

# 基于伪标签正则化损失的无监督行人重识别

贾洁茹<sup>1,2,3</sup>, 张硕蕊<sup>1,2,3</sup>, 钱宇华<sup>1,2,3</sup>, 阮秋琦<sup>4</sup>

(1. 山西大学大数据科学与产业研究院, 山西太原 030006; 2. 山西大学计算机与信息技术学院, 山西太原 030006;  
3. 山西省机器视觉与数据挖掘工程研究中心, 山西太原 030006; 4. 北京交通大学信息科学研究所, 北京 100044)

**摘要:** 无监督行人重识别旨在不需要行人身份标签的情况下, 将查询的行人图像与候选集中的行人图像相匹配. 目前主流的无监督行人重识别方法通常先利用聚类算法生成伪标签, 然后利用伪标签训练深度神经网络. 然而由于模型初始表征能力不足和聚类算法的局限性等, 伪标签中会引入大量噪声, 严重误导模型优化过程, 导致模型性能退化. 为了减轻伪标签噪声的影响, 本文提出了一种新的伪标签正则化损失函数, 用伪标签的置信度分数和样本相似度对伪标签噪声进行约束. 具体来说, 本文首先提出了一种聚类引导的注意力机制, 根据伪标签与聚类中心的语义相关程度来估计伪标签的置信度, 以此来识别噪声标签并给正确标签分配更多的权重, 有效降低伪标签噪声在总体损失函数中的作用. 同时, 为了充分利用伪标签的判别能力, 本文利用伪标签进行在线软样本挖掘, 构建 mini-batch 中的正负样本对并为每个正负样本对计算一个连续的权重分数. 通过将以上两种权重引入到对比损失中, 本文提出的伪标签正则化损失函数可以有效抑制伪标签噪声的影响, 减轻标签噪声对训练过程的影响, 提高模型的准确性和鲁棒性. 在多个公开行人数据集上的实验结果验证了本文方法的有效性, 在 Market1501、DukeMTMC-reID 和 MSMT17 数据集上 mAP 分别达到了 85.9%、75.1% 和 29.3%.

**关键词:** 行人重识别; 无监督学习; 伪标签噪声; 对比学习; 聚类优化

**基金项目:** 国家自然科学基金 (No.62106133)

**中图分类号:** TP391.41

**文献标识码:** A

**文章编号:** 0372-2112(2024)05-1743-16

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

**DOI:** 10.12263/DZXB.20230828

## Unsupervised Person Re-Identification with Pseudo Label Regularization Loss

JIA Jie-ru<sup>1,2,3</sup>, ZHANG Shuo-ru<sup>1,2,3</sup>, QIAN Yu-hua<sup>1,2,3</sup>, RUAN Qiu-qi<sup>4</sup>

(1. Institute of Big Data Science and Industry, Shanxi University, Taiyuan, Shanxi 030006, China;

2. School of Information and Technology, Shanxi University, Taiyuan, Shanxi 030006, China;

3. Engineering Research Center for Machine Vision and Data Mining of Shanxi Province, Taiyuan, Shanxi 030006, China;

4. Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China)

**Abstract:** Unsupervised person re-identification aims to match query pedestrian images with images in the gallery without the need for identity labels. Currently, mainstream unsupervised person re-identification methods typically utilize clustering algorithms to generate pseudo-labels, which are subsequently exploited to train deep neural networks. However, due to the model's inferior representation ability at early stages and the limitations of the clustering algorithms, a vast of noise is inevitably introduced into the pseudo-labels, which seriously misleads the training process and impedes the model performance. In this paper, we propose a novel pseudo-label regularization loss (PLRL) to remedy the detrimental effect of pseudo-label noises. Concretely, firstly, this paper proposes a clustering-guided attention mechanism (CGA) to estimate the confidence of pseudo-labels based on the semantic relevance between pseudo-labels and clustering centers. The CGA score is able to identify noisy labels and assign more weight to correct labels, which effectively reduces the influence of pseudo-label noise in the overall loss function. Meanwhile, for the sake of fully utilizing the discriminative power of pseudo-labels, this paper performs soft sample mining using pseudo-labels, which constructs positive and negative sample pairs in mini-batches and calculates a continuous weight score for each pair. By incorporating the confidence of pseudo-labels and the similarity of samples into the contrastive loss, the newly designed pseudo-label regularization loss can effectively alleviate

the influence of pseudo-label noise in the training process, thereby improving the accuracy and robustness of the model. Experiments and ablation studies on multiple public datasets demonstrate its effectiveness and superiority, with the mAP on Market1501, DukeMTMC-reID, and MSMT17 datasets reaching 85.9%, 75.1%, and 29.3%, respectively.

**Key words:** person re-identification; unsupervised learning; pseudo label noise; contrastive learning; clustering refinement

**Foundation Item(s):** National Natural Science Foundation of China (No.62106133)

## 1 引言

行人重识别是指从多个非重叠摄像头中拍摄到的人的图像或视频集合中找到需要查询的人。它在智能视频监控中有着广泛的应用,例如罪犯搜索、多摄像头跟踪、失踪人口搜索等。行人重识别的核心任务是学习具有判别性的人物特征,从而将查询人物与候选集中的图像或视频中最匹配的人物关联起来。有监督的行人重识别方法已经取得了非常可观的成果<sup>[1-3]</sup>,但是它需要大量的人工和计算成本来注释数据,这使得它不能扩展到具有高成本标签的大规模数据集。因此在不需行人身份标签的情况下,直接学习图像中行人特征的无监督行人重识别受到了越来越多的关注。

目前主流的无监督学习方法(UnSupervised Learning, USL)<sup>[4-6]</sup>先使用聚类算法,如k-means<sup>[7]</sup>,基于密度的聚类算法(Density-Based Spatial Clustering of Applications with Noise, DBSCAN)<sup>[8]</sup>等为未标记的数据生成伪标签,再使用伪标签训练和更新模型,以上两步迭代进行直到模型收敛。这类方法的基本假设是同一个聚类簇中的样本具有相同的类别标签。然而,与人工标注的真实标签不同,聚类算法不能保证同类样本被赋予相同的身份标签,由此产生的伪标签不可避免地会包含噪声。这些错误的标签会误导模型学习错误的特征,并且噪声标签的误差在训练过程中不断累积,从而造成模型性能严重退化。尤其在训练初始阶段,模型的表征能力较弱,会引入大量的伪标签噪声,模型有较大的崩溃风险。如何处理伪标签噪声对模型训练的性能至关重要,为了减轻伪标签噪声的影响,研究者们进行了许多探索,例如文献[9]利用同步平均教学(Mutual Mean-Teaching, MMT)框架进行伪标签优化,利用更为鲁棒的软标签对伪标签进行在线优化。文献[10]采用自步学习(Self-Paced Contrastive Learning, SPCL),逐步生成更可靠的训练聚类。文献[11]提出聚类引导的非对称对比学习(Cluster-guided Asymmetric Contrastive Learning, CACL),对原始聚类进行聚类细化,保留相似度较高的样本,去除相似度较低的样本。难样本引导混合对比学习(Hard-sample guided Hybrid Contrast Learning, HHCL)<sup>[12]</sup>充分利用硬样本的信息,引入了硬实例对比损失,进一步挖掘了判别信息。

尽管以上方法取得了不错的进展,但仍未完全消除伪标签噪声的影响,其中的关键挑战在于:(1)如何

区分含有噪声的标签和干净的标签,即如何评估伪标签的正确程度;(2)如何充分地挖掘伪标签的判别能力。针对第一个问题,本文提出了一种聚类引导的注意力机制(Cluster Guided Attention, CGA),根据伪标签与聚类中心的语义相关程度来估计伪标签的置信度。其基本思想是,含有噪声的伪标签通常是由于聚类结果不准确和不稳定导致的,因此与聚类中心的语义相关程度较小。通过这种方式,可以有效地估计伪标签的正确程度。针对第二个问题,与以往为每个样本分配一个二进制分数(0或1)的样本挖掘方法不同,本文提出一种软样本挖掘方法(Soft Sample Mining, SSM),利用伪标签构造正负样本对,并为每个样本对计算一个连续的权重分数,以充分利用小批量中的所有样本。通过将以上两个方法有机结合,本文设计了一种新的伪标签正则化损失函数(Pseudo Label Regularization Loss, PLRL),可以有效改善伪标签质量,削弱伪标签噪声对训练过程的影响。综上所述,本文的主要贡献如下:

