

# 基于异构卷积神经网络集成的 无监督行人重识别方法

彭锦佳<sup>1</sup>, 王辉兵<sup>2</sup>

(1. 河北大学网络空间安全与计算机学院, 河北保定 071000; 2. 大连海事大学信息科学技术学院, 辽宁大连 116026)

**摘要:** 行人重识别旨在从不同的摄像头中识别目标行人的图像。由于不同场景之间存在域偏差, 在一个场景中训练好的重识别模型无法直接应用在另一个场景中。为克服该问题, 现有的无监督行人重识别方法倾向通过使用聚类算法获得伪标签, 再利用伪标签训练重识别模型。但是, 由于聚类结果是不准确的, 这类方法会引入大量噪声标签, 从而限制了模型的泛化能力。因此, 为减轻噪声伪标签的影响, 本文提出了一种基于异构卷积神经网络集成的无监督行人重识别方法。该框架不使用任何人工标记信息, 自动推测目标域中行人图像之间的关系, 并构建协作可信实例选择机制, 选择可信度高的伪标签用于模型的训练。通过设计双分支异构卷积神经网络学习判别能力强的多种行人特征, 并利用记忆单元存储训练过程中的全局特征, 减少因噪声标签在训练过程中产生的波动, 提高模型的鲁棒性。本文方法在多个公开行人数据集上进行了验证并得到了良好的实验结果。在 Market1501 和 DukeMTMC-reID 数据集上, mAP 分别达到了 85.4% 和 74.8%。

**关键词:** 行人重识别; 异构卷积神经网络; 协作可信实例选择; 噪声平滑; 自适应更新

**基金项目:** 国家自然科学基金(No.62002041); 河北大学高层次人才科研启动项目(No.521100221029)

**中图分类号:** TP391.4 **文献标识码:** A **文章编号:** 0372-2112(2023)10-2902-13

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.12263/DZXB.20220467

## An Unsupervised Person Re-Identification Method Based on Heterogeneous Convolutional Neural Networks Ensemble

PENG Jin-jia<sup>1</sup>, WANG Hui-bing<sup>2</sup>

(1. College of Cyberspace Security and Computer, Hebei University, Baoding, Hebei 071000, China;

2. School of Information Science and Technology, Dalian Maritime University, Dalian, Liaoning 116026, China)

**Abstract:** Person re-identification (re-ID) aims to identify a person's images across different cameras. However, the domain bias between different datasets makes it a challenge for re-ID models trained on one dataset to be adapted to another. A variety of unsupervised domain adaptation methods tend to transfer learned knowledge from one domain to another by optimizing with pseudo-labels. However, these methods introduce a large number of noisy labels through one-shot clustering, which hinders the retraining process and limits generalization. To mitigate the impact of noisy pseudo-labels, this paper proposes an unsupervised person re-identification method based on an ensemble of heterogeneous convolutional neural networks. The framework does not apply any manual labeling information, automatically infers the relationship between pedestrian images in the target domain, and a cooperative trusted instance selection mechanism is established to select pseudo-labels with high credibility. By constructing a dual-branch heterogeneous network, a variety of different pedestrian features are learned, and memory structures are designed to store the life-long features during the training stage, which could reduce the fluctuation of noise labels, and improve the robustness of the model. Comprehensive experimental results have demonstrated that our proposed method can achieve excellent performances on benchmark datasets. And mAP is increased to 85.4% and 74.8% on Market1501 and DukeMTMC-reID, respectively.

**Key words:** person re-identification; heterogeneous convolutional neural networks; collaborative trusted instance selection; noise smoothing; adaptive updating

**Foundation Item(s):** National Natural Science Foundation of China (No.62002041); Hebei University High-level

Scientific Research Foundation for the Introduction of Talent (No.521100221029)

## 1 引言

行人重识别的目标是在不同摄像头视域中检索属于同一身份的行人图像。由于其广泛的潜在应用(例如,犯罪侦查、智能零售),近年来重识别研究经历了爆炸式增长。但是,大多数现有的方法侧重于全监督式学习,即训练数据和测试数据来自同一场景(域)并具有标签,这类方法取得了显著的性能<sup>[1-4]</sup>。然而,由于拍摄环境(背景、光照和拍摄角度)不断动态变化,场景之间存在明显的域偏差。在一个场景(源域)中训练的重识别模型被直接应用到另一个场景(目标域)中时,性能会明显下降。除此之外,人工标记多个摄像头下相同的行人图像是极其困难的。因此,不依赖标签的无监督重识别得到了越来越多的关注。

相对于全监督式重识别任务,无监督重识别更具有挑战性。由于目标域的行人图像没有标签,一些方法<sup>[5,6]</sup>考虑利用生成对抗网络,将行人的信息从源域转移到目标域,从而减小域偏差,例如 ECN<sup>[5]</sup>和 HHL (Hetero-Homogeneous Learning)<sup>[6]</sup>等。但是,这类方法会受到生成图像的限制,无法真正生成与目标域图像风格分布一致的图像。并且,由于没有缺少目标域中的行人图像作为参考,学习到的重识别模型性能往往不高。还有一些方法<sup>[7,8]</sup>考虑使用自动分配标签的策略。这类方法通常利用聚类算法为无标签数据分配伪标签,再利用具有伪标签的数据重新训练和更新重识别模型。聚类阶段和训练阶段交替进行,使模型逐渐适应目标域,例如 SPCL (Self-Paced Contrastive Learning)<sup>[7]</sup>和 CycAs (Cycle Association)<sup>[8]</sup>。虽然基于伪标签的方法已经得到了显著的效果,但仍存在两个主要的限制:(1)由于当前重识别模型不是最佳特征提取器,尤其是初始阶段的重识别模型,得到的特征并不能很好地表达行人的特点;(2)利用聚类算法为数据分配的是存在大量噪声的不准确伪标签,这些标签会给模型的训练带来一定程度的影响和限制。

因此,为减轻噪声标签的影响并训练鲁棒性高的重识别模型,本文提出了一种基于异构卷积神经网络集成的无监督行人重识别框架(Heterogeneous Convolutional Neural Network Embedded framework, HCNNE)。该框架通过设计包含多个不同卷积神经网络的分支结构来抵抗和削弱噪声标签的影响。在该框架中,针对限制1中标签生成过程受特征表达的影响,本文设计了异构双分支网络,利用多个分支网络获取多视角特征,以全面描述行人图像,提高特征表达能力。针对限制2,本文设计了协作可信实例选择模块。在该模块中,充分利用多分支输出特征的异构性,以对方作为参考,选择在

当前分支中可信度较高的实例,提高伪标签的准确性。此外,本文还构建了记忆单元存储训练过程中输出的特征,并自适应更新记忆单元中的平均特征,以此平滑噪声标签对训练过程产生的影响。综上所述,本文的主要贡献如下。

(1)提出了一种基于异构卷积神经网络集成的无监督行人重识别框架HCNNE。该框架不使用任何人工标记信息,自动推测目标域中行人图像之间的关系。并且,利用记忆单元存储训练过程中的全局特征,减少因噪声标签在训练过程中产生的波动,提高模型的鲁棒性。

(2)设计了双分支异构卷积神经网络学习判别能力强的多种行人特征。在该网络中,除标准主干网络外,还利用正交正则化的特征学习网络减少特征冗余,提高特征表达能力。基于双分支结构,本文构建了协作可信实例选择策略,两个分支互相以对方为参考,有条件地过滤噪声标签,提高伪标签的可靠性。

## 2 国内外现状

### 2.1 无监督行人重识别

针对无监督行人重识别任务,现有的工作取得了一定的进展。与全监督行人重识别不同的是,无监督行人重识别任务不需要任何标签信息,可自动分配伪标签<sup>[4]</sup>。例如,PUL (Progressive Unsupervised Learning)<sup>[9]</sup>交替执行分配标签和在目标域中训练重识别模型以提高模型的适应能力。考虑到标签的准确性对重识别任务是非常重要的,因此,为得到更准确的标签,2019年,Lin等人<sup>[10]</sup>提出了一种自上而下的聚类方法用于从粗到细的分配标签,以得到细化后的标签信息。MMCL (Memory-based Multi-label Classification Loss)<sup>[11]</sup>利用相似性计算以及循环一致性预测伪标签,并将模型作为多分类问题进行训练。ICE (Inter-instance Contrastive Encoding)<sup>[12]</sup>使用成对相似度排序作为实例的硬标签,再把相似度分数作为软标签,以增强视图和原始图之间的一致性,降低模型对噪声扰动的敏感度。Li等人<sup>[13]</sup>从每个聚类后的集群中分离出一部分不可靠的数据点,从而保留可靠性高的数据用于模型训练,并提出集群自适应平衡损失训练具有不平衡伪标签的网络。考虑到一些特征易受图像颜色的影响,Li等人<sup>[14]</sup>试图抑制颜色的负面主导影响,提出了基于集群引导的非对称对比学习方法,利用实例级和集群级对比学习关于RGB空间和灰度空间中数据的判别性特征。Cho等人<sup>[15]</sup>考虑到细粒度局部上下文信息的重要性,提出了基于局部的伪标签细化框架,利用全局特征和局部特

