

基于AnntDreamBooth微调技术的深度长尾哈希检索研究

李一昊, 王直杰

东华大学信息学院, 上海

收稿日期: 2025年1月7日; 录用日期: 2025年2月10日; 发布日期: 2025年2月20日

摘要

本文聚焦于图像检索中的长尾分布问题, 提出了一种基于数据增强的深度哈希检索框架。传统图像检索技术多依赖人工或自动标注, 但由于文本描述与图像存在语义鸿沟, 加之数据量的不断增长, 检索效率和成本成为主要瓶颈。深度哈希技术通过深度神经网络自动提取图像特征并映射到二进制哈希码空间, 显著提高了检索效率。然而, 实际数据中长尾分布的普遍存在导致深度哈希模型对尾部类别表现较差, 尤其在大规模、不平衡数据集上难以达到理想效果。为解决上述问题, 本文引入扩散模型这一生成技术, 通过生成尾部类别样本来丰富数据多样性, 从而缓解数据分布不均的影响, 增强模型对尾部类别的识别能力。同时, 结合AttnDreamBooth微调技术, 优化合成数据生成的质量, 将其与深度哈希技术相结合, 对东华大学丝绸数据集进行实验, 取得了显著的性能提升。通过从数据层面平衡分布并在模型层面强化训练, 本文提出的方法有效应对了长尾分布对深度哈希检索的负面影响, 为图像检索领域提供了新思路与实践方向。

关键词

长尾分布, 哈希检索, 丝绸数据集, 数据增强, 扩散模型

Deep Long-Tail Hash Retrieval Based on AttnDreamBooth Fine-Tuning

Yihao Li, Zhijie Wang

College of Information Science and Technology, Donghua University, Shanghai

Received: Jan. 7th, 2025; accepted: Feb. 10th, 2025; published: Feb. 20th, 2025

Abstract

This paper focuses on the long-tail distribution problem in image retrieval and proposes a deep

文章引用: 李一昊, 王直杰. 基于 AnntDreamBooth 微调技术的深度长尾哈希检索研究[J]. 计算机科学与应用, 2025, 15(2): 23-32. DOI: 10.12677/csa.2025.152030

hashing retrieval framework based on data augmentation. Traditional image retrieval techniques often rely on manual or automatic annotation; however, the semantic gap between textual descriptions and images, coupled with the continuous growth of data volume, has made retrieval efficiency and cost major bottlenecks. Deep hashing technology leverages deep neural networks to automatically extract image features and map them into binary hash code space, significantly improving retrieval efficiency. Nevertheless, the widespread presence of long-tail distributions in real-world data leads to poor performance of deep hashing models on tail classes, especially on large-scale, imbalanced datasets. To address this issue, this paper introduces diffusion models as a generative technique to enrich data diversity by generating samples for tail classes, thereby mitigating the impact of uneven data distribution and enhancing the model's ability to recognize tail classes. Additionally, AttnDreamBooth fine-tuning technology is employed to optimize the quality of synthesized data. By integrating this approach with deep hashing techniques, experiments on the Donghua University silk dataset demonstrate significant performance improvements. By balancing distribution at the data level and strengthening training at the model level, the proposed method effectively addresses the negative effects of long-tail distribution on deep hashing retrieval, offering new insights and practical directions for the field of image retrieval.

Keywords

Long-Tail Distribution, Hash Retrieval, Silk Dataset, Data Augmentation, Diffusion Model

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着大数据时代的到来, 互联网中的信息数据以指数级速度增长, 其中图像资源占据了相当大的比例。因此, 如何在海量图像中实现快速而精准的检索, 成为当前研究领域的一个重要课题。图像检索技术旨在为用户提供便捷的以图搜图功能, 以应对这一挑战。

图像检索, 尤其是基于深度学习的图像检索技术[1], 近年来得到了广泛的关注与应用。传统的基于文本的图像检索[2]方法通常依赖人工标注或自动标注来建立图像与文本之间的关联。然而, 由于文本描述与图像之间存在语义鸿沟, 且随着数据量的增加, 人工标注的成本与计算资源需求大幅上升, 导致了检索效率的降低。为了解决这些问题, 深度哈希检索技术应运而生, 通过深度学习模型从图像中自动提取特征并进行有效的哈希编码, 实现了更加高效的图像检索。

深度哈希检索方法的基本理念是通过深度神经网络自动学习图像的低维特征表示, 并将其映射到二进制哈希码空间。在检索过程中, 图像的哈希码会与数据库中其他图像的哈希码进行比对, 从而显著提高了检索效率和准确度。尽管该方法在数据集较为平衡的情况下表现良好, 但在实际应用中, 图像数据常常呈现长尾分布, 即大部分样本集中在头部类别, 而尾部类别的样本则非常稀缺。长尾分布的存在使得深度哈希模型容易过度关注头部类别的训练, 进而导致在尾部类别的检索性能较差。

