



主办: 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2021
11
VOL.26

ISSN1006-8961
CN11-3758/TB





第26卷第11期（总第307期）
2021年11月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 吴一戎
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮 编 100190
电子信箱 jig@aircas.ac.cn
电 话 010-58887035
网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订 购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
(邮政信箱: 北京399信箱 邮编: 100048)
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

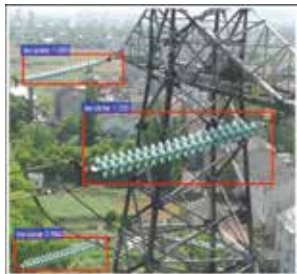
Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@aircas.ac.cn
Telephone 010-58887035
Website www.cjig.cn

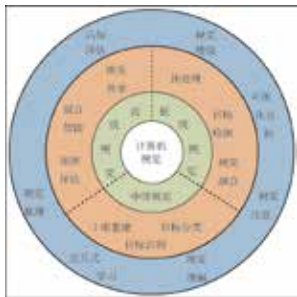
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元



面向改进尺度缩放网络的绝缘子识别(第2561页)



混合增强视觉认知架构及其关键技术进展(第2619页)



激光点云的稀疏体素金字塔邻域构建与分类(第2703页)

编者按 I

电力视觉前沿技术

输电线路部件视觉缺陷检测综述

赵振兵, 蒋志钢, 李延旭, 威银城, 翟永杰, 赵文清, 张珂 2545

面向改进尺度缩放网络的绝缘子识别

赵文清, 张海明, 徐敏夫 2561

最优知识传递宽残差网络输电线路螺栓缺陷图像分类

威银城, 金超熊, 赵振兵, 丁洁涛, 吕斌 2571

可变形NTS-Net的螺栓属性多标签分类

张珂, 何颖宣, 赵凯, 冯晓晗, 赵振兵, 马占宇 2582

嵌入双注意力机制的Faster R-CNN航拍输电线路螺栓缺陷检测

威银城, 武学良, 赵振兵, 史博强, 聂礼强 2594

轻量化航拍图像电力线语义分割

许刚, 李果 2605

综述

混合增强视觉认知架构及其关键技术进展

王培元, 关欣 2619

深度学习人脸特征点自动定位综述

徐亚丽, 赵俊莉, 吕智涵, 张志梅, 李劲华, 潘振宽 2630

图像处理和编码

多尺度特征复用混合注意力网络的图像重建

卢正浩, 刘丛 2645

图像分析和识别

自监督深度离散哈希图像检索

万方, 强浩鹏, 雷光波 2659

L-UNet: 轻量化云遮挡道路提取网络

许苗, 李元祥, 钟娟娟, 左宗成, 熊伟 2670

基于多尺度特征多对抗网络的雾天图像识别

陈硕, 钟汇才, 李勇周, 王师峰, 杨建刚 2680

图像理解和计算机视觉

结合动态图卷积和空间注意力的点云分类与分割

宋巍, 蔡万源, 何盛琪, 李文俊 2691

激光点云的稀疏体素金字塔邻域构建与分类

陶帅兵, 梁冲, 蒋腾平, 杨玉娇, 王永君 2703

计算机图形学

单幅植物叶片图像的3维重建

任非儿, 刘通, 杨龙 2713

医学图像处理

深度学习多部位病灶检测与分割

黎生丹, 柏正尧 2723

遥感图像处理

对抗学习遥感图像场景识别

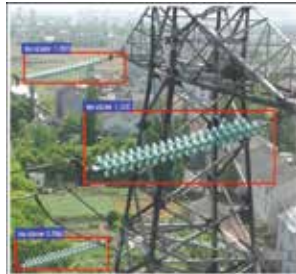
李彤, 张钧萍 2732

融合特征的卫星视频车辆单目标跟踪

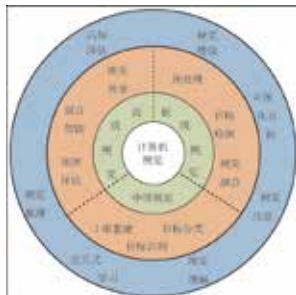
韩鸣飞, 李盛阳, 万雪, 轩诗宇, 赵子飞, 谭洪, 张万峰 2741

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Insulator recognition based on an improved scale-transferrable network(P2561)



Hybrid enhanced visual cognition framework and its key technologies(P2619)



Sparse voxel pyramid neighborhood construction and classification of LiDAR point cloud(P2703)

Frontier Technology of Power Computer Vision

Overview of visual defect detection of transmission line components

Zhao Zhenbing, Jiang Zhigang, Li Yanxu, Qi Yincheng, Zhai Yongjie, Zhao Wenqing, Zhang Ke 2545

Insulator recognition based on an improved scale-transferrable network

Zhao Wenqing, Zhang Haiming, Xu Minfu 2561

Image classification method of transmission line bolt defects using the optimal knowledge transfer wide residual network

Qi Yincheng, Jin Chaoxiong, Zhao Zhenbing, Ding Jietao, Lyu Bin 2571

Multi-label classification method of bolt attributes based on deformable NTS-Net

Zhang Ke, He Yingxuan, Zhao Kai, Feng Xiaohan, Zhao Zhenbing, Ma Zhanyu 2582

Bolt defect detection for aerial transmission lines using Faster R-CNN with an embedded dual attention mechanism

Qi Yincheng, Wu Xueliang, Zhao Zhenbing, Shi Boqiang, Nie Liqiang 2594

Research on lightweight neural network of aerial powerline image segmentation

Xu Gang, Li Guo 2605

Review

Hybrid enhanced visual cognition framework and its key technologies

Wang Peiyuan, Guan Xin 2619

Automatic facial feature points location based on deep learning: a review

Xu Yali, Zhao Junli, Lyu Zhihan, Zhang Zhimei, Li Jinhua, Pan Zhenkuan 2630

Image Processing and Coding

Multiscale feature reuse mixed attention network for image reconstruction

Lu Zhenghao, Liu Cong 2645

Image Analysis and Recognition

Self-supervised deep discrete hashing for image retrieval

Wan Fang, Qiang Haopeng, Lei Guangbo 2659

L-UNet: lightweight network for road extraction in cloud occlusion scene

Xu Miao, Li Yuanxiang, Zhong Juanjuan, Zuo Zongcheng, Xiong Wei 2670

Haze image recognition based on multi-scale feature and multi-adversarial networks

Chen Shuo, Zhong Huicai, Li Yongzhou, Wang Shizheng, Yang Jiangang 2680

Image Understanding and Computer Vision

Dynamic graph convolution with spatial attention for point cloud classification and segmentation

Song Wei, Cai Wanyuan, He Shengqi, Li Wenjun 2691

Sparse voxel pyramid neighborhood construction and classification of LiDAR point cloud

Tao Shuaibing, Liang Chong, Jiang Tengping, Yang Yujiao, Wang Yongjun 2703

Computer Graphics

3D reconstruction of a single plant leaf image

Ren Feier, Liu Tong, Yang Long 2713

Medical Image Processing

Multiorgan lesion detection and segmentation based on deep learning

Li Shengdan, Bai Zhengyao 2723

Remote Sensing Image Processing

Remote sensing image scene recognition based on adversarial learning

Li Tong, Zhang Junping 2732

Integrating multiple features for tracking vehicles in satellite videos

Han Mingfei, Li Shengyang, Wan Xue, Xuan Shiyu, Zhao Zifei, Tan Hong, Zhang Wanfeng 2741

