing

数据集	类别	图像	文本	粗粒	细粒
				度标	度标
				签	签
CIFAR-10 ^[15]	10	60 000		✓	
CIFAR-100 ^[15]	100	60 000		✓	
ImageNet ^[16]	1 000	1 281 167		✓	
MIRFlickr25k ^[17]	38	25 000	✓	✓	
NUS-WIDE ^[18]	81	269 498	✓	✓	
MS-COCO ^[19]	80	122 218	✓	✓	
CUB-200-2021 ^[20]	200	11 788			✓
Stanford Cars ^[21]	196	16 185			✓
FGVC-Aircraft ^[22]	100	10 000			✓
VegFru ^[23]	292	160 731			✓
Food-101 ^[24]	101	101 000			✓
NABirds ^[25]	555	48 562			✓

在无监督和常规全监督深度图像哈希中,常用的 6 个数据集包括 CIFAR-10^[15]、CIFAR-100^[15]、ImageNet^[16]、MIRFlickr25k^[17]、NUS-WIDE^[18]和 MS-COCO^[19]。其中,CIFAR-10 和 CIFAR-100 是包含 10 类和 100 类粗粒度标签的小图像集合;ImageNet 属于大规模图像检索数据集,包含来自 1 000 个类别的 1 281 167 张图像;MIRFlickr25k、NUS-WIDE 和 MS-COCO 为多模态数据集,前两者包含用户提供的标签,后者包含图像的文本描述。因此,MIRFlickr25k、NUS-WIDE 和 MS-COCO 也被广泛用于弱监督深度图像哈希的研究。

除了上述的粗粒度数据集,还有 6 个用于全监督细粒度图像哈希的图像数据集,包括 CUB-200-2011^[20]、Stanford Cars^[21]、FGVC-Aircraft^[22]、VegFru^[23]、Food-101^[24]和 NABirds^[25]。前 3 个数据集图像数量少于 2 万张,任务难度相对较低,但后 3 个为大规模数据集,任务难度更高。

1.3 评价指标

深度图像哈希方法主要关注计算复杂度与检索性能。其中,计算复杂度取决于编码器网络与哈希层的时间开销和空间开销。通常使用浮点运算次数和内存占用来衡量时间和空间成本。为保证比较的公平性,不同深度哈希方法在相同的时间和空间成本下评估检索性能,即采用相同的主

干网络和数据集设置。

在检索性能的评估方面,常用的评估指标包括:平均精度均值(mean average precision, mAP)、前 N 个返回样本的平均准确率($P@N$)、精度-召回(precision-recall, PR)曲线,以及汉明距离小于等于 2 范围内的平均准确率($P@H=2$)。这些评估指标的定义如下:

mAP 是所有查询图像的平均精度(average precision, AP)得分的均值。AP 的计算方式是在与查询图像相关的每一个位置上,取该位置的准确率,最后求其平均值。其定义为:

$$AP = \frac{1}{R} \sum_{r=1}^R P(r)I(r) \quad (5)$$

其中: $P(r)$ 表示前 r 个返回结果的准确率, $I(r)$ 表示第 r 个返回结果是否相关,若相关则为 1,否则为 0; R 表示与查询图像相关的图像总数。

$P@N$ 被定义为所有查询图像的前 N 个返回结果的平均准确率。

PR 曲线表示在不同召回率水平下的准确率,是评估整体检索性能的重要指标。

$P@H=2$ 表示所有查询图像在汉明距离不超过 2 的范围内的平均准确率。

对于上述评估指标,前三种属于汉明排序标准,而最后一种属于哈希查找标准。对于汉明排序标准,返回的样本根据与查询图像的汉明距离进行排序;而在哈希查找标准中,则根据汉明距离构建一个查找表,该表包含所有与查询图像的汉明距离不超过 2 的图像。

1.4 网络结构

现有的深度图像哈希方法通常采用多种主流神经网络架构作为编码器,包括 AlexNet^[26]、VGG^[27]、ResNet^[28]以及 Vision Transformer(ViT)^[29]。其中,AlexNet 和 VGG 是早期的卷积神经网络架构,在现代大规模视觉任务中的特征表达能力相对有限。ResNet 通过引入残差连接有效缓解了深层网络的梯度消失问题,显著提升了模型深度与表征能力,在各类深度图像任务中表现出色。而 ViT 则摒弃了传统的卷积操作,转而采用自注意力机制,能够显示建模图像块之间的长距离依赖关系,在多项视觉任务中取得了优异性能。

2 深度图像哈希的方法介绍

2.1 无监督深度图像哈希

无监督深度图像哈希的目的是从无标注数据中捕捉图像的结构信息以生成哈希码,为缺少标

注信息的图像检索任务提供了有效解决方案。近年来,相关研究提出了多种无监督深度图像哈希方法,主要包括重构、局部相似性保持和自监督方法,具体方法分类如表 2 所示。

表 2 高质量会议与期刊中发表的无监督深度图像哈希方法总结

Tab. 2 Summary of existing unsupervised deep hashing methods published in high-quality conferences and journals

方法	重构		局部相似性保持		自监督	
	自动编码器	生成对抗网络	相似性构建	图哈希	图卷积网络	实例聚类
Liong 等 ^[30]	✓					
SGH ^[31]	✓					
DVB ^[32]	✓					
TBH ^[33]	✓					
BinGAN ^[34]		✓				
HashGAN ^[35]		✓				
GMANs ^[36]		✓				
UADH ^[37]		✓				
SSDH ^[38]			✓			
MLS ³ RDUH ^[39]			✓			
DATE ^[40]			✓			
UDMSH ^[41]			✓			
Xi 等 ^[42]			✓			
DistillHash ^[43]			✓			
CIMON ^[44]			✓			
UHSCM ^[45]			✓			
SADH ^[46]				✓		
DU3H ^[47]			✓		✓	✓
DeepBit ^[48]						✓
NSH ^[49]			✓			✓
HAMAN ^[50]			✓			✓
WCH ^[51]			✓			✓
FSCH ^[52]			✓			✓
CIBHash ^[53]						✓
Hu 等 ^[54]						✓
Zhang 等 ^[55]						✓
PLDH ^[56]						✓
JMPH ^[57]						✓
CUDH ^[58]						✓
UDKH ^[59]						✓
UCSH ^[60]						✓
DSCH ^[61]			✓			✓
RCSH ^[62]			✓			✓

2.1.1 重构

重构作为无监督深度图像哈希中早期提出的方法,可分为基于自动编码器的重构和基于生成对抗网络(generative adversarial network, GAN)的重构,它们的具体介绍如下:

(1) 自动编码器

基于自动编码器的重构是训练一个编码器,使其输出包含足够量信息的哈希码,以便通过解码器重建输入的图像特征。编码器的目标是从输入特征中生成包含关键信息的哈希码,而解码器则根据哈希码重构输入特征。在一个批量 D^M 中,编码器和解码器通过最小化如下的重构损失函数进行训练:

$$\min_{E,G} \sum_{i=1}^M \mathcal{L}_{re}(f_i, \hat{f}_i) \quad (6)$$

其中, E 表示编码器, G 表示解码器, $\mathcal{L}_{re}(\cdot, \cdot)$ 表示重构损失, f_i 是样本 x_i 的输入特征, \hat{f}_i 是解码器估计得到的特征。估计特征 \hat{f}_i 和编码器 E 以及解码器 G 的函数关系,可表示为:

$$b_i = E(f_i) \quad (7)$$

$$\hat{f}_i = G(b_i) \quad (8)$$

其中, b_i 是由编码器提取的哈希码。

文献[30]采用传统的自编码器,通过哈希码重建输入数据。随机生成哈希^[31](stochastic generative hashing, SGH)提出了一种基于最小描述长度原理的生成模型用于哈希码学习,此外还设计了一种分布式随机梯度下降算法,用于估计二值随机神经元的梯度。深度变分二值化^[32](deep variational binaries, DVB)引入变分自编码器以建模数据的结构统计信息,从而学习鲁棒的二值哈希码。双瓶颈哈希^[33](twin-bottleneck hashing, TBH)提出了一种双瓶颈结构的 Wasserstein 自编码器变体,其中一个瓶颈用于捕捉高层次的数据结构,另一个瓶颈则用于保持所捕捉到的数据结构。

(2) 生成对抗网络

除了自编码器,近年来的无监督哈希研究还采用 GAN 来学习鲁棒的哈希码。与基于自编码器的方法不同,基于 GAN 的方法通常包括生成器、判别器以及编码器。生成器从随机分布中采样噪声生成图像,旨在“欺骗”判别器;判别器的目标是区分生成图像与真实图像;而编码器则将图像投影为哈希码。通过生成器和判别器的对抗训练,模型能够挖掘数据的整体分布特征。

二进制生成对抗网络^[34](binary generative

adversarial network, BinGAN)提出了一种基于 GAN 的哈希架构,并引入正则项来保证高维特征空间中样本对之间的相关性在哈希空间中保持一致。哈希生成对抗网络^[35](hashing generative adversarial network, HashGAN)设计了协同损失,强制编码器对生成器的合成图像输出与原图一致的哈希码,同时引入哈希损失以学习低熵、频率均衡、比特独立的一致性哈希表示。生成式多对抗网络^[36](generative multi adversarial networks, GMANs)提出了双判别对抗训练策略,使所学习的哈希码不仅能保留数据分布信息,而且能接近理想的二值化表示。无监督对抗哈希^[37](unsupervised adversarial hashing, UADH)则进一步扩展了这一思路,通过无监督条件下联合建模数据结构与分布信息,实现了更稳定的哈希学习。

这类方法充分发挥了 GAN 在无监督特征建模中的能力,有效提升了哈希码的判别性与泛化能力,成为近年来无监督深度哈希研究的一个重要方向。

2.1.2 局部相似性保持

局部相似性保持是无监督深度哈希中的一种常见策略,和重构方法不同,它的目的是将数据的潜在语义结构嵌入汉明空间。在该类方法中,主要包含以下三种学习机制:

(1)相似性构建

该策略的目的是在高维特征空间中构建可靠的数据关系,以指导深度哈希模型的学习过程。基于语义结构的无监督深度哈希^[38](semantic structure-based unsupervised deep hashing, SSDH)使用从预训练深度特征中获得的两个半高斯分布估计数据样本之间的成对相似性。基于流形的局部语义相似性结构重建深度无监督哈希^[39](manifold based local semantic similarity structure reconstructing deep unsupervised hashing, MLS³RDUH)则在特征空间的近邻图上应用随机游走,构建流形相似度矩阵,并设计了 logcosh 哈希损失函数,以在汉明空间中保持预定义的流形相似性。此外,基于分布的相似性结构重建^[40](distribution-based similarity structure reconstruction, DATE)开发了一种基于分布发散的度量方式计算相似性,使用基于球散度的非参数双样本检验方法来建模数据分布。为挖掘细粒度的语义结构,无监督深度多相似性哈希^[41](unsupervised deep multi-similarity hashing, UDMSH)被提出,分别通过基于全卷积网络(fully convolutional network, FCN)的注意力网络和卷积

神经网络(convolutional neural network, CNN)构建来自全局感知和空间感知表示的多层次相似性。类似地, Xi 等^[42]设计了一种蒸馏损失函数,将全局特征中的相关性知识迁移到哈希码中。

预定义的相似性中可能包含大量噪声,错误的信息对会在哈希模型的学习过程中导致梯度的误差累积。为解决这一问题,一些研究尝试利用多种技术对噪声相似性进行优化。蒸馏哈希^[43](distilled hashing, DistillHash)提出了一种基于贝叶斯的哈希框架,通过贝叶斯最优分类器在概率假设下修正噪声样本对,从而获得精炼的数据对集合,并使用贝叶斯学习框架从这些数据对中学习哈希码。全面相似性挖掘与一致性学习^[44](comprehensive similarity mining and consistency learning, CIMON)则利用谱聚类优化噪声相似性,并通过两个半高斯分布的累积分布函数获取对应的置信矩阵。基于优化后的相似性与置信矩阵, CIMON 使用加权语义一致性损失和实例级对比损失训练深度哈希模型。

近年来,视觉-语言预训练模型在各类视觉任务中展现出显著优势。受此启发,基于语义概念挖掘的无监督哈希^[45](unsupervised hashing with semantic concept mining, UHSCM)提出利用视觉-语言预训练模型从图像中挖掘高层次语义概念,并构建细粒度相似性矩阵。在此基础上, UHSCM 引入对比损失以使哈希码之间的相似性与所获得的细粒度语义相似性对齐。

(2)图哈希

图哈希问题在无监督浅层哈希方法中已被广泛研究。该方法通过图模型捕捉样本之间的成对相似性,并对图模型进行特征值分解,从而获得二值哈希码。尽管该方法在浅层模型中表现良好,但由于优化困难,在无监督深度哈希中较少被研究。

相似性自适应深度哈希^[46](similarity-adaptive deep hashing, SADH)是首个在无监督深度哈希框架中处理非平滑离散哈希问题的工作。该方法提出了一种基于乘子交替方向乘子法的二值码优化算法,从训练集中学习离散哈希码,然后利用这些哈希码来训练深度哈希模型。同时,该方法根据已学习的深度哈希模型的特征表示自适应地更新图模型,从而有效捕捉语义结构,学习具有判别性的哈希码。

(3)图卷积网络

深度无监督混合相似哈达玛哈希^[47](deep unsupervised hybrid-similarity Hadamard hashing, DU3H)解决了传统无监督哈希方法中存在的哈