为了有效应对长尾分布问题, 近期的研究者们提出了多种解决方案。传统的数据增强方法在一定程度上通过增加尾部类别样本的多样性来缓解这一问题。然而, 单纯的数据增强仍然无法完全解决长尾分布带来的影响, 尤其是在复杂的视觉任务中, 如何提高模型对尾部类别的辨识能力仍然是一个亟待解决的挑战。

近年来, 扩散模型[3] (Diffusion Models)作为一种生成模型, 已经在多个视觉任务中展现出强大的能

力, 尤其是在图像生成与编辑方面。与传统生成模型相比, 扩散模型能够通过逐步优化过程生成更为细致和精确的图像, 从而在复杂场景下提供更高质量的图像生成与编辑。特别是在处理长尾分布数据时, 扩散模型能够通过生成多样化的尾部类别样本来平衡数据分布, 增强模型的泛化能力。

受扩散模型的启发, 本文提出一种基于数据增强的长尾哈希检索方法, 以缓解长尾数据带来的挑战。

论文的主要贡献包括:

本文提出了一种基于数据增强的深度哈希检索框架。该框架通过微调现有训练数据的生成模型, 实现个性化的合成数据生成。随后, 利用生成的合成数据对检索模型进行预训练, 并针对长尾分布数据进行均衡化处理。其次, 首次将 AttnDreambooth 微调技术结合哈希检索应用于东华大学丝绸数据集上并取得较好的效果, 这样, 既从数据层面也从模型层面有效缓解了长尾分布对深度哈希检索模型训练的负面影响。

2. 相关工作

本文旨在通过数据增强方法解决深度长尾哈希检索问题。鉴于当前鲜有研究专注于深度哈希检索领域中的长尾分布问题, 本节对针对图像分类任务中长尾分布数据的主流方法进行了详细调研, 为解决深度哈希检索的长尾分布挑战提供了理论参考和实践依据。

常见的数据增强方法是通过对原始图像进行一系列变换, 生成新的图像样本, 从而有效扩充训练集的过程。常见的数据增强方法包括如图像的翻转、平移、旋转、缩放或裁剪等, 旨在通过不同的空间变换增加图像的多样性。同时, 数据增强还可以通过调整图像的颜色属性来增加变化, 如改变亮度、对比度、饱和度, 或通过通道混洗等方式。尽管几何变换和颜色调整等简单的数据增强方法可以增加数据集的多样性, 提升模型对常见类别的学习效果, 但它们往往无法有效弥补长尾分布中尾部类别的不足的问题。

当下, 深度学习中对于长尾分布数据集的处理方法可分为五种: 重采样、重加权、特征强化、解耦训练及复杂方法, 例如, Chawla 等人于 2002 年提出了 SMOTE [4] 算法, 旨在解决特征强化问题。该算法通过生成合成样本来平衡数据集, 并在使用朴素贝叶斯网络进行实验时取得了较好的效果。然而, SMOTE 算法主要针对的是少数类样本的不平衡问题, 适用范围有限。2020 年, Kang 等人 [5] 提出了解耦长尾学习的方法, 该方法将表示学习与分类器学习分开进行。首先在原始长尾数据上学习具有良好泛化能力的特征表示, 然后通过类平衡学习技术微调分类器, 以应对长尾分布问题。Lawrence 等人在重加权领域提出显式调节先验概率并得到了出色的成功 [6], 但该方法需要特定领域的专业知识和先验信息来调整先验概率, 这些信息可能难以获取或存在不确定性, 从而影响该方法在实际应用场景中的适用性。伴随人工智能在图像识别领域的普及, 缓解数据集呈长尾态的解决方案陆续提出。部分文献基于头尾知识迁移思想 [7], 利用头部类丰富、完整的训练资源指导尾部类的学习。另外, 一些研究主张将长尾数据集划分为多个分布更平衡的子集, 通过知识蒸馏 [8]、集成学习 [9] [10] 等方法提升模型鲁棒性与泛化能力。长尾学习与其它领域交叉的过程中, 逐渐融合了度量学习 [11]、层次学习 [12]、因果推断 [13]、对抗学习 [14]、元学习 [15]、提示学习 [16] 等先进算法, 尽管这些方法在缓解长尾分布数据对模型训练的影响方面表现出一定的效果, 但生成的虚拟合成样本质量较低时, 会对模型的整体性能产生显著影响。低质量的合成样本可能无法有效代表尾部类别的特征, 甚至可能引入噪声, 从而导致模型的泛化能力下降。