征之间的互补关系来降低标签噪声的影响。Zhang 等人<sup>[16]</sup>提出了一种在训练的迭代过程中使用聚类共识矩阵改进伪标签的策略,引入时序嵌入以优化噪声伪标签。在上述方法中,大部分都只考虑了生成图像的一种标签。但是在现实生活中,对行人的介绍可以由多角度的描述构成。因此,只生成一种标签限制了学习网络对图像特征的学习。

行人图像的时空信息可通过摄像头的位置和拍摄时间获得,并不需要人工标记。因此,一些学者利用该信息构建重识别模型。例如 Xuan 等人<sup>[17]</sup>提出了动态交叉熵损失用于处理噪声数据,并结合相机感知元学习算法降低摄像头之间的域偏差。Yang 等人<sup>[18]</sup>利用相机内图像的相似度生成图像的伪标签,再将每个样本在不同类别上的分类概率作为新的特征向量进行模型的学习,以缓解相机之间的分布差异,生成更可靠的伪标签。Wang 等人<sup>[19]</sup>考虑到摄像头角度变化会引起较大的类内方差,提出将每个聚类后的集群依据摄像头信息拆分为多个小类,并结合摄像头内和摄像头间的对比学习,提高对行人的判别能力。摄像头信息虽然可以提高行人图像之间的联系,但是现有的一部分数据集或者真实应用场景中的一些杂乱图像可能不具有该信息。所以,本文方法没有考虑加入摄像头约束。

考虑到行人数据的隐私问题,Zhuang 等人<sup>[20]</sup>设计了分布式联邦学习框架训练行人重识别模型,通过边云联合优化解决了多个边缘设备上行人重识别数据的异构性问题。Wu 等人<sup>[21]</sup>在 2021 年提出了一种联邦行人重识别的新范式,通过同时学习多个隐私保护的本地模型来构建可通用的全局模型。每个客户端由一个用于视觉特征提取的特征嵌入网络和一个用于分类的映射网络组成,并且在每个本地客户端中,额外使用本地专家来规范本地客户端模型的训练过程。数据增强也是重识别任务常用的一种技术手段。例如,Chen 等人<sup>[22]</sup>提出了一种基于网格的视角生成器,将网格投影作为参考生成新的行人图像,从而扩充数据集,得到多样性的行人图像。基于联邦学习的重识别方法虽然保护了行人图像的隐私,但是客户端和服务端需要频繁地通信以及多个客户端同时训练,会导致模型收敛慢,训练时间长。

## 2.2 记忆字典

较新的几项基于动态记忆字典的无监督视觉表示学习研究取得了较好的效果<sup>[23-25]</sup>。记忆字典依靠在训练阶段不断更新自身存储的特征,平滑噪声引起的波动,以促进无监督学习。因此,一些工作将记忆字典引入无监督行人重识别方法中,利用记忆结构中的加权特征更好地学习判别能力强的模型<sup>[5,11,26]</sup>。在训练过程中,记忆单元中的特征会根据属于同一类别的实例特

征进行更新。但是,由于实例数量的不平衡,每个类别的更新进度不一致。例如,若某一类别具有的实例数量较多,则相应的内存单元的更新速度会比具有较少实例的类别快。此外,由于利用聚类得到的伪标签是不准确的,并且对同一簇内的实例分配相同的伪标签,这些具有伪标签的实例在更新阶段对记忆单元的更新是平等的,因此,噪声标签很容易污染记忆单元中存储的特征。所以,本文讨论了几种现有的记忆字典更新方法,并尝试利用标签权重字典更新记忆单元,以保持记忆单元特征的纯粹性。

综上所述,为未标记的样本分配伪标签是无监督重识别中重要的一个环节。并且,现有的无监督重识别模型的准确率远低于一些基于真实标签的全监督模式的重识别模型。因此,本文重点关注如何充分利用伪标签之间的关系,削弱噪声标签对模型训练的影响。

## 3 异构卷积神经网络集成的无监督行人重识别方法

### 3.1 方法概述

如图 1 所示,为减弱噪声标签信息对重识别模型训练的影响,异构卷积神经网络集成的无监督行人重识别框架 HCNNE 主要包括:(1)利用集成异构卷积神经网络提取多样性特征,并通过聚类算法得到各自对应的伪标签信息;(2)利用协作可信实例选择策略挑选可靠标签,并构造标签权重字典;(3)利用具有可靠标签的数据集训练集成异构卷积神经网络,并通过记忆存储单元降低噪声标签对训练过程产生的影响,以训练判别能力高的重识别模型;(4)最后生成重识别模型,并再次转到第一步,利用该模型提取特征,进入下一次迭代过程。详细过程如图 1 所示。

### 3.2 集成异构卷积神经网络

目标域的实际类别数未知,而大多数现有的方法都只利用行人某一个角度的特征进行聚类,以得到伪标签。但是,仅仅一个标签是无法对行人进行全面描述的。例如,“该行人是穿着红色的裙子、背着双肩包的女性”使用了三个标签形容该行人的主要特点。所以,在 HCNNE 中构造了一种集成异构卷积神经网络,旨在从多角度描述行人,以增加特征的多样性。

(1)网络总体结构。本文构建的集成异构卷积神经网络具有两个分支,如图 1 所示,分别为基础的主干网络分支和正则化网络分支。基础分支用于提取行人图像的全局深度特征,而正则化分支用于提取冗余较少的正则化特征。对于每个分支的输出特征,在 HCNNE 中都建立了对应的记忆单元。该单元在训练过程中,利用行人特征不断进行更新,从而可存储各个分支的全程特征。该特征则被用于损失计算,从而平滑噪声标签

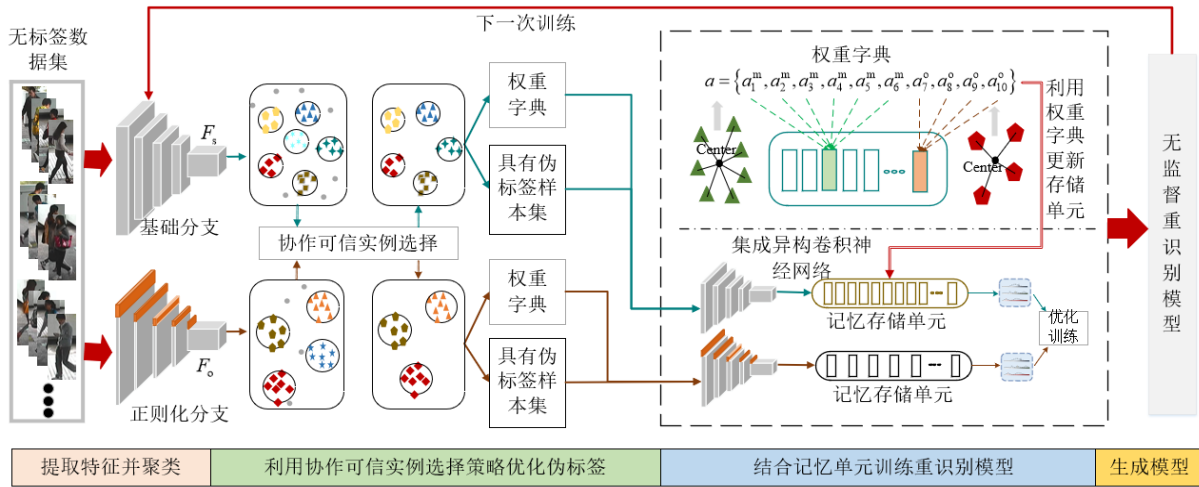


图1 基于异构卷积神经网络集成的无监督行人重识别方法框架图

产生的波动,提高模型的鲁棒性.

