

• 信息科学 •

# 基于弱监督深度学习的图像检索技术研究

胡琦瑶<sup>1</sup> 杨皓文<sup>1</sup> 王佳欣<sup>1</sup> 彭先霖<sup>2</sup> 汪思嘉<sup>1</sup>

(1. 西北大学 信息科学与技术学院, 陕西 西安 710127; 2. 西北大学 艺术学院, 陕西 西安 710127)

**摘要:** 哈希算法结合卷积神经网络是一种常用的有效图像检索方法, 这种有监督的学习方式要求大量人工标注的标签, 耗费巨额的人力成本, 且这种标签具有较强的专家制定性, 在图像深层次语义的挖掘上能力有限。针对这一问题, 该文提出了基于弱监督(Weakly-supervised)深度卷积神经网络的哈希函数框架, 该框架包括弱监督的预训练和有监督的微调两个阶段。通过利用弱监督标签信息来学习语义感知哈希函数, 将标签表示为语义词向量。根据语义间的相似与否来训练网络, 设计的目标函数可以使语义相似的图像间的汉明距离较小, 语义不相似的汉明距离相对较远。实验结果表明, 该文提出的图像检索方法与目前先进的图像检索方法相比, 精度提高了2%~6%。

**关键词:** 特征提取; 弱监督; 深度学习; 图像检索

**中图分类号:** TP391.4

**DOI:** 10.16152/j.cnki.xdxbr.2020-05-013 开放科学(资源服务)标识码(OSID):



## Research on image retrieval based on weakly-supervised deep learning

HU Qiyao<sup>1</sup>, YANG Haowen<sup>1</sup>, WANG Jiaxin<sup>1</sup>, PENG Xianlin<sup>2</sup>, WANG Sijia<sup>1</sup>

(1. School of Information Science and Technology, Northwest University, Xi'an 710127, China;

2. School of Art, Northwest University, Xi'an 710127, China)

**Abstract:** Hashing algorithm combined with convolutional neural network has become a popular and effective image retrieval method. However, this supervised learning method requires a large number of user-tagged labels, which consumes a large amount of labor costs. Unfortunately, these labels are set by experts formulately and have limited ability to mine deep semantics of images. To solve the problem, a novel hashing learning framework based on deep convolution neural network is proposed, which consists of two stages: weakly supervised pre-training and supervised fine-tuning. Semantic-aware hash functions are learned by leveraging the weakly supervised tag information, and labels are represented as semantic word vectors. And then, the network is trained according to the similarity between semantics. The objective function can make the Hamming distance between the semantic similar images small, while the Hamming distance between the semantic dissimilar images relatively far. Empirical results on real world datasets show that when it is integrated with state-of-the-art deep hashing methods, the performance increases by 2%~6%.

**Key words:** feature extraction; weakly supervised; deep learning; image retrieval

收稿日期: 2020-01-15

基金项目: 国家自然科学基金资助项目(61903300); 陕西省科技计划重点项目(2018ZDXM-GY-186)

作者简介: 胡琦瑶, 女, 陕西西安人, 从事图像处理与机器学习研究。

近年来,基于内容的图像搜索技术(content based image retrieval, CBIR) 较基于文本的图像检索技术(text based image retrieval, TBIR) 有了显著的突破<sup>[1]</sup>,同时深度学习的发展使得卷积神经网络可以更好地提取图像高层次特征<sup>[2]</sup>。国内互联网公司百度为提升图像检索的精度,研发了百度识图引擎<sup>[1]</sup>,这是一种利用深度学习进行图像检索的方法,通过卷积神经网络学习到高层的图像特征,进而实现图像检索的功能。但是,该方法有一些缺点:①采用有监督的学习方法要求大量有标注的训练数据,这需要耗费巨额的人力成本来进行标注工作;②人工标注的标签具有较强的专家制定性,其通常会受研究问题和方向的限制,在图像深层次内容的理解和图像间的潜在相似性关系的发掘上能力有限。



弱标签

show, rescue, museum,  
truck, fire, la, tv,  
firetruck, dodge, 51,  
squad, emergency

强标签

museum, truck, fire,  
tv, 51



portrait, people, house, japan,  
youth, photography, nikon,  
okinawa, d200, guest, iso1600,  
50mm18, hostile, mariogallucci

people, youth, okinawa, d200,  
guest, iso1600, 50mm18

图 1 弱标注标签与人工标签示意图

Fig. 1 Weakly-annotated label and manual label

由于这种用户标注的标签主观性强,且数据中有垃圾图像和相关性弱的标签<sup>[3]</sup>,这些都是含义不清晰的非视觉性语义标签。同时,没有语义结构的标签数据不能描述语义结构之间的复杂关系<sup>[4]</sup>,一些与对应图片的关联度不高的标签不能直接用于神经网络的训练。需要提前对标签进行过滤,除去与图像视觉内容不符的标签,只留下视觉性语义标签。