中图法分类号: P23 文献标识码: A 文章编号: 1006-8961(2021)11-2659-11

论文引用格式: Wan F, Qiang H P and Lei G B. 2021. Self-supervised deep discrete hashing for image retrieval. Journal of Image and Graphics, 26(11): 2659-2669 (万方, 强浩鹏, 雷光波. 2021. 自监督深度离散哈希图像检索. 中国图象图形学报, 26(11): 2659-2669) [DOI:10.11834/jig.200212]

自监督深度离散哈希图像检索

万方^{1*}, 强浩鹏², 雷光波¹

1. 湖北工业大学计算机学院, 武汉 430068; 2. 武汉理工大学理学院, 武汉 430070

摘要: **目的** 基于深度学习的图像哈希检索是图像检索领域的热点研究问题。现有的深度哈希方法忽略了深度图像特征在深度哈希函数训练中的指导作用, 并且由于采用松弛优化, 不能有效处理二进制量化误差较大导致的生成次优哈希码的问题。对此, 提出一种自监督的深度离散哈希方法 (self-supervised deep discrete hashing, SSD-DH)。**方法** 利用卷积神经网络提取的深度特征矩阵和图像标签矩阵, 计算得到二进制哈希码并作为自监督信息指导深度哈希函数的训练。构造成对损失函数, 同时保持连续哈希码之间相似性以及连续哈希码与二进制哈希码之间的相似性, 并利用离散优化算法求解得到哈希码, 有效降低二进制量化误差。**结果** 将本文方法在 3 个公共数据集上进行测试, 并与其他哈希算法进行实验对比。在 CIFAR-10、NUS-WIDE (web image dataset from National University of Singapore) 和 Flickr 数据集上, 本文方法的检索精度均为最高, 本文方法的准确率比次优算法 DPSH (deep pairwise-supervised hashing) 分别高 3%、3% 和 1%。**结论** 本文提出的基于自监督的深度离散哈希的图像检索方法能有效利用深度特征信息和图像标签信息, 并指导深度哈希函数的训练, 且能有效减少二进制量化误差。实验结果表明, SSDDH 在平均准确率上优于其他同类算法, 可以有效完成图像检索任务。

关键词: 深度学习; 图像检索; 哈希学习; 自监督; 离散优化

Self-supervised deep discrete hashing for image retrieval

Wan Fang^{1*}, Qiang Haopeng², Lei Guangbo¹

1. School of Computing, Hubei University of Technology, Wuhan 430068, China;

2. College of Science, Wuhan University of Technology, Wuhan 430070, China

Abstract: **Objective** Hashing techniques have attracted much attention and are widely applied in the nearest neighbor search for image retrieval on large-scale datasets due to the low storage cost and fast retrieval speed. With the great development of deep learning, deep neural networks have been widely incorporated in image hashing retrieval, and existing deep learning-based hashing methods demonstrate the effectiveness of the end-to-end deep learning architecture for hashing learning. However, these methods have several problems. First, these existing deep hashing methods ignore the guiding role of deep image feature information in training deep hashing functions. Second, most deep hashing methods are to solve a relaxed problem first to simplify the optimization involved in a binary code learning procedure and then quantize the solved continuous solution to achieve the approximate binary solution. This optimization strategy leads to a large binary quantization error, resulting in the generation of suboptimal hash codes. Thus, to solve these two problems, a self-supervised deep discrete hashing method (SSDDH) is proposed in this study. **Method** The proposed SSDDH consists of two steps. First, using

收稿日期: 2020-06-17; 修回日期: 2020-11-20; 预印本日期: 2020-11-27

* 通信作者: 万方 20021026@hbut.edu.cn

基金项目: 中央高校基本科研业务费专项资金资助 (2019IB010); 湖北省教育厅指导性项目 (B2019049)

Supported by: Fundamental Research Funds for the Central University (2019IB010); Guiding Project of Hubei Provincial Department of Education (B2019049)

matrix decomposition, the binary hash code is obtained by solving the self-supervised loss function composed of the deep feature matrix extracted by the convolutional neural network and the image label matrix. The obtained binary hash code is used as the supervision information to guide the training of deep hash function. Second, a pair-wise loss function is constructed to maintain the similarity between the hash codes generated by deep hash function while maintaining the similarity between these hash codes and binary hash codes. The discrete optimization algorithm is used to solve the optimal solution of the objective function, thus effectively reducing the binary quantization error. **Result** Several experiments are conducted on three public datasets to validate the performance of the proposed algorithm. The first experiment compares the mean average precision (mAP) values of different existing hash methods on different hash code lengths, including unsupervised methods, supervised shallow methods, and supervised deep methods. The performance experimental results show that the mAP of our method SSDDH achieves the best performance in all cases with different values of the code length. On the CIFAR-10 and NUS-WIDE (web image dataset from National University of Singapore) datasets, the mAP of SSDDH is 3% higher than the next highest method named DPSH (deep pairwise-supervised hashing). On the Flickr dataset, SSDDH is also 1% higher than the highest method DPSH. The second experiment involves the CIFAR-10 dataset. The precision recall (PR) curves of DPSH and SSDDH are plotted. Query result comparison shows the PR curves of DPSH and SSDDH with 48-bit hash codes on CIFAR-10, and our SSDDH remarkably outperforms its competitor. SSDDH and DPSH are also compared in terms of the accuracy of the top 20 returned images when the hash code length is 48 bits. The result of the experiment is visualized for easy observation. We also found that the retrieval performance of SSDDH is considerably higher than that of DPSH. Experiment 3 is designed for parameter sensitivity analysis of SSDDH. Here, a parameter is used, while the others are fixed. Our method is insensitive to the parameters. This finding relatively demonstrates the robustness and effectiveness of the proposed method. Experiment 4 is conducted on CIFAR-10 when the hash code length is 48 bits to explore the difference between DPSH and SSDDH in time complexity. At the later stage of model training, SSDDH performance is better than DPSH at the same time consumption. **Conclusion** Considering that the existing deep hash methods ignore the guiding role of deep image feature information in the training of deep hash function and have the problem of large binary quantization error, this study proposes a self-supervised deep discrete hashing method named SSDDH. The deep feature matrix extracted by the convolutional neural network and the image label matrix are used to obtain the binary hash codes and make the binary hash codes the supervised information to guide the training of deep hash function. The similarity between the hash codes generated by deep hash function and the similarity between these hash codes and binary hash codes are maintained by constructing a pair-wise loss function. The binary quantization error is effectively reduced using the discrete cyclic coordinate descent. Comparison with several existing methods on three commonly used public datasets proves that this method is more efficient than the existing hash retrieval method. Future work lies in two aspects: First, focus will be on learning better fine-grained representation with more effectively. Second, semi-supervised regularization will be applied to our framework to make full use of the unlabeled data. Both will be employed to boost the image retrieval accuracy further. Third, our current approach will be extended to cross-modal retrieval, such as given a text query, to obtain all semantic relevant images from the database.