生成模型因其生成自然、多样性、高保真的图像而闻名, 如 Stable diffusion [17]、LDM [17]、Imagen [18] 和 GLIDE [19] 等。特别地, 这些基础模型是现成且公开可用的。因此, 他们为一些数据稀缺场景提供了可能的解决策略, 即利用高质量的合成数据充当训练集来解决数据短缺问题。例如, N Ruiz [20] 等人提出了 DreamBooth。该方法通过对现有的预训练生成模型 Stable Diffusion 进行微调, 能够实现个性化和定

制化的图像生成。Joonghyuk 等人提出的 Textual Inversion [21] 是一种用于文本生成图像模型的技术，通过合成数据对尾部类别数据进行补充，最后使用原始数据进行微调，使数据更加均衡。

3. 基于数据增强的哈希检索模块

3.1. 经验性实验

为了深入研究长尾分布数据对深度哈希检索的影响，本节采用了均衡数据集 ImageNet 的子集 ImageNette [22]，通过构建长尾分布数据集，观察其对深度哈希检索性能的影响。我们选择了目前较为先进的检索框架 Orthohash [23] 进行训练。为了模拟现实中常见的长尾分布情况，图 1(a) 展示了将 ImageNette 训练集按照头部和尾部类别的不平衡比例进行重新构建，以形成长尾分布的训练集。通过与原始均衡数据集的对比，我们旨在更直观地观察长尾分布对哈希检索任务性能的影响。本节中，检索训练的初始化模型选择了在 ImageNet1k 上预训练的 ResNet50 模型。

基于均衡以及长尾分布的 ImageNette 数据集的深度哈希检索结果如图 1(b) 所示。

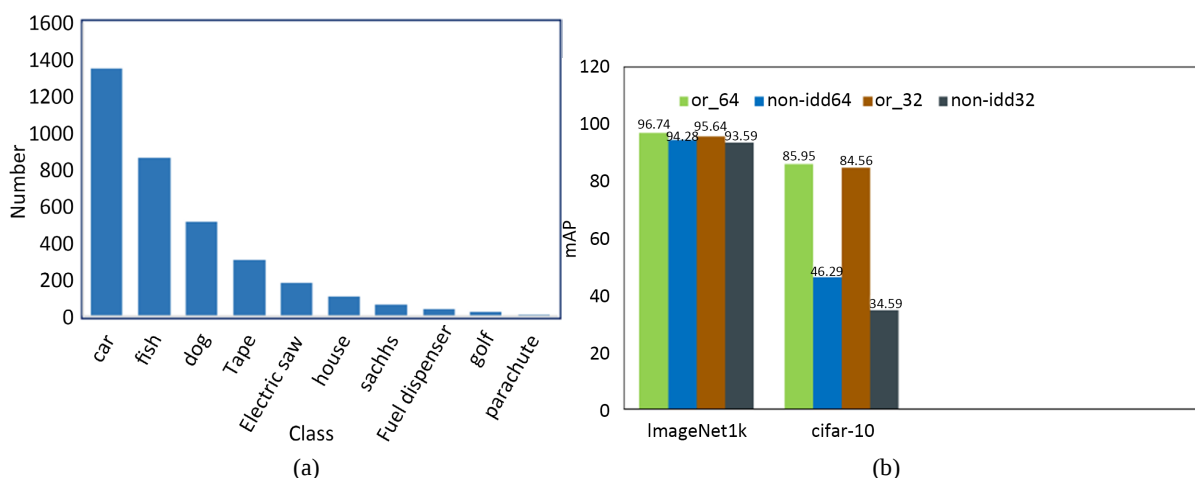


Figure 1. Distribution of long-tail data and corresponding search results; (a) ImageNette’s long-tail data distribution; (b) Empirical experimental results

图 1. 长尾数据分布以及相应的检索结果；(a) ImageNette 的长尾数据分布；(b) 经验性实验结果

在采用初始化训练的情况下，均衡数据和长尾数据上的哈希检索准确性差距不大。这是因为 ImageNette 本身是 ImageNet1k 的子集，其类别分布和数据特征与 ImageNet1k 非常相似。因此，即使在训练过程中使用了长尾分布的模拟数据，模型已经具备了一定的泛化能力，可以较好地处理长尾数据和均衡数据之间的差异，从而导致在这两种数据分布下的检索准确性差距较小。为了更有效地评估长尾分布的影响，检索训练的初始化模型选择了 cifar-10 上预训练的 ResNet50 模型进行。如图 1(b) 所示，在更换了预训练数据集的情况下，均衡数据和长尾数据上的哈希检索准确性有明显差距的长尾分布，ImageNette 训练的深度哈希检索模型精度平均下降约 40%，此外可以看出，64 比特位的哈希编码检索模型精度下降约 30%，32 比特位的哈希编码检索准确度下降了约 45%。