几种基于 CNN 的最新哈希方法<sup>[5-8]</sup>提出通过学习二进制图像哈希码实现图像检索。文献[5]提出了一种基于 CNN 的无监督散列方法 DeepBit,它学习了一组非线性投影函数来计算压缩二进制码。文献[7]介绍了深度监督哈希(DSH)方法,该方法将成对的相似图像和不相似

随着网络时代的迅速发展,微信、微博等社交媒体应用的日常化,越来越多的人在网络上发表大量的媒体信息,这些由用户提供的数据称为弱标签数据,它们可以很容易地获得,且适用于图像检索模型的训练。弱标签就是类似这种“大众分类”的语言,用户可以用相近的或模糊的任意词汇来进行标注。与一般分类算法所采用的强标签不同,它不是由专家制定的,其结构并不严谨,是一种无结构的语义标注,能够展现更广的大众意识。如图 1 显示的是 Flickr 数据集中的一些用户提供的标签和人工标注的标签,如“horse”,“person”,“feet”,“box”是人工标注的强标签,如“actor”,“blur”,“argentina”是用户标注的弱标签。用户标注的标签蕴含了更丰富、更深层的潜在语义信息,能够更好地表征图像间的相似关系<sup>[1]</sup>。

图像作为训练输入,并使每个图像的输出为二进制码。文献[8]提出另一种有监督的深度哈希方法,即深度正则相似性比较哈希方法(DRSCH),利用 CNN 和基于三元组的目标函数进行哈希函数学习。

文献[9]提出保留语义的有监督深度哈希算法(SSDH),将哈希函数构造为深度学习网络中的一个潜在层,通过最小化在分类错误上定义的目标函数来学习二进制码。文献[10]提出弱监督多模态哈希方法(WMH),使用标签进行哈希学习,是一种具有线性哈希函数的具体哈希方法,WMH 在文本模式上直接使用标签来计算图像的相似度,会容易受到噪声和模糊性问题的影响。文献[11]在快速 RCNN 中的区域提议网络

(RPN) 后增加一个完全连接层,以学习检测到的区域提议的二进制散列码。

本文提出了一种基于弱监督深度学习的图像检索方法,以大众用户产生的大量弱标签图像作为数据进行训练。该框架包括弱监督的预训练和有监督的微调两个阶段。首先,采用语义相似度函数对图像对的语义相似度进行判定。其次,由于图像内容的高层次语义分析不同于简单的分类问题<sup>[12]</sup>,本文抛弃简单的 softmax 损失函数,设计了一种能够反映图像语义距离的损失函数作为卷

积神经网络的最高输出层激活函数,这样可以通过弱标签间的语义关系学习图像间的相似性关系,以此来训练卷积神经网络。最后,由于卷积神经网络的输出是连续值,本文采用哈希编码将输出结果二值化。结合神经网络与哈希编码方法不但可以提取到图像的高层特征,还能将其映射到低维的二进制哈希码上,利用其对神经网络模型进行优化,可以有效地应用于图像检索功能。图2是本文的技术方案图。

