

ADH: 一种基于注意力机制的深度 哈希图像检索方法*

张 浩¹, 陆雯茜^{2**}

(1. 昆明冶金高等专科学校 计算机信息学院, 云南 昆明 650061; 2. 玉溪市气象局 气象台, 云南 玉溪 653100)

【摘 要】在图像近邻搜索领域, 哈希技术作为一种高效的数据处理方法, 在大规模多媒体检索应用中展现了显著优势. 它通过将高维数据映射为紧凑的二进制编码, 从而大幅提升搜索效率. 然而, 哈希编码的有效性高度依赖于高维数据的相似度评估, 而深度学习驱动的哈希算法尽管取得了显著进展, 仍面临随着卷积神经网络 (CNN) 层数增加而导致的图像特征信息逐渐丢失的问题, 这不仅降低了哈希编码的准确性, 也限制了其在实际应用中的效果. 为应对这一挑战, 提出了一种创新性的深度哈希模型——注意力机制深度哈希 (Attention Deep Hashing, ADH). 该模型巧妙地结合了注意力机制, 精确定位并增强图像中对相似度计算至关重要的特征区域, 同时为不同特征分配相应的权重. 此方法有效缓解了深层网络结构带来的信息衰减问题, 并进一步提升了哈希编码的表达能力. 为了验证 ADH 模型的有效性, 在公开的 CIFAR-10 数据集上进行了广泛的训练与测试实验. 结果表明, 与传统的深度哈希方法相比, ADH 模型将检索结果的平均精度提高至 0.768, 证明了注意力机制在优化深度哈希编码中的关键作用. 此外, 研究表明, ADH 模型具有良好的泛化能力, 为未来高维数据高效检索研究提供了新的思路与技术参考.

【关键词】深度学习; 深度哈希; 注意力机制; 图像搜索; 卷积神经网络

【中图分类号】 TP391 **【文献标志码】** A **【文章编号】** 1674-5639 (2025) 06-0099-08

DOI: 10.14091/j.cnki.kmxyxb.2025.06.013

随着大数据时代的到来, 海量高维数据的高效相似性检索需求不断增强, 基于哈希编码的图像检索技术因而受到广泛关注^[1]. 在高维数据相似性检索中, 核心任务是根据查询数据在特征空间中快速找到最相似的实例. 为实现这一目标并进一步提升检索效率, 哈希方法被广泛应用于高维数据处理^[2,3], 其基本原理是将原始数据特征映射为固定长度的二进制编码向量, 并通过计算海明距离衡量数据间的相似度.

哈希技术通常通过特定计算方法将高维数据映射为二进制编码. 常见方法可分为传统哈希和可学习哈希两类.

1) 传统哈希 (非数据依赖哈希): 局部敏感哈希 (Locality Sensitive Hashing, LSH)^[4]是代表性方法之一, 通过随机线性映射将相似数据映射为相近的二进制向量, 被广泛用于高维数据检索. 改进方法, 如 KLSH^[5]引入核函数以处理高维核数据, Super-bit LSH^[6]与 Non-metric LSH^[7]等也提升了 LSH 的泛化能力. 然而, 这类方法依赖人工选择哈希函数, 通常需要较长的二进制编码向量以保证检索精度, 存在一定局限性.

2) 可学习哈希 (数据依赖哈希): 该方法通过训练数据直接学习哈希函数, 可进一步划分为监督和

无监督方法两大类. 无监督方法通常在无需标签信息的情况下, 通过优化特征结构来获得哈希编码. 例

* [收稿日期] 2025-01-17

[作者简介] 张浩, 男, 云南昆明人, 昆明冶金高等专科学校高级工程师, 研究方向为深度学习.

** [通信作者] 陆雯茜, 女, 云南昆明人, 玉溪市气象局工程师, 研究方向为人工智能、深度学习, E-mail: Lu240711@163.com.

[基金项目] 云南省气象局基层台站气象科技创新与能力提升计划项目 (STIAP202311).