Key words: deep learning; image retrieval; hash learning; self-supervised; discrete optimization

0 引言

随着网络社交媒体和移动网络等技术的蓬勃发展,每天都会产生海量的多媒体数据,基于内容的图像检索技术(content-based image retrieval, CBIR)受到了广泛关注,而基于哈希的方法因其较低的存储成本与时间复杂度成为图像检索领域的热点。哈希算法在编码过程中,利用特定的哈希函数将检索系

统提取的图像的高维特征映射成长度固定的二进制码,大幅降低了空间复杂度;在检索过程中,利用按位的与或运算的高效性,利用汉明距离(Hamming distance)计算得到实例间的相似度,提高了检索速度(Yan等,2018;刘冶等,2016;Shen等,2020)。

通常,哈希方法根据是否依赖数据划分为数据无关方法和数据相关方法(Jiang等,2018)。数据无关的哈希方法通过随机投影,将数据点从原始特征空间投影到汉明空间,得到的哈希函数与训练数据

分布无关。典型方法有 Andoni 和 Indyk (2006) 提出的局部敏感哈希 (locality-sensitive hashing, LSH) 以及 Kulis 和 Grauman (2009) 提出的 LSH 的扩展算法——核局部敏感哈希 (kernelized locality sensitive hashing, KLSH)。LSH 使用从高斯分布中采样的一组随机生成的超平面, 将原始高维数据投影到超平面上, 最后将投影结果作为哈希函数的输出进行阈值化。但仅在生成较长的哈希码时, 数据无关的方法才能获得较好的检索性能 (Liu 等, 2012)。与之相比, 数据相关的方法在采用较小的哈希码长度时可以取得更好的检索性能。

根据是否引入监督信息, 数据相关的哈希方法分为无监督哈希方法和监督哈希方法。无监督哈希方法使用数据点的特征信息得到哈希函数, 且不需要任何外界提供的语义或者标签信息。这类方法旨在保持训练数据之间在原空间中的相似程度。代表算法有 Weiss 等人 (2008) 提出的谱哈希 (spectral hashing, SH)、Gong 等人 (2013) 提出的迭代量化 (iterative quantization, ITQ)、Jin 等人 (2014) 提出的密度敏感哈希 (density sensitive hashing, DSH) 以及 Liu 等人 (2014) 提出的离散图哈希 (discrete graph hashing, DGH)。SH 从图分割的角度考虑, 求解图拉普拉斯矩阵的特征向量, 并对其二值化, 从而得到相应的二进制码。ITQ 算法使用主成分分析方法对数据进行降维操作, 引入旋转矩阵, 最小化哈希码和原始数据之间的量化损失, 从而提高了哈希检索的性能。由于很多哈希方法采用松弛的优化策略, 导致当哈希码位数增加时, 算法检索能力下降。DGH 通过离散优化算法降低二进制量化误差, 并采用锚点图捕获数据集中隐含的近邻结构。

相比无监督方法, 引入了监督信息的哈希方法在实践应用中达到更高的检索精度。早期的监督哈希方法主要包括 Wang 等人 (2010) 提出的哈希的顺序投影学习法 (sequential projection learning for hashing, SPLH)、Liu 等人 (2012) 提出的基于核的监督哈希算法 (supervised hashing with kernels, KSH)、Zhang 等人 (2014) 提出的潜在因子哈希 (latent factor hashing, LFH) 和 Shen 等人 (2015) 提出的离散监督哈希 (supervised discrete hashing, SDH)。KSH 通过最小化相似图像之间的哈希码的距离, 通过最大化不相似图像之间的距离生成哈希码。LFH 将数据点映射到一个潜在子空间中, 在该子空间中嵌入

标签信息以保持语义结构。考虑到离散约束, SDH 不仅保留了语义结构, 而且不需要任何松弛即可离散地学习哈希码。但是, 这些方法中图像特征提取过程和哈希编码过程通常是独立进行的, 可能导致生成的哈希码与特征之间不能很好地兼容 (Jiang 和 Li, 2017)。

随着深度学习在图像分类 (Simonyan 和 Zisserman, 2015; Krizhevsky 等, 2012)、目标检测 (Liu 等, 2016b; Ren 等, 2017) 和语义分割 (Noh 等, 2015; He 等, 2020) 等领域表现出强大性能, 更多深度哈希方法相继提出, 主要包括 Xia 等人 (2014) 提出的基于卷积神经网络 (convolutional neural networks, CNN) 的哈希 (convolution neural network hashing, CNNH)、Liu 等人 (2016a) 提出的深度监督哈希 (deep supervised hashing, DSH)、Zhu 等人 (2016) 提出的深度哈希网络 (deep hashing network, DHN)、Li 等人 (2016) 提出的深度成对监督哈希 (deep pairwise-supervised hashing, DPSH) 以及 Cao 等人 (2018) 提出的哈希生成对抗网络 (hashing generative adversarial network, HashGAN)。CNNH 需要两个步骤得到图像的哈希码, 首先构造训练数据集的相似矩阵得到近似哈希码, 然后利用 CNN 学习哈希函数得到哈希码。DSH、DHN、DPSH 和 HashGAN 为端到端的方法, 即特征学习与哈希编码同时进行。DPSH 通过构造成对交叉熵损失函数得到较好的哈希码。HashGAN 借鉴 GAN (Goodfellow 等, 2014) 的思想, 从真实图像和生成模型合成的多种图像中学习紧凑的二进制哈希码。深度哈希方法将图像特征提取过程和哈希编码过程统一在同一个网络中, 很好地解决了早期哈希方法中存在的兼容性问题 (王志明和张航, 2019; Lu 等, 2020)。

现有的深度哈希方法通常直接采用松弛策略生成连续的哈希码, 然后利用符号函数 (或阈值分割) 将连续的哈希码转换成二进制码。这类方法虽然便于实现, 但会带来很大的量化损失, 从而生成低质量的哈希码。对此, 本文提出一种自监督的深度离散哈希方法 (self-supervised deep discrete hashing, SSDH), 网络结构如图 1 所示。首先, 通过矩阵分解, 利用深度特征矩阵和图像标签矩阵直接生成二进制哈希码, 并作为网络的自监督信息指导深度哈希函数的训练。其次, 构建两个相似性矩阵分别用来保持连续哈希码之间的相似性以及连续哈希码与离散哈希

码之间的相似性。同时,采用离散优化算法学习哈希码,减少了在连续哈希码到二进制码转化过程中的量

化误差。与其他哈希方法的对比实验结果验证了本文提出的 SSDDH 在图像检索领域的先进性。

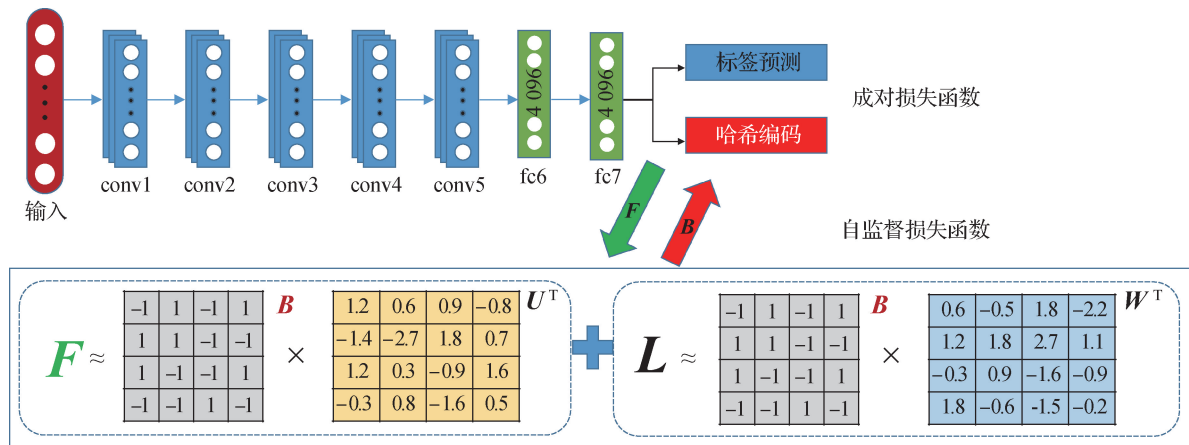


图1 自监督深度离散哈希网络框架

Fig. 1 The framework of self-supervised deep discrete hashing

1 深度成对监督哈希

在深度成对监督哈希 (DPSH) (Li 等, 2016) 中, 给定的哈希码矩阵 $B = \{b_i\}_{i=1}^n$, 其中 $b_i = \{-1, +1\}^{1 \times k}$ 。根据 LFH (Zhang 等, 2014), 定义成对标签相似性矩阵 $S = \{s_{ij}\}$ 的概率似然函数 p 为

$$p(s_{ij} | B) = \begin{cases} \sigma(\Omega_{ij}) & s_{ij} = 1 \\ 1 - \sigma(\Omega_{ij}) & s_{ij} = 0 \end{cases} \quad (1)$$