(1)提出了一种基于聚类引导的注意力机制,根据样本与伪标签的语义相关程度估计伪标签的置信度,有效降低伪标签噪声在总体损失函数中的作用,并利用伪标签进行软样本挖掘,为每个正负样本对计算一个连续的权重分数,以充分利用伪标签的判别能力;

(2)提出伪标签正则化损失函数,将聚类引导的注意力权重和软样本挖掘分数引入到对比损失中,利用伪标签的置信度和样本相似度从两方面对伪标签噪声进行约束,有效减少伪标签噪声对训练过程的影响,提高模型的准确性和鲁棒性。

## 2 相关研究

### 2.1 无监督行人重识别

有监督学习(supervised learning)利用有标签的数据,能够提取有较强区分能力的特征,可以在行人重识别任务中取得较高的准确率。但随着数据集规模越来越大,数据标注代价也愈发昂贵,而且有监督的学习通过标签训练得到的模型往往只能学到一些特定知识,不能充分利用数据本身丰富的信息,无法学习到对图像数据的通用表征,难以迁移到其他环境下的数据上。因此,近年来基于无监督学习的行人重识别方法越来越受到研究者的关注。

目前主流的 USL 主要采用基于聚类的伪标签法,具体做法是先用聚类算法对无标签的数据图像特征进行聚类,生成伪标签,再用该伪标签监督网络训练过程,以上两步循环直至收敛. 该类方法可以取得较好的识别精度,因此得到了广泛应用,例如渐进式无监督学习(Progressive Unsupervised Learning, PUL)<sup>[4]</sup>是一个迭代的过程,利用预训练模型提取特征并执行聚类,选择可靠样本,然后微调模型并逐步改进特征表示,通过模型和聚类的交替优化逐渐提高在无标签数据上的性能. 文献[5]提出自划分的方法(Self-Similarity Grouping, SSG),从特征的全局到局部去挖掘目标域行人图像之间的相似度,将目标域行人特征划分为全身、上半身和下半身三个集合. 为每个集合中的特征分配伪标签,并通过微调预训练模型不断优化行人之间的相似性. 文献[6]提出的渐进优化自训练框架(Self-Training with Progressive Augmentation, PAST)包括保守阶段和促进阶段,前者通过基于排名的三元组损失选择可靠的标签样本,后者应用 Softmax 交叉熵损失函数来训练模型,使模型能够将不同的类别区分开来,两阶段交替进行逐步增强模型泛化能力. 文献[13]提出摄像位置移动适应模块和异常值渐进回收策略(CAMERA Shift adaptation module and Outlier progressive Recycling strategy, CASTOR),在聚类前利用预训练的相机分类网络计算两个样本之间的相机相似度,将相机相似度从特征相似度中减去,聚类后使用指数衰减的方式加权相机相似度的值,以自适应地弥补相机间分布变化对聚类的不良影响,全面提高伪标签的质量. Li 等人<sup>[14]</sup>采用 CSA 模块自适应地通过在每个相机上进行局部聚类并将局部结果用于优化相机间聚类,实现了更精确可靠的伪标签训练. 实时内存更新策略(Real-Time Memory updating, RTMem)<sup>[15]</sup>提出两种对比损失,即样本到实例和样本到聚类损失,以调整样本与每个聚类以及不属于任何其他聚类的所有离群值之间的关系.

## 2.2 伪标签噪声

尽管基于聚类的方法展现出了良好的性能,但聚类过程中会不可避免地引入噪声标签,伪标签的质量会直接影响模型的学习效果. 为了降低伪标签噪声的影响, PurifyNet<sup>[16]</sup>使用联合细化标注方法,逐步调整预测逻辑来优化神经网络,充分利用错误标签,并采用硬感知实例重加权策略,减少噪声标签的负面影响,更关注具有正确标签的难样本,从而有效解决有监督和无监督中的标签噪声问题. 文献[17]提出在线共同完善标签的方法(CO-REfining, CORE),使用大学习率和自标签精炼策略训练预制模型,同时优化网络参数和标签. MMT<sup>[9]</sup>利用离线精炼的“硬”伪标签和在线精炼的“软”伪标签,以协同训练的方式对网络进行优化,在目

标域中进行伪标签提炼. SPCL<sup>[10]</sup>借鉴自步学习的思想,先从最可靠的聚类开始,再逐步增加聚类,由简入难生成更可靠的集群. CACL<sup>[11]</sup>保留相似度较高的样本,去除相似度较低的样本,将每个聚类划分为多个子聚类,当子聚类中所有样本到原始聚类中其他样本的平均距离小于原始聚类内部样本的平均距离时,该子聚类才会被保留. 文献[18]提出一种新的无监督混合对比范式,采用多粒度聚类集成策略(Multi-Granularity Clustering Ensemble based Hybrid Contrastive Learning, MGCE-HCL),用来描述正样本的置信度,并且提出了一种优先级加权的混合对比损失,更好地容忍伪正样本中的噪声. 文献[19]提出了一种基于不确定度评估策略的网络(Uncertainty-guided Noise Resilient Network, UNRN),根据不确定度作用于 ID 分类损失、三元损失和对比损失来减少一些不可靠伪标签的影响. 文献[20]提出了一种基于异构卷积神经网络集成(Heterogeneous Convolutional Neural Network Embedded framework, HCNNE)的框架,自动推测目标域中行人图像之间的关系,减少因噪声标签在训练过程中产生的波动,并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.

与之前工作不同,本文提出了一个简单而有效的方法,通过聚类引导的注意力机制来估计伪标签的可靠性,使用伪标签进行软样本挖掘,并在伪标签正则化损失中利用它们来防止模型拟合标签噪声,从而减轻伪标签噪声的负面影响. 之前大多数方法(如 PurifyNet<sup>[16]</sup>、CORE<sup>[17]</sup>、MMT<sup>[9]</sup>等)对噪声标签的处理主要依赖于模型本身的学习能力或者固定的策略,例如增大学习率,自标签精炼策略等,其能够处理的噪声程度有限. 而本文使用聚类引导的注意力机制,可动态地评估伪标签的可靠性,这使得我们的方法相比传统方法更能有效地降低噪声标签的影响. 此外,之前的方法(例如 SPCL<sup>[10]</sup>、CACL<sup>[11]</sup>等)在处理样本挖掘时,基于固定的或相对静态的规则,如自步学习、子聚类等,非常依赖预先设定的策略和参数. 而本文方法则基于伪标签进行“软样本挖掘”,这提供了一种更为灵活和自适应的策略,能够更好地处理各种情况.

## 2.3 对比学习

在以上基于聚类的算法中,每次迭代后类簇中心和伪标签都会发生变化. 为了应对这一问题,本文利用对比学习中构建记忆字典的方式动态适应簇的变化. 这种对比学习方法在无监督行人重识别任务中也取得了广泛的应用,例如 SPCL<sup>[10]</sup>提出了一个基于混合记忆模型的对比学习框架,利用多种不同形式的类别原型提供混合监督,以实现对所有训练数据的充分挖掘. Cluster-Contrast<sup>[21]</sup>通过 DBSCAN<sup>[8]</sup>聚类算法生成伪标

签,然后利用这些数据生成簇类级别的内存字典,之后利用目标域上的特征与簇类级别的内存字典中特征之间的损失来优化模型,在保证效果的同时,又极大降低了计算资源的使用。 $M^3L$ <sup>[22]</sup>采用基于内存的模块,以非参数方式实现识别损失,并能防止元优化过程中由于传统参数方式造成的不稳定优化。 $DSCE$ <sup>[23]</sup>通过维护一个特征存储器来存储所有图像特征,动态构建新的类中心应聚类变化并利用动态中心,提出动态对称交叉熵损失来减轻噪声样本引起的负面影响。 $HHCL$ <sup>[12]</sup>在Cluster-Contrast基础上提出了一个混合对比学习框架,这个框架结构将聚类级损失与实例级损失相结合,应用了聚类质心对比损失,以确保网络以更稳定的方式更新,同时,引入硬实例对比损失,进一步挖掘判别信息,学习区分容易混淆的样本。本文以HHCL网络结构为基线模型,在其基础上引入PLRL损失函数,给正确标签赋予更多的权重并充分挖掘了伪标签的判别能力,有效地提升了模型性能。

本文提出的聚类引导的注意力机制受文献[24]中CAA(Class Aware Attention)的启发,但与其不同的是:(1)CAA适用于有监督学习,通过估计图像与其类别标签的语义关系来识别异常值,其中类上下文向量为全连接层的训练参数,而本文提出的CGA用于无监督学习,通过计算图像特征与聚类中心之间的相似性来衡量伪标签的正确性;(2)本文利用CGA设计了新的伪标签正则化损失函数,通过对伪标签噪声进行正则约束,

