计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2020.20190476
57(4):888-896,2020

最小熵迁移对抗散列方法

卓君宝[1,2] 苏 驰[3] 王树徽[1] 黄庆明[1,2]
[1](智能信息处理重点实验室(中国科学院计算技术研究所) 北京 100190)
[2](中国科学院大学计算机科学与技术学院 北京 100049)
[3](数字视频编解码技术国家工程实验室(北京大学) 北京 100871)
(junbao.zhuo@vipl.ict.ac.cn)

Min-Entropy Transfer Adversarial Hashing

Zhuo Junbao[1,2], Su Chi[3], Wang Shuhui[1], and Huang Qingming[1,2]
[1](*Key Laboratory of Intelligent Information Process* (*Institute of Computing Technology*, *Chinese Academy of Sciences*), *Beijing* 100190)
[2](*School of Computer Science and Technology*, *University of Chinese Academy of Sciences*, *Beijing* 100049)
[3](*National Engineering Laboratory for Video Technology* (*Peking University*), *Ministry of Education*, *Beijing* 100871)

Abstract Owing to its storage and retrieval efficiency, hashing is widely applied to large-scale image retrieval. Most of existing deep hashing methods assume that the database in the target domain is identically distributed with the training set in the source domain. However, in practical applications, such assumption is so strict that there exists considerable domain discrepancy between source and target domain. To address such cross-domain image retrieval problem, some research works introduce domain adaptation techniques into image retrieval methods. The goal is to enhance the generalization ability of the learned hashing function. However, the learned Hash codes lack discrimination and domain-invariance in existing cross-domain hashing methods. We propose semantic preservation module and min-entropy loss to tackle these issues. We simply construct a classification sub-network as semantic preservation module to fully utilize labels in source domain. Semantic information encoded in labels can be passed to hashing learning network, which encourages learned Hash codes to contain more semantic information and discriminativity. As for unlabeled target domain samples, the entropy of their classification responses characterizes the confidence of classifier. Ideal target classification responses should tend to be one-hot vectors which minimizes the entropy. Therefore, we add minimization entropy loss to our model. Minimizing the entropy of classification responses of target samples aligns the distribution between source and target domain in classifier responses space. Therefore, the learned Hash codes tend to be more domain-invariant. With the semantic preservation module and min-entropy loss, we construct an end-to-end deep neural network for cross-domain image retrieval. Extensive experiments show the superiority of our model over existing state-of-the-art methods.

收稿日期:2019-07-08;**修回日期**:2019-10-16
基金项目:国家自然科学基金项目(61672497,U163621);国家"九七三"重点基础研究发展计划基金项目(2015CB351802);中国科学院前沿科学重点研究项目(QYZDJ-SSW-SYS013)
This work was supported by the National Natural Science Foundation of China (61672497, U163621), the National Basic Research Program of China (973 Program) (2015CB351802), and the Key Research Program of Frontier Sciences of CAS (QYZDJ-SSW-SYS013).
通信作者:王树徽(wangshuhui@ict.ac.cn)

Key words cross-domain image retrieval; transfer learning; hashing method; adversarial learning; deep learning; domain adaptation

摘 要 散列算法具有高效的存储和查询特性,被广泛应用于大规模的图像检索.大多数现有的深度散列方法都基于独立同分布的假设,即训练集(源域)和测试集(目标域)的分布一致.然而在现实应用中,源域和目标域往往存在较大的差异,即跨域检索.因此有些研究工作开始将跨域识别的方法引入到跨域检索中,以增强所学散列函数的泛化性.现有跨域检索方法仍存在散列码的判别力不足和域不变能力不足2个问题.提出语义保持模块和最小熵损失来解决这2个问题.语义保持模块是1个分类子网络,该模块可以充分利用源域的类别标注信息,并将该语义信息传递给散列学习子网络使得学习到的散列码包含更多的语义信息,即增强判别力.此外,对于无标注的目标域,熵表征目标域样本的分类响应的集中程度,理想的散列码经过语义保持模块后得到的分类响应应该集中于某一个类别,即最小熵状态.引入最小熵损失促使目标域样本与源域样本在类别响应这一空间上分布更加对齐,进而使得散列码更具域不变性.通过引入语义保持模块和最小熵损失,在现有方法的基础上构建了端到端的跨域检索网络,并在2个数据集上进行了大量实验,与领域内现有主要模型进行了详尽的对比,实验证明所提模型取得了更优的性能.

关键词 跨域图像检索;迁移学习;散列方法;对抗学习;深度学习;域适配

中图法分类号 TP391

大数据时代的到来,网络上涌现了大量的高维图像数据.图像检索越来越受到学术界和工业界的关注.随着深度学习的普及,深度散列方法[1-7]也备受关注,其性能远远超过了传统的无监督方法[8-10]和基于浅层模型的有监督方法[11-13].然而深度学习方法往往需要大量的标注信息,搜集这些标注信息往往耗费巨大人力物力.此外,大多数现有的深度学习方法都基于独立同分布的假设,即训练集(源域)和测试集(目标域)的分布一致.然而在现实应用中,源域和目标域往往存在较大的差异.因此利用有标注的数据集(源域)并迁移到相关的无标注目标域[14-19]受到极大的关注和发展.然而在图像检索领域,跨域迁移学习的研究处于起步阶段,仍待继续研究.跨域图像检索的难点在于目标域无标注且与源域存在较大域间差异,这种差异往往导致在带标注源域上训练好的模型应用于目标域时检索性能大幅度下降.如何学习具有判别力和域不变的散列码是跨域图像检索的重点.