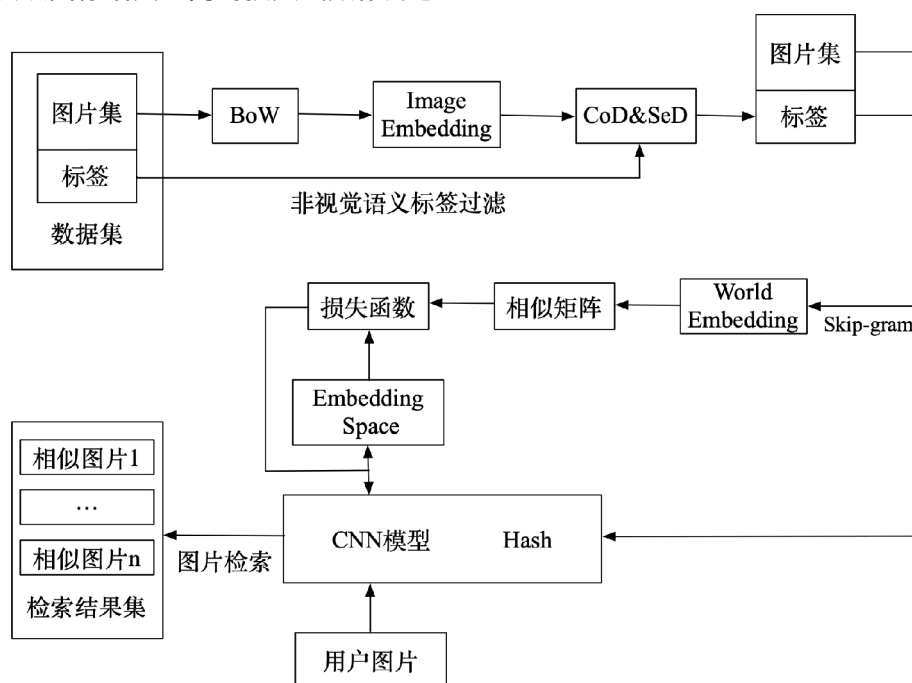


图2 技术方案图

Fig. 2 Technical plan

## 1 模型优化方法与模型结构

### 1.1 模型优化方法

图像的弱标签存在语义模糊、类别划分不清晰的问题,不能直接用于有监督的学习。本文提出弱监督的训练方式,将神经网络最后一层的输出看成一个嵌入空间,设计了一个损失函数,使相似弱标签的图像对的向量间距离更近,使不相似弱标签的图像对的向量间距离更远。与直接利用弱标签来训练模型相比,能够更好地利用神经网络学习到高层的特征,训练出更优的网络模型。

本文采用 Skip-gram 特征向量计算弱标签的相似度关系<sup>[13]</sup>。首先,图像的每个标签用 Skip-gram 语义特征向量来表示。采用基于图结构的

相似度计算方法 (SimRank), 语义相似度计算公式为

$$s(i, j) = \frac{1}{mn} \sum_{k=1}^m \sum_{l=1}^n \text{sim}(v_{ik}, v_{jl}) \quad (1)$$

式(1) 计算图像  $i$  和  $j$  间的语义相似度。它的基本思想是: 如果两个图像相似, 那么跟它们相关的图像标签应该也相似。 $v_{ik}$  表示图像  $i$  的第  $k$  个标签的语义特征向量, 类似的  $v_{jl}$  表示图像  $j$  的第  $l$  个标签的语义特征向量,  $m$  和  $n$  分别表示两张图像的标签总数量。 $\text{sim}(x, y)$  是一种相似度计算函数, 利用 Jaccard 系数来计算相似度, 值越接近于 1, 说明两个向量越相似。

$$\text{sim}(x, y) = \frac{xy}{\|x\|^2 + \|y\|^2 - xy} \quad (2)$$

利用式(1) 可以同时计算图像中所有标签的语义相似度, 与仅计算一部分标签相比, 在计算时

能够有效地减弱语义表述较弱的标签中的噪声带来的影响。同时也能很好地利用同义词、近义词距离相近的语义特征,对于相近图像在计算时会得到较高的相似度。

与常用的 softmax 函数进行分类目标不同<sup>[14]</sup>,本文设计了一种衡量图像弱标签间相似度关系的损失函数,这种基于弱监督学习的损失函数替代了简单的 softmax 函数作为神经网络的最高输出层。损失函数如下:

$$L = \sum_{(i,j) \in \Omega} \|f_i - f_j\|_2 + \sum_{(i,j) \in \Lambda} \max(0, \lambda - \|f_i - f_j\|_2) \quad (3)$$

式(3)中,用 2-范数来度量两个向量间的差异,

$$\|f_i - f_j\|_2 = \sqrt{\sum_{i=1}^n \sum_{j=1}^m (f_i - f_j)^2} \quad (4)$$

如果判断两两图像对为相似,则对应的  $(i, j) \in \Omega$ ,相反若判断两两图像对为不相似时,对应的  $(i, j) \in \Lambda$ 。 $\max(0, \sim)$  可以保持大于 0 的数不变,并将小于 0 的数归 0。 $\max$  函数中,不相似弱标签的图像对向量  $f_i$  与  $f_j$  之间的欧式距离与  $\lambda$  作差,只有当其大于  $\lambda$  时会得到负数,此时  $\max$  函数值为 0,就可以通过最小化损失函数来优化目标函数的正则化项,防止模型为了迎合训练集而过于复杂,造成过拟合的情况,从而提高模型的泛化能力。

实验首先计算了图像对的语义相似度,将其划分为两部分,分别代表相似的图像集与不相似的图像集。用  $(i, j, 1)$  来表示相似图集,用  $(i, j, 0)$  来表示不相似图集。再分别将这两部分输入到神经网络中,通过计算损失函数对应的这两部分来计算欧式距离和。那么减小判断为相似的两两图像对之间的欧式距离,同时增大判断为不相似的两两图像对间的欧式距离可以达到训练的目的,进而实现图像检索。即本文设计的损失函数可以达到拉近相似弱标签图像对的向量间的距离,增大不相似弱标签图像对的向量间的距离。实验中,损失函数的  $\lambda$  设为 1。图 3 是图片间的距离示意图。