由此得出结论，长尾分布的数据集会严重影响哈希检索模型的精度。但是当预训练模型包含头部类和尾部类的大部分知识时，预训练的哈希模型不仅具有强大的泛化以及鲁棒性，而且已经包含下游数据的特征信息，因此提高了检索模型在“尾部”类别数据上的性能表现。这种方法为应对长尾分布数据对哈希模型训练的挑战提供了一个潜在的解决方案：通过使用包含上下游知识的类别平衡数据对深度哈希模型进行预训练，可以提前将数据的特征信息融入模型中，以降低长尾分布数据对训练过程的负面影响。

尽管如此, 尽管实践已经证明, 预训练模型结合下游数据特征可以减轻长尾分布数据的影响, 但长尾分布数据本身仍然面临类别样本数量不足的问题。此外, 后续处理的不平衡数据仍然可能对深度哈希模型的性能造成影响。

3.2. 深度哈希检索模型

为了提高预训练模型对下游数据的鲁棒性, 并避免后续哈希检索训练中类别不均衡带来的影响, 本节提出了基于数据增强的深度哈希检索框架, 其中包括基于个性化生成模型的数据生成模块、初始化检索模型预训练模块和长尾分布数据均衡化模块。如图 2 所示, 在个性化数据生成模块中, 利用稳定的个性化图像生成技术 AttnDreamBooth [24]生成个性化的合成数据。在初始化检索模型预训练模块中, 利用生成的类别均衡的个性化合成数据训练深度哈希检索模型。在长尾分布数据均衡化模块中, 通过合成数据对类别不均衡数据进行均衡化处理, 构建“合成 + 真实”的均衡混合训练数据集, 并对预训练后的深度哈希检索模型进行训练。