深度适配散列(deep adaptive hashing, DAH)[20]首次将域适配的方法应用于跨域图像检索任务中,在学习散列码的同时,DAH引入最大均值差异(maximum mean discrepancy, MMD)[14-15]来度量域间差异,通过最小化MMD来学习域不变的散列码.此后迁移对抗散列(transfer adversarial hashing, TAH)[21]提出将跨域识别中经典的域对抗网络[19]应用到跨域图像检索中.TAH通过引入一个域分类器来判别源域和目标域的分布是否一致,采用对抗思想促使所学的源域与目标域散列码分布趋于一致,进而学习到域不变的散列码,取得了当前最好检索性能.

然而现有深度跨域图像检索方法仍然存在2个问题:1)在学习散列码时,标注信息仅仅被用于构建2个样本是否相似的监督信息去指导网络的学习,忽略了标注信息的语义信息.这使得所学的散列码的判别力不足,造成检索性能的瓶颈.2)现有的分布对齐方法学习域不变特征的能力仍然不足,使得将源域学习得到的散列函数应用于目标域时性能仍然有较大下降.

针对上面2个问题,我们提出语义保持模块和最小熵损失来改进现有深度跨域图像检索方法.首先,我们在散列特征后再引入一个分类子网络,通过源域的标注信息来训练该分类子网络并将语义信息反传给生成散列特征的子网络,有效保持了散列码的判别力.此外,在目标域上,由于没有语义标注信息,无法像源域那样引入监督信息.因此我们引入最小化目标域样本的类别响应分布的熵来促使目标域样本的类别响应能够集中在某个类别上.最小熵损失有效增强了散列码的泛化能力.

基于TAH模型以及我们所提的语义保持和最小熵损失,我们构建了一个新的可端到端训练的跨域

图像检索网络.由于语义保持采用的多类别的交叉熵损失也是一种熵,因此我们称所提的模型为最小熵迁移对抗散列(min-entropy transfer adversarial hashing, METAH).我们在2个数据集上进行了大量实验,与领域内现有主要模型进行了详尽的对比,实验证明了所提模型取得了更优的性能,证明了所提语义保持模块和最小熵损失的有效性.

1 相关工作

跟我们工作相关的2个任务分别是跨域识别和基于散列的图像检索.我们将从这2个任务进行相关工作的阐述.

1.1 跨域识别

跨域识别又称为域适配(domain adaptation, DA),跨域识别已经得到很大的发展,这里我们只回顾和我们方法比较相关的深度域适配方法.

深度域混淆网络(deep domain confusion, DDC)[14]基于AlexNet架构,其在 fc_7 层上使用单核的MMD来度量域间的差异,通过最小化MMD来使域间差异减小从而学到域不变的特征.深度适配网络(deep adaptation network, DAN)[15]则在多个全连接层上使用多核的MMD来度量域间差异,进一步加强特征迁移能力.深度相关对齐(deep correlation alignment, DCORAL)[17]和深度无监督卷积域适配(deep unsupervised convolutional domain adaptation, DUCDA)[16]则用源域特征协方差和目标域特征协方差之间的差值矩阵范数来度量域间差异,从而减小深度网络的特征分布距离以学习域不变的特征.群体匹配差异(population matching discrepancy, PMD)[22]则是对源域和目标域间的样本计算最优匹配,将匹配的样本对间的距离进行累加来表征域间差异,通过最小化PMD来学习域不变的特征.

生成对抗网络(generative adversarial network, GAN)[23]的提出让基于特征的跨域迁移方法又有了新的突破.域对抗网络[19]、对抗判别域适配(adversarial discriminative domain adaptation, ADDA)[24],和条件对抗域适配(conditional adversarial domain adaptation, CADA)[25],都是利用对抗思想将目标域特征空间向源域特征空间靠近,从而让目标域的特征可以适配源域的特征分类器.

1.2 散列方法

散列方法是经典的研究方向,主要包括无监督散列[8-10]和有监督散列[1-7,11-13].这里只回顾和我们的方法比较相关的有监督深度散列方法.

卷积神经网络散列(convolutional neural network hashing, CNNH)[4]采用2阶段策略:1)先学散列码;2)学习一个深度网络将图像映射到所学的散列码.深度神经网络散列(deep neural network hashing, DNNH)[5]改进了CNNH,不采用2阶段训练策略而是同时学习图像特征和散列函数,这种端到端训练方式能够更加充分利用深度网络的特征学习和函数拟合能力.深度散列网络(deep hashing network, DHN)[6]进一步优化DNNH,通过引入交叉熵损失和量化损失来保持相似度和约束量化误差.散列网络(HashNet)[7]则解决了符号函数的病态梯度问题,直接优化符号函数,HashNet是单域图像检索最好的方法.深度语义排序散列(deep semantic ranking hashing, DSRH)[26]则考虑了多标签图像间的语义相似度.

2 最小熵迁移对抗散列方法

假定给定1个由目标域 $\mathcal{T}$ 的图像构成的数据库 $X_{\mathcal{T}}=\{\boldsymbol{X}_k^{\mathcal{T}}\}_{k=1}^m$ 和1个由源域 $\mathcal{S}$ 带标注的图像构成的训练集 $X_{\mathcal{S}}=\{\boldsymbol{X}_k^{\mathcal{S}},\boldsymbol{y}_k\}_{k=1}^n$,其中 $\boldsymbol{y}_k$ 是类别标注,是1个向量,某一维值为1指明属于该类别.这2个域存在较大的域间差异,可建模成分布不一致:$P(X_{\mathcal{S}})\neq P(X_{\mathcal{T}})$,这也是跨域检索的难点.跨域检索的目标是运用有标注信息的 $X_{\mathcal{S}}$ 和无标注的 $X_{\mathcal{T}}$ 学习1个散列函数 f,使得该散列函数能够很好泛化到目标域,即给定目标域的查询集 $X_q^{\mathcal{T}}$ 中的1个查询 q,在目标域数据库 $X_{\mathrm{DB}}^{\mathcal{T}}$ 检索到图像与查询 q 是语义相关的.散列函数 $f:\mathbb{R}^d\rightarrow\{-1,1\}^b$ 将图像 $\boldsymbol{X}_i^{\mathcal{S}}$ 映射到1个 b 维的散列码 $\boldsymbol{h}_i^{\mathcal{S}}=f(\boldsymbol{X}_i^{\mathcal{S}})$.由于 $X_{\mathcal{S}}$ 是有标注的,我们可以根据其类别标签构造相似矩阵 $\boldsymbol{S}=(s_{ij})$ 来学习散列函数.$s_{ij}=1$ 表示 $\boldsymbol{X}_i^{\mathcal{S}}$ 与 $\boldsymbol{X}_j^{\mathcal{S}}$ 是语义相关的.