(2)正交正则化网络.在减少复杂网络带来的参数量升高的同时,又要提取与基础网络不同的特征,因此,HCNNE在基础网络上加入了正交正则化约束.在不增加任何参数的前提下,通过约束基础网络中各层的权重正交性有效减少提取特征的信息冗余.具体地说,在正则化网络中,HCNNE通过约束  $W^T W - I$  的特征值保证卷积参数的正交性,该过程可表达为

$$\|\beta(W^T W - I)\|_2^2 \quad (1)$$

其中,  $\beta(\cdot)$  表示矩阵的特征值,  $\|\cdot\|_2^2$  表示 2 范数的平方和,  $W$  表示参数矩阵,  $I$  为单位矩阵.通过最小化式(1)可有效缩小  $W^T W$  最大特征值和最小特征值之间的距离,从而提升各层权重矩阵的正交性.为减少对内存和时间的需求,HCNNE使用幂迭代法来替代特征值分解求解  $W^T W - I$  的特征值.在幂迭代法中,随机初始化变量  $v$ ,将  $W^T W - I$  左乘  $v$ ,利用所得到的向量更新  $v$ ,不断重复该过程,可得到  $W^T W - I$  的特征向量.在一般的迭代过程中,只需执行两次更新,即可收敛,并求得对应矩阵的特征值.该过程可表示为

$$u \leftarrow (W^T W - I)v \quad (2)$$

$$v \leftarrow (W^T W - I)u \quad (3)$$

$$\beta(W^T W - I) \leftarrow \frac{\|v\|}{\|u\|} \quad (4)$$

(3)模型协同训练.在训练过程中,使用基础网络和正则化网络在目标数据集上分别提取特征,得到两组特征  $F_s$  和  $F_o$  后,将通过聚类算法获得两组具有伪标签的数据集  $U_a$  和  $U_b$ ,再利用  $U_a$  和  $U_b$  训练基础网络和正则化网络.但是,需要注意的是,并不是仅使用  $U_a$  训练基础网络,使用  $U_b$  训练正则化网络,为了共享获得多样性特征,  $U_a$  和  $U_b$  都要经过两个网络进行特征提取操作,如图2所示,并且更新对应网络记忆存储结构中的

特征,提高特征的表达能力.因此,本文的总体损失可表示为

$$L = L_{sa} + L_{oa} + \lambda_1(L_{sb} + L_{ob}) + \lambda_0 L_o \quad (5)$$

其中,  $\lambda_0$  和  $\lambda_1$  为超参数;  $L_{sa}$  表示数据集  $U_a$  中的数据经过基础网络计算的损失;  $L_{sb}$  表示数据集  $U_b$  中的数据经过基础网络计算的损失;  $L_{oa}$  和  $L_{ob}$  分别表示数据集  $U_a$  和  $U_b$  中的数据经过正则化网络计算的损失.  $L_{sa}$ ,  $L_{sb}$ ,  $L_{oa}$  和  $L_{ob}$  计算方式相似,以  $L_{sa}$  为例,对于输入图像  $g$ ,首先计算  $g$  的特征与保存在记忆存储结构中的特征  $M$  之间的余弦相似度,然后利用  $\log$ -Softmax 函数计算出类别的预测概率,  $L_{sa}$  可表示为

$$L_{sa} = \log \frac{\exp(M^T \cdot f_g) / \tau}{\sum_{j=0}^N \exp(M_j^T \cdot f_g) / \tau} \quad (6)$$

其中,  $\tau$  是超参数;  $N$  为样本数.同理,利用上述公式可求得损失  $L_{sb}$ ,  $L_{ob}$ .  $L_o$  表示约束正则化网络正交性的损失,根据上一节的分析,可利用下式进行计算:

$$L_o = \|\beta(W^T W - I)\|_2^2 \quad (7)$$

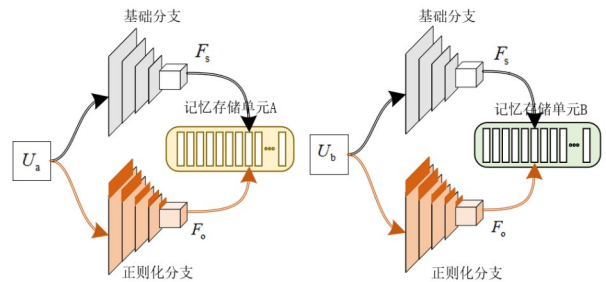


图2 异构集成网络训练过程示意图

### 3.3 协作可信实例选择

基于聚类的无监督重识别方法的性能主要受伪标签质量的限制.由于聚类结果是不准确的,不可避免地

会产生错误标签. 因此, 为了减少错误标签的影响, 本节提出协作可信实例选择策略, 用于在目标域中选择具有可靠伪标签的样本, 从而提高模型学习跨域判别表示的能力. 首先使用聚类算法对两个分支输出的特征分别进行聚类, 可得到两组具有不同伪标签的数据集:

$$C_s = \text{Clustering}(U, F_s) \quad (8)$$

$$C_o = \text{Clustering}(U, F_o) \quad (9)$$

其中,  $U$  是无标签数据集,  $F_s$  表示基础网络输出的特征,  $F_o$  表示正则化网络分支输出的特征,  $C_s$  和  $C_o$  为聚类获得的簇信息.

具体来说, 利用集成的异构卷积神经网络得到两组重识别模型后,  $U$  经过这两组模型输出相应的特征, 再通过聚类算法得到两种聚类结果. 得到聚类结果后, 先以正则化网络输出的特征聚类结果为参考, 基础网络聚类结果利用以下指标来衡量一个实例的不确定性:

$$u(x_i) = \frac{|I_s \cap I_o|}{I_s} \in [0, 1] \quad (10)$$

其中,  $I_s$  和  $I_o$  分别表示利用基础网络和正则化网络得到的特征聚类后包含  $x_i$  的簇. 同样, 对于正则化网络, 以基础网络聚类结果为参考, 利用以下指标衡量:

$$u(x_i) = \frac{|I_s \cap I_o|}{I_o} \in [0, 1] \quad (11)$$

$I_s$  和  $I_o$  重叠越多,  $u(x_i)$  越大.  $u(x_i)$  可以衡量  $x_i$  的可靠程度. 在每次迭代训练中, 设置阈值  $\theta \in [0, 1]$ , 将  $u(x_i)$  与  $\theta$  进行比较, 选择  $u(x_i) > \theta$  的实例. 如图 3 所示,  $A$  和  $B$  分别表示基础分支和正则化分支的聚类结果. 图 3(a) 为不可靠的簇. 因为若阈值设置较大, 无论以  $A$  和  $B$  中任何一个簇为参考, 结果可能都无法超过阈值. 图 3(b) 为以  $B$  为参考时,  $A$  中可靠的簇, 在图 3(b) 中, 只需要  $A$  和  $B$  中的重合样本的数量在  $A$  中占比超过了设置的阈值, 对于  $A$  来说就是可靠的簇. 图 3(c) 同图 3(b), 只需要  $A$  和  $B$  中的重合样本的数量在  $B$  中占比超过了设置的阈值, 对于  $B$  来说就是可靠的簇.

### 3.4 记忆单元更新策略

记忆单元存储结构具有键结构和值结构, 值结构存储的是网络输出的特征, 而键结构存放该特征对应的标签信息. 以一个分支为例, 首先利用聚类后的特征初始化记忆存储模块:

$$M_c \leftarrow U \left( \text{Avg} \left( \sum_{f \in B_c} f \right) \right) \quad (12)$$

其中,  $U(\cdot)$  表示  $L_2$  正则化函数; Avg 表示均值操作;  $f$  表示实例特征向量;  $B_c$  表示聚类后类别标记为  $c$  的簇;  $M_c$  表示记忆单元中标记为  $c$  的值结构存储的特征. 记忆

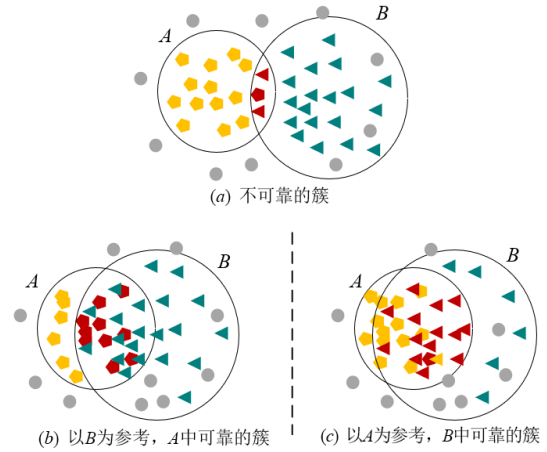


图 3 协作可信实例选择策略示意图

单元的更新方法有多种, 包括一般更新策略、难样本更新策略以及权重更新策略. 不同策略的基础都是使用同一类别中图像特征更新存储模块中的特征.