有效降低伪标签噪声对训练过程的负面影响,从而提升模型的表征能力和鲁棒性。

### 3 基于伪标签正则化的无监督行人重识别

#### 3.1 模型概述

如图1所示,本文所采用的无监督行人重识别模型主要包括以下流程:(1)利用ImageNet上预训练的ResNet50<sup>[25]</sup>模型 $\Phi_\theta$ 提取训练图像 $X = \{\mathbf{x}_i\}_{i=1}^N$ 的初始特征,并通过聚类算法将训练集中的图像聚成 $C$ 个不同的类别,并给每个聚类类别赋予一个标签,即初始伪标签;(2)利用提取到的初始特征构建聚类级内存字典 $M_c$ 和实例级内存字典 $M_i$ ,分别用于存储样本所属的簇类的平均值和每个簇类中的最难样本(离聚类中心最远的样本特征);(3)使用聚类引导的注意力机制计算样本与聚类中心的语义相关程度CGA分数 $a_i$ ,给样本重新分配权重,并利用伪标签进行在线软样本挖掘(SSM),将mini-batch中的所有样本构造成正负样本对计算每对样本的SSM分数 $s_{ij}^+$ 和 $s_{ij}^-$ ,结合SSM与CGA分数构建伪标签正则化损失函数 $L_{PLRL}$ ,在第3.2节中分别对上述内容进行了详细介绍;(4)利用内存字典聚类级内存字典 $M_c$ 和实例级内存字典 $M_i$ 分别计算非参数化的聚类级损失函数 $L_c$ 和实例级损失函数 $L_i$ ;(5)使用伪标签正则化损失函数、聚类级损失函数以及实例级损失函数三种损失函数共同监督网络更新;(6)更新实例级内存字典 $M_i$ 与聚类级内存字典 $M_c$ 。

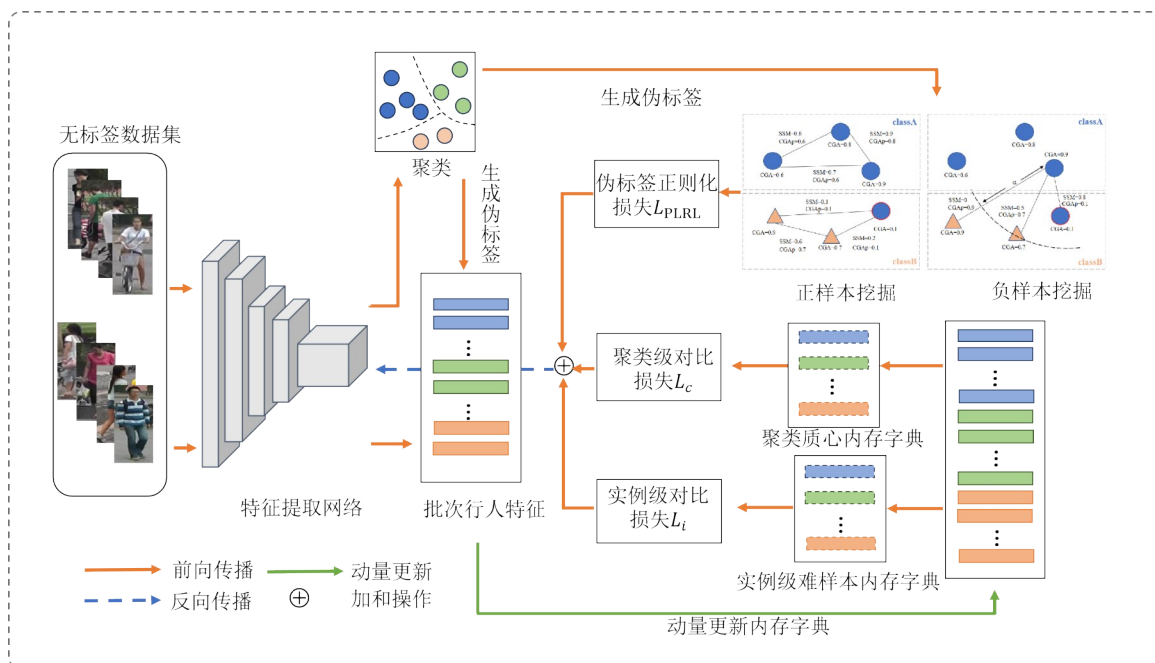


图1 基于伪标签正则化损失的无监督行人重识别算法整体流程示意图

### 3.2 伪标签正则化损失函数

#### 3.2.1 聚类引导的注意力机制

伪标签噪声指的是聚类产生的伪标签中存在的错误或噪声,通常是由于模型有限的表征能力、类别数目未知、聚类算法本身的局限性等因素导致的。伪标签噪声会对模型的训练和性能产生负面影响,因为它们会引导模型学习错误的特征,并且噪声标签的误差在训练过程中不断累积,从而造成模型性能严重退化。在行人重识别这一应用场景中,聚类噪声可能会导致模型将不同的行人特征错误地聚合在一起,从而导致模型无法学习到有判别力的行人特征。因此,如何有效地处理伪标签噪声对模型最终的性能有着至关重要的影响。

为了改善伪标签质量,本文提出了一种新的聚类引导的注意力机制,根据每个图像特征与其所在类的聚类中心的相关程度来估计该图像伪标签的置信度,以此来识别伪标签噪声并给正确标签分配更多的权重,有效降低伪标签噪声在总体损失函数中的作用。其基本思想是,含噪声的伪标签通常包含的是错误的或者不准确的类别标签,因此它与其所在的类簇中心的语义相关程度较低。若图像特征与其聚类中心的相似性越低,则该图像的伪标签含噪声的可能性越大。该方法是由聚类伪标签引导的,我们称之为聚类引导的注意力机制(Cluster Guided Attention, CGA),如图2所示,不同的形状代表不同的类别,当图像因为聚类噪声而被错误分类时,那么CGA分数就会非常低。

聚类注意力得分CGA的具体计算过程如下。通过预训练模型提取图像的初始特征后,通过聚类算法将训练集中的图像聚成不同的类别,并构造聚类级内存字典 $M_c$ ,用于存储聚类中心 $c_1, c_2, \dots, c_c$ ,聚类中心通过计算聚类后每个类中所有特征的平均值得来的, $C$ 为当前迭代中聚聚类簇的数量。CGA的基本思想是,与聚类中心的语义相关程度更大的图像特征,其伪标签的置信度更高,将获得更高的注意力权重,其计算方法如式(1)所示。

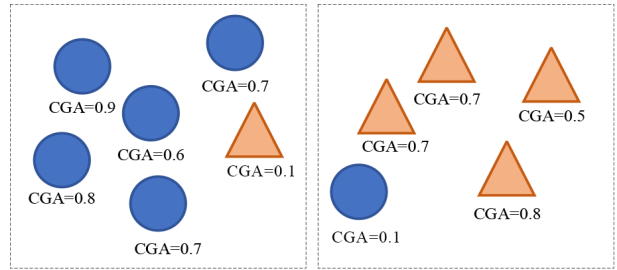


图2 聚类引导的注意力机制

$$a_i = \frac{\exp(\mathbf{f}_i^T \mathbf{c}_{y_i})}{\sum_{k=1}^C \exp(\mathbf{f}_i^T \mathbf{c}_k)} \quad (1)$$

其中, $a_i$ 表示图像 $x_i$ 的CGA得分; $\mathbf{f}_i$ 为图像 $x_i$ 经过编码器得到的初始特征向量; $\mathbf{c}_{y_i}$ 表示图像 $x_i$ 聚类后所在的聚类中心; $\mathbf{c}_k$ 为第 $k$ 类的特征向量。若图像 $x_i$ 的伪标签中噪声较大,那么它与类簇中心的语义相关程度较低,其CGA得分 $a_i$ 就很低。

#### 3.2.2 软样本挖掘策略

同时,为了充分利用伪标签的判别能力,我们使用软挖掘策略(SSM),利用伪标签在线构造正负样本对,并为每个样本对计算一个连续的权重分数,具体过程如图3所示。首先根据伪标签构建当前 mini-batch 中的正样本对 $(x_i, x_j) \in P$ ,并计算正样本对特征向量之间的欧式距离 $d_{ij} = \|\mathbf{f}_i - \mathbf{f}_j\|_2$ 。为了有效检验伪标签正确与否,我们利用高斯函数将欧式距离转换为SSM分数,如式(2)所示。

$$s_{ij}^+ = \exp(-d_{ij}^2 / \sigma_{SSM}^2) \quad (2)$$

其中, $\sigma_{SSM}$ 是控制SSM分数分布的超参数。如果当前伪标签是正确的,由其构建的正样本对的欧式距离较小,SSM分数较高;反之若伪标签中含噪声,则由其构建的正样本对间的欧式距离较大,SSM分数较低,并且正样本间距离越大,SSM分数越低。通过这种方式,可以有效减轻伪标签噪声的影响。