2.1 模型框架

如图1所示,我们的模型采用孪生网络结构,且上下2个子网络权值共享.子网络基于AlexNet,该网络由 $conv_1\sim conv_5$ 共5层卷积层和 $fc_6\sim fc_8$ 共3层全连接层构成.我们用1个输出为 b 的全连接层 fc_{hashing} 替换 fc_8 用于学习散列函数 f.由于 fc_{hashing} 难以学习到离散的输出,因此我们对其放宽了限制,即约束 fc_{hashing} 输出[-1,1]的连续值.为了能将所学的散列函数泛化到目标域,我们在 fc_{hashing} 后经过梯度取反层(图1中2个沙漏所表示),之后引入1个

域分类器 ad_net.该域分类器的作用在于促使 fc_{hashing} 学到域不变的散列特征.此外,为了更好地保持散列码的语义信息,我们在 fc_{hashing} 后增加语义保持模块,这里我们用 1 层全连接层 fc_{cls} 来构建语义保持模块,它将散列码映射到类别空间.我们称所提方法为最小熵迁移对抗散列(METAH).

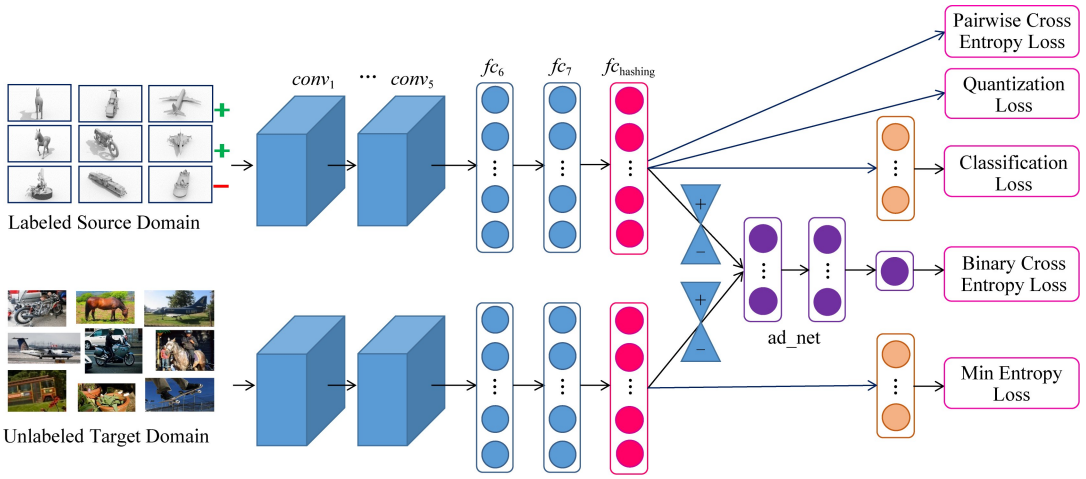


Fig. 1 Framework of the proposed method

图 1 所提方法的结构图

2.2 学习散列函数

我们采用经典的最大后验估计(maximum a posterior, MAP)来使得所学的散列函数能够保持成对样本间的相似度或不相似度.

给定由源域标注信息所构建的相似性矩阵 $\boldsymbol{S}=(s_{ij})$,对于所要学习的散列码矩阵 $\boldsymbol{H}^{\mathcal{S}}=(\boldsymbol{h}_1^{\mathcal{S}},\boldsymbol{h}_2^{\mathcal{S}},\cdots,\boldsymbol{h}_n^{\mathcal{S}})$,其最大后验估计取对数为

$$\ln p(\boldsymbol{H}^{\mathcal{S}}|\boldsymbol{S})\propto\ln p(\boldsymbol{S}|\boldsymbol{H}^{\mathcal{S}})p(\boldsymbol{H}^{\mathcal{S}})=\sum_{s_{ij}}\ln p(s_{ij}|\boldsymbol{h}_i^{\mathcal{S}},\boldsymbol{h}_j^{\mathcal{S}})p(\boldsymbol{h}_i^{\mathcal{S}})p(\boldsymbol{h}_j^{\mathcal{S}}),\tag{1}$$

其中,$p(\boldsymbol{S}|\boldsymbol{H}^{\mathcal{S}})$ 是似然函数,$p(\boldsymbol{H}^{\mathcal{S}})$ 是先验分布.条件概率 $p(s_{ij}|\boldsymbol{h}_i^{\mathcal{S}},\boldsymbol{h}_j^{\mathcal{S}})$ 计算为

$$p(s_{ij}|\boldsymbol{h}_i^{\mathcal{S}},\boldsymbol{h}_j^{\mathcal{S}})=\begin{cases}\sigma(sim(\boldsymbol{h}_i^{\mathcal{S}},\boldsymbol{h}_j^{\mathcal{S}})), & s_{ij}=1,\\ 1-\sigma(sim(\boldsymbol{h}_i^{\mathcal{S}},\boldsymbol{h}_j^{\mathcal{S}})), & s_{ij}=0,\end{cases}\tag{2}$$