式中, $\Omega_{ij} = \frac{1}{2} b_i b_j^T$, $\sigma(\Omega_{ij}) = \frac{1}{1 + e^{-\Omega_{ij}}}$, $s_{ij} = 1$ 代表两幅图像相似, $s_{ij} = 0$ 代表两幅图像不相似。两幅图像的相似程度越高, 二者对应的哈希码之间的汉明距离越小, 此时成对相似性等于 1 的概率越大。

通过计算 S 中成对标签的负对数似然, DPSH 得出以下优化问题

$$\min_B J_1 = -\log p(S | B) = -\sum_{s_{ij} \in S} \log p(s_{ij} | B) = -\sum_{s_{ij} \in S} (s_{ij} \Omega_{ij} - \log(1 + e^{\Omega_{ij}})) \quad (2)$$

通过最小化上述优化问题, 可以使得两个相似点之间的汉明距离尽可能小, 同时使两个不相似点之间的汉明距离尽可能大。由于哈希码矩阵 B 是离散的, 导致问题式 (2) 优化困难。DPSH 采用松弛策略, 将问题式 (2) 改写为

$$\min_{B, H} J_2 = -\sum_{s_{ij} \in S} (s_{ij} \Theta_{ij} - \log(1 + e^{\Theta_{ij}})) + \eta \sum_{i=1}^n \|b_i - h_i\|_F^2 \quad (3)$$

式中, $H = \{h_i\}_{i=1}^n$ 表示深度哈希函数生成的连续哈希码, $\Theta_{ij} = \frac{1}{2} h_i h_j^T$, $\|\cdot\|_F$ 表示矩阵的 F 范数, η 为超参数。然而 DPSH 依旧存在二进制量化较大的问题。对此, Li 等人 (2017) 提出深度监督离散哈希 (deep supervised discrete hashing, DSDH), 通过引入标签矩阵, 使生成的哈希码对应所属的分类, 得到如下问题

$$\min_{B, H} J_3 = -\sum_{s_{ij} \in S} (s_{ij} \Theta_{ij} - \log(1 + e^{\Theta_{ij}})) + \mu \|L - BW^T\|_F^2 + v \|W\|_F^2 \quad (4)$$

式中, L 为标签矩阵, μ 和 v 为超参数。对于问题式 (4), DSDH 采用离散循环坐标下降算法 (discrete cyclic coordinate decent, DCCD) (Shen 等, 2015) 进行优化求解, 有效减少了二进制量化误差。

2 本文方法

2.1 自监督损失函数

大多数深度哈希方法直接利用卷积神经网络生成连续的哈希码, 没有考虑利用卷积神经网络提取的图像特征可以生成离散哈希码。本文通过深度特征矩阵和图像标签矩阵直接生成二进制哈希码并作为网络的自监督信息, 指导深度哈希函数的训练。利用矩阵分解, 得到 $F \approx BU^T$, 其中 F 表示卷积神经网络提取的深度特征矩阵, 该特征数据为图 1 中 fc7 层的输出。同时, 为了使生成的

哈希码 $\mathbf{B}$ 具有可分性, 图像的标签矩阵与哈希码矩阵应满足 $\mathbf{L} \approx \mathbf{B}\mathbf{W}^T$, 由此得到自监督损失函数 J_4 , 具体为

$$\min_{\mathbf{B}, \mathbf{U}, \mathbf{W}} J_4 = \|\mathbf{F} - \mathbf{B}\mathbf{U}^T\|_F^2 + \|\mathbf{L} - \mathbf{B}\mathbf{W}^T\|_F^2$$

$$\text{s. t. } \mathbf{B} \in \{-1, +1\}^{n \times k} \quad (5)$$

通过最小化式(5), 可以直接得到离散哈希码矩阵 $\mathbf{B}$ 的最优解, 作为自监督信息指导深度哈希函数的训练。

2.2 成对损失函数

为保证生成高质量的哈希码, 定义成对标签相似矩阵 $\mathbf{S}'$ 为

$$\mathbf{S}' = 2\mathbf{S} - \mathbf{1}$$

即 $s'_{ij} = 1$ 表示两幅图像相似, $s'_{ij} = -1$ 表示两幅图像不相似。对于一对哈希码 $\mathbf{b}_i$ 和 $\mathbf{b}_j$ 而言, 汉明距离 $\text{dist}(\mathbf{b}_i, \mathbf{b}_j)$ 与内积 $\langle \mathbf{b}_i, \mathbf{b}_j \rangle$ 存在良好的线性关系, 即 $\text{dist}(\mathbf{b}_i, \mathbf{b}_j) = \frac{1}{2}(k - \langle \mathbf{b}_i, \mathbf{b}_j \rangle)$ 。由于哈希码距离可以反映图像的相似关系, 因此可得 $ks'_{ij} \approx \mathbf{b}_i \mathbf{b}_j^T$ 。通过最小化成对标签相似矩阵 $\mathbf{S}$ 的负对数似然, 可以保持连续哈希码之间的相似性。利用成对标签相似矩阵 $\mathbf{S}'$ 保持连续哈希码和离散哈希码之间的相似性, 得到成对损失函数 J_5 , 具体为

$$\min_{\mathbf{B}, \mathbf{H}} J_5 = - \sum_{s'_{ij} \in \mathbf{S}'} (s'_{ij} \Theta_{ij} - \log(1 + e^{\Theta_{ij}})) +$$

$$\sum_{s'_{ij} \in \mathbf{S}'} (\mathbf{h}_i \mathbf{b}_j^T - ks'_{ij})^2 + \alpha \|\mathbf{B} - \mathbf{H}\|_F^2 +$$

$$\beta \|\mathbf{H}\mathbf{1}\|_F^2 + \gamma \|\hat{\mathbf{L}} - \mathbf{L}\|_F^2$$

$$\text{s. t. } \mathbf{B} \in \{-1, +1\}^{n \times k}$$

$$\mathbf{H} \in [-1, +1]^{n \times k} \quad (6)$$

式中, $\hat{\mathbf{L}}$ 为预测的标签矩阵, α, β 和 γ 为设置的超参数。

显然, 最小化式(6)中的前两项, 可以使相似图像的哈希码的距离更小, 而不相似图像的距离更大。其中 $\sum_{s'_{ij} \in \mathbf{S}'} (\mathbf{h}_i \mathbf{b}_j^T - ks'_{ij})^2$ 保持了离散哈希码与连续哈希码之间的相似性。通过优化式(6)的第3项 $\alpha \|\mathbf{B} - \mathbf{H}\|_F^2$, 可以减小二进制哈希码 $\mathbf{B}$ 与连续哈希码 $\mathbf{H}$ 的量化误差。根据信息论中熵越大, 可以携带的信息越多的理论, 式(6)中的第4项 $\beta \|\mathbf{H}\mathbf{1}\|_F^2$ 用来平衡二进制码的均匀分布以增加信息容量。当每一位的值等于 +1 或 -1 的概率为 50% 时, 即可携带更多的信息, 因此, 哈希码中所有位的平均值要尽可能接近 0。此外, 在 SSDDH 网络的最后一层中, 除