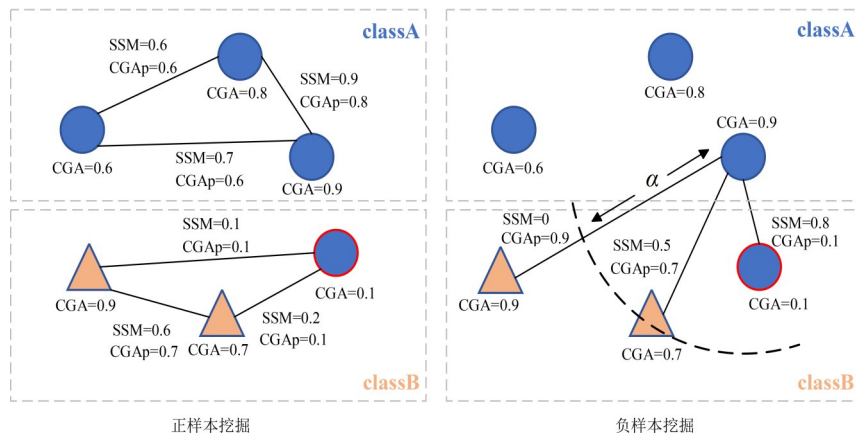


图3 CGA与SSM联合作用示意图

同样,我们根据伪标签构建负样本对 $(\mathbf{x}_i, \mathbf{x}_j) \in N$ ,并计算样本对间的欧式距离.由于负样本对数量较大,为了忽略大部分对学习没有贡献的样本对,本文将距离大于预定义边界的负样本对的SSM分数设为0,因为它们对损失和梯度的计算没有贡献.每个负样本对的SSM分数 $s_{ij}^-$ 计算方法如式(3)所示,对于距离在预定边界 $\alpha$ 内的负样本对,若其欧式距离很小,代表该样本对为难负样本对,SSM为难负样本分配更高的分数,可以有效提高样本挖掘和度量学习的效率.

$$s_{ij}^- = \max(0, \alpha - d_{ij}) \quad (3)$$

### 3.2.3 CGA与SSM联合作用

注意力引导的注意力权重CGA和软样本挖掘分数SSM都可以在一定程度上抑制伪标签中的噪声,为了进一步提升算法性能,本文将二者联合作为样本对的权重,从伪标签的置信度和样本相似度两个角度联合对伪标签噪声进行约束.对于样本对 $(\mathbf{x}_i, \mathbf{x}_j) \in P \cup N$ ,其权值 $w_{ij}$ 由SSM和CGA共同计算得来, $w_{ij}^+$ 和 $w_{ij}^-$ 分别代表正、负样本对的权重,如式(4)和式(5)所示,其中, $a_{ij}$ 定义为 $\min(a_i, a_j)$ ,取 $a_i$ 和 $a_j$ 中的最小值,运算符\*表示卷积运算.

$$w_{ij}^+ = s_{ij}^+ * a_{ij} \quad (4)$$

$$w_{ij}^- = s_{ij}^- * a_{ij} \quad (5)$$

通过将二者联合作用,可以更加有效地减少对噪声样本的关注,降低噪声值对模型训练的影响.如图3所示,红色边界为错误分类的噪声样本,在正样本挖掘中,本文使用SSM分数来给距离更近的图像正对分配更高的分数.同时,基于聚类的注意力机制利用CGAp分数来识别噪声值(红色边界),其中CGAp是指样本对中两张图像中的CGA分数的最小值.在负样本挖掘中,对于难负样本,如噪声值(红色边界,ClassA类别的样本错误地被分类为ClassB),SSM会因其与ClassA中样本相似度较高为其分配更高的挖掘分数,此时CGA通过计算得出该样本并不属于ClassB并为其分配较低分数,从而帮助忽略SSM挖掘出的难负样本中的异常值.

### 3.2.4 伪标签正则化损失函数

通过将以上两种权重分数引入到对比损失中,本文设计了一种新的伪标签正则化损失函数(Pseudo Label Regularization Loss, PLRL).传统对比损失函数 $L$ 的定义如式(6)所示:

$$L(W, (Y, \mathbf{X}_1, \mathbf{X}_2)) = \frac{1}{K} \sum_{k=1}^K Y D_w^2 + (1 - Y) \max(\alpha - D_w, 0)^2 \quad (6)$$

其中, $D_w(\mathbf{X}_1, \mathbf{X}_2) = \|\mathbf{X}_1 - \mathbf{X}_2\|_2$ ,代表两个样本特征的欧式距离; $Y$ 为两个样本是否匹配的标签; $\alpha$ 为设定的阈值; $K$ 为训练样本数量.

本文将样本对权值 $w_{ij}$ 引入到传统的对比损失中,

具体计算方式如式(7)~(9)所示,式(7)为正样本对的伪标签正则化损失函数,式(8)为负样本对的伪标签正则化损失函数,式(9)为总体的伪标签正则化损失函数.通过最小化伪标签正则化损失,有效削弱了伪标签噪声在总体损失计算中的作用,并且可以拉近由伪标签挖掘得到的正样本对距离,将挖掘得到的负样本距离推远至超过预先设定的边界,从而充分利用了伪标签的判别能力.

$$L_{\text{PLRL}}(P) = \frac{1}{2} \frac{\sum_{(\mathbf{x}_i, \mathbf{x}_j) \in P} w_{ij}^+ d_{ij}^2}{\sum_{(\mathbf{x}_i, \mathbf{x}_j) \in P} w_{ij}^+} \quad (7)$$

$$L_{\text{PLRL}}(N) = \frac{1}{2} \frac{\sum_{(\mathbf{x}_i, \mathbf{x}_j) \in N} w_{ij}^- \max(0, \alpha - d_{ij})^2}{\sum_{(\mathbf{x}_i, \mathbf{x}_j) \in N} w_{ij}^-} \quad (8)$$

$$L_{\text{PLRL}}(P, N) = L_{\text{PLRL}}(P) + L_{\text{PLRL}}(N) \quad (9)$$

## 3.3 损失函数

### 3.3.1 聚类对比损失

聚类对比损失函数由输入和存储在聚类内存字典中的聚类中心计算得来,如式(10)所示:

$$L_c = -\log \frac{\exp(\mathbf{q} \cdot \mathbf{c}_+ / \tau)}{\sum_{i=1}^C \exp(\mathbf{q} \cdot \mathbf{c}_i / \tau)} \quad (10)$$

其中, $C$ 表示每个epoch中聚类类簇的数量; $\tau$ 是温度超参数; $\mathbf{c}_i$ 表示类别 $i$ 的聚类中心;使用聚类中质心作为聚类级特征向量来计算查询实例 $\mathbf{q}$ 与所有聚类之间的距离; $\mathbf{c}_+$ 是实例特征 $\mathbf{q}$ 所属类的聚类中心特征.

### 3.3.2 实例级对比损失

实例级损失函数由输入和存储在实例级内存字典中的难样本计算得来,该实例级内存字典中存储的难样本是每个类中离类中心最远的样本.如式(11)所示,这种计算方式是基于所有的数据,包括了 $C-1$ 个类的负样本数据,考虑了样本与其他类之间的关系:

$$L_i = -\log \frac{\exp(\mathbf{q} \cdot \mathbf{z}_i / \tau)}{\sum_{i=1}^C \exp(\mathbf{q} \cdot \mathbf{z}_i / \tau)} \quad (11)$$

其中, $\mathbf{z}_i$ 表示与 $\mathbf{q}$ 属于同一个类中,具有最小余弦相似度的样本(即难样本)的特征; $\mathbf{z}_i$ 表示类别 $i$ 中的难样本特征表示.

### 3.3.3 总体损失函数

本文中采用的损失函数结合了聚类损失和实例级损失以及新提出的伪标签正则化损失,总体损失函数如式(12)所示:

$$L = \mu L_c + (1 - \mu) L_i + \gamma L_{\text{PLRL}} \quad (12)$$

其中,参数 $\mu$ 是用来控制聚类级损失和实例级损失在总体损失中的权重;参数 $\gamma$ 用于调节 $L_{\text{PLRL}}$ 在最终对比损

失中的权重分配.