如, 迭代量化 (ITQ)^[8] 通过交替最优学习正交变换以减少量化误差; 稀疏乘积量化 (SPQ)^[9] 则在相同思想下进一步利用 K-means 对特征子空间进行聚类量化, 从而得到稀疏向量形式的编码。相比之下, 监督学习哈希方法在训练中引入相似性标签, 通过约束样本, 使相似样本具有较小海明距离, 不相似样本具有较大海明距离, 来学习更具判别性的哈希函数, 如 KSH^[10, 11]。

近年来, 深度学习技术在各领域取得突破, 基于深度学习的可学习哈希方法迎来了新的发展机遇^[12, 13]。在众多技术中, 卷积神经网络 (CNN) 凭借其在图像处理方面的卓越表现, 成为构建深度哈希框架的理想选择^[14]。CNN 能够直接作用于图像像素值, 提取丰富、深层且具有高度区分性的特征信息, 从而避免传统方法特征提取不足或信息丢失的问题。借助端到端的训练机制, 深度哈希模型不仅能捕捉图像的本质特征, 还能将其映射为紧凑且有效的二进制编码, 显著提升了图像检索的效率与准确性。

卷积神经网络哈希 (CNNH)^[15] 是最早提出的深度学习哈希方法, 包括特征提取和哈希编码两个阶段。随后, 深度哈希网络 (DHN)^[16]、深度量化网络 (DQN)^[17]、实例相似深度哈希 (ISDH)^[18] 等端到端方法相继被提出, 这些方法能够直接输出二进制哈希编码, 进一步提高了检索性能与计算效率。

尽管基于深度学习的哈希方法在图像检索中取得了显著进展, 但是其仍存在一些挑战。深度哈希目标是利用哈希编码尽可能表达图像或者高维数据的关键信息。然而随着 CNN 网络深度增加, 长距离依赖问题逐渐凸显, 特征在逐层传递中可能出现衰减或丢失, 尤其是在网络后层, 难以准确聚焦于关键特征, 从而影响哈希编码的判别能力。因此, 研究者们开始探索能够增强特征选择性和表示能力的机制^[19], 其中注意力机制的引入被认为是解决上述问题的有效途径。

注意力机制最早应用于自然语言处理领域^[20], 其核心思想是根据特征的重要性分配不同权重, 从而更有效地建模长距离依赖并减轻信息丢失问题。据此, 本文提出了一种新颖的深度哈希模型——Attention Deep Hashing (ADH)。该模型将注意力机制融合于深度哈希框架中, 以增强特征表达能力并进一步提升检索表现。

本文的主要贡献如下:

- 1) 提出了一种端到端的深度哈希框架, 该框架基于卷积神经网络, 能够同时学习特征表示二进制哈希编码。
- 2) 创新性地引入注意力机制, 有效缓解了深层神经网络中因网络加深导致的信息衰减与特征丢失问题。
- 3) 基于公开数据集 CIFAR-10 对所提模型进行了训练与评估, 验证其在图像检索任务中的有效性。

本文结构如下: 首先介绍模型架构以及相应的训练与优化方法; 随后展示实验结果并进行分析; 最后给出全文总结与未来展望。

1 方法

1.1 ADH 网络架构

在图像相似度检索领域, 对于给定的查询数据 x , 系统需要在数据集 Y 中快速找到与其相似的数据实例。为实现高效、准确的相似度检索, 本文提出了一种端到端的深度哈希网络模型如图 1 所示。该模型主要由三个部分组成: 主干网络、注意力模块和二进制编码模块。

1) 主干网络: ADH 模型的主干网络由深度卷积神经网络和全连接层构成, 其主要功能是提取输入图像的高层语义特征。主干网络与其他模块之间保持松耦合设计, 以便灵活替换不同的特征提取结构。本文选用 ResNet^[21] 作为主干网络, 用于学习具有判别性的图像特征表示。