其中,$sim(\boldsymbol{h}_i^{\mathcal{S}},\boldsymbol{h}_j^{\mathcal{S}})$ 是散列码 $\boldsymbol{h}_i^{\mathcal{S}}$ 和 $\boldsymbol{h}_j^{\mathcal{S}}$ 的相似度,$\sigma(\cdot)$ 为概率函数.我们沿用文献[21]的方法,即:

$$\begin{cases}sim(\boldsymbol{h}_i^{\mathcal{S}},\boldsymbol{h}_j^{\mathcal{S}})=\dfrac{b}{1+\|\boldsymbol{h}_i^{\mathcal{S}}-\boldsymbol{h}_j^{\mathcal{S}}\|^2},\\ \sigma(x)=\tanh(\alpha x).\end{cases}\tag{3}$$

将式(2)和式(3)代入最大后验估计式(1)可以得到以下损失:

$$L_{\text{P}}=\sum_{s_{ij}}\left(\ln\left(\frac{2}{1+\exp\left(2\alpha\dfrac{b}{1+\|\boldsymbol{h}_i^{\mathcal{S}}-\boldsymbol{h}_j^{\mathcal{S}}\|^2}\right)}\right)+s_{ij}\times\ln\left(\frac{\exp\left(2\alpha\dfrac{b}{1+\|\boldsymbol{h}_i^{\mathcal{S}}-\boldsymbol{h}_j^{\mathcal{S}}\|^2}\right)-1}{2}\right)\right).\tag{4}$$

一旦学习得到 $\boldsymbol{h}_i^{\mathcal{S}}$ 后,我们可以采用符号函数将 $\boldsymbol{h}_i^{\mathcal{S}}$ 转化为散列码,即 $\hat{h}_i^{\mathcal{S}}=\text{sgn}(\boldsymbol{h}_i^{\mathcal{S}})$.该转化会带来较大的量化误差,因此需引入量化损失来控制量化误差,具体为

$$L_{\text{Q}}=\sum_{s_{ij}}s_{ij}\times(\|\,|\boldsymbol{h}_i^{\mathcal{S}}|-\mathbf{1}\|_1+\|\,|\boldsymbol{h}_j^{\mathcal{S}}|-\mathbf{1}\|_1),\tag{5}$$

其中,$\mathbf{1}=(1,1,\cdots,1)$ 是全为 1 的 b 维向量,$|\cdot|$ 是绝对值函数,即将输入向量的每个元素取绝对值.

2.3 对抗分布对齐

在跨域图像检索中,由于源域和目标域存在较大分布差异,且目标域没有标注信息,将源域训练好的模型应用于目标域时会造成性能大幅下降.因此我们需要在学习散列码的同时缩小域间差异,使得所学习的散列码是域不变的.

域对抗网络[19]是一个典型的域分布对齐方法,发展至今,域对抗方法具有良好的理论保证且在跨域识别达到较优的性能.本文也采用域对抗思想来减小域间差异.域对抗思想是引入一个域分类器,其作用是区分样本特征来自源域还是目标域.由于我们训练时知道样本来自于源域或目标域,域分类器可以通过这种标注来训练,即最小化二分类交叉熵损失.而另一方面,我们希望所学到的特征是域分类器区分不开的,即最大化二分类交叉熵损失.从分布拟合的角度来看,域分类器用于区分 2 个域的分布,而特征生成器则拉近 2 个域的分布,进而减小域间差异.

记 G_{F} 为散列特征生成器即 $conv_1\sim conv_5$,

$fc_6 \sim fc_7$, fc_{hashing}所组成的子网络,其可训练参数为$\boldsymbol{\theta}_{\text{F}}$.域分类器$G_{\text{D}}$的可训练参数记为$\boldsymbol{\theta}_{\text{D}}$.则域对抗网络的损失为

$$L_{\text{D}}(\boldsymbol{\theta}_{\text{F}},\boldsymbol{\theta}_{\text{D}})=\frac{1}{n}\sum_{\boldsymbol{x}_i^{\mathcal{S}}\in X_{\mathcal{S}}}\ln G_{\text{D}}(G_{\text{F}}(\boldsymbol{X}_i^{\mathcal{S}}))+\frac{1}{m}\sum_{\boldsymbol{x}_j^{\mathcal{T}}\in X_{\mathcal{T}}}\ln(1-G_{\text{D}}(G_{\text{F}}(\boldsymbol{X}_j^{\mathcal{T}}))). \tag{6}$$

式(6)是二分类的交叉熵损失.对抗学习则是寻求损失函数$L_{\text{D}}(\boldsymbol{\theta}_{\text{F}},\boldsymbol{\theta}_{\text{D}})$的鞍点:

$$\begin{cases}\hat{\boldsymbol{\theta}}_{\text{F}}=\arg\max\limits_{\boldsymbol{\theta}_{\text{F}}}L_{\text{D}}(\boldsymbol{\theta}_{\text{F}},\boldsymbol{\theta}_{\text{D}}),\\ \hat{\boldsymbol{\theta}}_{\text{D}}=\arg\min\limits_{\boldsymbol{\theta}_{\text{D}}}L_{\text{D}}(\boldsymbol{\theta}_{\text{F}},\boldsymbol{\theta}_{\text{D}}).\end{cases} \tag{7}$$

求解式(7)需要分开优化,且这种方式训练比较困难.因此我们也采用梯度取反层[19]来实现对抗学习,具体方法是引入梯度取反层(图1沙漏所示),操作为

$$\begin{cases}R_\eta(\boldsymbol{x})=\boldsymbol{x},\\ \dfrac{\mathrm{d}R_\eta}{\mathrm{d}\boldsymbol{x}}=-\eta\boldsymbol{I},\end{cases} \tag{8}$$

其中,$\boldsymbol{I}$是单位阵.梯度取反层正向计算时其输出保持不变,而反向传播时将原梯度取反并乘以η.