了 k 个用于哈希编码的结点外, 添加了 c 个用于预测图像标签的结点。通过最小化式(6)中第5项 $\gamma \|\hat{\mathbf{L}} - \mathbf{L}\|_F^2$, 确保学习到更准确的图像特征表示矩阵 $\mathbf{F}$ 。

2.3 目标损失函数以及算法求解

如上所述, SSDDH 的总体目标函数定义为

$$\min_{\theta, \mathbf{B}, \mathbf{U}, \mathbf{W}} J = J_4 + J_5 =$$

$$\|\mathbf{F} - \mathbf{B}\mathbf{U}^T\|_F^2 + \|\mathbf{L} - \mathbf{B}\mathbf{W}^T\|_F^2 -$$

$$\sum_{s'_{ij} \in \mathbf{S}'} (s'_{ij} \Theta_{ij} - \log(1 + e^{\Theta_{ij}})) +$$

$$\|\mathbf{H}\mathbf{B}^T - k\mathbf{S}'\|_F^2 + \alpha \|\mathbf{B} - \mathbf{H}\|_F^2 +$$

$$\beta \|\mathbf{H}\mathbf{1}\|_F^2 + \gamma \|\hat{\mathbf{L}} - \mathbf{L}\|_F^2$$

$$\text{s. t. } \mathbf{B} \in \{-1, +1\}^{n \times k} \quad (7)$$

式中, θ 为卷积神经网络中的参数。通过优化式(7), 可以生成任何图像的哈希码。但是由于哈希码矩阵 $\mathbf{B}$ 是离散的, 导致式(7)是非凸的, 求解很困难。对此, 通过采用迭代优化算法, 每次学习一个参数并固定其他参数, 同时采用离散算法求解矩阵 $\mathbf{B}$ 。模型的迭代优化步骤如下:

1) 更新参数 θ , 固定其他参数的值不变, 则式(7)可以改写为

$$\min_{\theta} J = - \sum_{s'_{ij} \in \mathbf{S}'} (s'_{ij} \Theta_{ij} - \log(1 + e^{\Theta_{ij}})) +$$

$$\|\mathbf{H}\mathbf{B}^T - k\mathbf{S}'\|_F^2 + \alpha \|\mathbf{B} - \mathbf{H}\|_F^2 +$$

$$\beta \|\mathbf{H}\mathbf{1}\|_F^2 + \gamma \|\hat{\mathbf{L}} - \mathbf{L}\|_F^2 \quad (8)$$

式(8)可以利用随机梯度下降法 (stochastic gradient descent, SGD) 通过反向传播对参数 θ 进行更新。

2) 更新参数 $\mathbf{U}$, 固定其他参数的值不变, 则式(7)可以改写为

$$\min_{\mathbf{U}} J = \|\mathbf{F} - \mathbf{B}\mathbf{U}^T\|_F^2 \quad (9)$$

通过式(9)计算关于 $\mathbf{U}$ 的导数, 当其导数等于 0 时, 可得 $\mathbf{U}$ 的最优解为

$$\mathbf{U} = \mathbf{F}^T \mathbf{B} (\mathbf{B}^T \mathbf{B})^{-1} \quad (10)$$

3) 更新参数 $\mathbf{W}$, 固定其他参数的值不变, 则式(7)可以改写为

$$\min_{\mathbf{W}} J = \|\mathbf{L} - \mathbf{B}\mathbf{W}^T\|_F^2 \quad (11)$$

通过式(11)计算关于 $\mathbf{W}$ 的导数, 并令其等于 0, 可得 $\mathbf{W}$ 的最优解为

$$\mathbf{W} = \mathbf{L}^T \mathbf{B} (\mathbf{B}^T \mathbf{B})^{-1} \quad (12)$$

4) 更新参数 $\mathbf{B}$, 固定 $\theta, \mathbf{U}$ 和 $\mathbf{W}$ 的值, 则式(7)可以改写为

$$\begin{aligned} \min_B J = & \|F - BU^T\|_F^2 + \|L - BW^T\|_F^2 + \\ & \|HB^T - kS'\|_F^2 + \alpha \|B - H\|_F^2 \\ \text{s. t. } & B \in \{-1, +1\}^{n \times k} \end{aligned} \quad (13)$$

对于式(13),引入离散循环坐标下降算法(DC-CD)(Shen等,2015),则有

$$\begin{aligned} \|F - BU^T\|_F^2 &= \text{tr}(BU^TUB^T) + \text{tr}(BQ_1^T) + c \\ \|L - BW^T\|_F^2 &= \text{tr}(BW^TWB^T) + \text{tr}(BQ_2^T) + c \\ \|HB^T - kS'\|_F^2 &= \text{tr}(BH^THB^T) + \text{tr}(BQ_3^T) + c \\ \alpha \|B - H\|_F^2 &= \text{tr}(BQ_4^T) + c \end{aligned}$$

式中, $Q_1 = -2FU$, $Q_2 = -2LW$, $Q_3 = -2k(S')^TH$, $Q_4 = -2\alpha H$, c 是与 B 无关的常数。因此,式(13)可以改写为

$$\begin{aligned} \min_B J = & \text{tr}(BU^TUB^T) + \text{tr}(BW^TWB^T) + \\ & \text{tr}(BQ_1) + \text{tr}(BQ_2) + \\ & \text{tr}(BQ_3) + \text{tr}(BQ_4) + c \\ \text{s. t. } & B \in \{-1, +1\}^{n \times k} \end{aligned} \quad (14)$$

为求解式(14),分别用 $B_{*r}, U_{*r}, W_{*r}, H_{*r}$ 表示 B, U, W, H 的第 r 列, $(Q_1)_{*r}, (Q_2)_{*r}, (Q_3)_{*r}, (Q_4)_{*r}$ 表示 Q_1, Q_2, Q_3, Q_4 的第 r 列, $\hat{B}$ 表示缺少 B_{*r} 的 B , $\hat{U}$ 表示缺少 U_{*r} 的 U , $\hat{W}$ 表示缺少 W_{*r} 的 W , $\hat{H}$ 表示缺少 H_{*r} 的 H 。

为求解 B_{*r} ,将式(14)改写为

$$\begin{aligned} \min_{B_{*r}} J(B_{*r}) = & \text{tr}(B_{*r}[2U_{*r}^T\hat{U}\hat{B}_r^T + 2W_{*r}^T\hat{W}\hat{B}_r^T + \\ & 2H_{*r}^T\hat{H}\hat{B}_r^T + (Q_1)_{*r}^T + (Q_2)_{*r}^T + \\ & (Q_3)_{*r}^T + (Q_4)_{*r}^T]) \\ \text{s. t. } & B_{*r} \in \{-1, +1\}^n \end{aligned} \quad (15)$$

因此,可得式(15)的最优解为

$$B_{*r} = -\text{sign}(2\hat{B}_r[\hat{U}_{*r}^TU_{*r} + \hat{W}_{*r}^TW_{*r} + \hat{H}_{*r}^TH_{*r}] + (Q_1)_{*r} + (Q_2)_{*r} + (Q_3)_{*r} + (Q_4)_{*r}) \quad (16)$$

综上所述,本文提出的 SSDDH 算法的具体步骤如下:

输入:图像数据 X 、标签矩阵 L 、相似性矩阵 S 和 S' 。

输出:卷积神经网络参数 θ 和哈希码矩阵 B 。

初始化:参数 θ, U, W, B 。批处理大小为 $bs=64$,单次迭代更新中图像网络迭代次数 $t=5$,总迭代次数 $T=180$ 。

for $iter = 1, 2, \dots, T$ do;

for $i = 1, 2, \dots, t$ do;