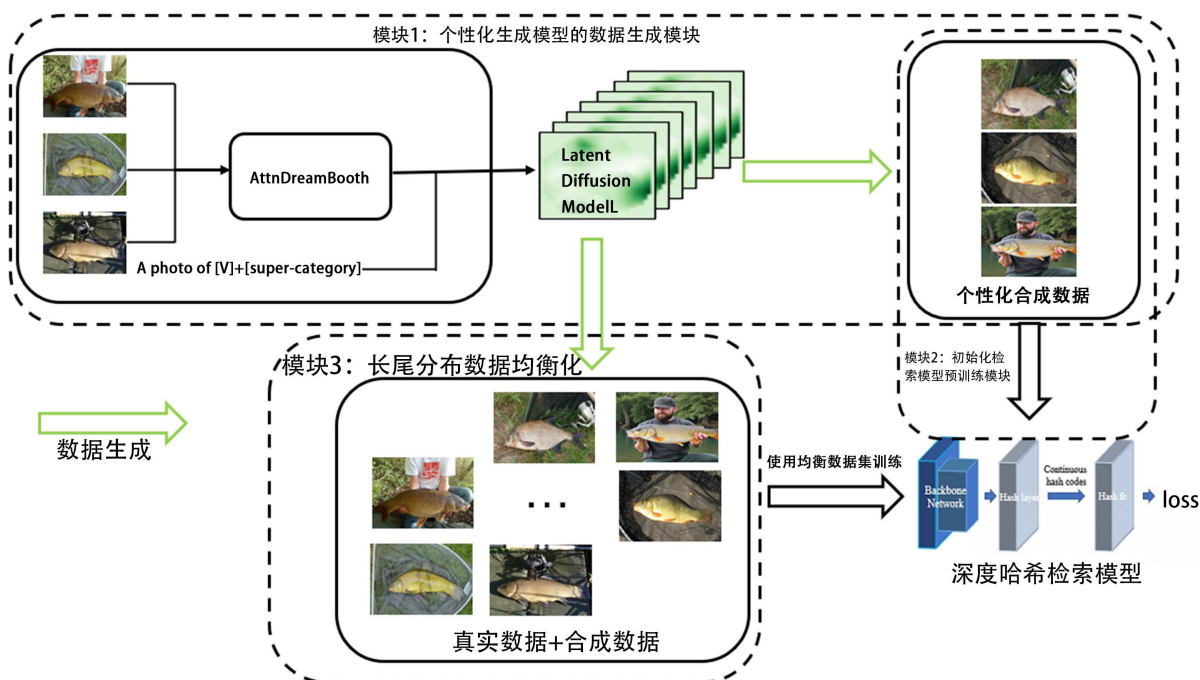


Figure 2. Deep hash retrieval framework based on data enhancement

图 2. 基于数据增强的深度哈希检索框架

3.3. AttnDreamBooth 微调技术

AttnDreamBooth 微调方法由三个训练阶段组成。

在第 1 阶段, 优化了新概念的文本嵌入, 使其嵌入与现有标记对齐。

在第 2 阶段, 对交叉注意力层进行微调以细化注意力图。

在第 3 阶段, 对整个 U-net 进行微调以捕获主题身份。为了增强注意图的学习, 引入了一个正则化项, 以鼓励[V]和[超类别]的注意图之间的相似性。这个正则化项有两个目的。首先, 由于 V 及其超类别属于同一对象范畴, 超类别标记的注意图可以作为新概念的参考。其次, 由于[V]和[超类别]在整合新提示时是一起用来描述新概念的, 所以[V]和[超类别]的注意图应该指向同一个区域。形式上, 对于 16 个注意力图 $\{M_1, M_2, \dots, M_{16}\}$ 来自 16 个不同的交叉注意层, 我们将[V]和[超类别]的注意图值的均值和方差

的平方差最小化如下:

$$L_{reg} = \lambda_{\mu} \left[\mu(M_{1:16}^V) - \mu(M_{1:16}^{category}) \right]^2 + \lambda_{\sigma} \left[\sigma^2(M_{1:16}^V) - \sigma^2(M_{1:16}^{category}) \right]^2 \quad (1)$$

其中 $\mu(M_{1:16})$ 和 $\sigma^2(M_{1:16})$ 分别表示 16 个注意图中所有值的均值和方差, [V] 和 [超类别] 分别表示新概念及其超类别的标记。AttnDreamBooth 通过将标识符(V) (超类别) 与用户提供的自己进行绑定, 从而微调模型的权重, 具体来说, 本节利用下游任务中的数据作为 AttnDreamBooth 的训练集, 使用包含类别信息以及特定标识符的提示词和于提示词相关的超类别, 来一同微调生成模型。

3.4. 基于个性化生成模型的数据生成模块

本章框架的核心目的是通过合成数据对初始化的检索模型进行预训练, 并对长尾分布数据集进行均衡化处理。因此, 合成数据的质量在整个过程中扮演着至关重要的角色。一个直观的方法是采用“文本 - 图像”生成技术, 将描述性文本提示词(例如“A photo of [object]”, 其中[object]表示合成数据的类别)输入到生成模型(如 Latent Diffusion Model)中, 以生成高质量的合成图像。然而, 对于细粒度数据或具有多样化风格的图片, 单纯的文本提示往往无法准确传达目标图像的特征, 导致生成结果无法真实反映目标内容, 产生偏离预期的效果。

然而, 对于丝绸这种具有独特风格特征的图片, 常规的文本 - 图像生成技术往往难以通过简单的提示词准确传达其细腻纹理和复杂图案。当前许多个性化生成模型如 Textual Inversion 容易出现过拟合概念的情况对于 text embedding 做注入, DreamBooth 容易忽视概念对于原本的扩散模型做微调。为了解决这些问题, 本节采用了一种个性化微调扩散模型技术: AttnDreamBooth, 该技术可以保留用户提供的数据集的风格特征, 生成真实且多样化的合成数据。

3.5. 初始化检索模型预训练模块

基于本章 3.2 节的经验性实验及相应结论, 本节通过 AttnDreamBooth 个性化生成模型为长尾分布数据的下游数据进行扩充, 使初始化检索模型具备丰富的上下游数据信息, 本节为每个类别生成 4000 张合成图片, 从而构成丰富数据信息的均衡化合成数据。

3.6. 长尾分布数据均衡化模块

为了进一步减少长尾分布对哈希检索模型的影响, 本节从不平衡的数据集出发, 通过高质量的个性化图像生成技术 AttnDreamBooth 为长尾分布的训练集进行数据增强, 使用不同的随机种子生成个性化合成数据, 并使均衡化后每个类别的“合成 + 真实”数量保持一致。

4. 实验

4.1. 数据集

为验证本文提出算法的正确性和可靠性, 本节融合东华大学丝绸数据集和细粒度数据集 ImageNet 的子集 ImageNette 作为我们的数据集。其中, 本文实验中的 ImageNette 选取 10 个类别, 为了模拟长尾分布的数据场景, 将不平衡常数(“头部”类别数量和“尾部”类别数量的比值)设置为 48。丝绸数据集选取 20 个类别进行实验, 如图 3 所示, 因其本身即为长尾数据集, 因此不做调整。