因此,域对抗网络的损失为

$$L_{\text{D}}=\frac{1}{n}\sum_{\boldsymbol{x}_i^{\mathcal{S}}\in X_{\mathcal{S}}}\ln G_{\text{D}}(G_{\text{F}}(R_\eta(\boldsymbol{X}_i^{\mathcal{S}})))+\frac{1}{m}\sum_{\boldsymbol{x}_j^{\mathcal{T}}\in X_{\mathcal{T}}}\ln(1-G_{\text{D}}(G_{\text{F}}(R_\eta(\boldsymbol{X}_j^{\mathcal{T}})))). \tag{9}$$

2.4 语义信息保持

现有跨域图像检索工作中,标注信息仅仅被用于判断2个样本是否相似,其本身的语义信息却被忽略了.而在测试阶段,语义信息则是判断被检索到的结果与查询q是否相关的唯一依据.因此我们提出在fc_{hashing}后面引入语义保持模块,我们用一层分类层fc_{cls}来构建语义保持模块.语义保持模块将$\boldsymbol{h}_i^{\mathcal{S}}$映射到分类响应$\boldsymbol{c}_i^{\mathcal{S}}$.引入分类损失:

$$L_{\text{C}}=\sum_{i=1}^{n}\langle\boldsymbol{y}_i,\ln\boldsymbol{c}_i^{\mathcal{S}}\rangle, \tag{10}$$

其中,$\langle\cdot,\cdot\rangle$是内积.引入该分类损失大大增强了散列特征的判别力,增强了模型的泛化能力.

2.5 最小熵

在目标域中,一个理想的散列码在经过fc_{cls}后得到的分类响应应该集中于某一类上.由于目标域没有标注,我们无法知道目标域样本应该属于哪一类,因此我们通过最小熵来促使目标域样本分类响应集中于某一类上.熵的计算为

$$L_{\text{E}}=-\sum_{i=1}^{n}\langle\boldsymbol{c}_i^{\mathcal{T}},\ln\boldsymbol{c}_i^{\mathcal{T}}\rangle. \tag{11}$$

源域由于有标注信息,其样本的分类响应往往集中在所标注的类别上;而目标域由于存在域间差异,其在分类响应上往往不够集中.最小熵能够在语义层减小源域和目标域的域间差异,进而影响特征层,使得特征层的域间差异也相应减小,即增强了散列码的域不变能力具有更强的泛化能力.

2.6 总损失

综合2.2~2.5节,我们采用最终目标损失来训练所提的最小熵迁移对抗散列方法:

$$L=\lambda_{\text{P}}L_{\text{P}}+\lambda_{\text{Q}}L_{\text{Q}}+\lambda_{\text{D}}L_{\text{D}}+\lambda_{\text{C}}L_{\text{C}}+\lambda_{\text{E}}L_{\text{E}}, \tag{12}$$

其中,λ_*是一些控制各个损失间平衡的超参数.

3 实验与结果

我们在2个常用数据集上进行了大量实验,并与领域内现有主要模型进行了详尽的对比.实验证明了所提模型取得了更优的性能.

3.1 实验设置与对比算法

NUS-WIDE是一个跨模态检索常用的数据集,其包含269 648个文本-图像对.该数据集标注了81个语义概念用于测试检索算法的性能.为了公平比较,我们沿用文献[5-8]的设定,只在出现频率最高的21个语义概念所涵盖的195 834张图像上做实验.查询集包含2 100张图像,训练集包含10 000张图像,剩下的作为被检索的数据库.

VisDA-2017是一个跨域识别常用的数据集,其包含2个域,源域由CAD模型渲染生成的12类图像构成,记为Syn,目标域是在COCO上选取的相应类别的子集,记为Real.由于该数据集域间差异比较大,我们构建2种设定:1)查询集和数据库都采用Real域而带标注训练集为Syn域,记为Syn→Real;2)查询集和数据库都采用Syn域而带标注训练集为Real域,记为Real→Syn.

我们在汉明距离小于2(Hamming radius 2)的检索结果上计算平均精度均值(mean average precision, MAP)作为评测性能.对比算法包括局部敏感散列(locality sensitive hashing, LSH)[8]、谱散列(spectral hashing, SH)[9]、迭代量化(iterative quantization, ITQ)[10]等传统无监督方法;核散列(kernel supervised hashing, KSH)[12]、监督离散散列(supervised discrete hashing, SDH)[13]等有监督浅层模型;CNNH[4], DNNH[5], DHN[6], HashNet[7]

等单域有监督深度模型以及传递散列网络(transitive hashing network, THN)[27],TAH[21]等跨模态或者跨域的有监督深度检索方法.为了更好地验证和分析我们所提的语义信息保持模块和最小熵的有效性,我们构造了 2 个变体 METAH-e 和 METAH.其中 METAH-e 中 $\lambda_E=0$,即不加最小熵损失.