#### (1) 一般更新策略

一般的记忆单元更新策略采用无差别更新模式, 即使用同一类别中所有的图像特征按照一定的规则更新存储模块中的特征, 例如, 对于属于类别  $c$  的样本  $x_i$ , 其对应的记忆存储特征  $M_c$  可更新为

$$M_c \leftarrow \gamma M_c + (1 - \gamma) \cdot f_i \quad (13)$$

其中,  $\gamma \in [0, 1]$  是更新参数.  $\beta = 0$  表示丢弃内存字典中的原始特征,  $\gamma = 1$  表示放弃更新内存字典. 再利用记忆单元模块中的特征对模型训练过程进行约束. 这种更新方法缺少对噪声图像的容错能力, 很容易被带有噪声标签的图像特征污染.

#### (2) 难样本更新策略

文献[27]提出了一种难样本更新策略. 在模型训练过程中, 每个批次中具有  $P \times H$  张图像, 即  $P$  个行人, 每个行人具有  $H$  张图像. 在每次更新记忆单元的过程中并不使用所有的图像, 仅挑选每个行人中最难识别的实例, 以便在短时间内快速更新特征. 对于行人  $x_i$ , 其对应的记忆单元特征的更新过程可表示为

$$q_{\text{hard}} = \arg \min q \cdot M_i, q \in Q^i \quad (14)$$

$$M_c = \gamma \cdot M_c + (1 - \gamma) \cdot q_{\text{hard}} \quad (15)$$

其中,  $q_{\text{hard}}$  表示在当前的批处理集合中与特征  $M_c$  相似度最小的正样本;  $\gamma \in [0, 1]$  是更新参数;  $Q^i$  表示在当前样本集合中与行人  $x_i$  具有相同标签的样本集合.

#### (3) 权重更新策略

基于权重的更新策略衡量了具有相同伪标签的样本之间的可信度, 建立了样本的权重字典. 与中心损失类似, 本方法计算聚类中心与同一类别中样本之间的距离. 因此, 同一簇中的样本具有不同的权重并构建权重字典  $a$ , 利用  $a$  更新记忆存储结构.

在聚类之后,未标注的样本被聚类为  $N$  个簇,可表示为  $C = \{c_1, c_2, \dots, c_r, \dots, c_N\}, r=1, 2, \dots, N$ . 在训练过程中,对于一个批次的图像,计算  $x_i$  与其所在簇中心  $c_r$  的距离以及与其对应的记忆单元中的特征  $M_c$  的距离作为行人  $x_i$  的更新权重,过程可表示为

$$d_i = \text{Eu}(f_i, c_r) + \text{Cosine}(f_i, M_c) \quad (16)$$

其中,  $f_i$  是  $x_i$  的特征. 对一个批操作处理内的  $K$  张图像按照式(16)计算得到对应的距离,将属于  $c_r$  的所有样本的距离进行组合可得到集合  $d$ ,并对其中的元素按降序排列,并保存为  $d_c$ . 因此,对于属于  $c_r$  的样本,可通过  $d$  和  $d_c$  计算权重:

$$a_r = \left(1 - \frac{d}{\max(d)}\right) \cdot [\psi_{dc}(d)/(K+1)] \quad (17)$$

其中,  $\psi_{dc}(\cdot)$  是  $d$  在  $d_c$  中的索引. 由此,可得到所有图像的权重,并构成权重字典. 在每次迭代中,可利用该权重字典对记忆单元存储结构中的特征进行更新. 例如,对于属于类别  $c$  的样本,其对应的记忆存储特征  $M_c$  可更新为

$$M_c \leftarrow \gamma M_c + (1-\gamma) \cdot a \cdot f \quad (18)$$

其中,  $\gamma \in [0, 1]$  是更新参数. 基于上述内容的介绍,基于异构卷积神经网络集成的无监督行人重识别方法整体过程可归纳为算法 1.

#### 算法 1 基于异构卷积神经网络集成的无监督行人重识别方法

输入: ImageNet 预训练模型  $E_{\text{init}}$ , 目标域无标签数据集  $U$

输出: 目标域行人重识别模型  $E = [E_s, E_o]$

1. 初始化集成异构网络中的两个分支模型  $E_s^{(0)} = E_o^{(0)} = E_{\text{init}}$
2. For  $t=0$  to epoch:
3. 利用  $E_s^{(t)}$  和  $E_o^{(t)}$  分别生成  $U$  中图像对应的特征  $F_s$  和  $F_o$ ;
4. 利用 DBSCAN 处理  $F_s$  和  $F_o$ , 并分配伪标签, 分别得到数据集  $U_s$  和  $U_o$ ;
5. 利用协作可信实例选择策略(式(10)和式(11)), 得到优化后数据集  $U_s$  和  $U_o$ ;
6. 建立  $U_s$  和  $U_o$  中实例对应的权重字典;
7. 利用式(5)对应的损失和具有权重字典的  $U_s$  和  $U_o$  训练  $E_s^{(t+1)}$  和  $E_o^{(t+1)}$ ;
8. 利用 3.4 节中的基于权重的更新策略(式(16)~(18))更新记忆单元.
9. End

## 4 实验结果与分析

### 4.1 数据集介绍

Market-1501<sup>[28]</sup>中共有 1 501 个行人的 32 668 张图像. 训练集具有 751 个行人, 共 12 936 张图像. 图库集有 750 人, 共 19 732 张图像. 测试集有 750 人, 共 3 368 张图像.

DukeMTMC-reID<sup>[29]</sup>中包含 1 812 个行人的图像. 训练集包含 702 个行人, 共 16 522 张图像. 查询集包含

702 个人的 2 228 张图像. 图库集包含 1 110 个人的 17 661 张图像. 在所有的图像中, 有 1 404 个人至少出现在 2 个摄像头下.

MSMT17<sup>[30]</sup>来源于真实场景中不同摄像头的监控视频. 数据集包含 4 101 个行人的 126 441 幅图像. 这些图像由 15 个摄像头捕获并由 Faster RCNN 进行检测. 训练集由 1 041 个行人的 32 621 张图像构成. 查询集和图库集分别包含 3 060 个行人的 11 659 和 82 161 张图像. 因此, 与其他数据集相比, MSMT17 更接近真实场景, 更具有挑战性. 表 1 列出了各数据集中训练集、查询集和图库集的构成.

表 1 各数据集中训练集、查询集和图库集的构成 单位:张

数据集	训练集	查询集	图库集
Market1501	12 936	3 368	19 732
DukeMTMC-reID	16 522	2 228	17 661
MSMT17	32 621	11 659	82 161

### 4.2 评价方法

本文在 Market-1501<sup>[28]</sup>, DukeMTMC-reID<sup>[29]</sup> 和 MSMT17<sup>[30]</sup>数据集上进行实验, 并采用累积匹配特征曲线 (Cumulative Match Characteristic, CMC) 以及平均查准率 (mean Average Precision, mAP) 来评价实验性能. 其中, 累积匹配特征曲线表示查询图像出现在排序后图库列表中的概率.

### 4.3 实验设置

本节实验的实验平台为 Pytorch<sup>[31]</sup>, 使用的显卡型号为 NVIDIA GTX A6000. 所有实验的重识别模型参数均利用 Adam 算法进行优化. 在模型训练阶段, 基础网络为 ResNet50 网络, 两个分支都先利用 ImageNet 上的预训练模型进行初始化. 动量因子和批处理操作大小分别设置为 0.9 和 96. 在三个数据集上, 初始学习率均为 0.000 35. 在聚类阶段, 最小聚类样本数均设置为 4. 对于 Market1501<sup>[18]</sup>数据集, 基础网络聚类设置密度半径为 0.5, 正则化网络为 0.4. 在 DukeMTMC-reID<sup>[19]</sup>数据集上, 两个分支上的密度半径分别设置为 0.6 和 0.5. 在 MSMT17<sup>[30]</sup>上, 两个分支上的密度半径分别设置为 0.7 和 0.6. 在测试时, 除图 4 中显示的不同分支的准确率分析图外, 其他结果均为两个分支利用连接操作得到的融合特征的测试结果. 并且所有涉及双分支结构的方法中, 均先使用基础分支输出的特征更新记忆字典, 再使用正则化分支输出的特征更新记忆字典.