2) 二进制编码模块: 该模块由若干全连接层构成, 通过非线性映射将主干网络提取的特征转换到哈希空间, 随后, 通过 softmax 函数 (或其他离散化策略) 将特征编码为二进制向量 $\{b_1, b_2, \dots, b_n\}$ 。

3) 注意力机制模块: 注意力机制模块是本模型的核心设计, 旨在解决深度哈希网络中随着卷积神经网络深度加深, 特征在信息传递过程中逐渐衰弱或丢失的问题。具体而言, 该模块能够自适应地关注对哈希编码生成至关重要的特征区域或通道, 从而有效减少信息丢失并提升哈希表示的判别性。

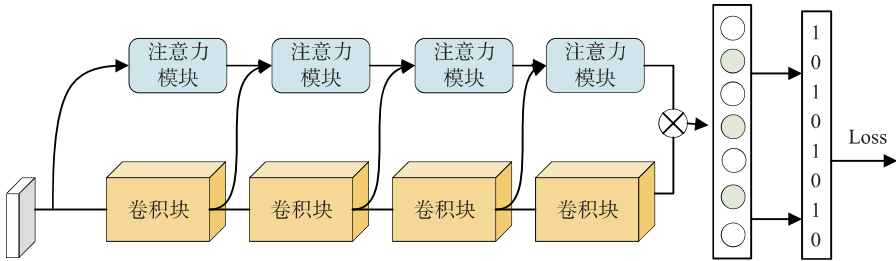


图1 ADH 网络模型架构

1.2 注意力机制结构

本文提出的注意力机制结构主要由两部分组成：长距离注意力和短距离注意力。核心思想在于，将两者生成的注意力权重向量进行逐元素融合（相乘），生成新的综合注意力权重向量，从而实现特征重要信息的有效传递，最终作用于输出层。如图2所示，首先，在短距离注意力模块中，输入特征图通过卷积操作实现降维。为了使输出权重维度与二进制编码向量一致，注意力模块添加了全连接层。通过注意力模块，特征间的长距离依赖关系得以量化。具体过程可以表示为公式（1）。

$$y_i = \text{sigmoid}\left(\frac{1}{c}f(x_i)w^T\right). \tag{1}$$

其中 $x_i \in R^c$ 表示输入特征图像素， f 对特征图做点卷积操作， c 是特征图的维度，卷积操作后的结果通过权重矩阵 w^T 进行微调，最后通过 sigmoid 函数将值映射到 $[0, 1]$ 之间。对于两个注意力模块通过函数 g 进行融合， $Y = g(y_i, y'_i)$ 函数 g 计算 y_i 和 y'_i 之间的相关度。

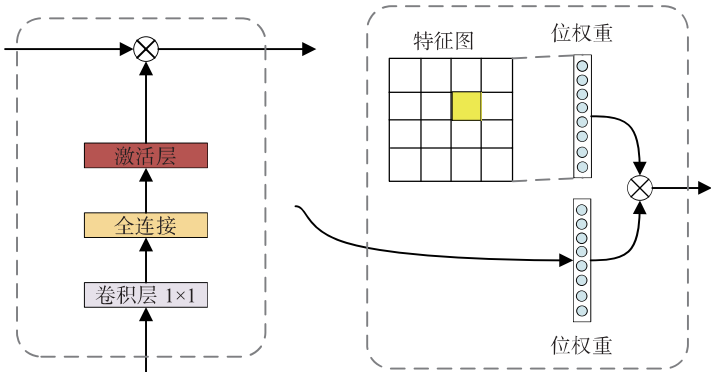


图2 注意力模块结构

1.3 损失函数

本文提出的深度哈希模型通过监督学习方法进行训练，借助注意力深度哈希模型，将图像映射为紧凑的一组二进制编码，同时保持高级语义信息。为了使得二进制编码向量有效表达图像的特征，必须满足以下两个条件：一是学习到的二进制编码应能在不同类别之间具有足够的区分度；二是同一类别的样本应具有较高的相似度。总体而言，训练目标是最大化相似图像之间的相似度，同时最小化不相似图像之间的相似度，从而实现语义一致性的高效哈希表示。