4.2. 实验设置

本文选择操作系统为

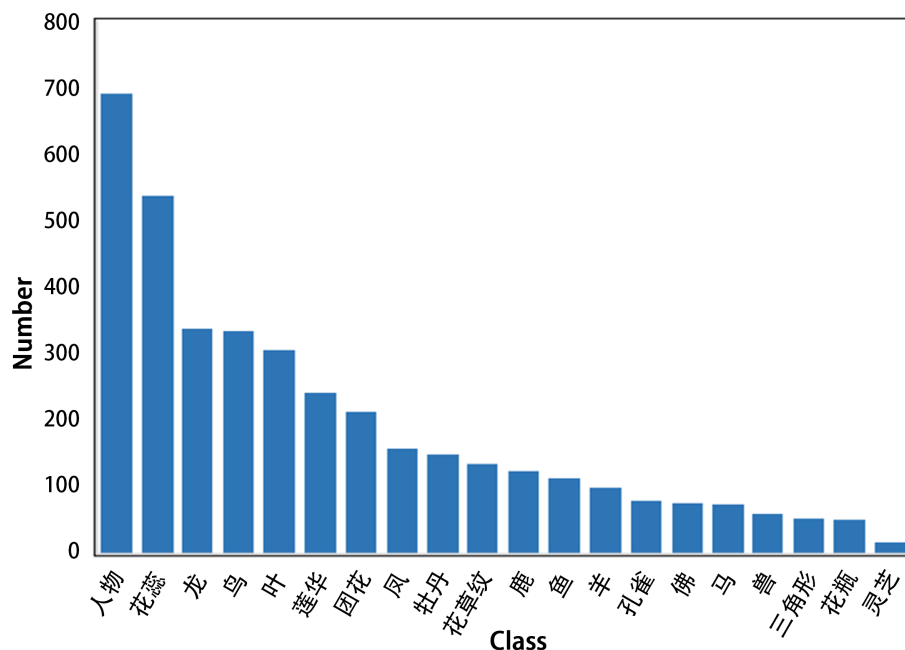


Figure 3. Long-tail data distribution of the Silk dataset

图 3. 丝绸数据集的长尾数据分布

Window10 (64 位), CPU 为 Intel(R) Core(TM) i7-9700K CPU @ 3.60 GHz, 内存为 12.0 GB, GPU 为 NVIDIA GeForce RTX 3060 的远程服务器进行部署。本文选择深度学习框架为 Pytorch1.10.1, torchvision0.12.2, Python 版本为 3.6.12。

考虑到本章的方法无需改变哈希检索模型的内部结构以及训练哈希函数的算法过程, 并且为了保证实验对比的公平性, 本章的实验均在较为先进的基于余弦相似度的单目标深度哈希检索框架 (Orthohash [5]) 上进行训练。本章的哈希检索训练设置也与 Orthohash 保持一致。具体来说, 检索训练的初始化模型选择了在 Cifar-10 上预训练的 ResNet50 模型, 选择随机初始化的 ResNet50、AlexNet 和 ViT 作为哈希检索模型的 backbone。选择 Adam 作为模型训练的优化器, 并使用余弦调度策略来促进学习率的衰减, 初始学习率设定为 3×10^{-4} 。本节实验中检索结果的评价指标也与目前的深度哈希检索工作保持一致, 即选择平均精度(mAP)作为检索结果的主要衡量标准, 同时分别在 32 以及 64 比特位的哈希码上进行验证。

4.3. 结果

本节展示了在均衡分布、长尾分布以及基于数据增强框架下, ImageNette 数据集和丝绸数据集的检索效果。如表 1 所示, 本章提出的方法在长尾分布的细粒度数据集(如 ImageNette)和丝绸数据集上均取得了显著的成果。通过对比均衡分布与长尾分布训练数据, 发现二者在深度哈希检索精度上存在明显差异。具体而言, 长尾分布数据显著影响了深度哈希检索模型的训练效果。在哈希码长度为 64 位的情况下, AlexNet、ResNet50 和 ViT 模型在长尾分布数据上的检索精度分别为 32.51%、39.16%和 45.83%, 与在均衡数据集上分别为 72.57%、79.15%和 87.13%的表现相比, 分别下降了 40.06%、39.99%和 41.30%。这一结果表明, 长尾分布对深度哈希检索性能造成了严重影响, 丝绸数据集也呈现类似问题。

然而, 本章提出的方法能够显著减小长尾分布训练集对哈希检索性能的影响。与均衡数据集的原始训练效果相比, AlexNet、ResNet50 和 ViT 在 64 位哈希编码下的检索精度分别仅下降了 4.29%、1.72%和 0.95%, 这一差距几乎与均衡数据集下的检索效果相当。在 ImageNette 数据集中, 相较于 AlexNet、ResNet50