### 4.4 对比实验

#### 4.4.1 与最新方法对比

首先, 为验证提出方法的有效性, 本文将提出的方法与最新的无监督行人重识别方法进行了比较, 表 2、表 3 和表 4 分别是在 Market1501<sup>[28]</sup>, DukeMTMC-

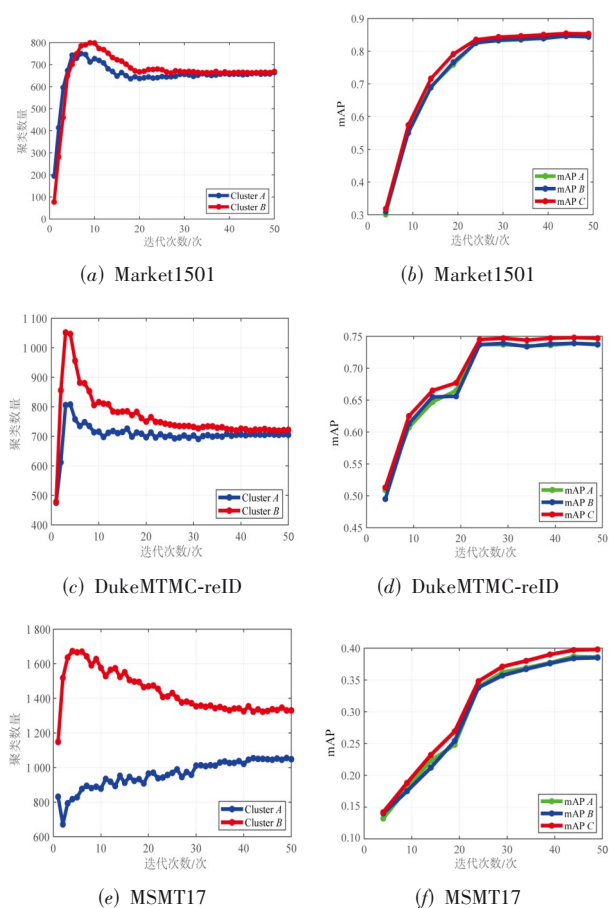


图4 迭代过程中两个分支聚类数量变化以及对应mAP的变化结果  
reID<sup>[29]</sup>, MSMT17<sup>[30]</sup>数据集上的对比结果,加粗数据表示最优结果.在不同的数据集中,本文提出的方法均取得了最好的结果.尤其是与最新方法CCU50<sup>[27]</sup>相比,本文提出的方法在3个数据集上的mAP分别提高了2.8%,2%以及6.5%.与BUC(Bottom-Up Clustering)<sup>[10]</sup>,SSL(Softened Similarity Learning)<sup>[32]</sup>和UGA(Unsupervised Graph Association)<sup>[33]</sup>等无监督行人重识别方法相比,HCNNE得到了较高的准确率,特别是与本文方法思想相似的SSL<sup>[32]</sup>.与SSL<sup>[32]</sup>不同的是,HCNNE除生成多种标签外,还构建了协作可信实例选择策略,进一步减少明显的噪声标签,降低对聚类方法结果的依赖性,并且利用记忆单元存储训练过程的平均特征,这也使本文的方法在所有数据集上的准确率都有较大的提高.

与MMT(Mutual Mean Teaching)<sup>[37]</sup>,ECN<sup>[5]</sup>等无监督域自适应重识别方法相比,利用HCNNE训练得到的重识别模型性能有了很大的提升.尤其需要注意的是,HCNNE是不依赖任何有标签数据的无监督框架,而在MMT<sup>[37]</sup>以及ECN<sup>[5]</sup>中都利用了源域图像优化重识别模型.例如,ECN<sup>[5]</sup>利用了生成对抗网络平滑了源域和

表2 Market1501数据集上的实验结果对比 单位:%

方法	源域	mAP	Rank-1	Rank-5	Rank-10
BUC <sup>[10]</sup>	None	38.3	66.2	79.6	84.5
SSL <sup>[32]</sup>	None	37.8	71.7	83.8	87.4
MMCL <sup>[11]</sup>	None	45.5	80.3	89.4	92.3
HCT <sup>[34]</sup>	None	56.4	80.0	91.6	95.2
CycAs <sup>[8]</sup>	None	64.8	84.8	—	—
ADClu+ <sup>[35]</sup>	Duke	68.3	86.7	94.4	96.5
UGA <sup>[35]</sup>	None	70.3	87.2	—	—
NRMT <sup>[36]</sup>	Duke	71.7	87.8	—	—
MMT <sup>[37]</sup>	MSMT17	75.6	89.3	95.8	97.5
JVTC+ <sup>[38]</sup>	Duke	67.2	86.8	—	—
SPCL <sup>[7]</sup>	None	73.1	88.1	95.1	97.0
CCU50 <sup>[27]</sup>	None	82.6	93.0	97.0	98.1
GLT <sup>[39]</sup>	Duke	79.5	92.2	96.5	97.8
DR <sup>[40]</sup>	Duke	78.0	90.9	96.4	97.7
C <sup>3</sup> AB <sup>[13]</sup>	None	77.0	90.0	96.4	97.6
RLCC <sup>[16]</sup>	None	77.7	90.8	96.3	97.5
CACL <sup>[14]</sup>	None	80.9	92.7	97.4	98.5
PPLR <sup>[15]</sup>	None	81.5	92.8	97.1	98.1
Ours	None	<b>85.4</b>	<b>93.3</b>	<b>97.6</b>	<b>98.7</b>

表3 DukeMTMC-reID数据集上的实验结果对比 单位:%

方法	源域	mAP	Rank-1	Rank-5	Rank-10
BUC <sup>[10]</sup>	None	27.5	47.4	62.6	68.4
SSL <sup>[32]</sup>	None	28.6	52.5	63.5	68.9
ECN <sup>[5]</sup>	Market	40.4	63.3	—	—
MMCL <sup>[11]</sup>	None	51.4	72.4	82.9	85.0
HCT <sup>[34]</sup>	None	50.7	69.6	83.4	87.4
CycAs <sup>[8]</sup>	None	60.1	77.9	—	—
ADClu+ <sup>[35]</sup>	Market	54.1	72.6	82.5	85.5
UGA <sup>[33]</sup>	None	53.3	75.0	—	—
NRMT <sup>[36]</sup>	Market	62.2	77.8	—	—
MMT <sup>[37]</sup>	Market	65.1	78.9	88.8	92.5
JVTC+ <sup>[38]</sup>	Market	66.5	80.4	—	—
SPCL <sup>[7]</sup>	Market	68.8	82.9	90.1	92.5
CCU50 <sup>[27]</sup>	None	72.8	85.7	92.0	93.5
GLT <sup>[39]</sup>	Market	69.2	82.0	90.2	92.8
DR <sup>[40]</sup>	Market	67.7	82.1	90.1	92.5
RLCC <sup>[16]</sup>	None	69.2	83.2	91.6	93.8
C <sup>3</sup> AB <sup>[13]</sup>	None	67.0	81.2	90.3	92.6
CACL <sup>[14]</sup>	None	69.6	82.6	91.2	93.8
Ours	None	<b>74.8</b>	<b>85.8</b>	<b>92.2</b>	<b>93.9</b>

目标域之间的域偏差.MMT<sup>[37]</sup>利用源域的图像训练重识别模型,并使用该模型参数初始化目标域的网络.这也说明了本文提出的HCNNE方法对新场景具有一定的鲁棒性,可自动探索出样本的内在联系.