### 3.3.4 内存字典更新策略

本文在计算损失函数后采用了动量更新的方法,用于更新实例级内存字典  $M_i$  与聚类级内存字典  $M_c$ . 在更新聚类级内存字典时,通过不断调整聚类中心,使模型能够适应不断变化的数据分布和特征从而提高聚类的效果和性能. 聚类级内存字典更新的计算公式为

$$c^i \leftarrow mc^i + (1-m)c^{-i} \quad (13)$$

其中,  $c^i$  表示类别  $i$  存储在聚类内存字典  $M_c$  中的聚类中心;  $c^{-i}$  是 mini-batch 中第  $i$  个类实例特征的平均值;  $m$  是动量更新因子.

同样,为了保证内存更新的一致性,使用动量更新的方法更新实例级内存字典中的实例特征,本文按式(14)的方式更新实例内存库,

$$z^i \leftarrow mz^i + (1-m)z^{-i} \quad (14)$$

其中,  $z^i$  是类别  $i$  存储在实例级内存字典  $M_i$  中的难样本;  $z^{-i}$  是 mini-batch 中第  $i$  个类中距离聚类中心最远的样本特征.

本文算法的总体流程可归纳为算法 1.

#### 算法 1 基于伪标签正则化损失的无监督行人重识别

输入: 无标签数据集  $X$ , ImageNet 预训练模型  $\Phi_\theta$ , 迭代次数  $T$ , 训练批量大小  $B$

输出: 行人重识别模型  $\Phi$ .

1. for epoch = 1 to  $T$  do
2. 通过  $\Phi(\cdot; \theta)$  从  $X$  中提取特征  $f$ ;
3. 使用 DBSCAN 算法将  $f$  聚类为  $C$  个簇;
4. 初始化聚类级内存字典和实例级内存字典;
5. for batchsize = 1 to  $B$  do
6. 软样本挖掘, 根据伪标签划分正负样本对;
7. 计算 CGA 分数和 SSM 分数(式(1)-(3));
8. 计算伪标签正则化损失函数、聚类级损失函数和实例级损失函数(式(9)-(11));
9. 反向传播以更新模型  $\Phi_\theta$ ;
10. 更新聚类级内存字典  $M_c$  和实例级内存字典  $M_i$ ;
11. end
12. end

## 4 实验与结果

### 4.1 数据集和评价标准

本文选用 Market1501<sup>[26]</sup>、DukeMTMC-reID<sup>[27]</sup> 和 MSMT17<sup>[28]</sup> 这三个公开数据集进行实验. 数据集具体内容如表 1 所示.

Market-1501 数据集在清华大学校园中采集, 夏天拍摄, 在 2015 年构建并公开. 它包括由 6 个摄像头(其中 5 个高清摄像头和 1 个低清摄像头)拍摄到的 1 501 个行人、32 668 个检测到的行人矩形框. 每个行人至少

表 1 各数据集中训练集、查询集和候选集的图像数量

数据集	训练集	查询集	候选集
Market1501 <sup>[26]</sup>	12 936	3 368	19 732
DukeMTMC-reID <sup>[27]</sup>	16 522	2 228	17 661
MSMT17 <sup>[28]</sup>	32 621	11 659	82 161

由 2 个摄像头捕获到, 并且在一个摄像头中可能具有多张图像. 训练集有 751 人, 包含 12 936 张图像, 平均每个人有 17.2 张训练数据; 测试集有 750 人, 包含 19 732 张图像, 平均每个人有 26.3 张测试数据.

DukeMTMC-reID 数据集是 2017 年发布的一个大规模行人重识别图片数据集, 由在美国杜克大学校园中 8 个不同位置采集的高分辨率相机拍摄而成, 包括 1 812 个行人的 36 411 张图像. 其中训练集包含 702 个行人, 共 16 522 张图像. 查询集包含 702 个人的 2 228 张图像, 候选集中包含 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, 其余 408 个行人只出现在单个摄像机下. 另外, 该数据集提供了行人属性的标注.

MSMT17 数据集采用了安放在校园内的 15 个摄像头网络, 其中包含 12 个户外摄像头和 3 个室内摄像头. 为了采集原始监控视频, 在一个月里选择了具有不同天气条件的 4 天. 每天采集 3 h 的视频, 涵盖了早上、中午、下午三个时间段. 因此, 总共的原始视频时长为 180 h. 训练集包含 1 041 个行人共 32 621 个包围框, 而测试集包括 3 060 个行人共 93 820 个包围框. 对于测试集, 11 659 个包围框被随机选出来作为 query, 而其他 82 161 个包围框作 gallery. MSMT17 数据集是一个更接近真实场景且更复杂的行人重识别数据集.

本文采用 2 个评价指标来评估在不同数据集上的行人重识别算法性能: Rank- $k$  为搜索结果中最靠前(置信度最高)的  $k$  张图有正确结果的概率; mAP(mean Average Precision) 为平均精度均值, 是将多分类任务中的平均精度 AP(Average Precision) 求和再取平均. AP 的值就是通过预测分析得出的实验结果中精确率(precision)和召回率(recall)来精确绘制一个 P-R 曲线的面积. 精确率是检验正确的个数跟检测为正的比值, 即检测为正的结果中有多少个检测是正确的. 召回率表示的是检测正确的个数跟实际为正的个数的比值, 即实际为正在检测过程中有多少检测是正确的. 精确率、召回率计算方法如式(15)和式(16)所示.

$$\text{precision} = \frac{TP}{TP + FP} \quad (15)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (16)$$

其中, TP 代表实际和检测都为正的个数; FP 表示检测为正, 但实际为负的个数; FN 表示检测为负, 但实际为正的个数. 综上所述 mAP 的计算方法如式(17)所示.

$$\text{mAP} = \frac{\sum \text{AP}}{n} \quad (17)$$

## 4.2 实验设置

训练过程中,采用 ResNet-50 作为特征提取器的骨干,用 ImageNet 上预训练的参数初始化模型. 为了进一步提高模型的表征能力,在 Layer 3 之后添加 RGA<sup>[29]</sup> 模块,将全局平均池化层的特征作为最终的行人特征. 与 k-means<sup>[7]</sup> 算法相比,基于密度的 DBSCAN<sup>[8]</sup> 聚类算法不需要预先给定聚类数量,可以发现任意形状的聚类,适合的数据分布范围更广,因此本文选取 DBSCAN 作为聚类算法. 每个轮次的开始,使用 DBSCAN 进行聚类,生成伪标签,聚类半径(邻域内两个样本之间的最大距离)设置为 0.45,核心点上的最小邻居数设置为 4. 对于数据集输入图像的大小设为 256×128.

对于训练图像,我们执行随机水平翻转、随机裁剪和随机擦除等数据增强策略. 每个 mini-batch 包含 128 个图像. 使用 Adam 优化器更新梯度,初始学习率设为  $3.5 \times 10^{-4}$ ,权重衰减设为  $5 \times 10^{-4}$ . 总共训练 60 个 epoch,每 20 个 epoch 将学习率降低到之前值的 1/10. 将式(2)中控制 SSM 分数分布的超参数  $\sigma_{\text{SSM}}$  设置为 0.4,式(3)中计算负样本对 SSM 分数使用的预设边界值  $\alpha$  设为 1.2,式(10)和式(11)中的温度参数  $\tau$  设为 0.05,式(12)中计算总体损失函数时控制聚类级损失和实例级损失权重  $\mu$  设为 0.5,控制  $L_{\text{PLRL}}$  权重的参数  $\gamma$  设为 0.5. 本文在第 4.5 节中进行了一系列参数实验,并对实验结果进行了详细讨论,得出上述参数配置可以在实验中表现出最优的效果.

## 4.3 实验结果和对比

为了检验本文提出方法的有效性,本文在 Market1501、DukeMTMC-reID 以及 MSMT17 三个数据集上进行了对比实验,将本文提出的方法与该领域近几年的主流方法进行了比较. 本文主要研究的是纯无监督的行人重识别问题,因此在进行实验对比时多数都是与纯无监督方法进行对比,包括 BUC<sup>[30]</sup>、SSL<sup>[31]</sup>、MMCL<sup>[32]</sup>、HCT<sup>[33]</sup>、CycAs<sup>[34]</sup>、UGA<sup>[35]</sup>、SPCL<sup>[10]</sup>、Cluster-Contrast<sup>[21]</sup>、HHCL<sup>[12]</sup>以及 HCNNE<sup>[20]</sup>. 域迁移无监督方法 UDA(Unsupervised Domain Adaptation, UDA)作为无监督方法中的一种,可以利用有标签的源域数据训练网络,相比于纯无监督方法可以取得更好的结果,在对比实验中本文也将该方法作为比较对象,以充分展示本文算法性能的优越性. 本文选取的域迁移无监督方法包括 ECN<sup>[36]</sup>、MMCL<sup>[32]</sup>、MMT<sup>[9]</sup>以及 SPCL<sup>[10]</sup>. 在 MSMT 数据集上测试时添加了文献[37]与文献[38]中的方法进行对比.