和 ViT 在均衡数据集上分别达到的 72.57%、80.15%和 87.13%的检索精度,本章方法分别达到了 68.28%、78.43%和 86.18%。这些实验结果表明,无论是在 64 位还是 32 位哈希码下,本方法均能有效减轻长尾分布数据对深度哈希检索模型训练的影响。

Table 1. Experimental comparison
表 1. 实验对比

训练方法	BackBone	ImageNette		丝绸数据集	
		64	32	64	32
均衡数据	AlexNet	72.57	71.01		
长尾数据		32.51	32.14	25.24	24.06
本章方法		68.28	67.49	54.48	53.36
均衡数据	ResNet50	80.15	78.48		
长尾数据		39.16	38.04	36.53	34.87
本章方法		78.43	76.45	62.26	60.47
均衡数据	ViT	87.13	85.67		
长尾数据		45.83	45.08	43.49	41.54
本章方法		86.18	84.21	69.12	68.26

4.4. 可视化实验



Figure 4. Comparison of synthetic data of different methods
图 4. 不同方法的合成数据对比

本节主要通过对比合成数据的质量来讨论 AttnDreamBooth 的可行性,并讨论不同合成数据的质量对哈希检索模型准确度的影响。如图 4 所示,本节分别对比在 ImageNette 和丝绸数据集上的原始数据, DreamBooth 和 AttnDreamBooth 生成数据的视觉效果,可以看出, AttnDreamBooth 生成的合成数据更加贴近真实数据。

4.5. 对比实验

本节主要通过两个方面进行对比实验:首先探讨不同的生成模型产生的均衡数据在哈希检索模型上的性能,其次探讨了本方法在不同的哈希检索模型上的适用性。

由于本章方法旨在通过数据增强解决长尾分布对深度长尾哈希的影响,为了体现本文方法的实用性,本节使用主流的微调模型 DreamBooth 和 Textual Inversion 进行数据增强,对比其在哈希检索模型的性能,如表 2 所示,可以发现本章的数据增强策略较目前较为先进的图像生成模型有显著提升。

Table 2. mAP comparison results of personalized fine-tuning techniques**表 2.** 个性化微调技术性能对比实验 mAP 结果

训练数据	微调技术	BackBone	ImageNette		丝绸数据集	
			64	32	64	32
长尾数据	Textual Inversion	ResNet50	71.156	69.48	55.16	53.71
	DreamBooth	ResNet50	76.24	74.49	60.48	58.18
	本章方法	ResNet50	78.43	76.45	62.26	60.47

本节同时探讨了本章框架在 CSQ, DPN 和 Orthohash 上的性能如表 3 所示, 可以发现本章的数据增强策略运用在不同的哈希检索框架都有着显著提升效果。换句话说, 在面向深度长尾哈希问题中, 本章的框架可以作为一个解决的工具, 可以在不影响原始优化策略的前提下提升深度哈希检索框架对于训练长尾分布数据的鲁棒性。

Table 3. Comparison experiment Map results of Hash retrieval model performance**表 3.** 哈希检索模型性能对比实验 Map 结果

训练方法	哈希检索方法	Backbone	ImageNette		丝绸数据集	
			64	32	64	32
长尾数据	CSQ	ResNet50	38.18	37.28	35.58	33.48
本章方法			76.48	74.81	60.84	59.49
长尾数据	DPN	ResNet50	37.64	36.72	34.15	32.76
本章方法			77.15	75.18	60.25	59.12
长尾数据	Orthohash	ResNet50	39.16	38.04	36.53	34.87
本章方法			78.43	76.45	62.26	60.47

5. 结束语

本文首次提出了一种基于 AttnDreamBooth 个性化生成模型微调的深度长尾哈希检索框架。该框架通过利用长尾分布数据作为训练集, 生成与图像特征匹配的个性化合成数据。并结合东华大学丝绸数据集, 解决了丝绸数据集中数据稀缺和长尾分布的问题, 从而提高模型在丝绸图像的深度哈希检索任务中对长尾数据的鲁棒性。