此外,与使用记忆单元结构平滑噪声影响的

表 4 MSMT17 数据集上的实验结果对比 单位:%

方法	源域	mAP	Rank-1	Rank-5	Rank-10
ECN <sup>[5]</sup>	Duke	10.2	30.2	41.5	46.8
MMCL <sup>[11]</sup>	None	11.2	35.4	44.8	49.8
TAUDL <sup>[41]</sup>	None	12.5	28.4	—	—
UTAL <sup>[42]</sup>	None	13.1	31.4	—	—
SPCL <sup>[7]</sup>	None	19.1	42.3	55.6	61.2
UGA <sup>[33]</sup>	None	21.7	49.5	—	—
MMT <sup>[37]</sup>	Market	24.0	50.1	63.5	69.3
CycAs <sup>[8]</sup>	None	26.7	50.1	—	—
CCU50 <sup>[27]</sup>	None	33.3	63.3	73.7	77.8
GLT <sup>[39]</sup>	Duke	27.7	59.5	70.1	74.2
DR <sup>[40]</sup>	Duke	26.9	55.0	68.4	73.2
RLCC <sup>[16]</sup>	None	27.9	56.5	68.4	73.1
CACL <sup>[14]</sup>	None	23.0	48.9	61.2	66.4
PPLR <sup>[15]</sup>	None	31.4	61.1	73.4	77.8
Ours	None	<b>39.8</b>	<b>67.9</b>	<b>78.3</b>	<b>82.1</b>

SPCL<sup>[7]</sup> 和 JVTC+ (Joint Visual and Temporal Consistency)<sup>[38]</sup>等方法相比,HCNNE 在 3 个数据集上的 mAP 也取得了较大的提高. 与 SPCL<sup>[7]</sup> 和 JVTC+<sup>[38]</sup>不同的是,HCNNE 利用权重字典衡量不同伪标签的重要程

度,有差别地更新记忆单元中的特征,有效并合理地保持了记忆单元内特征的纯度,降低了噪声样本的干扰.

如图 4 所示,本文给出了在训练重识别模型过程中,HCNNE 的 2 个分支聚类结果中簇数量、mAP 在 3 个数据集上详细的变化过程. “Cluster A”和“mAP A”分别表示基础网络分支的聚类结果和 mAP. “Cluster B”和“mAP B”分别表示正则化网络分支的聚类结果和 mAP. “mAP C”表示融合两个分支的特征作为整体特征得到的准确率. 从图 4 中的簇数量变化图可以看出,利用两个分支的特征获得的簇数量虽然开始时有很大的差别,但随着模型不断迭代更新,两种类别标签趋于相同. 这也表明早期训练过程中有很多嘈杂的标签,而利用两种标签互相训练可提高模型的鲁棒性.

#### 4.4.2 消融实验

本文利用多组消融实验验证方法中 4 个重要组成部分:(1)双分支网络结构对模型训练的影响;(2)协作可信实例选择策略;(3)正交正则化约束的有效性;(4)记忆单元更新策略的有效性. 表 5 为消融实验设置说明. 表 6 和表 7 是在 Market1501<sup>[28]</sup>和 MSMT17<sup>[30]</sup>数据集上的实验结果.

表 5 消融实验设置说明

方法	双分支结构	协作可信样本选择策略	正交正则化	更新策略
Baseline	×	×	×	×
Baseline+D	√	×	×	×
Baseline+D+S	√	√	×	×
Baseline+D+S+O	√	√	√	×
Baseline+D+S+O+W	√	√	√	√

表 6 在 Market1501 数据集上的消融实验结果 单位:%

方法	Market1501			
	mAP	Rank-1	Rank-5	Rank-10
Baseline	75.6	89.3	95.8	97.1
Baseline+D	82.8	92.4	97.3	98.0
Baseline+D+S	84.0	93.1	97.6	98.1
Baseline+D+S+O	85.1	93.6	97.7	98.5
Baseline+D+S+O+W	85.4	93.3	97.6	98.7

表 7 在 MSMT17 数据集上的消融实验结果 单位:%

方法	MSMT17			
	mAP	Rank-1	Rank-5	Rank-10
Baseline	29.6	56.6	68.2	72.9
Baseline+D	32.6	59.0	69.9	74.7
Baseline+D+S	36.8	64.7	75.4	79.9
Baseline+D+S+O	38.0	66.1	76.8	80.7
Baseline+D+S+O+W	39.8	67.9	78.3	82.1

#### (1) 双分支结构有效性分析

HCNNE 中的双分支结构主要是为了从多角度训练重识别模型,提高模型表达的多样性. 从表 6 和表 7 可以看出,在 Market1501 以及 MSMT17 数据集上,与“Baseline”相比,加入双分支结构后的“Baseline+D”,模型的准确率都有提高,例如 mAP 分别上升了 7.2% 和 3%. 也可体现出本文设计的双分支结构的特征表达能力较单分支更丰富.

#### (2) 协作可信实例选择策略的有效性

由于聚类结果是不准确的,所以会生成大量的噪声标签. 因此,为了减少错误标签的影响,HCNNE 利用协作可信实例选择策略提高模型学习跨域判别表示的能力,结果如表 6 和表 7 所示,与没有协作可信实例选择策略的“Baseline+D”相比,“Baseline+D+S”在 Market1501 以及 MSMT17 数据集上的 mAP 分别提升了 1.2% 和 4.2%. 这说明了协作可信实例选择策略有一定的过滤和平滑噪声标签的能力,可以选择出可信度较

高的伪标签,具有可靠标签的数据集更有利于重识别模型的训练.

除此之外,在 Market1501 数据集上对比了本文方法去掉协作可信实例选择策略后对结果的影响,如图 5 和图 6 所示. 其中,图 5 为 mAP 结果图,“w/S”表示使用协作可信实例选择策略的结果,“w/o S”表示没有使用协作可信实例选择策略的结果. 从图 5 可以清晰地看出,使用该策略的结果要优于不使用策略的方法. 图 6 显示了使用协作可信实例选择策略前后训练样本数量的变化. 其中,“B-Before”和“B-After”分别表示基础分支应用该策略前后的训练样本数,“O-Before”和“O-After”分别表示正则化分支应用该策略前后的训练样本数. 从图 6 也可以看出:(1)在使用该策略前后,两个分支的训练样本数都有明显的数量变化,尤其是在前 20 次迭代过程中,表明在初始训练中,包含了大量的噪声标签;(2)虽然训练样本多于选择策略后的样本,但是图 5 中不使用该策略的 mAP 却低于使用该策略的 mAP 值,说明虽然样本数多有利于模型多样化训练,但是噪声标签对模型的影响也非常大.

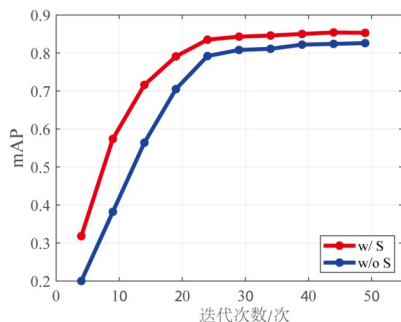


图 5 协作可信实例选择策略对 mAP 的影响结果

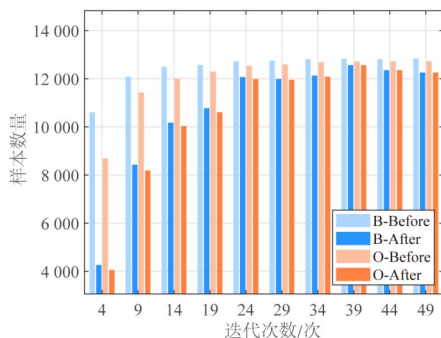


图 6 使用协作可信实例选择策略前后训练样本数量变化

### (3) 正交正则化的有效性

为了验证正交正则化约束的有效性,本文将“Baseline+D+S”与“Baseline+D+S+O”进行了对比,结果如表 6 所示. 在加入正则化约束后,对于 Market1501 数据集, mAP 和 Rank-1 分别提升了 1.1% 和 0.5%. 在 MSMT17 上, mAP 也有一定程度的提升. 结果表明,相比单一的

特征,加入正则化后模型的特征表达更多样性,也更有利于重识别任务.

### (4) 记忆单元更新策略的有效性

在本节中,首先将“Baseline+D+S+O”与“Baseline+D+S+O+W”进行了对比,从表 6 和表 7 中可以看出,利用权重字典更新记忆单元得到的准确率更高,也说明相比无差别的更新策略,基于权重的更新方法更有利于平滑噪声以及无监督重识别模型的训练. 此外,本节还分别利用 3.4 节中介绍的 3 种更新策略训练重识别模型,结果如图 7 所示. 从图 7 可以看出,在 MSMT17 数据集中,基于权重的记忆单元更新策略取得了最优的结果. 主要是由于这种更新方法充分利用了聚类结果中样本和簇中心的关系,相比单一更新策略,基于权重的方法能灵活地判断样本对应标签的重要性,从而减少记忆单元中的特征被污染的可能性.