本文设置了一个相似度阈值来判断两张图像相似与否。通过计算两张图的标签相似度<sup>[15]</sup>,可以得到  $[0, 1]$  范围的相似度指数,再按相似度的大小对两两图像对进行排序。采用随机的方法选取一个值作为相似度的阈值,通过肉眼观察其附

近的 100 组图像对,判断两张图是否相似,若无法判定,则再次选取一个新的阈值,直到能够直接区分出相似与否。本文实验选取的阈值为 0.4,即当相似度大于 0.4 时标记为  $(i, j, 1)$ ,小于 0.4 时标记为  $(i, j, 0)$ 。如图 4 中的 4 组图片,其中,前两张代表判定为相似的图片,第 3 张判定与前两张不相似。

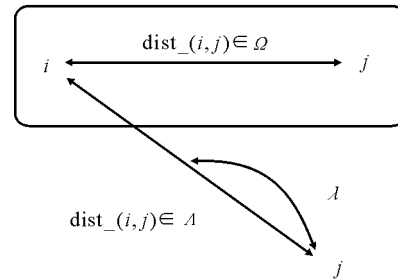


图 3 相似图片与不相似图片的距离示意图

Fig. 3 The distance between similar pictures and dissimilar images

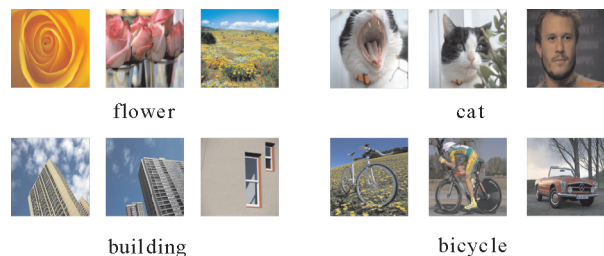


图 4 相似图片与不相似图片的对比图

Fig. 4 Comparison with similar images and dissimilar images

## 1.2 模型网络结构

弱监督哈希深度学习模型框架框架包括两个阶段:弱监督的预训练和有监督的微调,如图 5 所示。

神经网络层数越深,能够表达的函数就越复杂,但是,训练的结果更容易过拟合,且神经网络结构越复杂,训练的时间会越大<sup>[16]</sup>。

在弱监督的预训练阶段,最后一层的全连接层使用 sigmoid 激活函数将特征向量映射到  $[0, 1]$  的连续值,为哈希映射提供条件。并在最后一层全连接层分别输出 8, 16, 32, 64 个神经单元<sup>[17]</sup>,通过映射可以得到对应位数的哈希码。

本文将图像的特征向量哈希化可以有效提高图像检索的速度,卷积神经网络的训练得到的是连续的实数值,而哈希函数需要将输出表示为二进制码。首先,提取全连接层图像  $j$  的输出变量  $Y_j$ ,然后,设定一个阈值,将连续的实数输出二值化为二进制代码<sup>[18]</sup>。最后,对网络全连接层的输

出哈希化,参考 Kevin Lin 等人<sup>[19]</sup>提出的哈希化方法,通过式(5)可以得到  $Y_j$  的二进制代码输出:

$$H_j = \begin{cases} 1, & Y_j \geq 0.5 \\ 0, & Y_j < 0.5 \end{cases} \quad (5)$$