希码独立性和平衡性不足,以及构建相似度矩阵时忽视语义相似度差异的问题。首先,DU3H通过 K -means 聚类获取伪类别,将图像分配给采样自哈达玛矩阵的二进制正交码,有效保证哈希码的独立性质和平衡性。其次,文献[47]构建了混合相似度矩阵,显式区分高置信度和普通置信度的图像对,充分利用数据内在结构以提高相似度表示的准确性。

2.1.3 自监督

自监督学习利用从数据中提取的各种自监督信号来引导模型训练。在无监督深度哈希中,主要存在两种自监督学习策略:实例级方法和聚类级方法。

(1)实例级方法

这类方法将同一实例经过不同增强处理后的视图视为正样本对,而不同实例之间视为负样本对。哈希模型的训练目标是使正样本对的哈希码相互接近,负样本对的哈希码相互远离,使学习到的哈希码具备实例级别的判别能力。例如,深度二进制(deep bit, DeepBit)方法^[48]通过旋转输入图像的不同角度来构造正样本对,并最小化它们之间的汉明距离以训练深度哈希模型。与直接比较成对数据的哈希码不同,自然排序哈希^[49](naturally-sorted hashing, NSH)探索哈希码与潜在特征之间的相似性排序一致性,并引入 SortedNCE 损失函数实现这一目标。

然而,上述增强方式在多标签数据集上的性能提升有限,主要原因是多标签图像的语义在增强过程中发生了扭曲。为了解决这一问题,基于细粒度语义相似性挖掘的哈希^[50](hashing with fine-grained semantic similarity mining, HAMAN)利用局部图像块特征构建了细粒度相似度矩阵,以此生成伪标签,并设计了条件对比学习损失,在细粒度监督的引导下比较哈希码。与 HAMAN 类似,加权对比哈希^[51](weighted contrastive hashing, WCH)设计了新的互注意力模块以缓解该问题。而基于细粒度相似性保持的对比哈希^[52](fine-grained similarity-preserving contrastive hashing, FSCH)选择同时探索全局与局部的细粒度相似性一致性来学习哈希码。

此外,对比信息瓶颈哈希^[53](contrastive information bottleneck hashing, CIBHash)通过信息瓶颈机制学习概率二进制表示,它最大化二进制表示与相似标签之间的互信息,同时最小化二进制表示与潜在表示之间的互信息。这样,学习到的二进制表示不仅能够捕捉预测标签所需的关键

信息,还能减少图像中无关信息的负面影响。

(2)聚类级方法

与实例级方法不同,聚类级方法通过将相似的数据样本归为同一类别来构造伪类别标签。所获得的伪类别标签用于引导哈希模型从图像中挖掘高层次的语义信息。近年来,许多聚类级方法被提出,主要分为两阶段方法和端到端方法。

一般而言,两阶段方法首先使用 K -means 算法从整个数据集中获得伪类别标签,然后将哈希码分类到这些伪标签中^[54-56]。此外,联合多层感知机与伪标签的无监督哈希^[57](joint MLP and pseudo-label unsupervised hashing, JMPH)结合 Faster R-CNN^[63]网络和多头注意力模型,构造伪目标级标签,将这些目标语义嵌入哈希模型。

然而,此类方法的一个缺点是伪标签构建与深度哈希模型学习相互独立。伪标签来自预训练模型,基于预定义标签训练哈希模型限制了其捕获目标领域知识的能力,原因在于不同领域间的知识差异。为此,已有研究^[58-60]致力于将聚类与深度哈希整合到统一的学习框架中。

除了以上的实例级方法和聚类级方法,现有自监督哈希还会考虑两种方法的联合。例如,深度语义成分哈希^[61](deep semantic components hashing, DSCH)引入期望最大化框架,联合学习多个潜在语义成分(如共现关系和同源关系)及哈希码。具体而言,利用哈希码获取潜在语义成分:通过高斯混合模型建模共现关系,生成一组细粒度中心;通过聚类建模同源关系,生成一组粗粒度中心。这些潜在语义成分反过来也用于对哈希码进行分组。哈希码与潜在语义成分在期望最大化算法框架下进行迭代优化。

由于 K -means 算法需对整个数据集进行聚类分配,在大规模数据集上执行一次完整的前向传递计算复杂度较高。而且哈希码判别能力有限,直接对二进制哈希码进行聚类会导致聚类分配不可靠。为克服这些问题,关系一致性诱导的自监督哈希^[62](relational consistency induced self-supervised hashing, RCSH)在无监督深度哈希框架中引入在线聚类机制,对当前批次数据进行聚类分配。同时设计双原型对比损失,在潜在特征空间获得可靠的原型分配,并在汉明空间中保持这一分配的稳定性。

2.1.4 小结

综上所述,当前无监督深度图像哈希方法主要围绕重建模、局部结构保持与自监督信号利用三条技术路径展开。基于自动编码器或 GAN

的重构方法虽能有效捕捉数据分布,但其重建目标与哈希检索任务之间存在语义鸿沟,易导致哈希码判别性不足;局部相似性保持策略通过构建图结构或流形关系引入隐式语义,然而预定义的相似性矩阵常受噪声干扰,且对近邻参数敏感,影响模型鲁棒性;而自监督方法,尤其是对比学习范式,近年来展现出更强的语义挖掘能力,但其实例级增强在多标签场景下易失真,聚类级伪标签又面临分配不稳定与训练解耦的问题。总体而言,现有无监督方法仍难以在无标注条件下有效建模高层语义一致性,且缺乏统一的优化框架来协同处理离散约束、结构建模与语义对齐。未来研究可探索结合视觉-语言预训练先验、动态图学习机制或信息瓶颈理论,以提升无监督哈希的语义表达能力与泛化性能。

2.2 弱监督深度图像哈希

由于缺乏标签信息,无监督深度图像哈希方法在图像检索中的表现往往受限。然而,随着社交媒体网站的快速发展,用户提供的标签成为一种易于获取且成本低廉的附加信息源。这些标签作为图像的文本描述,蕴含了一定的语义信息。基于此,研究者们尝试利用图像及其对应的标签来学习哈希模型,这被称为弱监督深度图像哈希。现有的弱监督深度图像哈希方法主要分为三类:标签重构、标签优化和多模态知识蒸馏,具体分类如表 3 所示。

表 3 高质量会议与期刊中发表的弱监督深度图像哈希方法总结

Tab. 3 Summary of existing weakly supervised deep hashing methods published in high-quality conference and journal

方法	标签 重构	标签 优化	多模态知 识蒸馏
WDHT ^[64]	✓		
Guan 等 ^[65]	✓		
Zhu 等 ^[66]	✓		
Zhao 等 ^[67]	✓		
SCADH ^[68]		✓	
MGRN ^[69]		✓	
EWSH ^[70]		✓	
WSDHQ ^[71]		✓	
WSHRCA ^[72]		✓	
MKDH ^[73]			✓

2.2.1 标签重构

此类方法旨在从哈希码重建标签嵌入,使得学习到的哈希码尽可能保留标签的语义信息。基于标签嵌入的弱监督深度哈希^[64] (weakly supervised deep hashing using tag embeddings, WDHT) 利用词向量模型^[74] (word to vector, word2vec) 提取标签嵌入,再通过成对相似度损失和小批量铰链损失将哈希码与标签嵌入对齐。

为了减轻噪声标签的影响,Guan 等^[65] 提出不直接使用 word2vec 标签嵌入,而是通过学习语义向量来更好地表示图像。具体来说,语义向量与二进制哈希码通过稀疏编码算法联合学习,随后基于所学哈希码监督训练深度哈希模型。同时,Zhu 等^[66] 引入 $l_{2,1}$ 范数以降低噪声标签的影响。Zhao 等^[67] 则设计了一个语义网络以捕捉标签的重要性。

2.2.2 标签优化

由于用户提供的标签通常存在不完整、主观和噪声问题,一些研究致力于在学习深度哈希模型的同时解决标签问题。例如,可扩展深度哈希^[68] (scalable deep hashing, SCADH) 通过鲁棒主成分分析解决标签优化问题,并引入基于增强拉格朗日乘子的离散优化算法学习二进制哈希码。基于掩码的视觉-语义图推理网络^[69] (masked visual-semantic graph-based reasoning network, MGRN) 利用图注意力网络捕捉图像与标签间的依赖关系,从而通过重点关注相关标签而忽略无关标签,学习对齐的视觉-语义表示。基于这种联合表示,利用成对相似度损失和重建损失学习判别哈希码。

增强型弱监督哈希^[70] (enhanced weakly supervised hashing, EWSH) 提出统一学习框架,交替优化哈希码和标签。标签通过贪心算法最大化图像-标签相似度,实现添加相关标签与移除噪声标签。哈希码则通过保留优化后标签的成对相似度进行学习。弱监督深度超球量化^[71] (weakly-supervised deep hyperspherical quantization, WSDHQ) 提出结合标签信息的深度量化框架。它构建标签相关图以减少相似标签数量并增强其语义信息,再将优化后的标签嵌入与图像特征映射到同一语义超球面,并进行匹配。基于重构性跨模态注意力的弱监督哈希^[72] (weakly supervised hashing with reconstructive cross-modal attention, WSHRCA) 通过设计跨模态注意力模块,聚合内容相关的标签嵌入并抑制噪声标签嵌入。

2.2.3 多模态知识蒸馏

无论是标签重构方法还是标签优化方法,都仅依赖用户标签中所提供的文本信息,而忽略了图像自身所蕴含的丰富视觉内容,这在一定程度上限制了哈希模型在封闭词汇和开放词汇检索任务中的表现。为克服这一局限,多模态知识蒸馏哈希方法^[73] (multi-modal knowledge distillation hashing, MKDH)被提出。该方法利用对比语言-视觉预训练模型^[75] (contrastive language-image pre-training, CLIP)中融合的图片与文本知识,引导哈希网络学习具有更强鲁棒性的哈希码。

具体而言,该方法设计了一种双重注意力适配器结构,从 CLIP 教师网络中提取细粒度的图像与文本特征,生成跨模态联合表征。再通过知识蒸馏,实现多模态知识向学生哈希网络的有效迁移。此外,针对弱监督环境中伪负样本对可能带来的干扰,MKDH 还提出了一种加权策略。

2.2.4 小结

总体而言,弱监督深度哈希虽有效利用了低成本用户标签,但仍受限于标签噪声、稀疏性与主观性。标签重构方法依赖预训练语义嵌入,易受标签质量影响;标签优化策略虽能修正噪声,但常引入复杂优化流程且难以建模标注不确定性;而基于多模态知识蒸馏的方法(如 MKDH)借助 CLIP 等预训练模型提升了语义泛化能力,但其多模态建模仍局限于两层注意力结构,对细粒度跨模态交互的挖掘尚不充分,存在进一步改进空间。当前工作普遍缺乏对弱监督信号可靠性的系统建模,也较少关注跨场景迁移能力。未来可结合不确定性感知学习、轻量化蒸馏与动态标签校正机制,提升方法的鲁棒性与实用性。

2.3 全监督深度图像哈希

相较于无监督深度图像哈希(完全缺乏标签信息,仅依赖数据内在结构)和弱监督深度图像哈希(仅依赖含噪、稀疏或不完整的用户标签),全监督深度图像哈希能够充分利用训练集中完整的类别标注信息来指导哈希码的学习过程。这种机制使得模型能够在二进制嵌入空间中更精确地建模语义相似性,强制同类样本的哈希码汉明距离最小化,而异类样本的哈希码汉明距离最大化。因此,相比较无监督和弱监督方法,全监督深度哈希展现出显著优越的性能,广泛应用于各种检索任务中。现有全监督深度哈希方法可分为四类:相似度学习、类别标签学习、联合学习以及其他方法,具体的方法分类如表4所示。

表4 高质量会议与期刊中发表的全监督深度图像哈希方法总结

Tab.4 Summary of existing supervised deep hashing methods published in high-quality conferences and journals

方法	相似度学习	类别标签学习	联合学习	其他方法
CNNH ^[76]	✓			
HashNet ^[77]	✓			
DCH ^[78]	✓			
DAGH ^[79]	✓			
DSH ^[80]	✓			
MMHH ^[81]	✓			
Zhang 等 ^[82]	✓			
DTQ ^[83]	✓			
Lai 等 ^[84]	✓			
DOH ^[85]	✓			
DSRH ^[86]	✓			
He 等 ^[87]	✓			
ADSH ^[88]	✓			
DAPH ^[89]	✓			
SSDH ^[90]		✓		
GreedyHash ^[91]		✓		
JMLH ^[92]		✓		
DSEH ^[93]		✓		
LAH ^[94]		✓		
CSQ ^[95]		✓		
DPN ^[96]		✓		
OrthoHash ^[97]		✓		
Wang 等 ^[98]		✓		
SPRCH ^[99]			✓	
DIHN ^[100]				✓
Chen 等 ^[101]				✓
DIH ^[102]				✓
LTHNet ^[103]				✓
ACHNet ^[104]				✓
DDPHN ^[105]				✓
SPH ^[106]				✓
CIH ^[107]				✓
Zeng 等 ^[108]				✓
DSaH ^[109]				✓
sRLH ^[110]				✓
FPH ^[111]				✓
CFH ^[112]				✓
ExchNet ^[113]				✓
CSCE-Net ^[114]				✓
SwinFGHash ^[115]				✓
GDF-Net ^[116]				✓
FCAENet ^[117]				✓
SEMICON ^[118]				✓
FISH ^[119]				✓
DPAQ ^[120]				✓
A ² -NET ^[121]				✓
AGMH ^[122]				✓
CFBH ^[123]				✓