随机从 X 中挑选 bs 个图像组成一个批处

理数据输入网络训练并更新参数 θ ;

end for;

利用式(10)更新参数 U ;

利用式(12)更新参数 W ;

for $r = 1 \rightarrow k$ do;

利用式(16)更新 B_{*r} ;

end for;

end for。

2.4 样本外扩展

当训练过程结束得到深度哈希函数后,只能得到训练数据中图像的哈希码,仍然需要执行样本外扩展来预测未出现在训练集中图像的哈希码。对于任何图像 x_q ,SSDDH 可以预测其哈希码 b_q ,具体公式为

$$b_q = \text{sign}(F(x_q; \theta)) \quad (17)$$

式中, F 表示图 1 中的 SSDDH 的图像网络。

3 实验及结果

3.1 数据集与评价指标

实验在公开数据集 CIFAR-10 (Krizhevsky, 2009)、NUS-WIDE (web image dataset from National University of Singapore) (Chua 等, 2009) 和 Flickr (Huiskes 和 Lew, 2008) 上进行,对本文提出的方法进行评估。

CIFAR-10 包含 60 000 幅 32×32 像素的彩色图像,共 10 类,每幅图像属于其中的一类。NUS-WIDE 包含近 270 000 幅彩色图像,共 81 类,每幅图像至少属于其中一类,遵循 Lai 等人 (2015) 的实验设置,实验只使用其中与最常见的 21 类相关的图像数据。Flickr 包含从 Flickr 网站搜集的 20 015 幅图像,共 24 类,每幅图像至少属于其中一类语义标签。

使用平均准确率均值 (mean average precision, mAP) 作为评价指标。对于一个给定的查询图像 q ,首先计算其 AP (average precision),具体为

$$AP(q) = \frac{1}{Q} \sum_{r=1}^N P_q(r) \text{rel}(r) \quad (18)$$

式中, Q 表示数据库中与图像 q 相似的个数, N 是数据总数, $P_q(r)$ 表示前 r 个返回图像的准确率。当第 r 个返回图像与图像 q 相似时,则 $\text{rel}(r) = 1$, 否则 $\text{rel}(r) = 0$ 。计算所有查询图像的 AP 的平均值即可得到 mAP。

3.2 实验环境设置

实验环境为 Ubuntu18.04, CPU 为 E5-2670, 内存 64 GB, 显卡型号为 1080Ti 11 GB。遵循 DPSH 中的实验设置, 将数据集划分为训练集、测试集以及查询数据库。对于 CIFAR-10 数据集, 随机选择 5 000 个数据作为训练集, 1 000 个数据作为测试集, 剩余数据作为查询数据库。对于 NUS-WIDE, 在每一类中随机选取 500 个数据, 共计 10 500 个数据作为训练集, 随机选择 2 100 个数据作为查询数据, 剩余数据作为查询数据集。对于 Flickr, 随机选取 10 000 个数据作为训练集, 2 000 个数据作为测试集, 剩余数据作为查询数据库。

考虑到 AlexNet (Krizhevsky 等, 2012) 具有良好的特征表达能力, 将其作为基础网络结构, 如图 1 所示。同时, 利用一个包含 $m = k + c$ 个隐藏结点的全连接层替换 AlexNet 网络的最后一个全连接层, 其中 k 个结点用于哈希编码, c 个结点用于标签预测。除最后一层外, 其他层均采用 ReLU (rectified linear units) 作为激活函数。编码模块采用 tanh 作为激活函数, 标签预测模块采用 sigmoid 作为激活函数。

在实验开始时, 利用 ImageNet (Deng 等, 2010) 数据集上的预训练网络 CNN-F (Chatfield 等, 2014) 对 SSDDH 网络进行初始化。超参数 α , β 和 γ 均设置为 50, 学习率设置为 $10^{-5} \sim 10^{-9}$, 批处理大小设置为 64, 迭代次数设置为 180。在每一次迭代中设置更新网络参数 5 次, 再依次更新 U , W 和 B 。使用 pytorch 构造卷积神经网络, 每个实验结果为 3 次

实验的平均结果。

3.3 对比实验与分析

为了验证本文方法在图像检索中的性能, 选择 SH、ITQ、SDH、SH、ITQ、SDH、DSH、DHN、HashGAN 和 DPSH 等哈希方法分别在哈希码长度为 12 bit、24 bit、32 bit 和 48 bit 上进行对比实验, 结果如表 1 所示。其中 SH、ITQ 和 SDH 属于非深度哈希方法, SH 和 ITQ 为无监督哈希方法, SDH 为监督哈希方法, DSH、DHN、HashGAN 和 DPSH 属于深度哈希方法。

从表 1 可以看出, 监督哈希方法的精度普遍高于无监督哈希方法, 深度哈希方法的性能比非深度哈希方法更好。在数据集 CIFAR-10 上, SSDDH 的 mAP 有明显提升。与无监督方法 SH 和 ITQ 相比, 取得了近 60% 的增长。与未采用深度学习的监督哈希 SDH 相比, 取得了超过 40% 的增长。与深度哈希方法相比, 较 DSH 和 DHN 取得了超过 10% 的 mAP 增长, 较 HashGAN 和 DPSH, 在哈希码长度为 12 bit 和 24 bit 时性能提升明显, 在哈希码长度为 32 bit 和 48 bit 时增幅较小, 有 2% 的增长。在数据集 NUS-WIDE 上, SSDDH 在不同哈希码长度情况下均获得了最佳的检索性能。与 SH 和 ITQ 相比, 取得了超过 30% 的性能提升。与 SDH、DSH、DHN 和 Hash GAN 相比, 取得了 10%~20% 的增长。与 DPSH 相比, 增幅同样较小。在数据集 Flickr 上, SSDDH 相比其他方法同样取得了较好的性能提升。

表 1 不同哈希算法在 3 个数据集上的 mAP 结果
Table 1 mAPs results of different hashing methods on three datasets

方法	CIFAR-10				NUS-WIDE				Flickr			
	12 bit	24 bit	32 bit	48 bit	12 bit	24 bit	32 bit	48 bit	12 bit	24 bit	32 bit	48 bit
DPSH (Li 等, 2016)	0.713	0.727	0.744	0.757	0.752	0.790	0.794	0.812	0.837	0.844	0.856	0.863
HashGAN (Cao 等, 2018)	0.660	0.722	0.731	0.735	0.642	0.699	0.737	0.751	0.790	0.826	0.841	0.848
DHN (Zhu 等, 2016)	0.555	0.594	0.603	0.621	0.708	0.735	0.748	0.758	0.823	0.839	0.844	0.847
DSH (Jin 等, 2014)	0.616	0.651	0.66	0.675	0.548	0.551	0.558	0.562	0.805	0.835	0.842	0.845
SDH (Shen 等, 2015)	0.285	0.329	0.341	0.356	0.568	0.600	0.608	0.637	Nan	Nan	Nan	Nan
ITQ (Gong 等, 2013)	0.162	0.169	0.172	0.175	0.452	0.468	0.472	0.477	0.684	0.695	0.697	0.697
SH (Weiss 等, 2008)	0.127	0.128	0.126	0.129	0.454	0.406	0.405	0.400	0.645	0.651	0.65	0.646
本文	0.749	0.762	0.774	0.785	0.787	0.811	0.816	0.823	0.840	0.852	0.862	0.871