如表 2~4(最优结果使用粗体标注)以及图 4 中结果所示,在不同的数据集中,本文提出的方法均取得了较好的结果,在 Market1501、DukeMTMC-reID 以及 MSMT17 三个数据集上的 mAP 分别为 85.9%、75.1%和

29.3%,相比于基线方法分别提高了 2.9%、3.1% 和 3.0%. 该结果表明本文提出的聚类引导的注意力机制与伪标签正则化损失函数能够有效地降低伪标签噪声的影响,确保伪标签的准确性和可靠性,正确引导模型的训练过程,因而使得准确率有较大幅度的提高. 此外,表 2~4 中的数据显示,与最新的无监督行人重识别方法相比,本文提出的算法在 Market1501 与 DukeMTMC 上均取得了最好的结果, mAP 比之前的最好结果分别提高了 0.4%、0.3%,但在 MSMT17 数据集上与最新方法相比准确率较低,推测原因是 HCNNE<sup>[20]</sup> 在基础网络上加入了正交正则化约束,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,该策略在 MSMT17 这样大规模的数据集上性能提升更加明显,在之后的工作中我们会基于此问题进行改进. 尤其需要注意的是,本文的方法不依赖于任何有标签的源域数据,但性能优于利用源域数据的域迁移无监督方法 ECN<sup>[29]</sup>、MMCL<sup>[32]</sup>、MMT<sup>[9]</sup>等,进一步证明了本文方法的优越性.

表 2 Market1501 数据集上不同方法实验结果 单位: %

方法		Market1501				
		源域	mAP	Rank1	Rank5	Rank10
UDA	MMCL <sup>[32]</sup>	Duke	60.4	84.4	92.8	95.0
	MMT <sup>[9]</sup>	MSMT17	75.6	89.3	95.8	97.5
	SPCL <sup>[10]</sup>	MSMT17	77.5	89.7	96.1	97.6
USL	BUC <sup>[30]</sup>	None	38.3	66.2	79.6	84.5
	SSL <sup>[31]</sup>	None	37.8	71.7	83.8	87.4
	MMCL <sup>[32]</sup>	None	45.5	80.3	89.4	92.3
	HCT <sup>[33]</sup>	None	56.4	80.0	91.6	95.2
	CycAs <sup>[34]</sup>	None	64.8	84.8	—	—
	UGA <sup>[35]</sup>	None	70.3	87.2	—	—
	SPCL <sup>[10]</sup>	None	73.1	88.1	95.1	97.0
	Cluster-Contrast <sup>[21]</sup>	None	82.1	92.3	96.7	97.9
	HHCL <sup>[12]</sup>	None	83.0	92.7	97.3	98.1
	HCNNE <sup>[20]</sup>	None	85.4	93.3	97.6	98.7
	本文方法	None	<b>85.9</b>	<b>94.3</b>	<b>98.0</b>	<b>98.7</b>

## 4.4 消融实验

本文提出的伪标签正则化损失函数有两个重要的组成部分,分别是基于聚类引导的注意力机制 CGA 分数与样本软挖掘方法计算的 SSM 分数. 为了分析这两部分对总体性能的贡献并进一步验证它们的有效性,我们在相同的参数设置下,分别加入了上述两个部分,在不同数据集上进行了消融实验,结果如表 5~7 所示(最优结果使用粗体标注).

从表 5~7 可以看出,以 Market1501 数据集为例,与

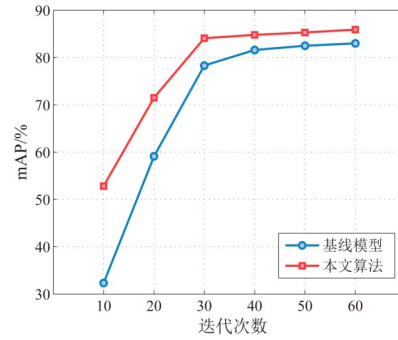
表3 DukeMTMC-reID数据集上不同方法实验结果比较 单位:%

方法		DukeMTMC-reID				
		源域	mAP	Rank1	Rank5	Rank10
UDA	MMCL <sup>[36]</sup>	Market	51.4	72.4	82.9	85.0
	MMT <sup>[9]</sup>	Market	65.1	78.9	88.8	92.5
	SPCL <sup>[10]</sup>	Market	68.8	82.9	90.1	92.5
USL	BU <sup>[30]</sup>	None	27.5	47.4	62.6	68.4
	SSL <sup>[31]</sup>	None	28.6	52.5	63.5	68.9
	MMCL <sup>[32]</sup>	None	51.4	72.4	82.9	85.0
	HCT <sup>[33]</sup>	None	50.7	69.6	83.4	87.4
	CycAs <sup>[34]</sup>	None	60.1	77.9	—	—
	UGA <sup>[35]</sup>	None	53.3	75.0	—	—
	SPCL <sup>[10]</sup>	None	65.3	81.2	90.3	92.2
	Cluster-Contrast <sup>[21]</sup>	None	72.6	84.9	91.9	93.9
	HHCL <sup>[12]</sup>	None	72.0	84.2	92.1	94.1
	HCNNE <sup>[20]</sup>	None	74.8	85.8	92.2	93.8
	本文方法	None	<b>75.1</b>	<b>86.3</b>	<b>92.4</b>	<b>94.7</b>

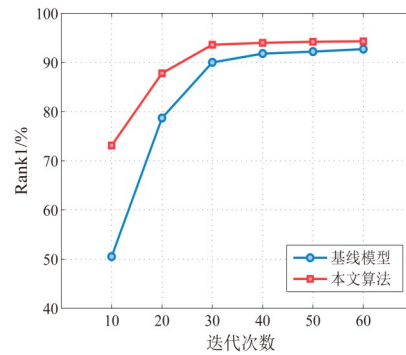
表4 MSMT17数据集上不同方法实验结果比较 单位:%

方法		MSMT17				
		源域	mAP	Rank1	Rank5	Rank10
UDA	ECN <sup>[31]</sup>	Duke	10.2	30.2	41.5	46.8
	MMT <sup>[9]</sup>	Market	24.0	50.1	63.5	69.3
	SPCL <sup>[10]</sup>	Market	26.8	53.7	65.0	69.8
USL	MMCL <sup>[32]</sup>	None	11.2	35.4	44.8	49.8
	TAUDL <sup>[37]</sup>	None	12.5	28.4	—	—
	UTAL <sup>[38]</sup>	None	13.1	31.4	—	—
	CycAs <sup>[34]</sup>	None	26.7	50.1	—	—
	UGA <sup>[35]</sup>	None	21.7	49.5	—	—
	SPCL <sup>[10]</sup>	None	19.1	42.3	55.6	61.2
	Cluster-Contrast <sup>[21]</sup>	None	27.6	56.0	66.8	71.5
	HHCL <sup>[12]</sup>	None	26.3	52.3	64.2	69.6
	HCNNE <sup>[20]</sup>	None	<b>39.8</b>	<b>67.9</b>	<b>78.3</b>	<b>82.1</b>
	本文方法	None	29.3	55.5	67.1	72.0

基线模型相比,加入CGA分数后,模型的mAP和Rank1准确率分别提高了1.9%和0.9%,说明本文提出的聚类引导的注意力分数可以有效地根据伪标签的置信度来重新分配标签权重,选择出置信度较高的可靠伪标签,因而能够有效地过滤和平滑伪标签噪声,从而提升模型性能.另外,加入SSM软样本挖掘策略后,模型的mAP和Rank1准确率分别提高了1.0%和0.1%,表明SSM可以充分挖掘伪标签的判别能力,与以往的硬样本挖掘方法只能选择一个mini-batch中的一个子集不同,SSM给mini-batch中的正负样本对分配一个连续的