2.3.1 相似度学习

相似度学习通过保留样本间的成对相似关系学习哈希编码。假设 $\mathbf{S} \in \{-1, 1\}^{N^r \times N^r}$ 为相似度矩阵, 当 $\mathbf{y}_i^T \mathbf{y}_j = 1$ 时, $s_{ij} = 1$, 否则 $s_{ij} = -1$ 。在一个批量 D^M 中, 其目标函数可以表述为:

$$\sum_{i,j=1}^M \mathcal{L}_{ps}(\hat{s}_{ij}, s_{ij}) \quad (9)$$

其中, $\hat{s}_{ij} = f(\mathbf{b}_i, \mathbf{b}_j)$ 表示哈希编码 \mathbf{b}_i 与 \mathbf{b}_j 之间的预测相似度, $\mathcal{L}_{ps}(\cdot, \cdot)$ 是用于衡量相似度差异的损失函数。一些具有代表性的设计如下:

(1) 重构损失

卷积神经网络哈希^[76] (convolutional neural network hashing, CNNH) 将全监督哈希与深度卷积神经网络相结合, 采用两阶段的范式, 包括哈希码学习和哈希网络学习。在哈希码学习阶段, 通过最小化重构损失学习哈希码, 如式(10)所示, 以保持样本间的成对相似性; 在哈希网络学习阶段, 利用所学二值码训练深度模型, 用于生成新样本的哈希码。

$$\mathcal{L}_{CNH} = \sum_{i,j=1}^M \left(s_{ij} - \frac{1}{L} \mathbf{b}_i^T \mathbf{b}_j \right)^2 \quad (10)$$

(2) 负对数似然损失

哈希网络^[77] (HashNet) 基于成对相似关系的二元逻辑 (logistic) 回归建模, 采用重加权的二元组负对数似然损失。给定哈希码 \mathbf{b}_i 和 \mathbf{b}_j , 它们之间的内积 Θ_{ij} 可表示为:

$$\Theta_{ij} = \frac{1}{2} \mathbf{b}_i^T \mathbf{b}_j \quad (11)$$

对应的重加权负对数似然损失可以表示为:

$$\mathcal{L}_{HashNet} = \sum_{i,j=1}^M \omega_{ij} [\ln(1 + e^{\Theta_{ij}}) - \alpha s_{ij} \Theta_{ij}] \quad (12)$$

其中, ω_{ij} 用于平衡相似对与不相似对的数量。同时, HashNet 在训练中逐步增大哈希网络输出层中 tanh 函数的温度参数, 使 tanh 函数逐渐逼近 sign 函数, 实现稳定的二值化学习。

与 HashNet 不同, 深度柯西哈希^[78] (deep Cauchy hashing, DCH) 改用柯西分布建模距离, 其负对数似然损失为:

$$L_{DCH} = \sum_{i,j=1}^M \omega_{ij} \left[s_{ij} \ln \left(\frac{d_{ij}}{\lambda} \right) + \ln \left(1 + \frac{\gamma}{d_{ij}} \right) \right] \quad (13)$$

其中, γ 是柯西分布的尺度参数, $d_{ij} = \frac{L}{2} [1 - \cos(\mathbf{b}_i, \mathbf{b}_j)]$ 表示基于余弦相似度的距离函数。此外, 为了缓解由连续松弛引起的量化误差, DCH 还设计了一个基于柯西分布的量化损失, 如下

所示:

$$\mathcal{L}_q = \sum_{i=1}^M \ln \left(1 + \frac{d(\|\mathbf{b}_i\|, \mathbf{1})}{\gamma} \right) \quad (14)$$

其中, $\mathbf{1} \in \mathbb{R}^L$ 是一个全 1 向量。

深度哈希网络是通过对小批量数据进行采样来更新的, 这导致剩余的训练数据未被充分利用。为克服这一缺点, 深度锚图哈希^[79] (deep anchor graph hashing, DAGH) 结合锚图学习, 引入回归项最小化连续编码与二值码距离, 目标函数为:

$$\min_{\mathbf{B}^{all}, \mathbf{U}} \sum_{i=1}^{N^r} \sum_{j=1}^{N^{bt}} a_{ij} \|\mathbf{b}_{:,i}^{all} - \mathbf{u}_j\|^2$$

$$\text{s. t. } \mathbf{b}_{:,i}^{all} \in \{-1, 1\}^L, \mathbf{b}_{:,i}^{all} \mathbf{1}_{N^r} = 0 \quad (15)$$

其中, $\mathbf{B}^{all} \in \{-1, 1\}^{L \times N^r}$ 表示所有训练数据的二进制哈希码, $\mathbf{b}_{:,i}^{all}$ 表示 \mathbf{B}^{all} 的第 i 列, $\mathbf{b}_{:,i}^{all}$ 表示其第 i 行, $\mathbf{1} \in \mathbb{R}^{N^r}$ 是一个全为 1 的向量, a_{ij} 是锚图 $A \in \mathbb{R}^{N^r \times N^{bt}}$ 中的第 (i, j) 个边的权重。 a_{ij} 的计算方式如下:

$$a_{ij} = \begin{cases} \frac{1}{|\mathcal{D}_i^+|} & s_{ij} = 1 \\ 0 & s_{ij} = 0 \end{cases} \quad (16)$$

其中, \mathcal{D}_i^+ 表示与 \mathbf{x}_i 相关的数据点集合。

(3) 最大间隔损失

最大间隔损失首先定义一个合适的边界, 然后利用该边界来区分相似样本与不相似样本。深度全监督哈希^[80] (deep supervised hashing, DSH) 能够最小化相似样本对的汉明距离, 同时当不相似样本对的汉明距离小于设定的间隔时, 最大化它们的距离。在一个批量 D^M 中, DSH 的目标函数被定义为:

$$\mathcal{L}_{DSH} = \sum_{i,j=1}^M \frac{1}{2} s_{ij} \|\mathbf{b}_i - \mathbf{b}_j\|^2 + \frac{1}{2} (1 - s_{ij}) \max(m - \|\mathbf{b}_i - \mathbf{b}_j\|^2) + \alpha (\|\|\mathbf{b}_i\| - 1\|_1 + \|\|\mathbf{b}_j\| - 1\|_1) \quad (17)$$

其中, m 是边界参数, α 是用于控制量化损失强度的参数, 而 $\|\cdot\|_1$ 表示向量的 $L1$ 范数。

与 DSH 不同, 最大间隔汉明哈希^[81] (maximum-margin hamming hashing, MMHH) 提出了一种最大间隔 t -分布损失函数, 关注位于汉明球外部的相似样本对:

$$\mathcal{L}_{mmh} = \sum_{i,j=1}^M w_{ij} s_{ij} \ln(1 + \max(0, D(\mathbf{b}_i, \mathbf{b}_j) - h)) + \sum_{i,j=1}^M w_{ij} (1 - s_{ij}) \ln \left(1 + \frac{1}{\max(H, D(\mathbf{b}_i, \mathbf{b}_j))} \right) \quad (18)$$

其中, w_{ij} 是用于缓解数据不平衡问题的样本对

$(\mathbf{b}_i, \mathbf{b}_j)$ 的权重, h 表示汉明球的半径, $d(\mathbf{b}_i, \mathbf{b}_j) = \frac{L}{2} [1 - \cos(\mathbf{b}_i, \mathbf{b}_j)]$ 用于衡量 \mathbf{b}_i 和 \mathbf{b}_j 之间的汉明距离。

除了上面的二元组间隔损失, 一些方法^[82-84] 利用三元组间隔损失来保持三元组之间的排序关系。给定一个三元组 $(\mathbf{x}_i, \mathbf{x}_i^+, \mathbf{x}_i^-)$, 基于三元组间隔损失的目标函数可以定义为:

$$\mathcal{L}_{\text{triplet}} = \max(m, d(\mathbf{b}_i, \mathbf{b}_i^+) - d(\mathbf{b}_i, \mathbf{b}_i^-)) \quad (19)$$

其中, m 表示正负样本对之间的间隔。

(4) 排序损失

排序损失通过探索数据样本之间的相对排序相关性来学习具有判别性的表示。深度序数哈希^[85] (deep ordinal hashing, DOH) 利用双分支网络分别提取局部空间与全局语义信息, 并融合二者学习基于排序的哈希函数:

$$\mathcal{L}_{\text{DOH}} = \frac{1}{2} \sum_{i,j=1}^M \left(\frac{1}{R} \sum_{r=1}^R (\mathbf{h}_i^r)^T \mathbf{h}_j^r - s_{ij} \right)^2 \quad (20)$$

其中

$$\mathbf{h}^r = \text{softmax}(f_{\text{fuse}}(\mathbf{v}^l, \mathbf{v}^g)) \quad (21)$$

函数 $f_{\text{fuse}}(\cdot, \cdot)$ 用于融合局部空间表示 \mathbf{v}^l 和全局表示 \mathbf{v}^g 。在式(20)中, 当与 L bit 的二值哈希码进行比较时, 长度 R 被设置为 $L/\log_2 K$ 。

基于深度语义排序的哈希^[86] (deep semantic ranking based hashing, DSRH) 则利用了数据对的排序列表中的顺序信息来学习深度哈希模型。给定一个样本 \mathbf{x}_i 及其对应的排序列表 $\{\mathbf{x}_j\}_{j=1}^M$, 排序损失定义为:

$$\mathcal{L}_{\text{sl}} = \sum_{j=1}^M \sum_{k:r_k < r_j} \max(D_{\text{H}}(\mathbf{b}_i, \mathbf{b}_k, \mathbf{b}_j) + \gamma, 0) \quad (22)$$

其中: r_k 和 r_j 定义了样本对 $(\mathbf{x}_i, \mathbf{x}_k)$ 与 $(\mathbf{x}_i, \mathbf{x}_j)$ 之间的相似性等级; $D_{\text{H}}(\mathbf{a}, \mathbf{b}, \mathbf{c}) = D_{\text{H}}(\mathbf{a}, \mathbf{b}) - D_{\text{H}}(\mathbf{a}, \mathbf{c})$, $D_{\text{H}}(\mathbf{a}, \mathbf{b})$ 表示向量 \mathbf{a} 与 \mathbf{b} 之间的汉明距离, γ 是一个间隔参数, 用于控制两对样本之间距离的最小差距。与 DSRH 类似, He 等^[87] 将排序评价指标 (如平均精度和归一化折损累计增益) 引入哈希函数的学习, 以提升检索结果的排序性能。

(5) 离散优化

对二值约束的松弛通常会引入较大的量化误差, 从而影响哈希编码的质量。一些方法通过直接求解离散优化问题来生成高质量二值哈希码。例如, 不对称深度监督哈希^[88] (asymmetric deep supervised hashing, ADSH) 和深度不对称成对哈希^[89] (deep asymmetric pairwise hashing, DAPH)。具体来说, 在查询样本 $\{\mathbf{x}_i\}_{i=1}^{N_q}$ 与数据库样本

$\{\mathbf{x}_i\}_{i=1}^{N_{\text{db}}}$ 之间构建一个不对称的相似性矩阵 $\mathbf{S} \in \{-1, +1\}^{N_q \times N_{\text{db}}}$, 目标函数可以表述为:

$$\begin{aligned} \mathcal{L}_{\text{asy}} &= \sum_{s_{ij} \in \mathbf{S}} \ell(f_s(\mathbf{u}_i, \mathbf{b}_j), s_{ij}) \\ \text{s. t. } \mathbf{U} &\in \{-1, +1\}^{L \times N_{\text{db}}} \end{aligned} \quad (23)$$

其中, \mathbf{U} 表示数据库样本的二值哈希码, \mathbf{u}_i 是 \mathbf{U} 的第 i 列, $f_s(\cdot)$ 用于衡量两个给定向量之间的相似性, $\ell(\cdot)$ 定义了损失函数。当固定松弛后的编码 \mathbf{B} 时, 二值哈希码 \mathbf{U} 可以通过求解一个二值二次规划问题来获得。

2.3.2 类别标签学习

多数全监督方法仅建模成对相似性, 忽略了类别标签蕴含的判别信息。类别标签学习方法则利用类别信息引导哈希码实现类内紧凑、类间可分。目前, 主要的类别标签学习方式有三种: 交叉熵损失、标签表示学习和哈希中心损失。

(1) 交叉熵损失

一种直接的方法是将分类作为学习目标。通常, 在哈希网络的顶部添加一个分类头, 并使用基于交叉熵的分类损失进行模型训练, 在一个批量 D^M 中, 该损失可以表示为:

$$\mathcal{L}_{\text{ce}} = -\frac{1}{M} \sum_{i=1}^M \ln y_i^T \hat{\mathbf{y}}_i \quad (24)$$

其中, \mathbf{y}_i 是样本 \mathbf{x}_i 的类别标签, $\hat{\mathbf{y}}_i$ 是模型对应的预测输出, 其定义为:

$$\hat{\mathbf{y}}_i = \text{softmax}(\mathbf{W}_c^T \mathbf{b}_i + c) \quad (25)$$

式中, \mathbf{W}_c 是分类器参数, \mathbf{b}_i 是样本 \mathbf{x}_i 的哈希码, c 是偏置项。

全监督语义保持深度哈希^[90] (supervised semantics-preserving deep hashing, SSDH) 将哈希函数实现为深度网络中的隐藏层, 将分类与检索任务统一, 联合优化交叉熵、量化及平衡分布损失, 但因放弃严格二值约束而引入量化误差。

而贪心哈希^[91] (GreedyHash) 采用直通梯度策略: 前向传播保留 sign 函数, 反向传播直接传递梯度。与 GreedyHash 不同, 仅最大化似然哈希^[92] (just-maximizing-likelihood hashing, JMLH) 利用变分信息瓶颈^[124] 设计随机编码器生成二值码, 并联合优化分类损失与 KL 散度正则项, 以保留熵信息。