参考文献

- [1] Chen, W., Liu, Y., Wang, W., *et al.* (2021) Deep Image Retrieval: A Survey.
- [2] Benavent, J., Benavent, X., de Ves, E., *et al.* (2010) Experiences at Image CLEF 2010 Using CBIR and TBIR Mixing Information Approaches. CLEF (Notebook Papers/LABs/Workshops).
- [3] Wijmans, J.G. and Baker, R.W. (1995) The Solution-Diffusion Model: A Review. *Journal of Membrane Science*, **107**, 1-21. [https://doi.org/10.1016/0376-7388\(95\)00102-i](https://doi.org/10.1016/0376-7388(95)00102-i)
- [4] Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P. (2002) SMOTE: Synthetic Minority Over-Sampling Technique. *Journal of Artificial Intelligence Research*, **16**, 321-357. <https://doi.org/10.1613/jair.953>
- [5] Kang, B., Xie, S., Rohrbach, M., *et al.* (2019) Decoupling Representation and Classifier for Long-Tailed Recognition.
- [6] Lawrence, S., Burns, I., Back, A., Tsoi, A.C. and Giles, C.L. (1998) Neural Network Classification and Prior Class Probabilities. In: *Lecture Notes in Computer Science*, Springer, 299-313. https://doi.org/10.1007/3-540-49430-8_15
- [7] Chu, P., Bian, X., Liu, S. and Ling, H. (2020) Feature Space Augmentation for Long-Tailed Data. In: *Lecture Notes in*

- Computer Science*, Springer, 694-710. https://doi.org/10.1007/978-3-030-58526-6_41
- [8] Zhang, S., Chen, C., Hu, X. and Peng, S. (2023) Balanced Knowledge Distillation for Long-Tailed Learning. *Neurocomputing*, **527**, 36-46. <https://doi.org/10.1016/j.neucom.2023.01.063>
 - [9] Sharma, S., Yu, N., Fritz, M. and Schiele, B. (2021) Long-Tailed Recognition Using Class-Balanced Experts. In: *Lecture Notes in Computer Science*, Springer, 86-100. https://doi.org/10.1007/978-3-030-71278-5_7
 - [10] Xiang, L., Ding, G. and Han, J. (2020) Learning from Multiple Experts: Self-Paced Knowledge Distillation for Long-Tailed Classification. In: *Lecture Notes in Computer Science*, Springer, 247-263. https://doi.org/10.1007/978-3-030-58558-7_15
 - [11] Cai, J., Wang, Y., Hsu, H., Hwang, J., Magrane, K. and Rose, C.S. (2022) LUNA: Localizing Unfamiliarity Near Acquaintance for Open-Set Long-Tailed Recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, **36**, 131-139. <https://doi.org/10.1609/aaai.v36i1.19887>
 - [12] Wu, J., Song, L., Zhang, Q., Yang, M. and Yuan, J. (2022) Forestdet: Large-Vocabulary Long-Tailed Object Detection and Instance Segmentation. *IEEE Transactions on Multimedia*, **24**, 3693-3705. <https://doi.org/10.1109/tmm.2021.3106096>
 - [13] Zhu, B., Niu, Y., Hua, X. and Zhang, H. (2022) Cross-Domain Empirical Risk Minimization for Unbiased Long-Tailed Classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, **36**, 3589-3597. <https://doi.org/10.1609/aaai.v36i3.20271>
 - [14] Wu, T., Liu, Z., Huang, Q., Wang, Y. and Lin, D. (2021) Adversarial Robustness under Long-Tailed Distribution. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 8655-8664. <https://doi.org/10.1109/cvpr46437.2021.00855>
 - [15] Ren, J., Yu, C., Ma, X., et al. (2020) Balanced Meta-Softmax for Long-Tailed Visual Recognition. *Advances in Neural Information Processing Systems*, **33**, 4175-4186.
 - [16] Dong, B., Zhou, P., Yan, S., et al. (2022) LPT: Long-Tailed Prompt Tuning for Image Classification.
 - [17] Rombach, R., Blattmann, A., Lorenz, D., Esser, P. and Ommer, B. (2022) High-Resolution Image Synthesis with Latent Diffusion Models. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 10674-10685. <https://doi.org/10.1109/cvpr52688.2022.01042>
 - [18] Saharia, C., Chan, W., Saxena, S., et al. (2022) Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. *Advances in Neural Information Processing Systems*, **35**, 36479-36494.
 - [19] Nichol, A., Dhariwal, P., Ramesh, A., et al. (2021) Glide: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models.
 - [20] Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M. and Aberman, K. (2023) Dreambooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 17-24 June 2023, 22500-22510. <https://doi.org/10.1109/cvpr52729.2023.02155>
 - [21] Gal, R., Alaluf, Y., Atzmon, Y., et al. (2022) An Image Is Worth One Word: Personalizing Text-to-Image Generation Using Textual Inversion.
 - [22] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015) ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, **115**, 211-252. <https://doi.org/10.1007/s11263-015-0816-y>
 - [23] Hoe, J.T., Ng, K.W., Zhang, T., et al. (2021) One Loss for All: Deep Hashing with a Single Cosine Similarity Based Learning Objective. *Advances in Neural Information Processing Systems*, **34**, 24286-24298.
 - [24] Pang, L., Yin, J., Zhao, B., et al. (2024) AttnDreamBooth: towards Text-Aligned Personalized Text-to-Image Generation.