为实现上述目标，本文提出了如公式（2）所示的损失函数。

$$\min L(w) = \alpha \cdot L(w)_{\text{inter}} + \beta \cdot L(w)_{\text{intra}}. \tag{2}$$

其中 $\alpha \cdot L(w)_{\text{inter}}$ 为不同类别之间的损失，训练过程中需要最大化类与类之间的损失。 $\beta \cdot L(w)_{\text{intra}}$ 表示同类样本之间的损失，训练过程中需要最小化该损失。

具体来说，哈希编码的相似度计算方法有多种，例如余弦距离、海明距离、欧式距离等。其中海明距离因其计算简单且适用于二进制编码而被广泛采用，其值通过统计两条等长哈希编码在对应比特上的

差异数量获得, 具体如公式 (3) 所示.

$$D_h(\mathbf{b}_1, \mathbf{b}_2) = (K - \mathbf{b}_1^T \mathbf{b}_2) / 2. \quad (3)$$

其中 K 为编码向量长度, $\mathbf{b}_1, \mathbf{b}_2$ 分别为两个哈希编码向量, 通过内积的方式计算两个向量之间的相同位数.

对于给定输入图片数据 x_1, x_2 及其相似标签, 当给定 $y=0$ 则表示 x_1, x_2 为不相似样本, 反之当 $y=1$ 则表示 x_1, x_2 为相似样本, 通过 ADH 网络分别输出两个哈希向量 $\mathbf{b}_1, \mathbf{b}_2$, 其海明距离记作 $D_h(\mathbf{b}_1, \mathbf{b}_2)$. 本文设计了带阈值的损失, 如公式 (4) 所示.

$$L(\mathbf{b}_1, \mathbf{b}_2, y) = y \cdot D_h(\mathbf{b}_1, \mathbf{b}_2) + (1 - y) \cdot \max(0, m - D_h(\mathbf{b}_1, \mathbf{b}_2)). \quad (4)$$

其中 $m > 0$ 为预设的阈值. 该损失将相似对的距离最小化, 同时迫使不相似对的距离至少大于 m , 从而在哈希空间中同时保证聚类性与可分性.

2 实验

2.1 实验设置

本文在数据集 CIFAR-10^[22] 上对 ADH 模型进行了评估, 实验使用了 50 000 个训练数据和 10 000 个测试数据, 输入图像大小为 32×32 像素, 模型的训练和测试均基于 Pytorch 框架进行实现. 在初始化阶段, 主干网络采用 ResNet 架构并加载了 ImageNet^[23] 预训练参数, 网络的其余部分则采用了随机初始化. 训练过程中, 学习率范围设置为 $[1e-3, 1e-6]$, 所有实验均在 NVIDIA GeForce 1660 显卡上完成, 模型的训练大约需要 250 个 epoch, 训练时间约为 28 h. 为了评估不同长度的哈希码的性能, 实验中将二进制矢量的长度分别设置为 12 位、24 位、36 位和 48 位.

2.2 实验结果分析

为了验证 ADH 模型的有效性, 本文与 LSH^[24]、SH^[25]、ITQ^[8]、CCA-ITQ^[8]、MLH^[26]、BRE^[27]、KSH^[28]、CNNH^[29]、DLBHC^[30]、和 DNNH^[31] 方法比较, 分别测试了这些模型在数据集 CIFAR-10 上的表现, 测试的二进制编码长度为: 12 位、24 位、36 位和 48 位. 各方法的 mAP 值可见于表 1 和表 2. 具体来说, 首先使 ADH 与非数据依赖方法 LSH、SH、ITQ、CCA-ITQ、MLH、BRE 和 KSH 进行对比. 从表 1 中可以看出, ADH 模型与非数据依赖方法中表现最好的 KSH 方法比较: 当哈希编码为 12 位时, ADH 模型 mAP 值为 0.607, 较 KSH 方法的 0.295 提高了 0.312; 在 24 位编码条件下, ADH 的 mAP 值为 0.664, 相比 KSH 的 0.372, 提升了 0.292; 当编码长度为 36 位时, ADH 依然保持明显优势, 其 mAP 为 0.670, 比 KSH 的 0.402 高出 0.268; 而在 48 位编码设置下, ADH 的 mAP 进一步提升至 0.768, 相比于 KSH 的 0.417 提高了 0.351. 这些结果证明, 深度学习哈希函数和分类器可以提高检索性能.