(2) 标签表示学习

为了从图像标签中捕捉丰富的语义信息, 一些方法专注于学习标签嵌入, 使其与视觉特征对齐。通过这种方式, 所学习到的标签嵌入为视觉特征学习和哈希学习提供了互补的语义信息。

深度联合语义嵌入哈希^[93] (deep joint

semantic-embedding hashing, DSEH) 设计标签网络从二值标签向量中捕捉语义相关性, 并通过成对相似性损失与交叉熵损失联合对齐图像与标签网络的特征空间及汉明空间。

标签关注哈希^[94] (label-attended hashing, LAH) 则利用图卷积网络建模标签共现关系, 构建共现嵌入, 并通过多模态分解双线性模块融合视觉特征与标签嵌入, 最终采用柯西交叉熵与量化损失联合优化。

(3) 哈希中心损失

与学习标签嵌入的方法不同, 一些现有方法提出利用一组预先定义的、相互分离的哈希中心来学习具有类间可分性的哈希码。这类方法通常包括两个步骤: 首先, 在汉明空间中构建均匀分布且相互分离的哈希中心; 然后, 将哈希编码分配到对应的预定义哈希中心, 利用这些中心来引导哈希模型的学习过程。

中心相似性量化^[95] (central similarity quantization, CSQ) 是首个引入哈希中心损失的工作。它首先构造一个 $K \times K$ 的哈达玛矩阵, 具体的构造方式为:

$$\mathbf{H}_K = \begin{bmatrix} \mathbf{H}_{2^{n-1}} & \mathbf{H}_{2^{n-1}} \\ \mathbf{H}_{2^{n-1}} & -\mathbf{H}_{2^{n-1}} \end{bmatrix} = \mathbf{H}_2 \otimes \mathbf{H}_{2^{n-1}} \quad (26)$$

其中, $K = 2^n$, $\mathbf{H}_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$, $\mathbf{H}_1 = [1]$, \otimes 表示哈达玛积 (逐元素乘积)。令 $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_A]$ 表示一组哈希中心, 其中 A 为哈希中心的数量。若 $A \leq K$, 则哈希中心可以直接从上述哈达玛矩阵中采样; 若 $K < A \leq 2K$, 则通过组合两个哈达玛矩阵来构造哈希中心; 若 $A > 2K$ 或不是 2 的幂次, 哈希中心从伯努利分布中进行采样。

在构建哈希中心之后, CSQ 通过最小化哈希编码与其对应哈希中心之间的距离, 引入了一个中心相似性损失:

$$\mathcal{L}_{\text{CSQ}} = \frac{1}{M} \sum_{i=1}^M \frac{1}{L} \sum_{l=1}^L [c_{i,l} \ln b_{i,l} + (1 - c_{i,l}) \ln(1 - b_{i,l})] \quad (27)$$

其中, \mathbf{c}_i 是样本 \mathbf{x}_i 对应的哈希中心, $c_{i,l}$ 是 \mathbf{c}_i 的第 l 个元素。通过使用上述目标函数进行学习, 相似图像的哈希码会靠近其对应的哈希中心, 而远离其他哈希中心。

深度极化网络^[96] (deep polarized network, DPN) 则设计了一种极化损失函数, 能够同时最小化类间汉明距离并最大化类内汉明距离。该极化损失函数定义为:

$$\mathcal{L}_p = \frac{1}{M} \sum_{i=1}^M \max(m - \mathbf{b}_i^T \mathbf{c}_i, 0) \quad (28)$$

其中, $m \geq 1$ 表示一个间隔阈值, \mathbf{c}_i 表示为样本 \mathbf{x}_i 预定义的目标哈希中心。具体而言, 该方法设计了两种策略来获取预定义的目标中心: 随机分配和基于随机初始化的自适应更新。

为了学习高效的哈希编码, 现有方法通常使用多个目标函数, 导致训练策略复杂。为了解决这一问题, 正交哈希^[97] (orthogonality hashing, OrthoHash) 致力于设计一个单一的目标函数, 能够在减少量化误差的同时提升哈希码的判别能力。与前述方法类似, OrthoHash 也生成相互分离的哈希中心, 然后使用分类损失来最大化连续编码与预定义哈希中心之间的余弦相似性:

$$\mathcal{L}_{\text{OrthoHash}} = -\frac{1}{M} \sum_{i=1}^M \ln \frac{\exp(\mathbf{b}_i^T \mathbf{c}_i)}{\sum_{a=1}^M \exp(\mathbf{b}_i^T \mathbf{c}_a)} \quad (29)$$

此外, OrthoHash 在哈希层之后添加了一个批归一化层, 以使得哈希码分布更加均衡。

尽管这些基于哈希中心的方法能够取得令人满意的检索性能, 但它们忽略了一个问题: 在最坏的情况下, 所生成的哈希中心之间的汉明距离可能为零, 从而影响哈希模型的检索性能。因此, Wang 等^[98] 提出了一种基于 Gilbert-Varshamov 边界^[125] 的优化策略以克服这一局限性。具体来说, 该优化策略在生成哈希中心时施加约束, 使得任意两个哈希中心之间的最小距离由 Gilbert-Varshamov 边界确定, 从而保证哈希中心之间的可分性与分布合理性。

2.3.3 联合学习

和以上只考虑样本间的相似性关系或样本类别关系的哈希方法不同, 联合学习策略则是设计一个统一的目标哈希损失函数, 同时考虑样本对之间的相似性以及样本与类别之间的关系, 例如自步关系对比哈希^[99] (self-paced relational contrastive hashing, SPRCH)。具体而言, SPRCH 设计了一个统一的对比损失函数, 通过捕捉来自困难的数据对和数据与类别的关系中的有意义的判别信息来学习哈希码。在一个批量中, SPRCH 的损失函数为:

$$\mathcal{L}_{\text{SPRCH}} = -\frac{1}{M} \sum_{i=1}^M \ln \frac{P_i}{N_i + P_i} \quad (30)$$

其中:

$$P_i = \sum_{\mathbf{b}_j^+ \in \Phi_i^+} \alpha(\mathbf{b}_i, \mathbf{b}_j^+) \exp(S(\mathbf{b}_i, \mathbf{b}_j^+)/\tau) + \lambda_i^+ \sum_{\mathbf{c}_k^+ \in \Psi_i^+} \alpha(\mathbf{b}_i, \mathbf{c}_k^+) \exp(S(\mathbf{b}_i, \mathbf{c}_k^+)/\tau) \quad (31)$$

$$N_i = \sum_{b_j^- \in \Phi_i^-} \alpha(\mathbf{b}_i, \mathbf{b}_j^-) \exp(S(\mathbf{b}_i, \mathbf{b}_j^-)/\tau) + \lambda_i^- \sum_{c_k^- \in \Psi_i^-} \alpha(\mathbf{b}_i, \mathbf{c}_k^-) \exp(S(\mathbf{b}_i, \mathbf{c}_k^-)/\tau) \quad (32)$$

在式(31)和式(32)中: $S(\cdot, \cdot)$ 表示两个给定向量之间的余弦相似度; $\Phi_i^+ = \{\mathbf{b}_j^+ | \mathbf{y}_i^T \mathbf{y}_j > 0\}$ 和 $\Phi_i^- = \{\mathbf{b}_j^- | \mathbf{y}_i^T \mathbf{y}_j = 0\}$ 分别定义了相似样本对集合和不相似样本对集合; $\Psi_i^+ = \{\mathbf{c}_k^+ | y_{ik} = 1\}$ 和 $\Psi_i^- = \{\mathbf{c}_k^- | y_{ik} = 0\}$ 分别定义了相似类别中心集合和无关类别中心集合; λ_i^+ 和 λ_i^- 被设计用于缓解数据对之间以及数据与类别之间配对数量不平衡的问题; $\alpha(\cdot, \cdot)$ 作为自步因子,根据样本对的难易程度以及学习过程中的自步学习策略,自适应地为不同样本对分配权重。

2.3.4 其他

除上述的常规全监督深度图像哈希方法之外,近年来还涌现出一系列面向特定标注场景的全监督深度图像哈希方法。这些方法在传统语义保持的基础上,进一步考虑实际应用中的复杂需求,可大致归纳为三类前沿研究方向:全监督增量图像哈希、全监督长尾图像哈希和全监督细粒度图像哈希。这些方法分别致力于解决模型类别更新中的灾难性遗忘、标注数据类别分布的严重不平衡,以及细微语义差异的精确刻画等关键问题,显著拓展了全监督深度图像哈希的应用边界。

(1) 全监督增量图像哈希

全监督增量图像哈希的目的是持续学习深度哈希模型,以适应流式输入的训练数据,其中每个增量数据批次包含新样本和新类别。深度增量哈希网络^[100](deep incremental hashing network, DIHN)是首个能够为增量训练数据学习二值哈希编码的深度图像哈希方法。特别地,它设计了一个全监督增量图像哈希损失,用于保持新样本与旧样本之间的相似性关系,该损失可表示为:

$$\mathcal{L}_{\text{DIHN}} = \sum_{i=1}^m \sum_{j=1}^q ((\mathbf{u}_i^{\text{old}})^T \mathbf{b}_j^q - L_{s_{ij}})^2 + \sum_{i=m+1}^{m+n} \sum_{j=1}^q ((\mathbf{u}_i^{\text{new}})^T \mathbf{b}_j^q - L_{s_{ij}})^2$$

s. t. $D^{\text{all}} = [D^{\text{old}}; D^{\text{new}}] \in \{-1, +1\}^{L \times N^{\text{tr}}}$ (33)

其中, D^{all} 表示由旧数据 \mathbf{U}^{old} 和新数据 \mathbf{U}^{new} 组成的数据库集合的二值哈希码, \mathbf{b}_j^q 表示当前数据批次中查询样本的松弛哈希码, s_{ij} 表示样本 \mathbf{x}_i 与 \mathbf{x}_j 之间的相似性。然而,DIHN在更新深度哈希模型时需要保存所有训练数据的哈希码,这带来了较高的训练复杂度和存储开销。为了解决这一问题,Chen等^[101]提出不保存所有旧数据的哈希码,

而是通过新学习的深度哈希模型估计旧数据的特征表示,从而缓解模型遗忘问题。与上述依赖显式数据相似关系的增量方法不同,深度增量哈希^[102](deep incremental hashing, DIH)并不在线构建样本间相似性,而是利用类别标签与离线随机投影生成语义目标码,通过逐点回归实现高效增量学习。

(2) 全监督长尾图像哈希

大多数现有的全监督图像哈希方法都是为均衡数据集设计的,即训练集和测试集是独立同分布(independent and identically distributed, IID)采样的。然而,在实际应用中,许多数据集往往呈现不平衡的长尾分布,其中少数头部类别占据数据集的大部分,而其余尾部类别仅包含少量样本。这种不平衡的数据分布对尾部类别样本的学习带来了更大挑战。因此,在均衡分布下训练的图像哈希方法在处理长尾图像检索任务时表现不佳。

近年来,一些面向长尾分布图像检索的深度哈希方法被提出。例如,长尾哈希网络^[103](long-tail hashing network, LTHNet)引入了一个动态元嵌入模块和一个重加权分类损失,用于在长尾类别分布下学习深度哈希模型。该动态元嵌入模块来自数据丰富的头部类别的知识迁移到数据稀缺的尾部类别表示中,从而提升其判别能力。而重加权分类损失利用类别频率调整分类器的预测结果,这种基于类别级别的重加权方式可以通过引入先验分布缓解长尾数据带来的偏差。

进一步地,注意力引导对比哈希网络^[104](attention-guided contrastive hashing network, ACHNet)设计了一个跨注意力特征增强模块,以缓解哈希过程中带来的信息损失;同时提出了一种类别中心化对比损失,用于减轻数据丰富的头部类别对整体学习过程的偏倚影响,从而提升尾部类别的检索性能。

相比之下,双动态代理哈希网络^[105](dual dynamic proxy hashing network, DDPHN)在ACHNet的基础上进一步拓展了中心表示的层次,不仅使用了可学习的低维哈希类别中心,还引入了可学习的高维特征类别中心,联合建模语义层次的距离关系,从而更全面地指导哈希码的生成,显著提升尾部类别的表达能力。

而语义增强代理引导哈希^[106](semantic-enhanced proxy-guided hashing, SPH)不仅联合建模哈希类别中心与特征类别中心,还引入马氏距离度量来更准确地刻画中心与样本之间的关系,

增强类别中心与样本的语义表示,有效提升了模型对尾部类别语义边界的建模能力,从而进一步提升模型检索性能。

因果推理哈希^[107] (causal inference hashing, CIH) 则将因果推理引入长尾图像哈希,利用因果图分析长尾分布带来的偏差,并构造哈希中介,从哈希类别中心中挖掘对尾部类别有益的偏差,并通过因果干预和后门调整阻断有害偏差对尾部类别的影响,从而提高哈希模型对尾部类别的编码能力。

(3) 全监督细粒度图像哈希

全监督细粒度图像哈希的目的是从具有细粒度标签的子类别数据集中学习紧凑的哈希码。与针对粗粒度数据的常规全监督深度图像哈希相比,该任务更具挑战性。为了增强对不同细粒度类别之间细微差异的区分能力,现有的细粒度深度哈希方法更加注重提取具有判别性的视觉特征表示。现有细粒度方法可以分为以下五个方向:区域定位、多尺度特征、特征交互、掩码以及其他方法。