为了公平比较,我们的算法也是基于在 ImageNet 上与训练好的 AlexNet 上.我们采用 Caffe 框架来微调 $conv_1\sim conv_5$ 等卷积层和 $fc_6\sim fc_7$ 等全连接层.此外,我们多加 1 层全连接层 fc_{hashing}层,并在 fc_{hashing}层后接 2 个分支:1)2 层全连接层构成的子网络 ad_net 用于对抗学习;2)1 个分类全连接层 fc_{cls}构成的语义保持模块.fc_{hashing},ad_net,fc_{cls}等新加全连接层的学习率设置为 $conv_1\sim conv_5$,$fc_6\sim fc_7$ 学习率的 10 倍.整个优化过程采用冲量为 0.9 的小批量随机梯度下降法(stochastic gradient descent, SGD).1 次迭代对每个域随机抽取 64 张图像用于估计梯度.权重衰减设为 0.0005.在 3.3 节中的梯度取反层中的参数 η 的更新为:$\eta=2/(1+\exp(10i))$,其中 i 是当前迭代的次数.式(12)中的超参数设置如表 1 所示:

Table 1 Values for Hyper-Parameters

表 1 超参数设置

Tasks	λ_P	λ_Q	λ_D	λ_C	λ_E
Syn→Real	6.0	9.0	0.8	0.8	0.01
Real→Syn	10.0	0.8	1.0	0.8	0.01
NUS-WIDE→NUS-WIDE	1.0	0.3	0.8	0.8	0.01

3.2 定量实验结果

NUS-WIDE 上的平均精度均值如表 2 所示.我们可以看出:即使在训练集和测试集域间差异几乎不存在的情况下,我们的方法也能取得最优的结果.在散列码的长度分别为 48 b 和 64 b 的设定中,我们的方法比 TAH[21]的平均精度均值分别提高了 0.086 和 0.087.在 32 b 设定中,我们的方法提升很小,原因在于 32 b 维度太小,模型能力较弱,不能同时学习散列码和保持语义信息.而在 48 b 和 64 b 的设定中,语义信息保持(METAH-e)所带来的性能提升则非常显著.由于域间差异几乎不存在,METAH 相比于 METAH-e 性能提升很小甚至起负迁移的作用.

上述现象符合我们的预期,因为最小熵的作用在于减小域间差异,对于域间差异几乎不存在的设定中,强行减小域间差异反而会带来反作用.

Table 2 MAP Results Within Hamming Radius 2 on NUS-WIDE

表 2 NUS-WIDE 上汉明距离小于 2 的平均精度均值

Methods	Length of Hash Code		
	32 b	48 b	64 b
LSH[8]	0.453	0.323	0.255
KSH[12]	0.554	0.421	0.335
SH[9]	0.543	0.437	0.371
ITQ[10]	0.549	0.517	0.402
SDH[13]	0.571	0.517	0.499
CNNH[4]	0.601	0.587	0.535
DNNH[5]	0.622	0.611	0.585
DHN[6]	0.672	0.661	0.598
HashNet[7]	0.693	0.681	0.615
THN[27]	0.676	0.662	0.603
TAH[21]	0.729	0.692	0.680
METAH-e	**0.734**	**0.779**	0.768
METAH	0.730	0.778	**0.777**

Notes: METAH-e is a variant of METAH with $\lambda_E=0$; bold values indicate the best performance; underlined values indicate the second best performance.

VisDA-2017 上 Syn→Real 的平均精度均值如表 3 所示.在散列码的长度分别为 32 b,48 b,64 b 的设定中,我们的方法比 TAH 的平均精度均值分别提高了 0.048,0.101,0.101.可以看出由于 32 b 维度较小,METAH 的平均精度均值提升相比于 48 b 和 64 b 较小.由于该任务中源域和目标域存在较大差异,因而我们所提的最小熵作用更加明显.因此在 32 b,48 b,64 b 的设定中,METAH 相比于 METAH-e 的平均精度均值分别提升了 0.023,0.024,0.038.值得注意的是我们对不同长度的散列码都采用同样的超参数,而 TAH 各个设定的超参数都是通过交叉验证获得的,所以 METAH 具有更大的潜能.

VisDA-2017 上 Real→Syn 的平均精度均值如表 4 所示.相比于最好的对比算法 TAH[8],我们的方法 METAH 在散列码的长度分别为 32 b,48 b,64 b 的设定中平均精度均值分别提高了 0.001,0.008,0.071.相比于 Syn→Real,在 Real→Syn 上 METAH 的提升较小,原因可能是我们使用了和 Syn→Real 同样的超参数设置,即 $\lambda_C=0.8$,$\lambda_E=0.01$,然而对于跨域图像检索的设定,目标域是无标注的,因此通过交叉验证针对不同长度散列码去搜索最优的参数是不可取的,所以我们这里针对所有不同长度散列码都只用一套相同的超参数.此外,在散列码的长度分别为 32 b,48 b,64 b 的设定中,METAH 相比于

METAH-e 的平均精度均值分别提升了 0.008, 0.010, 0.003.证明了所提的最小熵在减小域间差异上的有效性.

Table 3 MAP Results Within Hamming Radius 2 on Syn→Real

表 3 Syn→Real 上汉明距离小于 2 的平均精度均值

Methods	Length of Hash Code		
	32 b	48 b	64 b
LSH[8]	0.092	0.083	0.071
KSH [12]	0.183	0.124	0.085
SH[9]	0.141	0.130	0.105
ITQ [10]	0.175	0.146	0.123
SDH[13]	0.238	0.29	0.212
CNNH[4]	0.254	0.238	0.230
DNNH[5]	0.276	0.252	0.243
DHN[6]	0.354	0.309	0.281
HashNet[7]	0.403	0.345	0.274
THN[27]	0.396	0.228	0.127
TAH[21]	0.423	0.433	0.404
METAH-e	0.438	0.510	0.467
METAH	**0.471**	**0.534**	**0.505**

Notes: METAH-e is a variant of METAH with $\lambda_E = 0$; bold values indicate the best performance; underlined values indicate the second best performance.