注: 加粗字体为各列最优结果, Nan 表示该方法在数据集上没有实验数据。

在查询结果方面,SSDDH 与 DPSH 在 CIFAR-10 测试数据集(哈希码长度为 48 bit)上的 PR (precision-recall) 曲线及二者对飞机、鸟、鹿、蛙和船类图像查询返回的前 20 个图像准确率对比如图 2 和图 3 所示,图 3 中红色边框表示返回的错误类别图像。由图 2 和图 3 可知,SSDDH 的 PR 曲线始终处于 DPSH 的上方。DPSH 和 SSDDH 在飞机和鸟类返回结果中均未出现错误类别图像,但 DPSH 在鹿、蛙和船类返回结果中分别出现了 6 幅、1 幅和 3 幅错误类别图像,而 SSDDH 仅在鹿和船类各返回 1 幅错误类别图像,由此可见 SSDDH 方法具有较好的检索性能。

为探究 α 、 β 和 γ 对 SSDDH 的性能影响,对 α 、 β 和 γ 进行参数灵敏度分析。实验在 CIFAR-10 上进行,哈希码长度为 48 bit,通过固定 2 个超参数,调整剩下的 1 个超参数,记录当前 mAP 值,结果如图 4

所示。可以看出,当 α 、 β 和 γ 分别在 $[10, 100]$ 、 $[1, 100]$ 和 $[0.1, 1\ 000]$ 取值时,SSDDH 对参数值变化不敏感,表明 SSDDH 具有较好的鲁棒性。

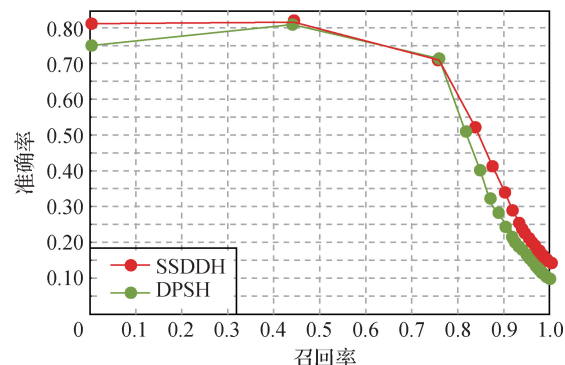


图2 DPSH 和 SSDDH 在 CIFAR-10 上的 PR 曲线
Fig. 2 The precision-recall curves of DPSH and SSDDH on CIFAR-10



图3 DPSH 和 SSDDH 的返回结果对比

Fig. 3 Comparison of returned results between DPSH and SSDDH ((a) DPSH; (b) SSDDH)

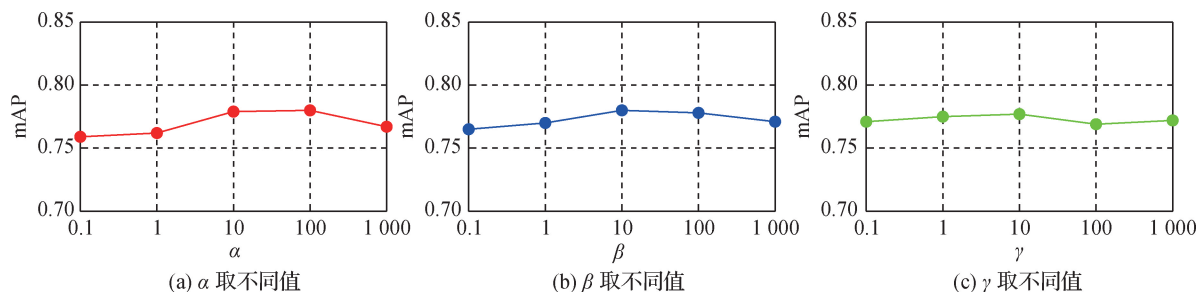


图4 SSDDH 在 α 、 β 和 γ 取不同值时的 mAP 值

Fig. 4 The mAP values of SSDDH when α , β and γ takes different values
((a) α takes different values; (b) β takes different values; (c) γ takes different values)

此外,本文通过对比 DPSH 和 SSDDH 在训练过程中 mAP 值的变化,探究二者的差异。由于 DPSH 和 SSDDH 采用成对训练方式,因此二者的时间复杂度为 $O(n^2)$ 。实验在 CIFAR-10 上进行,哈希码长度为 48 bit。为保证实验的公平性,设置 SSDDH 单次迭代更新的图像网络迭代次数为 1。图 5 为 DPSH 和 SSDDH 前 30 次迭代的 mAP 对比。可以看出,在训练初期,由于加入自监督损失和离散哈希约束项,SSDDH 的优化难度大于 DPSH,因此 mAP 低于 DPSH。在更新 10 次后,得益于自监督损失和离散哈希约束项的加入,SSDDH 的 mAP 值高于 DPSH。

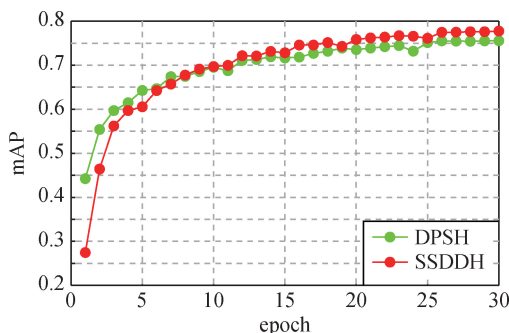


图5 DPSH 和 SSDDH 在前 30 次迭代中的 mAP 对比

Fig.5 Comparison of mAPs in the top-30 epochs between DPSH and SSDDH

4 结 论

考虑到现有深度哈希方法忽略了深度图像特征信息在深度哈希函数学习中的指导作用,且存在二进制量化误差较大问题,本文提出自监督的深度离散哈希 SSDDH,旨在通过利用卷积神经网络提取的深度特征矩阵和图像标签矩阵计算得出二进制哈希码,并将其作为监督信息指导训练深度哈希函数。通过构建成对损失函数,在保持连续哈希码之间相似性的同时,保持了连续哈希码与二进制哈希码之间的相似性。利用循环坐标下降法,离散求解二进制码,显著减小了二进制量化误差。通过与现有方法在 3 个常用公开数据集上的对比分析,验证了本文方法比现有哈希检索方法更高效。

但是,由于本文方法是基于图像数据,因此不能适用所有类型的数据。下一步的研究目标是将目前方法扩展到跨模态检索。如给定一段文本数据,检

索得到与之语义相似的图像、音频或者视频等数据,利用多种模态之间的复杂相关性提高检索性能。

参考文献 (References)

- Andoni A and Indyk P. 2006. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions//The 47th Annual IEEE Symposium on Foundations of Computer Science. Berkeley, UK; IEEE: 459-468 [DOI: 10.1109/FOCS.2006.49]
- Cao Y, Liu B, Long M S and Wang J M. 2018. HashGAN: deep learning to hash with pair conditional Wasserstein GAN//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA; IEEE: 1287-1296 [DOI: 10.1109/CVPR.2018.00140]
- Chatfield K, Simonyan K, Vedaldi A and Zisserman A. 2014. Return of the devil in the details: delving deep into convolutional nets//Proceedings of the British Machine Vision Conference. Nottingham, UK; BMVA Press: 1-12 [DOI: 10.5244/C.28.6]
- Chua T S, Tang J H, Hong R C, Li H J, Luo Z P and Zheng Y T. 2009. NUS-WIDE: a real-world web image database from national university of Singapore//Proceedings of the ACM International Conference on Image and Video Retrieval. Santorini, Greece; ACM: 1-9 [DOI: 10.1145/1646396.1646452]
- Deng J, Dong W, Socher R, Li L J, Li K and Li F F. 2010. ImageNet: a large-scale hierarchical image database//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA; IEEE: 248-255 [DOI: 10.1109/cvpr.2009.5206848]
- Gong Y C, Lazebnik S, Gordo A and Perronnin F. 2013. Iterative quantization: a procrustean approach to learning binary codes for large-scale image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35 (12): 2916-2929 [DOI: 10.1109/TPAMI.2012.193]
- Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y. 2014. Generative adversarial nets//Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Canada; MIT Press: 2672-2680
- He K M, Gkioxari G, Dollár P and Girshick R. 2020. Mask R-CNN. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(2): 386-397 [DOI: 10.1109/TPAMI.2018.2844175]
- Huiskes M J and Lew M S. 2008. The MIR Flickr retrieval evaluation//Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval. Vancouver, Canada; ACM: 39-43 [DOI: 10.1145/1460096]
- Jiang Q Y and Li W J. 2017. Deep cross-modal hashing//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA; IEEE: 3270-3278 [DOI: 10.1109/CVPR.2017.