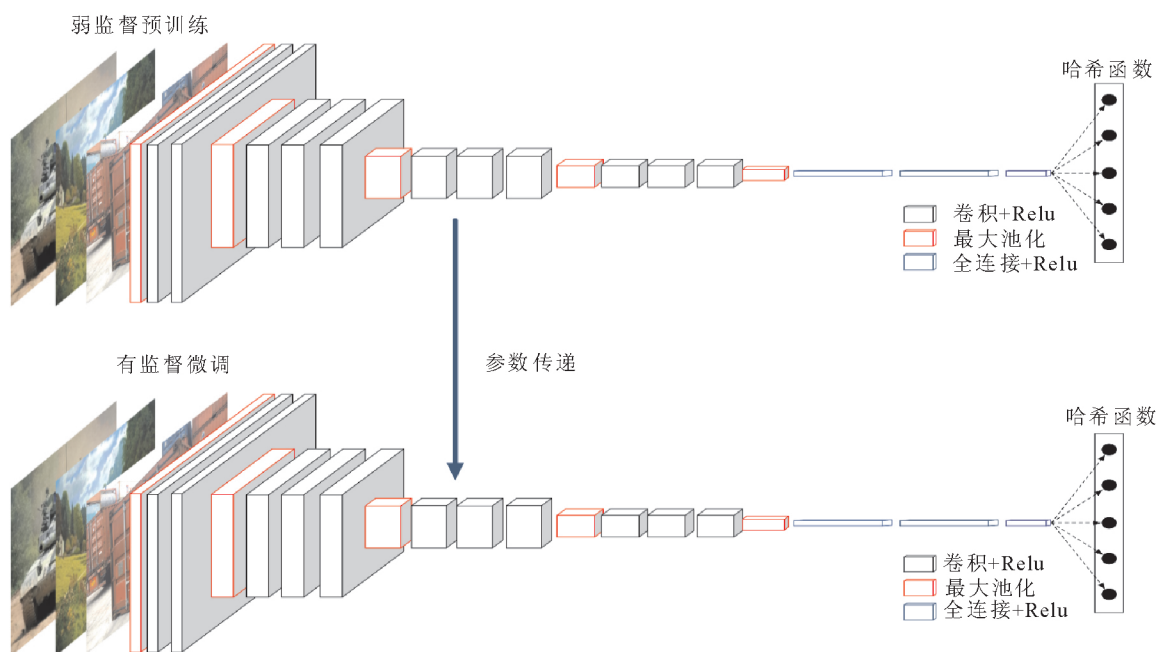


图5 弱监督哈希深度学习模型框架

Fig. 5 The proposed weakly-supervised deep hashing framework

## 2 实验与分析

### 2.1 数据集

本文使用 NUS-WIDE 弱标签数据作为数据集,每张图包含多个类别的标签。下载的数据集包含 269 648 张图像和由业余爱好者标记的 5 018 种标签,数据集还提供 81 个真值标签 (ground-truth) 和 500 维以 SIFT 特征描述的 BoW 词袋。实验将数据集划分为两部分,随机选取 5 000 幅图像用于测试,进行网络性能的评估,其余作为训练集,在预训练网络的基础上进行参数的微调。

### 2.2 数据预处理

对数据集中的图像进行预处理,考虑到图像检索所需要的存储空间,本文对图像归一化处理,压缩图像到 120\*120 像素大小。由于图像边缘部分可能包含大量无用信息的背景部分,先对图像去均值,使物体在图像中的占比基本相同,且位于图像中央,这样不但减少了计算量,同时避免了数据偏差过大带来的训练误差。

弱标签中包含一些与图像内容不相关的非视觉性语义标签,在对数据的预处理时,需要剔除这

些非视觉标签。本文先将图像表示为 SIFT 词袋形式<sup>[20]</sup>,计算每个标签对应图像集的 TF-IDF 向量,并计算向量的内聚和分离距离<sup>[21]</sup>。内聚性表示同一个标签下图像集之间的视觉相似性,分散性表示视觉内容相似的图像集与整个图像集的差别<sup>[22-23]</sup>。通过这两个距离判定标签是否具有视觉代表性,从而去除非视觉性语义标签。如图 6 为过滤后的结果。

过滤后的弱标签虽然已去除了非视觉性语义标签,但仍存在一词多义、近义词等问题,由于这些噪声的存在,不能够准确检索出两个相似的图像。所以,在训练神经网络之前需要先通过本文提出的语义相似度公式计算标签之间的相似度,利用弱标签间的相似度关系进行网络的监督学习。在计算语义相似度时用到了 Skip-gram 特征向量,它是由 Google 提出的一款生成词汇向量的工具 word2vec 生成的。

根据本文设计的语义相似度计算公式,对 Skip-gram 语义特征向量间的距离进行计算。当两个词汇的语义越相似,结果的数值就越大。表 1 列举了语义相似度计算的输出结果,搜索词汇“sunset”得到的结果有如“dusk”,“sun”,“sun-