区域定位:为了捕捉细粒度物体中更具判别性的局部区域,第一类方法是设计区域定位模块,以挖掘具有信息量的局部区域来生成高效的二值哈希码。例如,文献[108]提出了一种显著区域定位与哈希学习相互增强的方法。具体而言,该方法设计了一个区域定位模块用于生成候选区域,以及一个深度哈希模块用于将这些有信息量的区域编码为二值哈希码,并同时为区域定位模块生成标注信息。

与文献[108]使用候选区域的方法不同,深度显著性哈希^[109] (deep saliency hashing, DSaH) 通过定位显著区域来学习哈希编码。具体来说,该方法设计了一个基于全卷积网络的注意力网络来自动生成显著图像,并引入了一种新的显著性损失函数,引导网络寻找更具吸引力的区域。

与定位显著区域不同,子区域定位哈希^[110] (sub-region localized hashing, sRLH) 设计了一个子区域定位模块,用于发现与全局特征相似的 top- k 个不同的局部特征。除了该子区域定位模块,sRLH 还引入了中心损失和 Gram-Schmidt 正交化操作,以学习在类内紧凑、类间可分的哈希码。

多尺度特征:更大的尺度特征具有更大的感受野和更强的全局语义表达能力,而在较小尺度下,特征能够提供更丰富的局部细节。通过某种融合策略将不同尺度的特征进行结合,可以获得

包含丰富上下文信息的多尺度特征,从而帮助我们更准确地捕捉细粒度物体。

特征金字塔哈希^[111] (feature pyramid hashing, FPH) 首次引入垂直金字塔和水平金字塔结构,分别提取包含高层语义信息的高层特征和包含细微局部细节的低层特征。随后,将这两类特征进行组合以学习哈希码。

与 FPH 类似,相关性滤波哈希^[112] (correlation filtering hashing, CFH) 同样利用特征金字塔网络提取多尺度特征。不同的是,CFH 设计了一个相关性滤波模块来捕捉潜在的局部信息,其中使用标签嵌入对多尺度特征图执行相关性操作,以获取物体的判别性位置。随后,采用全局平均池化和特征拼接的融合策略对多级特征表示进行融合。

特征交互:第三个方向的划分依据是方法是否在局部区域内部进行特征交互。第三个方向的划分依据是判断方法是否在局部区域内进行特征交互。交换哈希^[113] (exchange hashing, ExchNet) 通过在正样本对中交换相同部分级别的特征来探索语义一致性。通过这种方式,局部特征被对齐,以从细粒度图像中捕捉共同的语义局部模式。跨尺度上下文提取哈希网络^[114] (cross-scale context extracted hashing network, CSCE-Net) 设计了一个注意力引导的信息提取模块,通过在全局特征与局部区域特征之间进行跨层次交互,以捕捉细粒度的上下文信息。

最近,基于 Swin Transformer 的全监督细粒度图像哈希网络^[115] (Swin Transformer-based fine-grained hashing network, SwinFGHash) 提出了一种基于 Transformer 的模块,用于建模局部特征之间的长距离依赖关系。具体而言,通过 Swin Transformer block 中的自注意力机制,对不同视觉块之间的交互进行建模,从而增强判别性特征的代表能力。此外,图特征聚合哈希^[116] (graph feature aggregation hashing, GDF-Net) 利用 K 近邻图聚合局部区域特征之间的上下文关系,以提升哈希码的判别性和泛化能力。

掩码:第四个方向专注于解决细粒度方法中长期存在的过拟合问题。由于大多数基于注意力机制的方法只能关注少数信息丰富的区域而忽略次要显著区域,许多思路开始利用注意力掩码或注意力丢弃策略来强制网络学习互补特征,以缓解过拟合问题。特征一致性驱动的注意力擦除网络^[117] (feature consistency driven attention erasing network, FCAENet) 利用一个选择性区域擦除模

块,在注意力掩码上擦除前 n 个值,使特征提取器对次判别性特征变得更加敏感和鲁棒。

基于抑制增强掩码的注意力与交互通道变换网络^[118] (suppression-enhancing mask based attention and interactive channel transformation, SEMICON)设计了一个抑制增强掩码模块,以获得与区域相关的注意力图。由于通道与细粒度局部部分高度相关,该方法开发了一个交互通道变换模块,通过自注意力机制捕捉不同通道间的相关性,以改进全监督细粒度图像哈希学习。

细粒度实例感知语义哈希^[119] (fine-grained instance-aware semantic hashing, FISH)设计了一个空间过滤模块,该模块结合了掩码机制,基于随机概率隐藏最显著的区域以突出更重要的区域。随后引入了特征过滤模块,用于进一步精炼提取出特征。

其他方法:除了以上四个方向,还有一些从不同角度出发的全监督细粒度图像哈希方法。

深度渐进非对称量化^[120] (deep progressive asymmetric quantization, DPAQ)利用因果推断来减少细粒度数据集偏差所带来的混杂效应。该方法设计了一个拼图生成器以生成多粒度的图像块,并将其用于构建因果干预中的混杂因子集合。同时,引入了一种渐进式的非对称量化损失,用于保留通过因果推断发现的多层次相似性。

此外,属性注意力网络^[121] (attribute-attention network, A²-NET)和属性引导多级哈希^[122] (attribute-guided multi-level hashing, AGMH)试图学习属性特定的视觉特征,以针对细粒度图像生成具有属性感知能力的哈希码。其中,A²-NET的核心是一个编码器-解码器模块,通过自一致性重建损失构建一个属性感知的哈希空间;而AGMH中基于自注意力机制的分步交互式外部注意力在发现类别相关且分布离散的对象属性方面起到了关键作用。

而因果特征到二值注入哈希学习^[123] (causal feature to binary-injected hash learning, CFBH)针对现有注意力机制易过度聚焦显著区域而导致模型过拟合的问题,提出了一种兼顾局部多样性与哈希码平衡性的新框架。具体来说,该方法在因果特征学习阶段引入因果推断,通过融合局部特征与原始图像信息,削弱网络对显著区域的观测偏差,从而引导模型关注更细微的判别性局部变化;在二值注入哈希学习阶段,设计了一个受dropout启发的二值噪声注入模块,有效抑制与显著特征强相关的哈希位过激活现象,使生成的哈

希码在汉明空间中更加去相关且分布均衡。

2.3.5 小结

综上所述,现有全监督深度图像哈希方法虽在语义建模精度上显著优于无监督与弱监督范式,但仍面临若干共性挑战。首先,绝大多数方法严重依赖高质量、均衡且静态的类别标签,在面对类别分布偏移(如长尾场景)或动态新增类别(增量学习)时,模型性能急剧下降,泛化能力受限。其次,不同技术路径存在明显权衡:相似度学习虽能精细刻画样本对关系,但计算复杂度高且对负样本采样敏感;类别标签学习通过中心引导提升判别性,却可能因哈希中心设计不当(如距离过近)而削弱类间可分性;联合学习虽兼顾成对与类别信息,但多关系优化易导致训练不稳定;而面向增量、长尾、细粒度等特定场景的方法,虽拓展了应用边界,却往往引入额外模块(如元嵌入、区域定位),显著增加模型复杂度与工程成本。此外,当前方法普遍缺乏对哈希码语义可解释性与跨域迁移能力的系统考量。未来研究需在保持高精度的同时,探索更鲁棒、轻量且自适应的全监督哈希框架,例如结合因果推理缓解数据偏差、设计统一的动态学习机制,或引入可解释性约束以增强哈希位与语义概念的对齐。

3 深度图像哈希的性能评估

为了分析比较当前深度图像哈希方法的检索性能,本文主要选择了粗粒度数据集 NUS-WIDE^[18]和 MS-COCO^[19]、细粒度数据集 CUB-200-2011^[20]和 FGVC-Aircraft^[22]参与实验。在实验过程中,本文将每个数据集分为训练集、查询集和数据库集三个部分,其中训练集用于学习哈希模型,查询集和数据库集用于评估哈希模型的检索性能。

对于 NUS-WIDE,从每个类别中选择 100 张图像,总共 2 100 张图像作为查询集,其余 193 734 张图像用作数据库集;从每个类别选择 500 张图像,总共 10 500 张图像组成训练集。

对于 MS-COCO,本文从官方数据集中随机选择 5 000 张图像作为查询集,其余 117 218 张图像作为数据库集,并从数据库集中随机选择 10 000 张图像用于训练。

对于 CUB-200-2011,一共有 11 788 张细粒度鸟类图片,可划分为 200 类。按照官方划分,本文选择 5 994 张测试图片作为查询集,5 794 张训练图片作为训练集和数据库集。

对于 FGVC-Aircraft,总共有 10 000 张细粒度

飞机图片,可以被分为 100 类。官方划分的 3 333 张测试图片用于查询,官方划分的 6 667 张测试图片用于训练和数据库集构建。

为了保证公平比较,所有图像哈希方法都采用预训练的 ResNet18^[28]作为主干网络。选择 ResNet18 主要是因为其作为图像哈希领域的标准主干网络,在保持残差结构优势的同时参数量适中、计算开销较低,已被广泛用于深度哈希研究的默认或基准架构。此外,对于无监督深度图像哈希和全监督细粒度深度图像哈希,选

择 mAP@all 作为评价指标;而对于弱监督和全监督深度图像哈希,则选择 mAP@5K 作为评价指标。

3.1 无监督深度图像哈希

在无监督深度图像哈希中,本文选择了 HashGAN^[35]、SSDH^[38]、CIBHash^[53]和 RCSH^[62]这 4 种代表性的哈希方法进行比较,它们的 mAP 性能指标如表 5 所示,在实验中,将 RCSH^[62]归类为自监督方法。从表 5 的结果可得到以下结论。

表 5 代表性的无监督深度哈希方法在 NUS-WIDE 和 MS-COCO 数据集上的 mAP@all 性能比较

Tab.5 Comparison of mAP@all performance of representative unsupervised deep hashing methods on NUS-WIDE and MS-COCO datasets

方法		NUS-WIDE				MS-COCO			
		16 bit	32 bit	48 bit	64 bit	16 bit	32 bit	48 bit	64 bit
重构	HashGAN ^[35]	40.8	43.1	43.9	44.8	37.1	37.3	37.5	37.8
局部相似性保持	SSDH ^[38]	57.1	58.2	58.9	59.4	47.0	47.9	48.6	51.2
自监督	CIBHash ^[53]	53.9	54.5	55.8	56.5	52.3	53.9	54.3	55.0
	RCSH ^[62]	59.1	60.3	61.2	60.9	52.7	54.3	55.2	56.2

首先,从不同类别的无监督深度哈希方法对比可以看出,基于自监督学习范式的方法在大多数情况下显著优于传统的重构型方法和局部相似性保持方法,这表明通过构建代理任务以挖掘数据内在结构的自监督机制,能够更有效地引导网络学习具有判别性的语义特征,从而生成语义一致性更强的哈希码。

其次,重构方法(如 HashGAN^[35])受限于像素级或特征级重建目标对纹理细节的过度关注,难以捕捉高层语义信息,导致所学哈希码的判别能力较弱。而局部相似性保持方法(如 SSDH^[38])虽引入了样本邻域结构约束,但其性能仍受限于手工设计的相似性度量或对局部结构建模的不充分性。相比之下,自监督方法(如 CIBHash^[53]和 RCSH^[62])在两个基准数据集上均展现出优越的性能,这说明通过构建自监督信号,能够有效模拟类别语义结构,在无真实标签的情况下甚至能够接近部分监督方法的表达能力。

综上所述,无监督深度图像哈希的检索性能在很大程度上取决于模型对数据内在语义结构的挖掘能力。只有设计合理的自监督任务以充分揭示样本间的全局语义关联与局部相似性约束,并实现两者的协同优化,才能有效提升哈希码的判别性与鲁棒性。未来的研究应进一步

探索更加精细的语义代理任务与更高效的对比学习架构。

3.2 弱监督深度图像哈希

3.2.1 不同哈希方法的性能比较

在弱监督深度图像哈希中,本文选择了 WDHT^[64]、MGRN^[69]、EWSH^[70]、WSHRCA^[72]、MKDH^[73]这 5 种具有代表性的方法进行比较,它们的 mAP 性能指标如表 6 所示。

从表 6 可知,相比较早期的 WDHT^[64]方法,MGRN^[69]、EWSH^[70]和 WSHRCA^[72]基于标签优化的方法在两个弱监督数据集上均展现出更优异的检索性能。这表明,弱监督数据集中所提供的用户标签存在不完整性和较高噪声的问题。而标签优化方法能够有效结合哈希模型学习到的图像表征,对原始标签进行修正与补全,从而挖掘出更契合图像语义的文本信息,有助于提升整体检索效果。

相较于仅依赖预训练词嵌入模型(如 word2vec)进行标签重构或优化的方法,基于多模态知识蒸馏的 MKDH^[73]方法在 NUS-WIDE 和 MS-COCO 上取得了更好的检索性能。这证明了从预训练视觉-语言模型 CLIP 中同时挖掘视觉和文本这两种模态知识的必要性,相比较依赖单一文本模态的知识,从两种模态知识中学

习的联合表示,可以提高弱监督哈希模型的检索性能。

表6 代表性的弱监督深度哈希方法在 NUS-WIDE 和 MS-COCO 数据集上的 mAP@5K 性能比较

Tab.6 Comparison of mAP@5K performance of representative weakly supervised deep hashing methods on NUS-WIDE and MS-COCO datasets