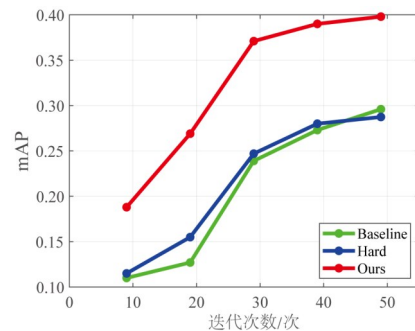


图 7 利用不同记忆单元更新方法训练得到的 mAP 的变化结果

### 4.4.3 参数分析

#### (1) 协作可信实例选择策略中的阈值分析

在协作可信实例选择策略中,需要设置相应的阈值限制得到的簇是否是可靠的. 因此,在本节中,将阈值  $\theta$  分别设置为 0.4, 0.5, 0.6, 0.7, 0.8 和 0.9, 得到对应的数据集训练重识别模型,结果如图 8 所示. 数据集选择 DukeMTMC-reID, 总体损失式(5)中的  $\lambda_0$  和  $\lambda_1$  都设置为 3. 从图 8 可以看出,当阈值设置为 0.9 时,模型取得了最优的结果 0.762, 而且不同的阈值对结果的准确性影响比较大,也间接说明了具有可靠标签的数据集对模型的重要性.

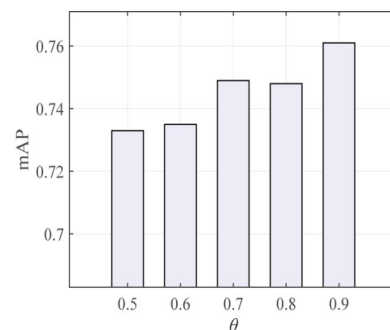


图 8 协作可信实例选择中不同阈值对 mAP 的影响结果

(2) 损失模型中的超参数分析

本文的训练损失函数主要由两部分组成,分别是ID损失和正则化约束损失,在式(5)中利用 $\lambda_0$ 平衡两者之间的关系.本节主要分析 $\lambda_0$ 对网络训练的影响,即正则化约束的影响.因此,在这组实验中,分别将 $\lambda_0$ 设为0,2,3,4和5,讨论不同 $\lambda_0$ 对结果的影响.除此之外,数据集为MSMT17,设置 $\lambda_1$ 为3,协作可信实例选择阈值 $\theta$ 为0.8,记忆字典更新策略为一般更新策略.图9显示了在不同的参数设置下得到的mAP变化图.从图9中可以看出,随着 $\lambda_0$ 的改变,mAP具有明显变化,并且在 $\lambda_0=3$ 时取得了最优的结果38.0%.也说明 $\lambda_0$ 的设置对网络训练的影响比较大.

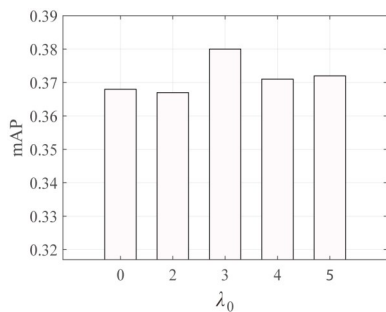


图9  $\lambda_0$ 变化对mAP的影响结果

(3) 记忆单元更新顺序分析

在3.2节中的模型协同训练过程中,基础分支的输出特征和正则化分支输出的特征都要输入到对应的记忆字典结构中进行损失计算并更新记忆单元中的特征.因此,本文将分析使用不同的更新顺序对网络训练的影响,表8给出了对应的结果.其中,如图2所示,“G”表示记忆单元G,“L”表示记忆单元L, $F_s$ 和 $F_o$ 分别表示基础分支输出特征和正则化分支输出的特征.从结果中可以看出,随着更新顺序的变化,mAP,Rank-1以及Rank-5有轻微的浮动.因此,更新顺序对本文的方法没有较大的影响.

表8 在Market1501数据集上使用不同顺序更新记忆单元的结果  
单位:%

方法	mAP	Rank-1	Rank-5	Rank-10
$G:F_s \rightarrow F_o$ $L:F_s \rightarrow F_o$	85.4	93.3	97.6	98.7
$G:F_s \rightarrow F_o$ $L:F_o \rightarrow F_s$	85.4	93.7	97.7	98.5
$G:F_o \rightarrow F_s$ $L:F_s \rightarrow F_o$	85.2	93.6	97.9	98.7
$G:F_o \rightarrow F_s$ $L:F_o \rightarrow F_s$	85.6	93.6	97.6	98.5

4.4.4 可视化

为了更好地说明本文方法的有效性,本文利用训练好的模型将DukeMTMC-reID数据集中部分图像通过GradCAM进行了可视化.图10显示了对应的结果,其中,图10(a)和(e)为原始图像,图10(b)(c)(d)分别为

使用ImageNet预训练模型、本文方法中的基础网络分支和正则化网络分支的模型生成的热图,图10(f)(g)(h)生成过程同图10(b)(c)(d).从图10可以看出,与ImageNet预训练模型相比,HCNNE中的基础网络分支和正则化网络分支更能提取到判别能力强的特征,例如背包、头部等.

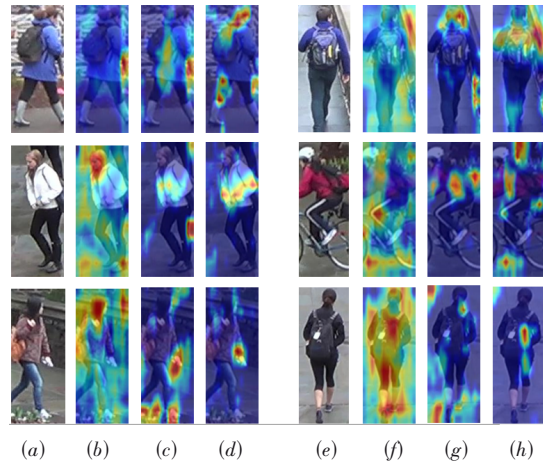


图10 利用不同模型对部分图像可视化结果

为了进一步说明本文提出方法的有效性,本文对在Market1501数据集上的部分检索结果进行了可视化.如图11所示,对于每行图像,顶部的数字表示检索到的图像与目标图像的相似性排名,数字越小表示与检索图像越相似.最左列图像为查询图像,而右侧的图像是前10个检索结果.绿色数字对应的图像是正确的检索结果,而其他图像是错误的结果.图11中前4行的

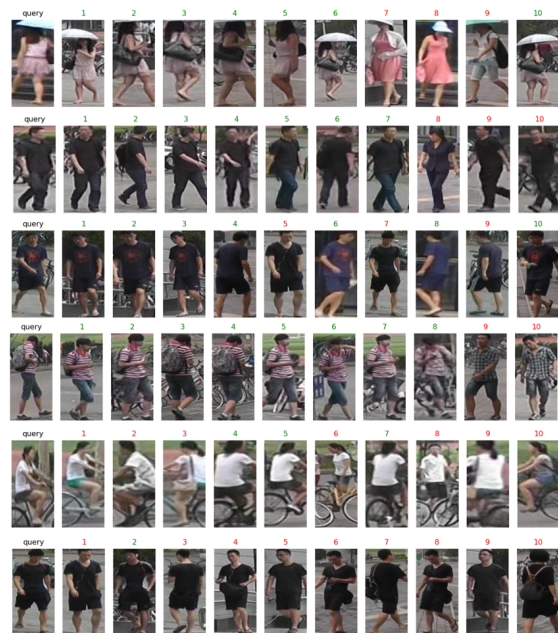


图11 利用本文模型在Market1501上的检索结果

检索结果较好,可以看出利用本文提出的方法训练得到的重识别模型对不同姿势、不同光照以及环境的图像具有很好的鲁棒性,这也说明了本文提出的方法是有效的.而后两行的检索结果中错误的示例多于正确的示例,这是由于在这两行中行人的穿着以及姿态非常相似,容易混淆,尤其是对于最后一行穿着黑色背心以及短裤的行人,在图片分辨率不高、难以进行面部识别的前提下,难以区分为不同的行人.