(a) mAP精度对比



(b) Rank1精度对比

图4 Market1501数据集上不同epoch下mAP和Rank1精度对比

表5 Market1501数据集上的消融实验 单位:%

Index	Settings		Market1501			
	SSM	CGA	mAP	Rank1	Rank5	Rank10
1	×	×	83.0	92.7	97.3	98.1
2	√	×	84.0	92.8	97.4	98.3
3	×	√	84.9	93.6	97.7	98.6
4	√	√	<b>85.9</b>	<b>94.3</b>	<b>98.0</b>	<b>98.7</b>

表6 DukeMTMC-reID数据集上的消融实验 单位:%

Index	Settings		DukeMTMC-reID			
	SSM	CGA	mAP	Rank1	Rank5	Rank10
1	×	×	72.0	84.2	92.1	94.1
2	√	×	73.9	85.5	92.2	94.7
3	×	√	74.4	86.1	92.1	93.9
4	√	√	<b>75.1</b>	<b>86.3</b>	<b>92.4</b>	<b>94.7</b>

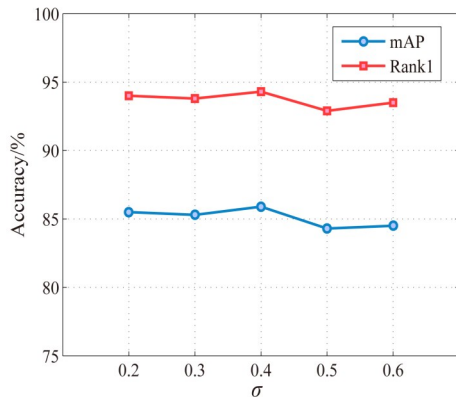
表7 MSMT17数据集上的消融实验 单位:%

Index	Settings		MSMT17			
	SSM	CGA	mAP	Rank1	Rank5	Rank10
1	×	×	26.3	52.3	64.2	69.6
2	√	×	27.4	52.5	64.6	69.9
3	×	√	27.6	56.0	66.8	71.5
4	√	√	<b>29.3</b>	<b>55.5</b>	<b>67.1</b>	<b>72.0</b>

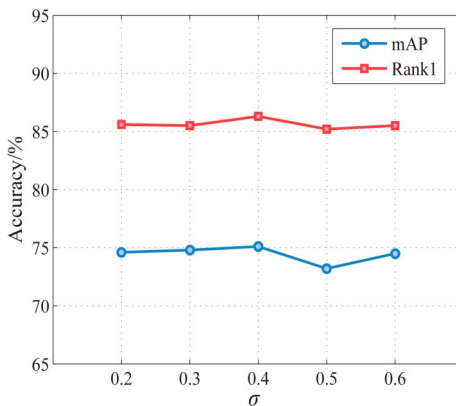
分数,从而充分利用伪标签的判别能力.此外,将CGA分数和SSM结合构造伪标签正则损失后,模型的性能均有了进一步的提升,表明这两部分改进都是必不可少的,它们的共同作用可以有效削弱伪标签噪声的影响,有助于模型性能的稳定提升.在DukeMTMC-reID和MSMT17数据集上可以得到类似的结论.

#### 4.5 参数分析

(1)正样本挖掘中控制SSM分数分布的超参数 $\sigma_{SSM}$   $\sigma_{SSM}$ 是控制SSM分数分布的参数,在本节中,将 $\sigma_{SSM}$ 分别设置为0.2、0.3、0.4、0.5、0.6,训练对应的重识别模型,得到的结果如图5所示.在数据集Market1501与DukeMTMC-reID上进行实验,从图5可以看出,当阈值设置为0.4时,mAP和Rank1准确率均达到了最高,模型取得了最优的结果.



(a) Market1501



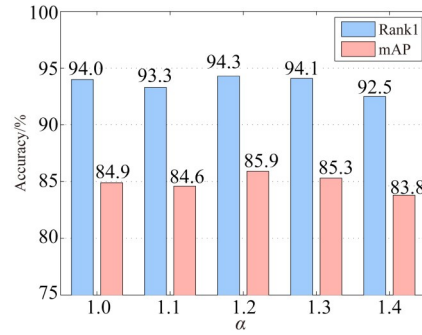
(b) DukeMTMC-reID

图5  $\sigma_{SSM}$ 不同取值对模型准确率的影响结果

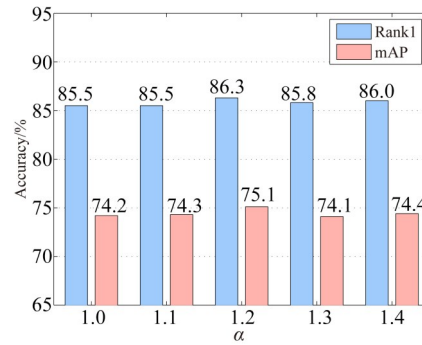
(2)负样本挖掘中SSM分数使用的预设边界值 $\alpha$

在计算过程中将更高的SSM分数分配给距离在预定边界 $\alpha$ 内的负配对,本文在实验中尝试了不同的 $\alpha$ 取值1.0、1.1、1.2、1.3和1.4,并在数据集Market1501和

DukeMTMC-reID上分别进行实验,通过训练重识别模型并在图6中进行可视化展示,观察 $\alpha$ 不同取值对模型性能的影响.实验结果表明,在本实验设置下,当阈值设置为1.2时,模型取得了最优的结果,这意味着在给定的数据集和实验条件下,使用 $\alpha=1.2$ 时,模型获得了最高的准确率.



(a) Market1501



(b) DukeMTMC-reID

图6  $\alpha$ 不同取值对模型准确率的影响结果

(3)温度超参数 $\tau$

在表8和表9(最优结果使用粗体标注)中研究了温度超参数 $\tau$ 的影响.较小的 $\tau$ 导致较低的熵,并且可能有助于在重识别中获得更好的结果.但如果 $\tau$ 值过小,如0.03,则会导致训练的崩溃.在本文的实验中,设置 $\tau=0.05$ ,这在所有数据集上都取得了很好的性能.

(4)控制聚类级损失和实例级损失权重 $\mu$

当 $\mu=0$ 时,损失函数只包含实例对比损失 $L_i$ 与 $L_{PLRL}$ ,

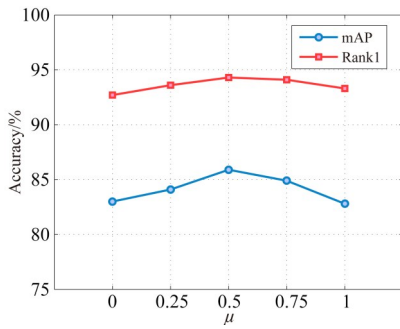
表8 Market1501数据集上计算 $\tau$ 值对实验的影响 单位:%

$\tau$	Market1501			
	mAP	Rank1	Rank5	Rank10
0.03	84.8	94.0	97.4	98.2
0.05	<b>85.9</b>	<b>94.3</b>	<b>98.0</b>	<b>98.7</b>
0.10	85.1	94.1	97.6	98.5
0.30	84.9	94.1	97.8	98.5

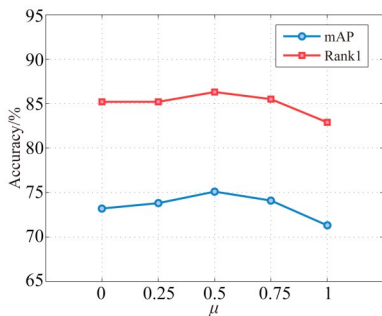
表 9 DukeMTMC-reID数据集上计算  $\tau$  值对实验的影响 单位:%

$\tau$	DukeMTMC-reID			
	mAP	Rank1	Rank5	Rank10
0.03	70.4	83.3	90.3	93.0
0.05	75.1	86.3	92.4	94.7
0.10	74.2	85.0	92.1	94.7
0.30	74.2	85.5	92.3	94.1

从图 7 发现模型在训练过程的早期收敛速度很慢,仅使用硬样本进行比较不利于学习广义特征和获得更好的聚类伪标签.相反,当  $\mu=1$  且仅使用聚类级损失  $L_c$  与  $L_{PLRL}$  时,虽然可以实现更快的收敛,失去了类内的多样性,仍然不利于网络学习更多的判别特征.当  $\mu=0.5$  时, mAP 和 Rank1 值均达到最高,可以看出两者结合明显提高了性能.