表 1 ADH 与传统方法 mAP 对比

方法	CIFAR-10			
	12 位	24 位	36 位	48 位
LSH	0.128	0.137	0.141	0.149
SH	0.132	0.128	0.136	0.132
ITQ	0.108	0.109	0.112	0.118
CCA-ITQ	0.165	0.196	0.209	0.217
MLH	0.184	0.199	0.205	0.209
BRE	0.159	0.163	0.169	0.172
KSH	0.295	0.372	0.402	0.417
ADH	0.607	0.664	0.670	0.768

为了进一步验证 ADH 方法的有效性, 表 2 展示了其与深度学习哈希方法 CNNH^[29]、DLBHC^[32]、DNNH^[31] 和 DSH^[33] 的 mAP 值对比. 从表 2 可见, ADH 方法效果优于绝大多数的深度学习哈希方法, 以

表现最好的 DSH 方法为例作比较: 当哈希编码设置为 24 位时, ADH 的 mAP 值为 0.664, 相比于 DSH 方法的 0.651 提高了 2.0%; 而在编码长度为 36 位条件下, ADH 的 mAP 达到 0.670, 比 DSH 的 0.660 提高了 1.5%; 进一步提高编码长度至 48 位, ADH 的 mAP (0.768) 较 DSH (0.675) 提高了 13.8%, 优势明显.

表 2 ADH 与深度学习方法结果对比

方法	CIFAR-10			
	12 位	24 位	36 位	48 位
CNNH	0.543	0.560	0.564	0.557
DLBHC	0.550	0.580	0.578	0.588
DNNH	0.571	0.588	0.590	0.590
DSH	0.616	0.651	0.660	0.675
ADH	0.607	0.664	0.670	0.768

此外, 本文通过 t-SNE^[34] 降维方法对 4 种哈希编码进行了可视化. 如图 3 所示, 每一种颜色代表一种方法, 4 个子图分别表示 12 位、24 位、36 位和 48 位编码效果. 从图中可以看出 4 种编码中 10 个类别间界限分布明显, 表明 ADH 方法生成的编码保留了原图像中包含的语义信息.