方法		NUS-WIDE				MS-COCO			
		16 bit	32 bit	48 bit	64 bit	16 bit	32 bit	48 bit	64 bit
标签重构	WDHT ^[64]	60.6	69.4	72.0	71.2	51.8	58.9	58.8	61.7
	MGRN ^[69]	69.5	73.7	75.0	76.3	57.0	59.8	63.2	65.1
标签优化	EWSH ^[70]	72.7	74.7	75.6	76.7	60.0	63.5	64.2	65.0
	WSHRCA ^[72]	70.0	75.4	76.1	77.4	60.0	63.9	66.3	67.5
多模态知识蒸馏	MKDH ^[73]	79.5	80.7	82.0	82.1	67.5	69.1	70.8	70.7

综上,弱监督哈希方法的性能提升主要得益于两个方面:一是通过标签优化策略缓解了用户标签的噪声和不完整问题,提升了标签的语义表达质量;二是多模态知识的引入突破了以往对单一文本模态的依赖,使模型能够从图像和文本的协同关系中学习更加丰富和精准的表达。

3.2.2 不同主干网络的性能比较

除方法策略的改进之外,底层特征表示能力的增强往往也是提升检索性能的核心驱动力。为了进一步分析方法增益与主干网络性能之间的关系,在弱监督深度图像哈希中,本文除了选用 ResNet18,还引入了 ViT^[29] 作为主干网络,在 MS-COCO 数据集上进行了对比实验,结果可见表 7。

实验结果表明,主干网络的特征提取能力与哈希检索精度呈正相关关系。在相同的哈希码长下,采用 ViT 作为主干网络的模型在各项指标上均普遍优于 ResNet18,这主要是因为 Transformer 架构强大的建模能力,能够捕捉更丰富的语义信息,从而生成判别性更强的二进制哈希码。然而,这种性能提升在不同算法框架下的表现存在差异。例如,WDHT 在采用 ViT 后性能提升尤为显著,甚至超过 EWSH,这可能是因为 WDHT 基于用户标签的简单而直接的监督信号能更有效地利用 ViT 输出高维全局特征,而 EWSH 所依赖的复杂标签优化机制在缺乏针对性适配时难以充分发挥 ViT 的表达能力。

表7 代表性的弱监督深度哈希方法在 MS-COCO 数据集上利用不同主干网络的 mAP@5K 性能比较

Tab.7 Comparison of mAP@5K performance of representative weakly supervised deep hashing methods using different backbones on MS-COCO datasets

方法		ResNet18				ViT			
		16 bit	32 bit	48 bit	64 bit	16 bit	32 bit	48 bit	64 bit
标签重构	WDHT ^[64]	51.8	58.9	58.8	61.7	55.6	66.1	71.2	74.9
	EWSH ^[70]	60.0	63.5	64.2	65.0	63.2	64.9	68.1	68.5
标签优化	WSHRCA ^[72]	60.0	63.9	66.3	67.5	61.7	70.8	73.8	74.1
	MKDH ^[73]	67.5	69.1	70.8	70.7	75.6	78.3	78.5	78.5

值得注意的是,尽管 ViT 带来了明显的增益,但 ResNet18 在计算开销与检索精度之间展现出了更好的平衡性,且其作为基准主干网络,能够更公平地反映哈希方法本身对性能的贡献,避免了因过度依赖超大规模预训练模型而掩盖算法核心设计的改进。因此,本文选择 ResNet18 作为主要主干网络,既保证了与主流研究的可比性,也确保

了实验设计的严谨性。

3.3 全监督深度图像哈希

3.3.1 不同哈希方法的性能比较

在全监督深度图像哈希中,本文选择了 DSH^[80]、HashNet^[77]、DCH^[78]、GreedyHash^[91]、CSQ^[95]、DPN^[96]、OrthoHash^[97]、SPRCH^[99] 这 8 种具有代表性的方法,它们的 mAP 性能指标如表 8

所示,相关实验结果来自文献[99]。

从结果可知:首先,类别标签学习方法的检索性能在大部分情况下都要优于相似度学习方法,这表明通过构建图像与类别中心之间的语义关系,模型能够更好地捕捉图像的判别性特征,从而提升哈希码的语义一致性。相较之下,相似度学习方法往往依赖于二元组或三元组样本对的相对关系,受限于监督信息的稀疏性,难以充分表达类别间的全局结构。其次,联合学习方法的检索性能优于相似度学习方法和类别标签学习方法,这表明数据间的相似度关系和数据与类别中心间的语义关系具有互补性。前者更关注局部样本之间的相对关系,而后者则强调全局类别结构的建模。通过将两者融合,联合学习方法能够同时兼顾局部相似性保持与全局语义表达,从而生成更具区分性和鲁棒性的哈希码。

综上所述,全监督深度图像哈希的性能高度依赖于对监督信号中多层次语义关系的挖掘与融合。值得注意的是,在实际应用中,细粒度图像检索作为全监督哈希的重要延伸场景,对模型区分

细微语义差异的能力提出了更高要求。

3.3.2 不同哈希方法的哈希码可视化分析

为了让检索结果更加直观,本文在全监督深度图像哈希检索实验中,补充了基于单标签数据集 CIFAR-100 的哈希码可视化分析,以展示不同哈希方法在汉明空间中的分布特性与聚类效果。对于单标签数据集 CIFAR-100,本文选择其 5 000 张官方训练图像作为数据库集,1 000 张官方测试图像作为训练集和查询集。

从表 8 可知,HashNet、CSQ 和 SPRCH 分别代表仅考虑样本间的相似性关系的相似度学习方法、仅考虑样本类别关系的类别标签学习方法,以及同时融合两种语义关系的联合学习方法。为进行可视化分析,本文在 CIFAR-100 上对这 3 种代表性哈希方法进行训练,并选取 CIFAR-100 数据库集中第 43 类、第 77 类和第 82 类样本,分别提取三种方法生成的 16 位哈希码,并采用 t -分布随机邻域嵌入 (t -distributed stochastic neighbor embedding, t -SNE) 方法将高维哈希码降维至二维空间进行可视化。

表 8 代表性的全监督深度哈希方法在 NUS-WIDE 和 MS-COCO 数据集上的 mAP@5K 性能比较^[99]

Tab. 8 Comparison of mAP@5K performance of representative fully supervised deep hashing methods on NUS-WIDE and MS-COCO datasets^[99]

方法		NUS-WIDE				MS-COCO				%
		16 bit	32 bit	48 bit	64 bit	16 bit	32 bit	48 bit	64 bit	
相似度学习	DSH ^[80]	74.6	79.0	79.1	79.4	63.6	66.4	69.5	70.3	
	HashNet ^[77]	82.0	84.5	85.4	85.6	67.1	68.9	70.4	71.6	
	DCH ^[78]	80.1	81.1	81.0	81.2	71.9	74.6	75.2	75.7	
类别标签学习	GreedyHash ^[91]	79.2	80.7	81.7	82.0	68.2	74.1	75.9	77.0	
	CSQ ^[95]	77.0	79.1	79.4	80.1	70.6	79.7	81.1	84.0	
	DPN ^[96]	79.2	80.4	79.9	80.9	71.7	78.9	81.5	82.5	
	Orthohash ^[97]	81.7	83.9	85.3	85.9	74.6	81.6	83.8	85.4	
联合学习	SPRCH ^[99]	83.7	84.9	85.6	85.9	83.0	85.6	86.4	86.8	

如图 3 所示,3 个代表性的哈希方法在哈希空间中的分布呈现出显著差异:HashNet 虽能对 3 类样本实现初步区分,但类间边界模糊,存在明显的交叉与混杂现象,反映出其仅依赖数据-数据成对相似性约束,在全局结构保持能力上存在局限;CSQ 在引入类别语义监督后,类内点云明显更加紧凑,然而类间仍存在显著重叠,尤其是第 43 类与第 77 类在中心区域相互渗

透,表明仅利用类别标签信息难以充分刻画复杂的类间关系;相比之下,SPRCH 通过联合建模样本间语义关系与样本类别语义关系,使 3 类样本在哈希空间中实现了高度分离——各簇内部紧密聚集,类间距离显著增大,其中第 43 类与第 77 类形成清晰分界,第 82 类则独立分布于另一区域,充分展现了其在特征判别性与结构一致性方面的综合优势。

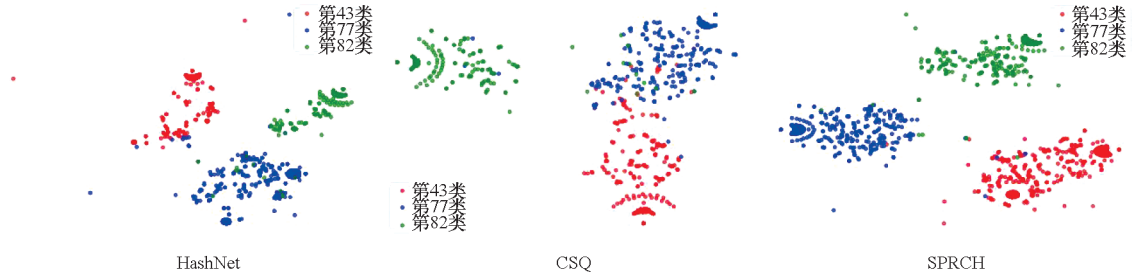


图3 不同全监督哈希方法的哈希码的 *t*-SNE 可视化结果

Fig. 3 The *t*-SNE visualization results of hash codes from different fully supervised hashing methods

综上所述,可视化结果直观反映了不同方法的哈希码在哈希空间中的聚类与分离能力,表明能有效挖掘多维度语义信息的方法,其生成的哈希码在哈希空间中具有更强的类内紧凑性和类间可分性。

3.3.3 不同细粒度哈希方法的性能比较

以上的实验分析都聚焦于常规的全监督深度

图像哈希方法,为深入评估现有细粒度哈希方法在细粒度图像检索任务中的表现,本文选择了 DSaH^[109]、sRLH^[110]、SEMICON^[118]、FISH^[119] 和 CFBH^[123] 这5种具有代表性的细粒度哈希方法,它们的 mAP 性能指标如表9所示,相关实验结果来自文献[123]。

表9 代表性的全监督细粒度深度哈希方法在 CUB-200-2011 和 FGVC-Aircraft 数据集上的 mAP@all 性能比较^[123]

Tab. 9 Comparison of mAP@all performance of representative full supervised fine-grained deep hashing methods on CUB-200-2011 and FGVC-Aircraft datasets^[123]

		CUB-200-211				FGVC-Aircraft				%
方法		16 bit	32 bit	48 bit	64 bit	16 bit	32 bit	48 bit	64 bit	
区域定位	DSaH ^[109]	50.5	54.7	66.0	67.02	55.1	68.2	71.7	70.5	
	sRLH ^[110]	62.8	69.9	71.0	71.9	80.2	83.1	83.9	83.9	
掩码	SEMICON ^[118]	39.3	51.5	62.4	69.9	51.4	60.8	74.1	76.3	
	FISH ^[119]	69.2	73.1	74.1	74.1	82.9	83.8	83.6	83.9	
因果推断	CFBH ^[123]	70.1	74.7	75.4	76.2	84.3	85.7	86.7	87.2	

从结果可见,sRLH、FISH 与 CFBH 在多数码长设置下显著优于其他方法。这主要得益于它们在特征学习或微调阶段引入了分类损失,有效增强了高维特征嵌入的判别性,从而为后续哈希编码提供了更优的语义基础,验证了联合优化特征表示与哈希函数对提升检索性能的重要性。值得注意的是,除 FISH 与 CFBH 外,多数方法在短码长(如 16 bit)下的性能明显弱于长码长,反映出其对高维连续特征到低维二值码的映射能力有限。FISH 通过双滤波掩码机制抑制与标签预测无关的信息,简化了特征到哈希码的转换过程,但其性能高度依赖标签预测的准确性;而 CFBH 则通过因果特征学习与二值噪声注入机制,在不依赖强标签先验的前提下,有效提升了哈希码的泛化能力与局部多样性建模能力。最终,CFBH 在两个数据集上均取得最优性能,尤其在 FGVC-Aircraft 上 64 bit 码长达到 87.2% 的 mAP,显著领

先于现有方法,表明其在捕捉细粒度局部差异与生成高效哈希码方面具有突出优势。

3.3.4 不同细粒度哈希方法的时间与存储开销分析

为评估全监督细粒度深度哈希方法的计算复杂度和存储需求,本文进一步比较了代表性方法在 CUB-200-2011 数据集上的时间与存储开销(见表10),相关实验结果来自文献[123]。所有实验均在相同硬件环境(NVIDIA RTX A6000)下进行,批量大小统一设为64。

从结果可见,不同方法在效率上存在显著差异。例如,sRLH 虽在检索性能上表现优异,但其训练阶段单批次耗时高达 428.33 ms,远高于其他方法,主要源于其复杂的子区域定位与中心损失联合优化机制。相比之下,基于掩码策略的 SEMICON 虽推理速度较快,但其模型存储占用达 19.82 MB,明显高于其余方法,反映出其多模块结构带来的额外参数负担。

表 10 代表性的全监督细粒度深度哈希方法在 CUB-200-2011 数据集上的时间和存储开销比较^[123]

Tab. 10 Comparison of the time and storage costs of representative full supervised fine-grained deep hashing methods on CUB-200-2011 dataset^[123]

方法		训练阶段		推理阶段	
		时间/ms	存储/MB	时间/ms	存储/MB
区域定位	DSaH ^[109]	173.60	11.45	24.40	11.45
	sRLH ^[110]	428.33	11.30	19.77	11.18
掩码	SEMICON ^[118]	89.55	19.82	53.93	19.82
	FISH ^[119]	80.44	11.28	21.21	11.18
因果推断	CFBH ^[123]	62.86	11.40	19.57	11.18