rise”, “sky”等语义较为相似的词汇,也可以得到如“plants”, “religion”等语义不相似的词汇。结

果证明本文的语义相似度计算方法能够较好地表征语义间的相似度。

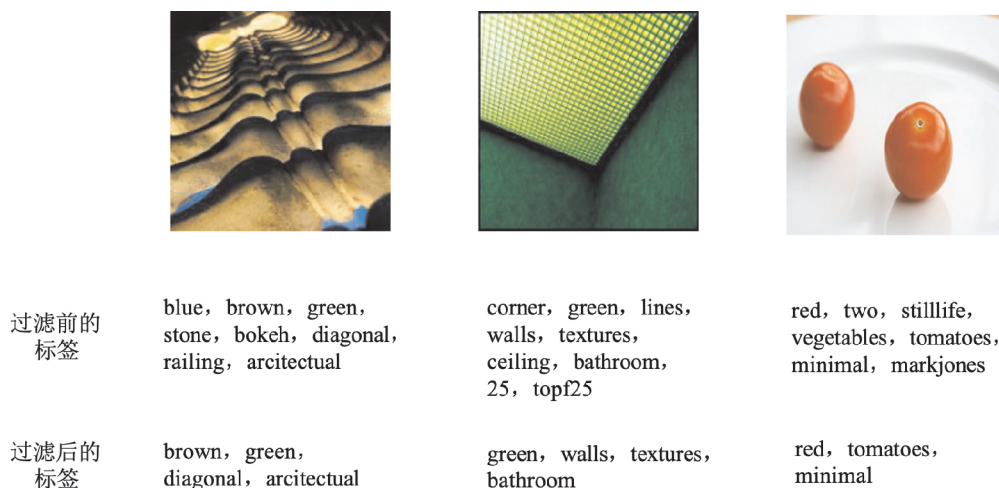


图 6 弱标注标签过滤前后的效果

Fig. 6 Weakly-labeled tags before and after filtering

### 2.3 实验结果

实验采用的 NUS-WIDE 数据集包含约 26 万张图像。采用的优化方法为 Mini-Batch 梯度下降法,设置 Mini-Batch 的大小为 64。卷积神经网络的卷积层和全连接层的参数以正态分布的随机值进行初始化<sup>[24]</sup>,并采用一个较小的学习率进行训练,优化方法采用随机梯度下降。表 2 为网络训练的各项参数。

本文采用 intel(R) Core(TM) i7-7700 CPU @ 3.60GHz 进行实验测试。与其他几种有监督学习方法和无监督学习方法的特征提取时间对比如图 7 所示。

表 1 与“sunset”语义相似与不相似距离表

Tab. 1 Semantic similarity and dissimilar distance table with "sunset"

搜索结果	语义距离
dusk	0.862
sun	0.791
sunrise	0.735
sky	0.687
plants	0.498
religion	0.362

表 2 训练网络模型使用参数表

Tab. 2 The parameter list of training network model

参数名	参数值
批大小	64
目标函数 $\lambda$	1
初始学习率	0.001
学习率衰减因子	0.1
迭代次数	120
初始动量参数	0.5

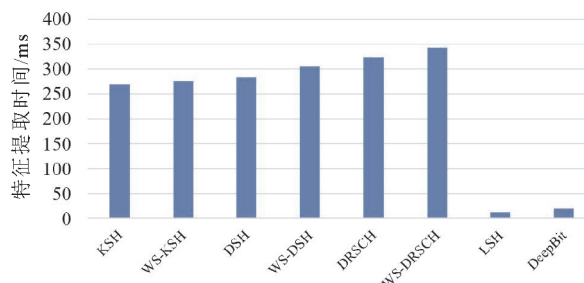


图 7 不同方法特征提取时间图

Fig. 7 Feature extraction time of such hashing methods

为了评估本文提出的弱监督哈希学习框架,我们把它应用到目前主流的几种有监督方法中,如 KSH<sup>[25]</sup>,DSH<sup>[7]</sup>和 DRSCH<sup>[8]</sup>。而使用了弱监督(weakly-supervised)预训练版本的这些方法分别被命名为 WS-KSH,WS-DSH 和 WS-DRSCH。

另外,还使用了两种代表性的无监督哈希方法作为基准,分别为 LSH(locality sensitive hashing)<sup>[26]</sup>和 DeepBit<sup>[5]</sup>。为了公平比较,对于基于 CNN 的方法(如 DeepBit,DSH,DRSCH 和对应的弱监督预训练版本),在 ImageNet 数据集上使用预训练的 VGG-16 网络<sup>[27]</sup>作为基础网络。对于其他的哈希方法,以预训练 VGG-16 网络的全连接层的输出层作为图像的输入特征。数据集中有 5 018 种人工标注的标签,经过弱监督的预训练后,将网络参数传递给有监督的 CNN 网络进行微调。这是一个通用的框架,因为任何有监督的哈希训练方法都可以用于微调。

本文提出的方法属于弱监督的方法。首先,由弱标签来判定图像间的相似度关系,将图像分

为两组相似或不相似的图像对集合,然后,利用本文设计的目标函数作为网络训练的约束,将弱标签的训练数据输入到神经网络中,训练模型的参数。

## 2.4 结果分析

本文用数据集集中的 81 个真值标签作为评测标准对精确度(Precision@R)进行评价。图 8 显示的是用 8,16,32,64 位哈希码进行输出的情况下,使用 Precision@500 进行评测的对比结果。图 9 显示的是在 100,200,400,1 000 个检索输出图像情况下,使用 Precision 在 64 位哈希码输出时进行评测的对比结果。图 10 显示的是以 64 位哈希码输出的 P-R 曲线。