## 5 总结

本文提出了一种基于异构卷积神经网络集成的无监督行人重识别方法 HCNNE. 该框架通过构建双分支异构网络学习多种不同的行人特征,并建立协作可信实例选择机制,选择可信度高的伪标签.大量的实验也从多个方面说明了本文提出的框架 HCNNE 及其各部分的有效性.此外,本文发现在重识别任务中涉及了行人的隐私问题.因此,在未来的工作中将主要探索在保护行人隐私的基础上,得到有效的行人重识别模型.

### 参考文献

- [1] TANG Z M, HUANG J. Harmonious multi-branch network for person re-identification with harder triplet loss[J]. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2022, 18(4): 98.
- [2] AHMED E, JONES M, MARKS T K. An improved deep learning architecture for person re-identification[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2015: 3908-3916.
- [3] 匡澄, 陈莹. 基于多粒度特征融合网络的行人重识别[J]. *电子学报*, 2021, 49(8): 1541-1550.
- [4] KUANG C, CHEN Y. Multi-granularity feature fusion network for person re-identification[J]. *Acta Electronica Sinica*, 2021, 49(8): 1541-1550. (in Chinese)
- [4] ZHANG G Q, ZHANG H W, CHEN Y H, et al. Close-set camera style distribution alignment for single camera person re-identification[J]. *Neurocomputing*, 2022, 486: 93-103.
- [5] ZHONG Z, ZHENG L, LUO Z M, et al. Invariance matters: Exemplar memory for domain adaptive person re-identification[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 598-607.
- [6] ZHONG Z, ZHENG L, LI S Z, et al. Generalizing a person retrieval model hetero- and homogeneously[C]//European Conference on Computer Vision. Cham: Springer, 2018: 176-192.
- [7] GE Y X, ZHU F, CHEN D P, et al. Self-paced contrastive learning with hybrid memory for domain adaptive object re-ID[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2020: 11309-11321.
- [8] WANG Z D, ZHANG J W, ZHENG L, et al. CycAs: Self-supervised cycle association for learning re-identifiable descriptions[C]//European Conference on Computer Vision. Cham: Springer, 2020: 72-88.
- [9] FAN H H, ZHENG L, YAN C G, et al. Unsupervised person re-identification: Clustering and fine-tuning[J]. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2018, 14(4): 83.
- [10] LIN Y T, DONG X Y, ZHENG L A, et al. A bottom-up clustering approach to unsupervised person re-identification[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 33(1): 8738-8745.
- [11] WANG D K, ZHANG S L. Unsupervised person re-identification via multi-label classification[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 10978-10987.
- [12] CHEN H, LAGADEC B, BREMOND F. ICE: Inter-instance contrastive encoding for unsupervised person re-identification[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2022: 14940-14949.
- [13] LI M K, SUN H, LIN C Q, et al. The devil in the tail: Cluster consolidation plus cluster adaptive balancing loss for unsupervised person re-identification[J]. *Pattern Recognition*, 2022, 129: 108763.
- [14] LI M K, LI C G, GUO J. Cluster-guided asymmetric contrastive learning for unsupervised person re-identification [J]. *IEEE Transactions on Image Processing*, 2022, 31: 3606-3617.
- [15] CHO Y, KIM W J, HONG S, et al. Part-based pseudo label refinement for unsupervised person re-identification [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2022: 7298-7308.
- [16] ZHANG X, GE Y X, QIAO Y, et al. Refining pseudo labels with clustering consensus over generations for unsupervised object re-identification[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 3435-3444.
- [17] XUAN S Y, ZHANG S L. Intra-inter camera similarity for unsupervised person re-identification[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 3435-3444.

- nition (CVPR). Piscataway: IEEE, 2021: 11921-11930.
- [18] YANG F X, ZHONG Z, LUO Z M, et al. Joint noise-tolerant learning and meta camera shift adaptation for unsupervised person re-identification[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 4853-4862.
- [19] WANG M L, LAI B S, HUANG J Q, et al. Camera-aware proxies for unsupervised person re-identification[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(4): 2764-2772.
- [20] ZHUANG W M, WEN Y G, ZHANG S. Joint optimization in edge-cloud continuum for federated unsupervised person re-identification[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM, 2021: 433-441.
- [21] WU G L, GONG S G. Decentralised learning from independent multi-domain labels for person re-identification [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(4): 2898-2906.
- [22] CHEN H, WANG Y H, LAGADEC B, et al. Joint generative and contrastive learning for unsupervised person re-identification[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 2004-2013.
- [23] BACHMAN P, HJELM R D, BUCHWALTER W. Learning representations by maximizing mutual information across views[C]//Proceedings of the 33rd International Conference on Neural Information Processing System. New York: Curran Associates Inc., 2019: 15535-15545.
- [24] HE K M, FAN H Q, WU Y X, et al. Momentum contrast for unsupervised visual representation learning[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 9726-9735.
- [25] HÉNAFF O J, SRINIVAS A, DE FAUW J, et al. Data-efficient image recognition with contrastive predictive coding[C]//Proceedings of the 37th International Conference on Machine Learning. New York: ACM, 2020: 4182-4192.
- [26] XIAO T, LI S, WANG B C, et al. Joint detection and identification feature learning for person search[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 3376-3385.
- [27] DAI Z Z, WANG G Y, YUAN W H, et al. Cluster contrast for unsupervised person re-identification[EB/OL]. (2021-03-22)[2022-04]. <https://arxiv.org/abs/2103.11568>.
- [28] ZHENG L, SHEN L Y, TIAN L, et al. Scalable person re-identification: A benchmark[C]//2015 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2016: 1116-1124.
- [29] RISTANI E, SOLERA F, ZOU R, et al. Performance measures and a data set for multi-target, multi-camera tracking[C]//European Conference on Computer Vision. Cham: Springer, 2016: 17-35.
- [30] WEI L H, ZHANG S L, GAO W, et al. Person transfer GAN to bridge domain gap for person re-identification [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 79-88.
- [31] PASZKE A, GROSS S, CHINTALA S, et al. Automatic differentiation in PyTorch[C]//NIPS 2017 Autodiff Workshop. Long Beach: NIPS, 2017: 1-4.
- [32] LIN Y T, XIE L X, WU Y, et al. Unsupervised person re-identification via softened similarity learning[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 3387-3396.
- [33] WU J L, LIU H, YANG Y, et al. Unsupervised graph association for person re-identification[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2020: 8320-8329.
- [34] ZENG K W, NING M N, WANG Y H, et al. Hierarchical clustering with hard-batch triplet loss for person re-identification[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 13654-13662.
- [35] ZHAI Y P, LU S J, YE Q X, et al. AD-cluster: Augmented discriminative clustering for domain adaptive person re-identification[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 9018-9027.
- [36] ZHAO F, LIAO S C, XIE G S, et al. Unsupervised domain adaptation with noise resistible mutual-training for person re-identification[C]//European Conference on Computer Vision. Cham: Springer, 2020: 526-544.
- [37] GE Y X, CHEN D P, LI H S. Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification[C]//2020 International Conference on Learning Representations. Addis Ababa: ICLR, 2020: 1-15.
- [38] LI J N, ZHANG S L. Joint visual and temporal consistency for unsupervised domain adaptive person re-identification[C]//European Conference on Computer Vision.

Cham: Springer, 2020: 483-499.

- [39] ZHENG K C, LIU W, HE L X, et al. Group-aware label transfer for domain adaptive person re-identification[C]// 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 5306-5315.
- [40] DAI Y X, LIU J, BAI Y, et al. Dual-refinement: Joint label and feature refinement for unsupervised domain adaptive person re-identification[J]. IEEE Transactions on Image Processing, 2021, 30: 7815-7829.
- [41] LI M X, ZHU X T, GONG S G. Unsupervised person re-identification by deep learning tracklet association[C]// European Conference on Computer Vision. Cham: Springer, 2018: 772-788.
- [42] LI M X, ZHU X T, GONG S G. Unsupervised tracklet person re-identification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 42(7): 1770-1782.

#### 作者简介



彭锦佳 女, 1990年出生, 河北保定人. 现于河北大学网络空间安全与计算机学院任职. 主要研究方向为重识别和图像处理.  
E-mail: pengjinjia@hbu.edu.cn



王辉兵 男, 1989年出生, 山东烟台人. 博士、副教授. 主要研究方向为机器学习与图像处理.  
E-mail: huibing.wang@dlmu.edu.cn