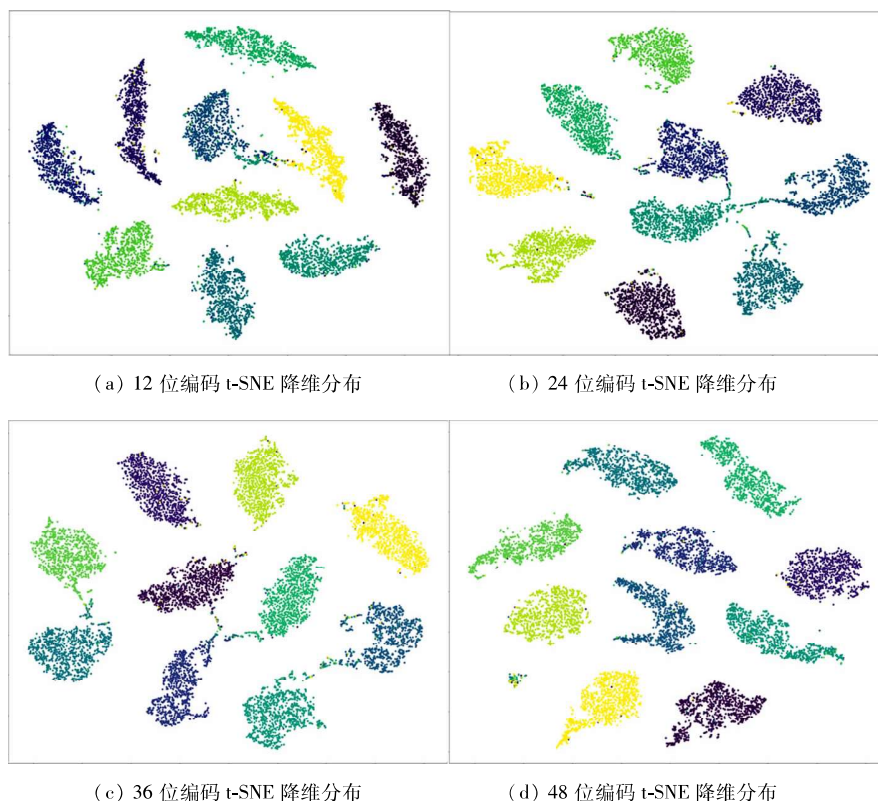


图 3 ADH 结果 t-SNE 降维分布

图 4 展示了本文提出的 ADH 方法在 CIFAR-10 数据集上的 Top 5 检索结果. 左侧列为输入的查询图片, 每张图片代表一个特定类别, 作为检索的参考; 右侧 5 列为与查询图片匹配的 Top 5 检索结果, 这些图片由 ADH 方法从索引库中自动选取, 按照匹配度从高到低排列. 该图展示了 CIFAR-10 数据集中 5 个类别的检索效果. 具体来说, 第一行展示了“飞机”类别, 查询图片为一架停放在机场的飞机, 检索结果包括不同背景和角度下的飞机, 所有结果与查询图片外观相近. 类似的, 第二行是“汽车”类别, 第三

行是“鸟类”类别，第四行是“猫”类别，第五行是“狗”类别，检索结果包含多种外观相似的图像，保持了较高的视觉一致性。整体来看，ADH 方法在视觉和语义上能够准确匹配查询图片与索引库中的图像，确保所有检索结果与查询图片类别一致，充分证明了该方法在图像索引和检索任务中的有效性和准确性。



图4 图片搜索结果前5个示例

3 总结

本文提出了一种创新的深度哈希图像检索方法——ADH 模型，其结合了注意力机制，通过在网络中引入长距离和短距离的注意力模块，精确地捕捉并强化关键特征，从而有效缓解了信息衰减和长距离依赖问题。ADH 模型的核心创新在于，通过端到端的深度学习框架，将图像特征与哈希编码进行一体化优化，并通过注意力机制动态地分配特征权重，以提高哈希编码的表达能力。

实验结果表明，ADH 模型在公开的 CIFAR-10 数据集上表现出色，相较于传统的哈希方法和其他深度学习哈希方法，ADH 模型的 mAP 显著较高，尤其在较长的哈希编码下，其性能优势更加明显。具体而言，当哈希编码长度为 48 位时，ADH 模型的 mAP 值达到 0.768，较 KSH 等传统方法提高值超过了 0.35，较深度学习方法如 DSH 提高了近 14%。这一结果证明了注意力机制在提升深度哈希编码准确性和检索效率中的关键作用。

总的来说，本文提出的 ADH 模型为深度哈希图像检索提供了一种新的解决方案，未来的研究将从以下几个方面进一步扩展：首先，可以探索不同的注意力机制变种，以进一步提升模型的性能；其次，考虑将 ADH 模型与其他深度学习技术结合；最后，可以考虑将该方法应用于更大规模的数据集，验证其在大规模图像检索中的实用性。

【参考文献】

- [1] 张雪凝, 刘兴波, 宋井宽, 等. 面向大规模图像检索的哈希学习综述 [J]. 软件学报, 2025, 36 (1): 79-106.