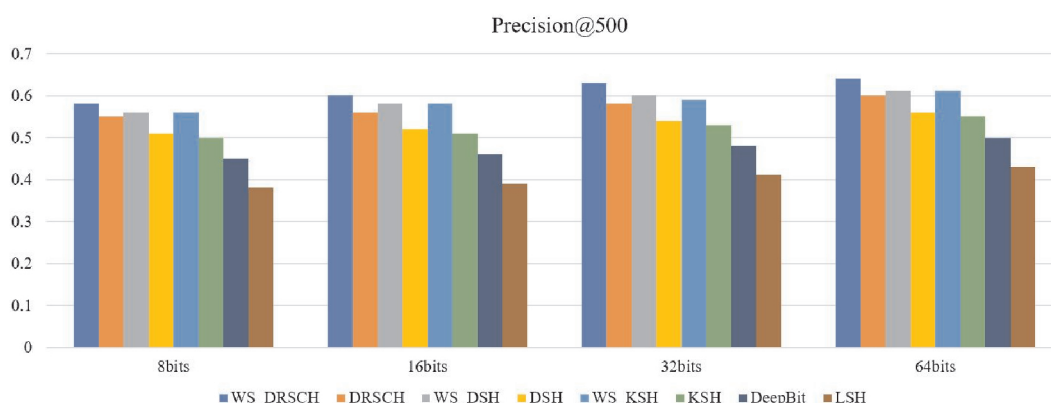


图 8 NUS-WIDE 数据集上具有不同位数的图像检索结果

Fig. 8 Image retrieval results(Precision@500) with different digits on the NUS-SIDS Dataset

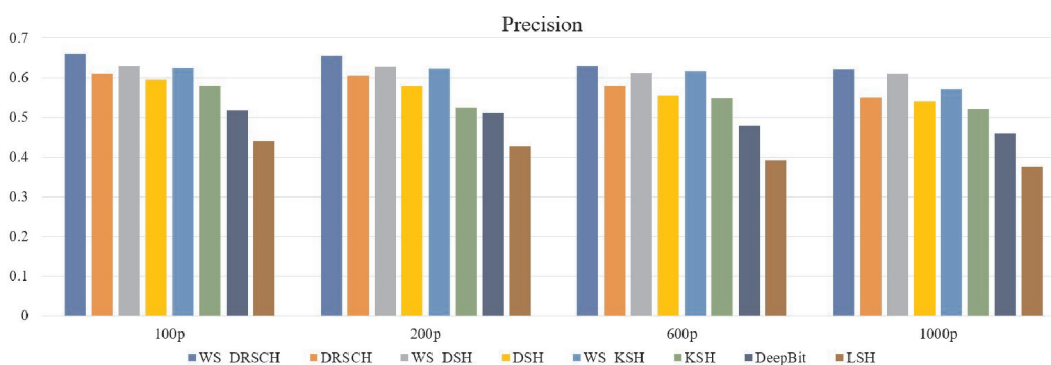


图 9 Precision 以 64 位哈希码输出的实验结果对比图

Fig. 9 Precision performance comparison with 64 hash bits

从实验结果可以看出,和无监督的哈希方法相比,有监督的方法总是可以达到更好的检索结果。WS-KSH,WS-DSH 和 WS-DRSCH 的性能均优于对应的有监督版本的方法。结果表明,弱监督框架可以通过利用用户标记的弱标记标签获得完全标记的图像数据集,从而提高检索性能。这

验证了本文的分析,弱标记的标签可以提供更多层次的语义信息,将有助于表征图像之间的细粒度相似关系,弱监督哈希学习算法可以更好地处理用户标签。

本文使用“大众分类”得到的弱标签训练网络,蕴含了更深层的潜在语义信息,弱标签的训练

针对性更弱,在学习的过程中会相互优化,使学习到的语义更全面,能够更好地表征图像间的相似关系。并且本文创新性地提出通过学习弱标签的语义相似度优化图像间的关系,解决了有监督学习中不能很好地用向量来描述图像相似度的问题,设计的目标函数可以很好地约束特征向量的提取,应用于图像检索领域明显提高了结果的精确度。

本文将卷积神经网络与哈希函数相结合,设计了一种可以度量相似度的损失函数来优化神经网络,通过网络提取 NUS-WIDE 数据库的高层次视觉特征,并通过哈希函数映射得到哈希码的表示,再通过有监督的方法进行微调,应用于图像检索领域,比常规的有监督学习方式更有效。

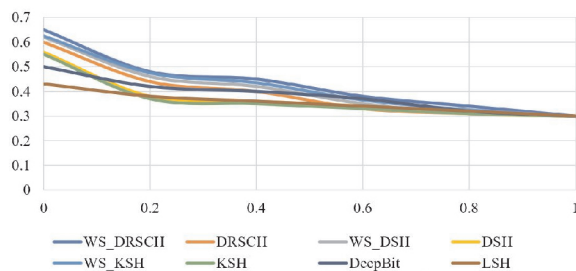


图 10 以 64 位哈希码输出的 P-R 曲线

Fig. 10 Precision-recall curve with 64 bits

### 3 结 语

针对目前主流的图像检索技术在数据集的获取和训练上精度不足的问题,本文提出了一种基于弱监督深度学习的图像检索方法。该方法采用互联网用户产生的弱标签数据进行训练,极大程度上降低了标注标签的人工成本。同时,该方法设计的语义相似度计算方法及网络的目标函数有效解决了弱标签中的噪声问题,并通过从弱标签向量间的语义关系来学习图像间的相似性关系,以此训练卷积神经网络,可以挖掘出更深层的语义信息,提升了检测的精度。实验结果证明,本文提出的图像检索方法的精度明显优于目前先进的其他算法。