- 348]
- Jiang Q Y, Cui X and Li W J. 2018. Deep discrete supervised hashing. *IEEE Transactions on Image Processing*, 27 (12): 5996-6009 [DOI: 10.1109/TIP.2018.2864894]
- Jin Z M, Li C, Lin Y and Cai D. 2014. Density sensitive hashing. *IEEE Transactions on Cybernetics*, 44(8): 1362-1371 [DOI: 10.1109/TCYB.2013.2283497]
- Krizhevsky A. 2009. Learning Multiple Layers of Features from Tiny Images. Technical Report TR-2009. University of Toronto
- Krizhevsky A, Sutskever I and Hinton G E. 2012. ImageNet classification with deep convolutional neural networks//*Proceedings of the 26th International Conference on Neural Information Processing Systems*. Montreal, Canada; MIT Press: 1097-1105
- Kulis B and Grauman K. 2009. Kernelized locality-sensitive hashing for scalable image search//*Proceedings of the 12th IEEE International Conference on Computer Vision*. Kyoto, Japan; IEEE: 2130-2137 [DOI: 10.1109/ICCV.2009.5459466]
- Lai H J, Pan Y, Liu Y and Yan S C. 2015. Simultaneous feature learning and hash coding with deep neural networks//*Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston, USA; IEEE: 3270-3278 [DOI: 10.1109/CVPR.2015.7298947]
- Li Q, Sun Z N, He R and Tan T N. 2017. Deep supervised discrete hashing//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach, USA; Curran Associates Inc.: 2482-2491
- Li W J, Wang S and Kang W C. 2016. Feature learning based deep supervised hashing with pairwise labels//*Proceedings of the 25th International Joint Conference on Artificial Intelligence*. New York, USA; AAAI: 1711-1717
- Liu H M, Wang R P, Shan S G and Chen X L. 2016a. Deep supervised hashing for fast image retrieval//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA; IEEE: 2064-2072 [DOI: 10.1109/CVPR.2016.227]
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y and Berg A C. 2016b. SSD: single shot MultiBox detector//*Proceedings of the 14th European Conference on Computer Vision*. Amsterdam, the Netherlands; Springer: 398-413 [DOI: 10.1007/978-3-319-46448-0_2]
- Liu W, Mu C, Kumar S and Chang S F. 2014. Discrete graph hashing//*Proceedings of the 27th International Conference on Neural Information Processing Systems*. Montreal, Canada; MIT Press: 3419-3427
- Liu W, Wang J, Ji R R, Jiang Y G and Chang S F. 2012. Supervised hashing with kernels//*Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition*. Providence, USA; IEEE: 2074-2081 [DOI: 10.1109/CVPR.2012.6247912]
- Liu Y, Pan Y, Xia R K, Liu D and Yin J. 2016. FP-CNN: a fast image hashing algorithm based on deep convolutional neural network. *Computer Science*, 43(9): 39-46, 51 (刘冶, 潘炎, 夏榕楷, 刘荻, 印鉴. 2016. FP-CNN: 一种基于深度卷积神经网络的快速图像哈希算法. *计算机科学*, 43(9): 39-46, 51) [DOI: 10.11896/j.issn.1002-137X.2016.9.007]
- Lu X Q, Chen Y X and Li X L. 2020. Siamese dilated inception hashing with intra-group correlation enhancement for image retrieval. *IEEE Transactions on Neural Networks and Learning Systems*, 31(8): 3032-3046 [DOI: 10.1109/TNNLS.2019.2935118]
- Noh H, Hong S and Han B. 2015. Learning deconvolution network for semantic segmentation//*Proceedings of 2015 IEEE International Conference on Computer Vision*. Santiago, Chile; IEEE: 1520-1528 [DOI: 10.1109/ICCV.2015.178]
- Ren S Q, He K M, Girshick R and Sun J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6): 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]
- Shen F M, Shen C H, Liu W and Shen H T. 2015. Supervised discrete hashing//*Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston, USA; IEEE: 37-45 [DOI: 10.1109/CVPR.2015.7298598]
- Shen Y M, Feng Y, Fang B, Zhou M L, Kwong S and Qiang B H. 2020. DSRPH: deep semantic-aware ranking preserving hashing for efficient multi-label image retrieval. *Information Sciences*, 539: 145-156 [DOI: 10.1016/j.ins.2020.05.114]
- Simonyan K and Zisserman A. 2015. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2020-05-17]. <http://arxiv.org/pdf/1409.1556.pdf>
- Wang J, Kumar S and Chang S F. 2010. Sequential projection learning for hashing with compact codes//*Proceedings of the 27th International Conference on International Conference on Machine Learning*. Haifa, Israel; Omnipress: 1127-1134
- Wang Z M and Zhang H. 2019. A fast image retrieval method based on multi-layer CNN features. *Journal of Computer-Aided Design and Computer Graphics*, 31(8): 1410-1416 (王志明, 张航. 2019. 融合多层卷积神经网络特征的快速图像检索方法. *计算机辅助设计与图形学学报*, 31(8): 1410-1416) [DOI: 10.3724/SP.J.1089.2019.17845]
- Weiss Y, Torralba A and Fergus R. 2008. Spectral hashing//*Proceedings of the 21st International Conference on Neural Information Processing Systems*. Vancouver, Canada; Curran Associates Inc.: 1753-1760
- Xia R K, Pan Y, Lai H J, Liu C and Yan S C. 2014. Supervised hashing for image retrieval via image representation learning//*Proceedings of the 28th AAAI Conference on Artificial Intelligence*. Québec City, Canada; AAAI: 2156-2162
- Yan C G, Xie H T, Yang D B, Yin J, Zhang Y D and Dai Q H. 2018. Supervised hash coding with deep neural network for environment perception of intelligent vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 19(1): 284-295 [DOI: 10.1109/TITS.2017.2749965]

Zhang P C, Zhang W, Li W J and Guo M Y. 2014. Supervised hashing with latent factor models//Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval. Gold Coast, Australia; ACM; 173-182 [DOI: 10.1145/2600428.2609600]

Zhu H, Long M S, Wang J M and Cao Y. 2016. Deep hashing network for efficient similarity retrieval//Proceedings of the 30th AAAI Conference on Artificial Intelligence. Phoenix, USA; AAAI; 2415-2421

作者简介



万方,男,讲师,硕士生导师,主要研究方向为计算机视觉和图像检索。

E-mail: 20021026@hbut.edu.cn

强浩鹏,男,硕士研究生,主要研究方向为深度学习和跨模态检索。E-mail: qianghaopengshuxuejd@whut.edu.cn

雷光波,女,讲师,主要研究方向为视觉 SLAM 和深度学习。

E-mail: 64694786@qq.com