- [2] CAO Z, LONG M, WANG J, et al. HashNet: deep learning to hash by continuation [C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy. IEEE, 2017: 5609-5618.
- [3] 姚佩昀, 于炯, 李雪, 等. 基于跨尺度 Vision Transformer 的深度哈希算法 [J]. 计算机应用研究, 2024, 41 (11): 3477-3483.
- [4] 冯小康, 彭延国, 崔江涛, 等. 基于最优排序的局部敏感哈希索引 [J]. 计算机学报, 2020, 43 (5): 930-947.
- [5] KULIS B, GRAUMAN K. Kernelized locality-sensitive hashing for scalable image search [C]//2009 IEEE 12th International Conference on Computer Vision. Kyoto, Japan. IEEE, 2010: 2130-2137.
- [6] JI J, LI J, YAN S, et al. Super-bit locality-sensitive hashing [C]//NIPS 12: Neural Information Processing Systems. Lake Tahoe, USA. Curran Associates Inc, 2012: 108-116.
- [7] MU Y, YAN S. Non-metric locality-sensitive hashing [C]//The Twenty-Fourth AAAI Conference on Artificial Intelligence. Atlanta, USA. AAAI Press, 2010: 539-544.
- [8] GONG Y, LAZEBNIK S, GORDO A, et al. Iterative quantization: a procrustean approach to learning binary codes for large-scale image retrieval [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35 (12): 2916-2929.
- [9] JANG Y K, CHO N I. Self-supervised product quantization for deep unsupervised image retrieval [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada. IEEE, 2021: 12065-12074.
- [10] QIN Q, WEI Z, HUANG L, et al. Deep top similarity hashing with class-wise loss for multi-label image retrieval [J]. Neurocomputing, 2021, 439: 302-315.
- [11] JIANG X, HU F. Multi-scale adaptive feature fusion hashing for image retrieval [J]. Arabian Journal for Science and Engineering, 2025, 50 (15): 12027-12036.
- [12] WANG J, ZHANG T, SONG J, et al. A survey on learning to hash [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40 (4): 769-790.
- [13] 赵永晖, 胡海根. 基于子空间关系学习的跨模态哈希检索方法 [J]. 计算机应用与软件, 2024, 41 (10): 304-313.
- [14] REDAOUI A, BELALIA A, BELLOULATA K. Deep supervised hashing by fusing multiscale deep features for image retrieval [J]. Information, 2024, 15 (3): 143.
- [15] SU S, ZHANG C, HAN K, et al. Greedy hash: towards fast optimization for accurate hash coding in CNN [C]//32nd Conference on Neural Information Processing Systems. Montréal, Canada. Curran Associates Inc, 2018: 806-815.
- [16] PASSALIS N, TEFAS A. Deep supervised hashing using quadratic spherical mutual information for efficient image retrieval [J]. Signal Processing: Image Communication, 2021, 93: 116146.
- [17] CHEN Y, LU X. Deep discrete hashing with pairwise correlation learning [J]. Neurocomputing, 2020, 385: 111-121.
- [18] CHAO Z, CHENG S, LI Y. Deep internally connected transformer hashing for image retrieval [J]. Knowledge-Based Systems, 2023, 279: 110953.
- [19] 郭璠, 张泳祥, 唐璠, 等. YOLOv3-A: 基于注意力机制的交通标志检测网络 [J]. 通信学报, 2021, 42 (1): 87-99.
- [20] 任欢, 王旭光. 注意力机制综述 [J]. 计算机应用, 2021, 41 (增刊1): 1-6.
- [21] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA. IEEE, 2016: 770-778.
- [22] KRIZHEVSKY A. Learning multiple layers of features from tiny images [EB/OL]. (2009-04-08)[2025-01-01]. <http://www.cs.utoronto.ca/~kriz/learning-features-2009-TR.pdf>.
- [23] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60 (6): 84-90.
- [24] GIONIS A, INDYK P, MOTWANI R. Similarity Search in High Dimensions via Hashing [C]// Proceedings of the 25th International Conference on Very Large Data Bases. Edinburgh, Scotland, UK. Morgan Kaufmann Publishers Inc. 1999: 518-529.
- [25] WEISS Y, TORRALBA A, FERGUS R. Spectral hashing [C]//Proceedings of the 22nd International Conference on Neural Information Processing Systems. Vancouver, Canada. Curran Associates Inc. 2008: 1753-1760.
- [26] NOROUZI M, FLEET D J. Minimal loss hashing for compact binary codes [C]//Proceedings of the 28th International Conference on International Conference on Machine Learning. Bellevue, Washington, USA. Omnipress. 2011: 353-360.
- [27] KULIS B, DARRELL T. Learning to hash with binary reconstructive embeddings [C]// Proceedings of the 23rd International Conference on Neural Information Processing Systems. Vancouver, Canada. Curran Associates Inc. 2009: 1042-1050.