### 参考文献:

- [1] 胡琦瑶. 基于弱监督深度学习的图像检索方法研究[D]. 西安: 西北大学, 2017.
- [2] LIU Y, SONG J K, ZHOU K, et al. Deep self-taught hashing for image retrieval[J]. IEEE Transactions on

Cybernetics, 2019, 49(6): 2229-2241.

- [3] QUEMADA J, GONZALO L, MAAREK Y, et al. Proceedings of the 18th international conference on world wide web[J]. Journal Danalyse Mathématique, 2009, 113(1): 197-225.
- [4] LI X R, SNOEK C G M, WORRING M. Learning social tag relevance by neighbor voting[J]. IEEE Transactions on Multimedia, 2009, 11(7): 1310-1322.
- [5] LIN K, LU J W, CHEN C S, et al. Learning compact binary descriptors with unsupervised deep neural networks[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 1183-1192.
- [6] YAN X, ZHANG L, LI W J. Semi-supervised deep hashing with a bipartite graph[C]//Twenty-Sixth International Joint Conference on Artificial Intelligence, 2017: 3238-3244.
- [7] 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). IEEE Computer Society, 2016: 2064-2072.
- [8] 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.
- [9] 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.
- [10] TANG J H, LI Z C. Weakly supervised multimodal hashing for scalable social image retrieval[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 28(10): 2730-2741.
- [11] XIONG C Z, SHAN Y M. Subject features and hash codes for multi-label image retrieval[C]//2018 IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS). Enshi: IEEE, 2018: 808-812.
- [12] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [13] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality[J]. Advances in Neural Information Processing Systems, 2013, 26: 3111-3119.
- [14] YU W, YANG K Y, YAO H X, et al. Exploiting the



- complementary strengths of multi-layer CNN features for image retrieval [J]. *Neurocomputing*, 2017, 237: 235-241.
- [15] LUO Y, YANG Y, SHEN F, et al. Robust discrete code modeling for supervised hashing [J]. *Pattern Recognition: The Journal of the Pattern Recognition Society*, 2018.
- [16] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. *Nature*, 2015, 521(7553): 436-444.
- [17] 杨斌, 钟金英. 卷积神经网络的研究进展综述 [J]. *南华大学学报(自然科学版)*, 2016, 30(3): 66-72.
- YANG B, ZHONG J Y. Review of convolution neural network [J]. *Journal Of University of South China( Science and Technology)* 2016, 30(3): 66-72.
- [18] SCHMIDHUBER J. Deep learning in neural networks: An overview [J]. *Neural Networks*, 2015, 61: 85-117.
- [19] 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: IEEE, 2015: 2160-2508.
- [20] FATEMEH A, MOHAMMAD R K. A new color feature extraction method based on dynamic color distribution entropy of neighborhoods [J]. *International Journal of Computer Science Issues*, 2012, 8(5): 42-48.
- [21] AIXIN S, SOURAV S, BHOWMICK, et al. Tag-based social image retrieval: An empirical evaluation [J]. *Journal of the Association for Information Science & Technology*, 2011, 62(12): 2364-2381.
- [22] HUANG C Q, YANG S M, PAN Y, et al. Object-location-aware hashing for multi-label image retrieval via automatic mask learning [J]. *IEEE Transactions on Image Processing*, 2018, 27(9): 4490-4502.
- [23] 王小鹏. 基于弱监督深度学习的图像检索技术与实现 [D]. 西安: 西北大学, 2018.
- [24] SALVADOR A, GIRO-I-NIETO X, MARQUES F, et al. Faster R-CNN features for instance search [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops ( CVPRW ). Las Vegas: IEEE, 2016: 9-16.
- [25] LIU W, WANG J, JI R R, et al. Supervised hashing with kernels [C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012: 2074-2081.
- [26] GIONIS A. Similarity search in high dimensions via hashing [C]// *Proceedings of the 25th VLDB Conference*, 2000: 518-529.
- [27] LI Z C, TANG J H. Weakly supervised deep matrix factorization for social image understanding [J]. *IEEE Transactions on Image Processing*, 2017, 26(1): 276-288.

(编辑 李 静)