Table 4 MAP Results Within Hamming Radius 2 on Real→Syn

表 4 Real→Syn 上汉明距离小于 2 的平均精度均值

Methods	Length of Hash Code		
	32 b	48 b	64 b
LSH[8]	0.145	0.122	0.063
KSH [12]	0.178	0.146	0.092
SH[9]	0.182	0.145	0.123
ITQ [10]	0.193	0.176	0.158
SDH[13]	0.388	0.339	0.277
CNNH[4]	0.568	0.530	0.445
DNNH[5]	0.564	0.551	0.503
DHN[6]	0.612	0.608	0.604
HashNet[7]	0.676	0.662	0.642
THN[27]	0.687	0.664	0.532
TAH[21]	0.695	0.784	0.761
METAH-e	0.678	0.782	0.829
METAH	**0.696**	**0.792**	**0.832**

Notes: METAH-e is a variant of METAH with $\lambda_E = 0$; bold values indicate the best performance; underlined values indicate the second best performance.

3.3 定性实验结果

我们在 VisDA-2017 数据集上进行了可视化实验.在 Syn→Real 任务中,我们在 Real 域随机选取了 1 个查询,并在 Real 域构成的数据库中进行检索,我们列出前 10 的检索结果,与 TAH, METAH-e 的对比结果如图 2 所示.虚线框指错误检索结果,实线框指正确检索结果.可以看出 METAH-e 和 METAH 的查询结果相比于 TAH 错误结果更少,证明了所提方法的有效性.在 Real→Syn 上的检索结果如图 3

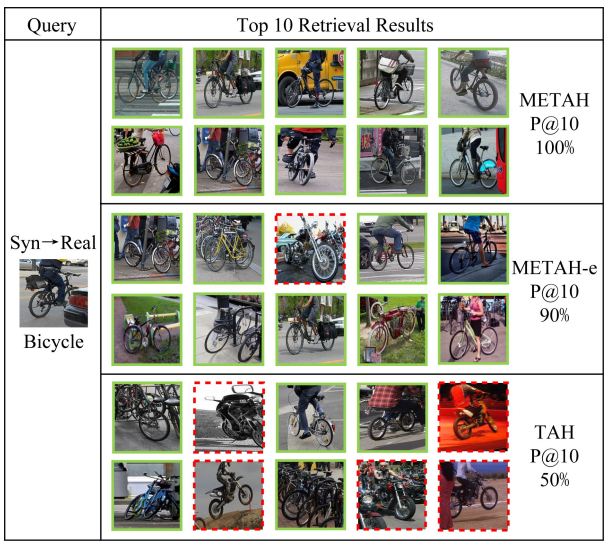


Fig. 2 Examples of top 10 retrieval images and P@10 in Syn→Real

图 2 在 Syn→Real 上前 10 检索结果和 P@10 值

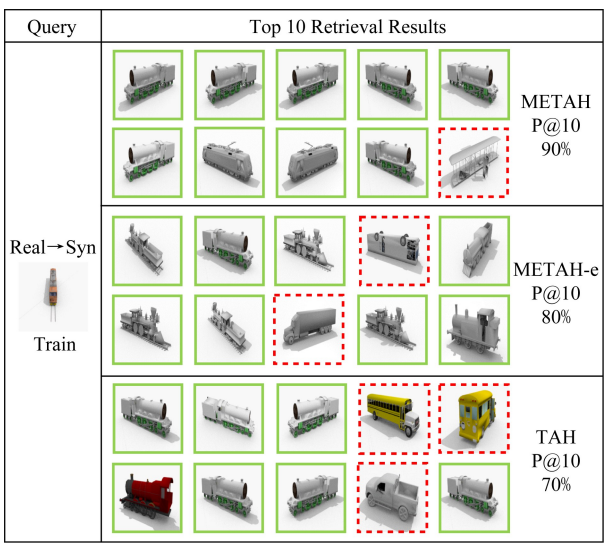


Fig. 3 Examples of top 10 retrieval images and P@10 in Real→Syn

图 3 在 Real→Syn 上前 10 检索结果和 P@10 值

所示,我们可以观察到相似的现象,即 METAH-e 和 METAH 的查询结果相比于 TAH 错误结果更少.

4 总 结

在本文中,我们针对现有深度跨域图像检索方法所学散列码判别力和域不变能力不足这 2 个问题,提出了语义保持模块和最小熵损失来改进现有的模型.语义保持模块能够使所学到的散列码包含更多的语义信息.最小熵能使目标域样本与源域样本在语义空间上分布更加对齐,使得散列码更具域不变性.大量的实验表明我们的模型相比于领域内主要模型取得了更优的性能,验证了所提改进技术的有效性.

参 考 文 献

[1] Liong V E, Lu Jiwen, Wang Gang, et al. Deep hashing for compact binary codes learning [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 2475-2483

[2] Li Wujun, Wang Sheng, Kang Wangcheng. Feature learning based deep supervised hashing with pairwise labels [C] //Proc of the 25th Int Joint Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2016: 1711-1717

[3] Liu Haomiao, Wang Ruiping, Shan Shiguang, et al. Deep supervised hashing for fast image retrieval [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 2064-2072

[4] Xia Rongkai, Pan Yan, Lai Hanjiang, et al. Supervised hashing for image retrieval via image representation learning [C] //Proc of the 25th AAAI Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2014: 2156-2162

[5] Lai Hanjiang, Pan Yan, Liu Ye, et al. Simultaneous feature learning and Hash coding with deep neural networks [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 3270-3278

[6] Zhu Han, Long Mingsheng, Wang Jianmin, et al. Deep hashing network for efficient similarity retrieval [C] //Proc of the 30th AAAI Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2016: 2415-2421