- [28] LIU W, WANG J, JI R, et al. Supervised hashing with kernels [C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI, USA. IEEE, 2012: 2074-2081.
- [39] XIA R, PAN Y, LAI H, et al. Supervised hashing for image retrieval via image representation learning [C]//Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence. Québec City, Québec, Canada. ACM, 2014: 2156-2162.
- [30] LIN K, YANG H F, HSIAO J H, et al. Deep learning of binary hash codes for fast image retrieval [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Boston, MA, USA. IEEE, 2015: 27-35.
- [31] LAI H, PAN Y, LIU Y, et al. Simultaneous feature learning and hash coding with deep neural networks [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA. IEEE, 2015: 3270-3278.
- [32] LIU X, ZHAO L, DING D, et al. Deep hashing with category mask for fast video retrieval [J]. ArXiv e-Prints, 2017: 08315.
- [33] LIU H, WANG R, SHAN S, et al. Deep supervised hashing for fast image retrieval [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA. IEEE, 2016: 2064-2072.
- [34] VAN-DER-MAATEN L, HINTON G. Visualizing data using t-SNE [J]. Journal of Machine Learning Research, 2008 (9): 2579-2605.

ADH: A Deep Hashing Image Retrieval Method Based on Attention Mechanism

ZHANG Hao¹, LU Wenxi²

- (1. School of Computer and Information Technology, Kunming Metallurgy College, Kunming, Yunnan, China 650061;
- 2. Yuxi Meteorological Observatory, Yuxi Meteorological Bureau, Yuxi, Yunnan, China 653100)

Abstract: In the field of image nearest-neighbor search, hashing technology has emerged as an efficient data processing method, demonstrating significant advantages in large-scale multimedia retrieval applications. It improves search efficiency by mapping high-dimensional data into compact binary codes. However, the effectiveness of hashing coding heavily depends on the similarity assessment of high-dimensional data. Although deep learning-driven hashing algorithms have made remarkable progress, they still face the problem of gradually losing image feature information as the layers of Convolutional Neural Networks (CNN) increase. This not only reduces the accuracy of hashing codes, but also limits their effectiveness in practical applications. To address this challenge, this paper proposes an innovative deep hashing model—Attention Deep Hashing (ADH). This model subtly integrates the attention mechanism to accurately locate and enhance the feature regions of the image that are crucial for similarity calculation, while assigning appropriate weights to different features. This approach effectively alleviates the information degradation issue caused by deep network structures and further improves the expressive ability of hashing codes. To validate the effectiveness of the ADH model, we conducted extensive training and testing experiments on the publicly available CIFAR-10 dataset. The experimental results show that compared to traditional deep hashing methods, the ADH model increases the average precision of retrieval results to 0.768, demonstrating the key role of the attention mechanism in optimizing deep hashing encoding. Additionally, the research indicates that the ADH model has good generalization ability, providing new ideas and technical references for future research on efficient high-dimensional data retrieval.

Key words: deep learning; deep Hashing; attention mechanism; image retrieval; convolutional neural networks

(责任编辑: 陈伟超)