值得注意的是,CFBH 在训练与推理阶段均展现出更高的效率:训练耗时仅 62.86 ms,推理耗时低至 19.57 ms,存储开销稳定在 11.18 MB。这表明 CFBH 并未引入额外的复杂模块,而是通过因果特征学习与二值噪声注入机制,在不增加模型规模的前提下有效挖掘局部判别性信息。因此,CFBH 不仅在检索精度上领先,同时在计算效率与存储需求方面也具备良好的实用性。

4 发展与展望

尽管深度哈希在图像检索任务中取得了显著进展,但是当前方法普遍存在难以适应数据动态变化、对多模态数据建模能力不足、特定领域应用能力有限等问题。针对这些挑战,未来有关深度哈希的研究可从以下几个方面展开:

1) 基于增量学习的可扩展深度哈希:如何使模型在不重新训练全量数据的前提下,吸收新类别信息并克服“灾难性遗忘”,是实现大规模扩展的关键挑战。未来的研究计划应重点关注保持拓扑结构的知识蒸馏技术,通过引入教师网络约束,确保新模型在更新参数时能够维持旧数据在汉明空间中的相对距离分布。同时,研究可适应的动态码增长策略,利用生成式回放技术合成旧样本特征,在有限存储成本下实现模型性能的持续演进,提升其在流式数据环境下的鲁棒性。

2) 多模态与跨模态哈希:随着多媒体技术的飞速发展,图像、文本、视频等模态呈现高度的语义互补性,如何打破异质模态间的“语义鸿沟”已成为研究高地。通过构建端到端的深度框架,将多源数据映射至统一的汉明空间,是实现高效跨模态检索的关键。近年来,利用对比语言-图像预训练模型(如 CLIP)所蕴含的跨模态先验知识,已成为提升哈希模型性能的主流趋势^[126-129]。未来研究应进一步探索如何利用超大规模预训练

模型的泛化能力,设计更具可扩展性、能适应动态跨域分布的多模态哈希方法。

3) 面向特定领域与实际场景的深度哈希:通用数据集与医学影像、遥感、工业探伤等特定领域之间存在巨大的分布差异,通用模型往往难以处理细粒度特征区分与标签稀缺问题。未来的研究应强调领域知识迁移与提示学习的结合,通过在预训练哈希模型中嵌入特定领域的规则或约束,提升模型对小样本数据的泛化能力。针对实际部署需求,需定制化设计符合特定行业数据分布的损失函数与采样策略,以构建更具实用价值的专业化检索系统。

4) 融合先进深度网络结构的深度哈希:现有方法对新型架构如 ViT 的利用仍显不足,如何在保证特征表达能力的同时降低计算复杂度是主要挑战。未来的研究应探索结构重参数化与局部-全局混合注意力机制在哈希模型中的集成。具体计划包括研发轻量化的 Transformer 哈希变体,通过减小参数量并优化自注意力计算开销,使深度哈希模型在保持强大全局建模能力的同时,能够高效处理高分辨率图像并适应更为复杂的真实分布。

5) 多任务深度哈希:现有的检索模型常独立于分类、检测等任务,导致语义信息利用率不高且泛化性受限。未来的研究应超越“哈希为中心”的范式,转向构建统一的多任务语义表示框架。在该框架中,哈希编码并非最终目的,而是与分类、定位、属性预测等任务并列的下游应用之一。通过共享主干网络与结构化的联合优化目标,模型可同时学习对多种任务均有判别力的通用特征表示。这种统一架构不仅能提升哈希性能,更能实现“一次训练、多任务输出”的高效部署,使学到的表示真正具备跨任务泛化能力,为实际智能系统提供灵活、紧凑且语义丰富的基础嵌入。

6)轻量高效的哈希优化方法:当前面向大规模数据检索的深度哈希模型普遍存在参数冗余度高、推理延迟大等问题,严重制约其在资源受限场景下的实际部署。为此,亟须发展轻量高效的哈希优化方法。一方面,可以探索贡献度感知的结构化剪枝技术,通过定量分析神经元对汉明空间拓扑布局的影响力,在大幅剔除冗余通道的同时规避映射空间的语义坍塌;另一方面,应该研发哈希感知知识蒸馏框架,设计跨架构的流形对齐损失,将大规模教师模型中复杂的非线性语义信息精准迁移至轻量化学生网络中。通过上述策略的协同优化,可以在维持检索精度的前提下,大幅降低哈希生成的计算开销。

参考文献 (References)

- [1] KULIS B, DARRELL T. Learning to hash with binary reconstructive embeddings [C]//Proceedings of the 23rd International Conference on Neural Information Processing Systems, 2009: 1042–1050.
- [2] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. *Nature*, 2015, 521(7553): 436–444.
- [3] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep learning [M]. Cambridge: MIT Press, 2016.
- [4] DING Y, ZHANG Z L, ZHAO X F, et al. Deep hybrid: multi-graph neural network collaboration for hyperspectral image classification [J]. *Defence Technology*, 2023, 23: 164–176.
- [5] YANG M, LING J, CHEN J M, et al. Discriminative semi-supervised learning via deep and dictionary representation for image classification [J]. *Pattern Recognition*, 2023, 140: 109521.
- [6] CHHAPARIYA K, BUDDHIRAJU K M, KUMAR A. A deep spectral-spatial residual attention network for hyperspectral image classification [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024, 17: 15393–15406.
- [7] DIWAN T, ANIRUDH G, TEMBHURNE J V. Object detection using YOLO: challenges, architectural successors, datasets and applications [J]. *Multimedia Tools and Applications*, 2023, 82(6): 9243–9275.
- [8] VARGHESE R, SAMBATH M. YOLOv8: a novel object detection algorithm with enhanced performance and robustness [C]//Proceedings of 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), 2024: 1–6.
- [9] WANG A, CHEN H, LIU L H, et al. YOLOv10: real-time end-to-end object detection [C]//Advances in Neural Information Processing Systems 37 (NeurIPS 2024), 2024: 107984–108011.
- [10] XU M D, ZHANG Z, WEI F Y, et al. Side adapter network for open-vocabulary semantic segmentation [C]//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023: 2945–2954.
- [11] ZHOU T F, WANG W G. Cross-image pixel contrasting for semantic segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, 46(8): 5398–5412.
- [12] ZHOU T F, WANG W G. Prototype-based semantic segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, 46(10): 6858–6872.
- [13] SUNG C, KIM W, AN J, et al. Contextrast: contextual contrastive learning for semantic segmentation [C]//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024: 3732–3742.
- [14] LUO X, WANG H X, WU D Q, et al. A survey on deep hashing methods [J]. *ACM Transactions on Knowledge Discovery from Data*, 2023, 17(1): 1–50.
- [15] KRIZHEVSKY A. Learning multiple layers of features from tiny images: TR-2009 [R]. Toronto: University of Toronto, 2009.
- [16] DENG J, DONG W, SOCHER R, et al. ImageNet: a large-scale hierarchical image database [C]//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009: 248–255.
- [17] HUISKES M J, LEW M S. The MIR flickr retrieval evaluation [C]//Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval, 2008: 39–43.
- [18] CHUA T S, TANG J H, HONG R C, 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, 2009: 1–9.
- [19] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context [C]//Computer Vision-ECCV 2014, 2014: 740–755.
- [20] WAH C, BRANSON S, WELINDER P, et al. Caltech-UCSD birds-200-2011 (CUB-200-2011): CNS-TR-2011-001 [R]. Pasadena: California Institute of Technology, 2011.
- [21] KRAUSE J, STARK M, DENG J, et al. 3D object representations for fine-grained categorization [C]//Proceedings of 2013 IEEE International Conference on Computer Vision Workshops, 2013: 554–561.
- [22] MAJI S, RAHTU E, KANNALA J, et al. Fine-grained visual classification of aircraft [EB/OL]. (2013–06–21) [2025–12–25]. <https://arxiv.org/abs/1306.5151>.
- [23] HOU S H, FENG Y S, WANG Z L. VegFru: a domain-specific dataset for fine-grained visual categorization [C]//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV), 2017: 541–549.
- [24] BOSSARD L, GUILLAUMIN M, VAN GOOL L. Food-101: mining discriminative components with random forests [C]//Computer Vision-ECCV 2014, 2014: 446–461.
- [25] VAN HORN G, BRANSON S, FARRELL R, et al. Building a bird recognition app and large scale dataset with citizen scientists: the fine print in fine-grained dataset collection [C]//Proceedings of 2015 IEEE Conference on Computer Vision and

- Pattern Recognition (CVPR), 2015: 595 – 604.
- [26] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]// Proceedings of the 26th International Conference on Neural Information Processing Systems-Volume 1, 2012: 1097 – 1105.
- [27] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015 – 04 – 10) [2025 – 12 – 25]. <https://arxiv.org/abs/1409.1556>.
- [28] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 770 – 778.
- [29] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16 × 16 words: Transformers for image recognition at scale [C]//Proceedings of International Conference on Learning Representations, 2021.
- [30] LIONG V E, LU J W, WANG G, et al. Deep hashing for compact binary codes learning [C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 2475 – 2483.
- [31] DAI B, GUO R Q, KUMAR S, et al. Stochastic generative hashing [C]//Proceedings of the 34th International Conference on Machine Learning-Volume 70, 2017: 913 – 922.
- [32] SHEN Y M, LIU L, SHAO L. Unsupervised binary representation learning with deep variational networks [J]. International Journal of Computer Vision, 2019, 127(11): 1614 – 1628.
- [33] SHEN Y M, QIN J, CHEN J X, et al. Auto-encoding twin-bottleneck hashing [C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 2815 – 2824.
- [34] ZIEBA M, SEMBERECKI P, EL-GAALY T, et al. BinGAN: learning compact binary descriptors with a regularized GAN [C]//Proceedings of Advances in Neural Information Processing Systems 31 (NeurIPS 2018), 2018: 1 – 11.
- [35] DIZAJI K G, ZHENG F, NOURABADI N S, et al. Unsupervised deep generative adversarial hashing network [C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 3664 – 3673.
- [36] WANG B N, LIU K, ZHAO J. Deep semantic hashing with multi-adversarial training [C]//Proceedings of the 27th ACM International Conference on Information and Knowledge Management, 2018: 1453 – 1462.
- [37] DENG C, YANG E K, LIU T L, et al. Unsupervised semantic-preserving adversarial hashing for image search [J]. IEEE Transactions on Image Processing, 2019, 28(8): 4032 – 4044.
- [38] YANG E K, DENG C, LIU T L, et al. Semantic structure-based unsupervised deep hashing [C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence, 2018: 1064 – 1070.
- [39] TU R C, MAO X L, WEI W. MLS³RDUH: deep unsupervised hashing via manifold based local semantic similarity structure reconstructing [C]//Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, 2020: 3466 – 3472.
- [40] LUO X, WU D Q, MA Z Y, et al. A statistical approach to mining semantic similarity for deep unsupervised hashing [C]//Proceedings of the 29th ACM International Conference on Multimedia, 2021: 4306 – 4314.
- [41] QIN Q B, HUANG L, WEI Z Q, et al. Unsupervised deep multi-similarity hashing with semantic structure for image retrieval [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(7): 2852 – 2865.
- [42] XI Z, WANG X M, CHENG P T. Unsupervised hashing retrieval via efficient correlation distillation [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(7): 3529 – 3541.
- [43] YANG E K, LIU T L, DENG C, et al. DistillHash: unsupervised deep hashing by distilling data pairs [C]// Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 2941 – 2950.
- [44] LUO X, WU D Q, MA Z Y, et al. CIMON: towards high-quality hash codes [C]//Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, 2021: 902 – 908.
- [45] TU R C, MAO X L, LIN K Q, et al. Unsupervised hashing with semantic concept mining [J]. Proceedings of the ACM on Management of Data, 2023, 1(1): 1 – 19.
- [46] SHEN F M, XU Y, LIU L, et al. Unsupervised deep hashing with similarity-adaptive and discrete optimization [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(12): 3034 – 3044.
- [47] ZHANG W Q, WU D Y, ZHOU Y, et al. Deep unsupervised hybrid-similarity Hadamard hashing [C]//Proceedings of the 28th ACM International Conference on Multimedia, 2020: 3274 – 3282.
- [48] LIN K, LU J W, CHEN C S, et al. Learning compact binary descriptors with unsupervised deep neural networks [C]// Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 1183 – 1192.
- [49] YU J G, SHEN Y M, WANG M H, et al. Learning to hash naturally sorts [C]//Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, 2022: 1587 – 1593.
- [50] MA Z Y, LUO X, CHEN Y J, et al. Improved deep unsupervised hashing with fine-grained semantic similarity mining for multi-label image retrieval [C]//Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, 2022: 1254 – 1260.
- [51] YU J G, QIU H M, CHEN D B, et al. Weighted contrastive hashing [C]//Computer Vision-ACCV 2022, 2023: 251 – 266.
- [52] CAO H, HUANG L, NIE J, et al. Unsupervised deep