[7] Cao Zhangjie, Long Mingsheng, Wang Jianmin, et al. HashNet: Deep learning to Hash by continuation [C] //Proc of IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2017: 5609-5618

[8] Gionis A, Indyk P, Motwani R. Similarity search in high dimensions via hashing [J]. International Journal on Very Large Data Bases, 1999, 99(6): 518-529

[9] Weiss Y, Torralba A, Fergus R. Spectral hashing [C] //Proc of the Advances in Neural Information Processing Systems. Cambridge, MA: MIT, 2009: 1753-1760

[10] Gong Yunchao, Lazebnik S. Iterative quantization: A procrustean approach to learning binary codes [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2011: 817-824

[11] Liu Wei, Wang Jun, Kumar S, et al. Hashing with graphs [C] //Proc of the 28th Int Conf on Machine Learning. New York: ACM, 2011: 1-8

[12] Liu Wei, Wang Jun, Ji Rongrong, et al. Supervised hashing with kernels [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2012: 2074-2081

[13] Shen Fumin, Shen Chunhua, Liu Wei, et al. Supervised discrete hashing [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 37-45

[14] Tzeng E, Hoffman J, Zhang Ning, et al. Deep domain confusion: Maximizing for domain invariance [J]. arXiv preprint, arXiv:1412.3474, 2014

[15] Long Mingsheng, Cao Yue, Wang Jianmin, et al. Learning transferable features with deep adaptation networks [C] //Proc of the 32nd Int Conf on Machine Learning. New York: ACM, 2015: 97-105

[16] Sun Baochen, Saenko K. Deep CORAL: Correlation alignment for deep domain adaptation [J]. arXiv preprint, arXiv:1607.01719, 2016

[17] Zhuo Junbao, Wang Shuhui, Zhang Weigang, et al. Deep unsupervised convolutional domain adaptation [C] //Proc of the 25th ACM Int Conf on Multimedia. New York: ACM, 2017: 261-269

[18] Zhuo Junbao, Wang Shuhui, Cui Shuhao, et al. Unsupervised open domain recognition by semantic discrepancy minimization [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 750-759

[19] Yaroslav G, Victor L. Unsupervised domain adaptation by backpropagation [J]. arXiv preprint, arXiv: 1409.7495v1, 2014

[20] Venkateswara H, Eusebio J, Chakraborty S, et al. Deep hashing network for unsupervised domain adaptation [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 5018-5027

[21] Cao Zhangjie, Long Mingsheng, Huang Chao, et al. Transfer adversarial hashing for Hamming space retrieval [C] //Proc of the 32nd AAAI Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2018: 6698-6705

[22] Chen Jianfei, Li Chongxuan, Ru Yizhong, et al. Population matching discrepancy and applications in deep learning [C] // Proc of Advances in Neural Information Processing Systems. Cambridge, MA: MIT, 2017: 6262-6272

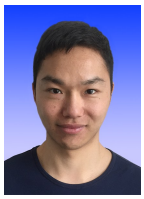
[23] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks [C] //Proc of Advances in Neural Information Processing Systems. Cambridge, MA: MIT, 2016: 2672-2680

[24] Tzeng E, Hoffman J, Saenko K, et al. Adversarial discriminative domain adaptation [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 7167-7176

[25] Long Mingsheng, Cao Zhangjie, Wang Jianmin, et al. Conditional adversarial domain adaptation [C] //Proc of Advances in Neural Information Processing Systems. Cambridge, MA: MIT, 2018: 1640-1650

[26] Zhao Feng, Huang Yongzhen, Wang Liang, et al. Deep semantic ranking based hashing for multi-label image retrieval [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 1556-1564

[27] Cao Zhangjie, Long Mingsheng, Wang Jianmin, et al. Transitive hashing network for heterogeneous multimedia retrieval [C] //Proc of the 31st AAAI Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2017: 81-87



Zhuo Junbao, born in 1991. PhD candidate. His main research interests include machine learning, deep learning and transfer learning.

Su Chi, born in 1983. PhD. His main research interests include computer vision, deep learning, content understanding.

Wang Shuhui, born in 1983. PhD. Associate professor. His main research interests include semantic image analysis, image and video retrieval and large-scale Web multimedia data mining.

Huang Qingming, born in 1965. PhD, professor, PhD supervisor. His main research interests include multimedia computing, image/video processing, pattern recognition, and computer vision.

《计算机研究与发展》征订启事

《计算机研究与发展》(Journal of Computer Research and Development)是中国科学院计算技术研究所和中国计算机学会联合主办、科学出版社出版的学术性刊物,中国计算机学会会刊.主要刊登计算机科学技术领域高水平的学术论文、最新科研成果和重大应用成果.读者对象为从事计算机研究与开发的研究人员、工程技术人员、各大专院校计算机相关专业的师生以及高新企业研发人员等.

《计算机研究与发展》于1958年创刊,是我国第一个计算机刊物,现已成为我国计算机领域权威性的学术期刊之一.并历次被评为我国计算机类核心期刊,多次被评为"中国百种杰出学术期刊".此外,还被《中国学术期刊文摘》、《中国科学引文索引》、"中国科学引文数据库"、"中国科技论文统计源数据库"、美国工程索引(Ei)检索系统、日本《科学技术文献速报》、俄罗斯《文摘杂志》、英国《科学文摘》(SA)等国内外重要检索机构收录.

国内邮发代号:2-654;国外发行代号:M603

国内统一连续出版物号:CN11-1777/TP

国际标准连续出版物号:ISSN1000-1239

联系方式:

100190 北京中关村科学院南路6号《计算机研究与发展》编辑部

电话:+86(10)62620696(兼传真);+86(10)62600350

Email:crad@ict.ac.cn

http://crad.ict.ac.cn