- hashing with fine-grained similarity-preserving contrastive learning for image retrieval [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024, 34(5): 4095–4108.
- [53] QIU Z X, SU Q L, OU Z J, et al. Unsupervised hashing with contrastive information bottleneck [C]//*Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, 2021: 959–965.
- [54] HU Q H, WU J X, CHENG J, et al. Pseudo label based unsupervised deep discriminative hashing for image retrieval[C]//*Proceedings of the 25th ACM International Conference on Multimedia*, 2017: 1584–1590.
- [55] ZHANG H F, LIU L, LONG Y, et al. Unsupervised deep hashing with pseudo labels for scalable image retrieval[J]. *IEEE Transactions on Image Processing*, 2018, 27(4): 1626–1638.
- [56] LIU H W, YIN M H, WU Z D, et al. PLDH: pseudo-labels based deep hashing[J]. *Mathematics*, 2023, 11(9): 2175.
- [57] CHAO L, LIU Y, YU H, et al. Joint MLP and pseudo-label unsupervised hash for image retrieval[C]//*Proceedings of the 2022 5th International Conference on Artificial Intelligence and Pattern Recognition*, 2023: 587–591.
- [58] GU Y F, WANG S D, ZHANG H F, et al. Clustering-driven unsupervised deep hashing for image retrieval [J]. *Neurocomputing*, 2019, 368: 114–123.
- [59] DONG X, LIU L, ZHU L, et al. Unsupervised deep K -means hashing for efficient image retrieval and clustering[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 31(8): 3266–3277.
- [60] ZHANG W Q, WU D Y, YANG C L, et al. Clustering and separating similarities for deep unsupervised hashing [C]//*Proceedings of ICASSP 2022 – 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022: 1655–1659.
- [61] LIN Q H, CHEN X J, ZHANG Q, et al. Deep unsupervised hashing with latent semantic components[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, 36(7): 7488–7496.
- [62] JIN L, LI Z C, PAN Y H, et al. Relational consistency induced self-supervised hashing for image retrieval[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2025, 36(1): 1482–1494.
- [63] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C]//*Proceedings of the 29th International Conference on Neural Information Processing Systems-Volume 1*, 2015: 91–99.
- [64] GATTUPALI V, ZHUO Y X, LI B X. Weakly supervised deep image hashing through tag embeddings [C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019: 10367–10376.
- [65] GUAN Z Y, XIE F, ZHAO W Q, et al. Tag-based weakly-supervised hashing for image retrieval [C]//*Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018: 3776–3782.
- [66] ZHU L, CUI H, CHENG Z Y, et al. Dual-level semantic transfer deep hashing for efficient social image retrieval[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 31(4): 1478–1489.
- [67] ZHAO W, XU C, GUAN Z Y, et al. TelecomNet: tag-based weakly-supervised modally cooperative hashing network for image retrieval [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(11): 7940–7954.
- [68] CUI H, ZHU L, LI J J, et al. Scalable deep hashing for large-scale social image retrieval[J]. *IEEE Transactions on Image Processing*, 2020, 29: 1271–1284.
- [69] JIN L, LI Z C, PAN Y H, et al. Weakly-supervised image hashing through masked visual-semantic graph-based reasoning[C]//*Proceedings of the 28th ACM International Conference on Multimedia*, 2020: 916–924.
- [70] WANG M, ZHOU W G, TIAN Q, et al. Deep enhanced weakly-supervised hashing with iterative tag refinement[J]. *IEEE Transactions on Multimedia*, 2022, 24: 2779–2790.
- [71] WANG J P, CHEN B, ZHANG Q, et al. Weakly supervised deep hyperspherical quantization for image retrieval [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, 35(4): 2755–2763.
- [72] DU Y C, WANG M, LU Z B, et al. Weakly supervised hashing with reconstructive cross-modal attention [J]. *ACM Transactions on Multimedia Computing, Communications and Applications*, 2023, 19(6): 1–19.
- [73] LU Z Y, JIN L, LI Z C, et al. Multi-modal knowledge distillation hashing based on CLIP for weakly supervised image retrieval[J]. *IEEE Transactions on Multimedia*, 2026, 28: 2996–3009.
- [74] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[EB/OL]. (2013–09–07) [2025–12–25]. <https://arxiv.org/abs/1301.3781>.
- [75] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision[C]//*Proceedings of the 38th International Conference on Machine Learning*, 2021: 8748–8763.
- [76] XIA R K, PAN Y, LAI H J, et al. Supervised hashing for image retrieval via image representation learning [C]//*Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014: 2156–2162.
- [77] CAO Z J, LONG M S, WANG J M, et al. HashNet: deep learning to hash by continuation[C]//*Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV)*, 2017: 5609–5618.
- [78] CAO Y, LONG M S, LIU B, et al. Deep Cauchy Hashing for hamming space retrieval [C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 1229–1237.
- [79] CHEN Y D, LAI Z H, DING Y J, et al. Deep supervised hashing with anchor graph[C]//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019: 9795–9803.

- [80] LIU H M, WANG R P, SHAN S G, et al. Deep supervised hashing for fast image retrieval [C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 2064 – 2072.
- [81] KANG R, CAO Y, LONG M S, et al. Maximum-margin hamming hashing [C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 8251 – 8260.
- [82] ZHANG R M, LIN L, ZHANG R, et al. Bit-scalable deep hashing with regularized similarity learning for image retrieval and person re-identification[J]. IEEE Transactions on Image Processing, 2015, 24(12): 4766 – 4779.
- [83] LIU B, CAO Y, LONG M S, et al. Deep triplet quantization[C]//Proceedings of the 26th ACM International Conference on Multimedia, 2018: 755 – 763.
- [84] LAI H J, PAN Y, LIU Y, et al. Simultaneous feature learning and hash coding with deep neural networks [C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 3270 – 3278.
- [85] JIN L, SHU X B, LI K, et al. Deep ordinal hashing with spatial attention[J]. IEEE Transactions on Image Processing, 2019, 28(5): 2173 – 2186.
- [86] ZHAO F, HUANG Y Z, WANG L, et al. Deep semantic ranking based hashing for multi-label image retrieval [C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 1556 – 1564.
- [87] HE K, CAKIR F, BARGAL S A, et al. Hashing as tie-aware learning to rank [C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 4023 – 4032.
- [88] JIANG Q Y, LI W J. Asymmetric deep supervised hashing[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32(1): 3342 – 3349.
- [89] SHEN F M, GAO X, LIU L, et al. Deep asymmetric pairwise hashing [C]//Proceedings of the 25th ACM International Conference on Multimedia, 2017: 1522 – 1530.
- [90] YANG H F, LIN K, CHEN C S. Supervised learning of semantics-preserving hash via deep convolutional neural networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(2): 437 – 451.
- [91] SU S P, ZHANG C, HAN K, et al. Greedy hash: towards fast optimization for accurate hash coding in CNN [C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems, 2018: 806 – 815.
- [92] SHEN Y M, QIN J, CHEN J X, et al. Embarrassingly simple binary representation learning [C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2019: 2883 – 2892.
- [93] LI N, LI C, DENG C, et al. Deep joint semantic-embedding hashing [C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence, 2018: 2397 – 2403.
- [94] XIE Y Z, LIU Y, WANG Y T, et al. Label-attended hashing for multi-label image retrieval [C]//Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, 2021: 955 – 962.
- [95] YUAN L, WANG T, ZHANG X P, et al. Central similarity quantization for efficient image and video retrieval [C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 3080 – 3089.
- [96] 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 Twenty-Ninth International Joint Conference on Artificial Intelligence, 2021: 825 – 831.
- [97] HOE J T, NG K W, ZHANG T Y, et al. One loss for all: deep hashing with a single cosine similarity based learning objective [C]//Proceedings of Advances in Neural Information Processing Systems 34 (NeurIPS 2021), 2021: 24286 – 24298.
- [98] WANG L D, PAN Y, LIU C, et al. Deep hashing with minimal-distance-separated hash centers [C]//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023: 23455 – 23464.
- [99] LU Z Y, JIN L, LI Z C, et al. Self-paced relational contrastive hashing for large-scale image retrieval [J]. IEEE Transactions on Multimedia, 2024, 26: 3392 – 3404.
- [100] WU D Y, DAI Q, LIU J, et al. Deep incremental hashing network for efficient image retrieval [C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 9061 – 9069.
- [101] CHEN W, LIU Y, PU N, et al. Feature estimations based correlation distillation for incremental image retrieval [J]. IEEE Transactions on Multimedia, 2021, 24: 1844 – 1856.
- [102] TIAN X, NG W W Y, XU H H. Deep incremental hashing for semantic image retrieval with concept drift [J]. IEEE Transactions on Big Data, 2023, 9(4): 1102 – 1115.
- [103] CHEN Y, HOU Y Q, LENG S, et al. Long-tail hashing [C]//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021: 1328 – 1338.
- [104] KOU X, XU C H, YANG X, et al. Attention-guided contrastive hashing for long-tailed image retrieval [C]//Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, 2022: 1017 – 1023.
- [105] JIANG Y, XIE H T, ZHANG L, et al. Dual dynamic proxy hashing network for long-tailed image retrieval [C]//Proceedings of the 31st ACM International Conference on Multimedia, 2023: 8942 – 8953.
- [106] XIE H T, JIANG Y, ZHANG L, et al. Semantic-enhanced proxy-guided hashing for long-tailed image retrieval [J]. IEEE Transactions on Multimedia, 2024, 26: 9499 – 9514.
- [107] JIN L, LU Z Y, LI Z C, et al. Causal inference hashing for long-tailed image retrieval [J]. IEEE Transactions on Image Processing, 2025, 34: 5099 – 5114.
- [108] ZENG H E, LAI H J, YIN J. Simultaneous region localization and hash coding for fine-grained image retrieval [EB/OL]. (2019 – 11 – 19) [2025 – 12 – 25]. <https://arxiv.org/abs/1911.08028>.
- [109] JIN S, YAO H X, SUN X S, et al. Deep saliency hashing for fine-grained retrieval [J]. IEEE Transactions on Image

- Processing, 2020, 29: 5336 – 5351.
- [110] XIANG X G, ZHANG Y J, JIN L, et al. Sub-region localized hashing for fine-grained image retrieval[J]. IEEE Transactions on Image Processing, 2022, 31: 314 – 326.
- [111] YANG Y F, GENG L B, LAI H J, et al. Feature pyramid hashing [C]//Proceedings of the 2019 on International Conference on Multimedia Retrieval, 2019: 114 – 122.
- [112] MA L, LI X, SHI Y, et al. Correlation filtering-based hashing for fine-grained image retrieval [J]. IEEE Signal Processing Letters, 2020, 27: 2129 – 2133.
- [113] CUI Q, JIANG Q Y, WEI X S, et al. ExchNet: a unified hashing network for large-scale fine-grained image retrieval[C]//Computer Vision-ECCV 2020, 2020: 189 – 205.
- [114] XUE X T, SHI J Y, HE X X, et al. Cross-scale context extracted hashing for fine-grained image binary encoding[C]//Proceedings of Machine Learning Research 189, 2022.
- [115] LU D, WANG J P, ZENG Z Y, et al. SwinFGHash: fine-grained image retrieval via Transformer-based hashing network[C]//Proceedings of the 32nd British Machine Vision Conference, 2021: 421.
- [116] SUN H, LANG W X, XU C, et al. Graph-based discriminative features learning for fine-grained image retrieval [J]. Signal Processing: Image Communication, 2023, 110: 116885.
- [117] ZHAO Q, WANG X, LYU S C, et al. A feature consistency driven attention erasing network for fine-grained image retrieval[J]. Pattern Recognition, 2022, 128: 108618.
- [118] SHEN Y, SUN X H, WEI X S, et al. SEMICON: a learning-to-hash solution for large-scale fine-grained image retrieval[C]//Computer Vision-ECCV 2022, 2022: 531 – 548.
- [119] CHEN Z D, LUO X, WANG Y X, et al. Fine-grained hashing with double filtering [J]. IEEE Transactions on Image Processing, 2022, 31: 1671 – 1683.
- [120] MA L, HONG H Y, MENG F M, et al. Deep progressive asymmetric quantization based on causal intervention for fine-grained image retrieval [J]. IEEE Transactions on Multimedia, 2024, 26: 1306 – 1318.
- [121] WEI X S, SHEN Y, SUN X H, et al. Attribute-aware deep hashing with self-consistency for large-scale fine-grained image retrieval[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(11): 13904 – 13920.
- [122] LU X, CHEN S K, CAO Y C, et al. Attributes grouping and mining hashing for fine-grained image retrieval [C]//Proceedings of the 31st ACM International Conference on Multimedia, 2023: 6558 – 6566.
- [123] XIANG X G, DING X H, JIN L, et al. Alleviating overfitting in hashing-based fine-grained image retrieval: from causal feature learning to binary-injected hash learning[J]. IEEE Transactions on Multimedia, 2024, 26: 10665 – 10677.
- [124] ALEMI A A, FISCHER I, DILLON J V, et al. Deep variational information bottleneck [C]//Proceedings of International Conference on Learning Representations, 2017.
- [125] VARSHAMOV R R. Estimate of the number of signals in error correcting codes[J]. Doklady Akademii Nauk SSSR, 1957, 117(5): 739 – 741.
- [126] ZHUO Y X, LI Y K, HSIAO J, et al. CLIP4Hashing: unsupervised deep hashing for cross-modal video-text retrieval [C]//Proceedings of the 2022 International Conference on Multimedia Retrieval, 2022: 158 – 166.
- [127] LI M Y, LI Y W, GE M Y, et al. CLIP-based fusion-modal reconstructing hashing for large-scale unsupervised cross-modal retrieval [J]. International Journal of Multimedia Information Retrieval, 2023, 12(1): 2.
- [128] LI J X, WONG W K, JIANG L, et al. CKDH: CLIP-based knowledge distillation hashing for cross-modal retrieval[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(7): 6530 – 6541.
- [129] LI F L, WANG B W, ZHU L, et al. Cross-domain transfer hashing for efficient cross-modal retrieval [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(10): 9664 – 9677.