(a) Market1501



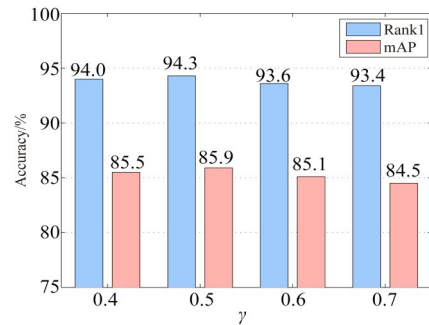
(b) DukeMTMC-reID

图 7  $\mu$  不同取值对模型准确率的影响结果

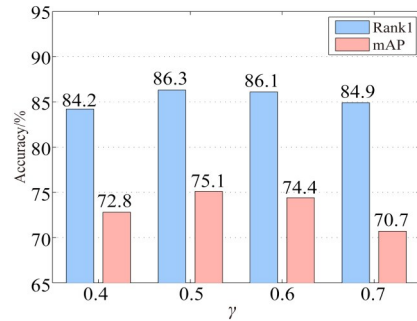
(5) 计算总体损失  $L$  时控制  $L_{PLRL}$  的参数  $\gamma$

在计算总体损失函数时,其中的参数  $\gamma$  控制伪标签正则化损失对最终对比损失的贡献,在实验中将  $\gamma$  分别设置为 0.4、0.5、0.6、0.7 来训练对应的重识别模型,得到的结果如图 8 所示.数据集选择 Market1501 与 DukeMTMC-reID,从图 8 中可以看出,当阈值设置为 0.5 时,模型取得了最优结果,而且不同的阈值对结果的准

确率影响比较大,也间接说明了识别噪声伪标签对模型的重要性.



(a) Market1501



(b) DukeMTMC-reID

图 8  $\gamma$  不同取值对 mAP 和 Rank1 的影响

4.6 可视化分析

(1) t-SNE 可视化分析

从 Market1501 中随机选择 10 个人,用 t-SNE 可视化他们在不同训练时期的特征.图 9 和图 10 分别展示了基线模型的结果以及本文方法训练的模型的结果,用不同的颜色来表示身份.可以发现随着两种设置的模型训练,同一身份的特征信息逐渐积累、丰富和完善,这证明本文的方法能够学习出具有辨别力的个体特征表示,而且本文提出的方法训练的模型可以生成更紧凑的特征簇,证明该方法可以减少噪声伪标签的影响,从而进一步促进优化.

(2) 检索结果可视化

为了证明本文提出方法的有效性,在 Market1501 数据集上对部分检索结果进行了可视化.结果如图 11 所示,第一列是待查询图像,其他列则是候选图像,候选图像顶部的数字表示该图像与查询图像的相似度排名.在候选图像中,红色数字表示不同行人,而绿色数字表示该候选图像与待查询图像是同一行人.从图 11 中前 4 行的检索结果可以看出,利用本文提出的方法训练得到的模型对于不同姿势、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

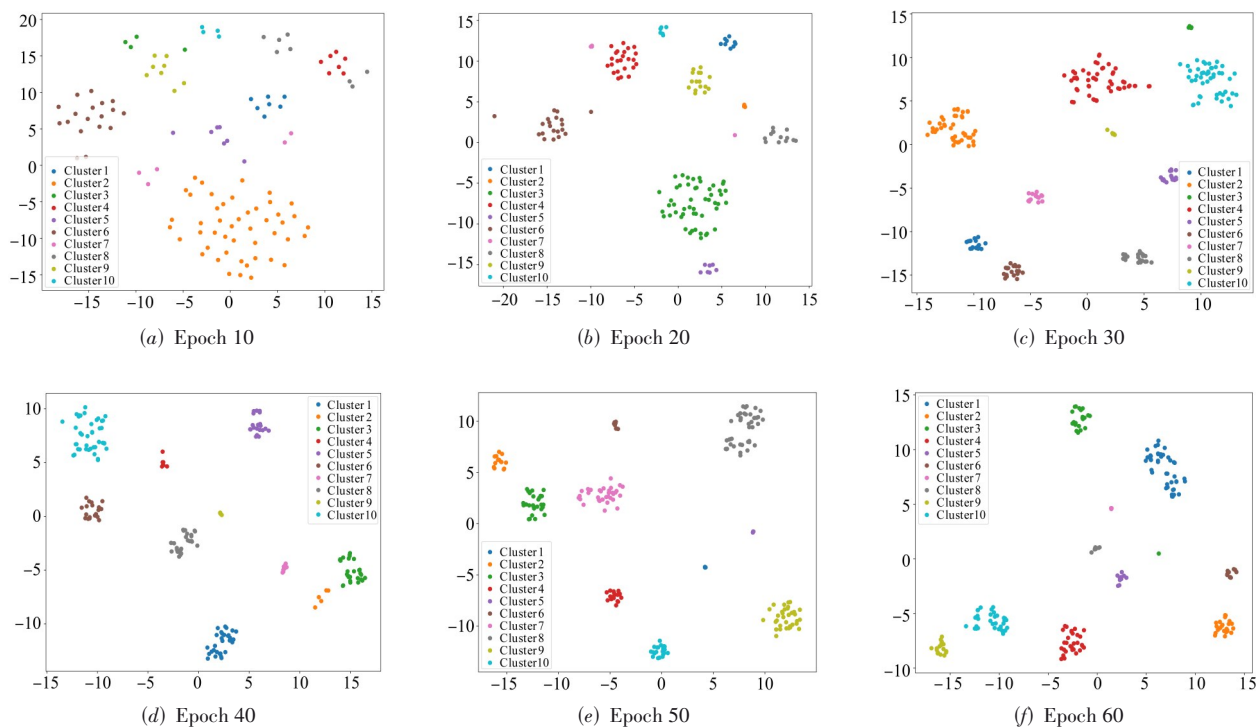


图9 基线方法训练的模型结果t-SNE可视化

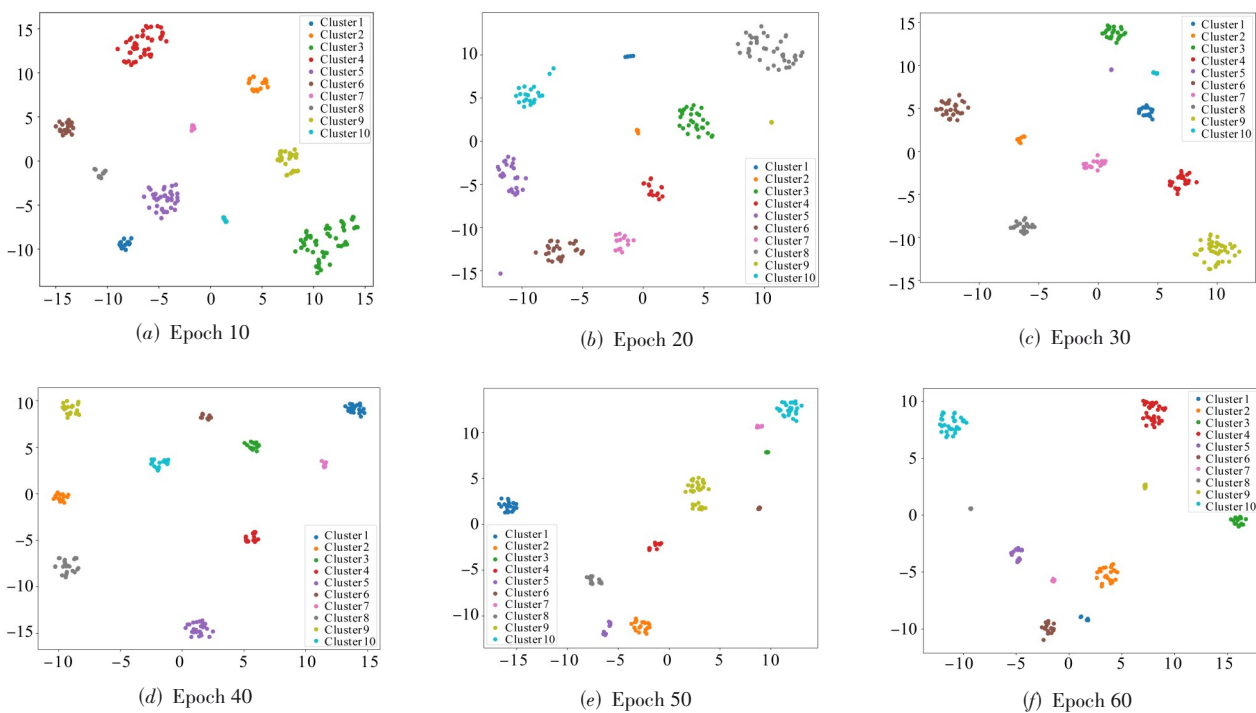


图10 本文方法训练的模型结果t-SNE可视化

有效的. 而后两行的检索结果中错误的示例较多,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,在图片分辨率不高的前提下,难以区分为不同的行人.

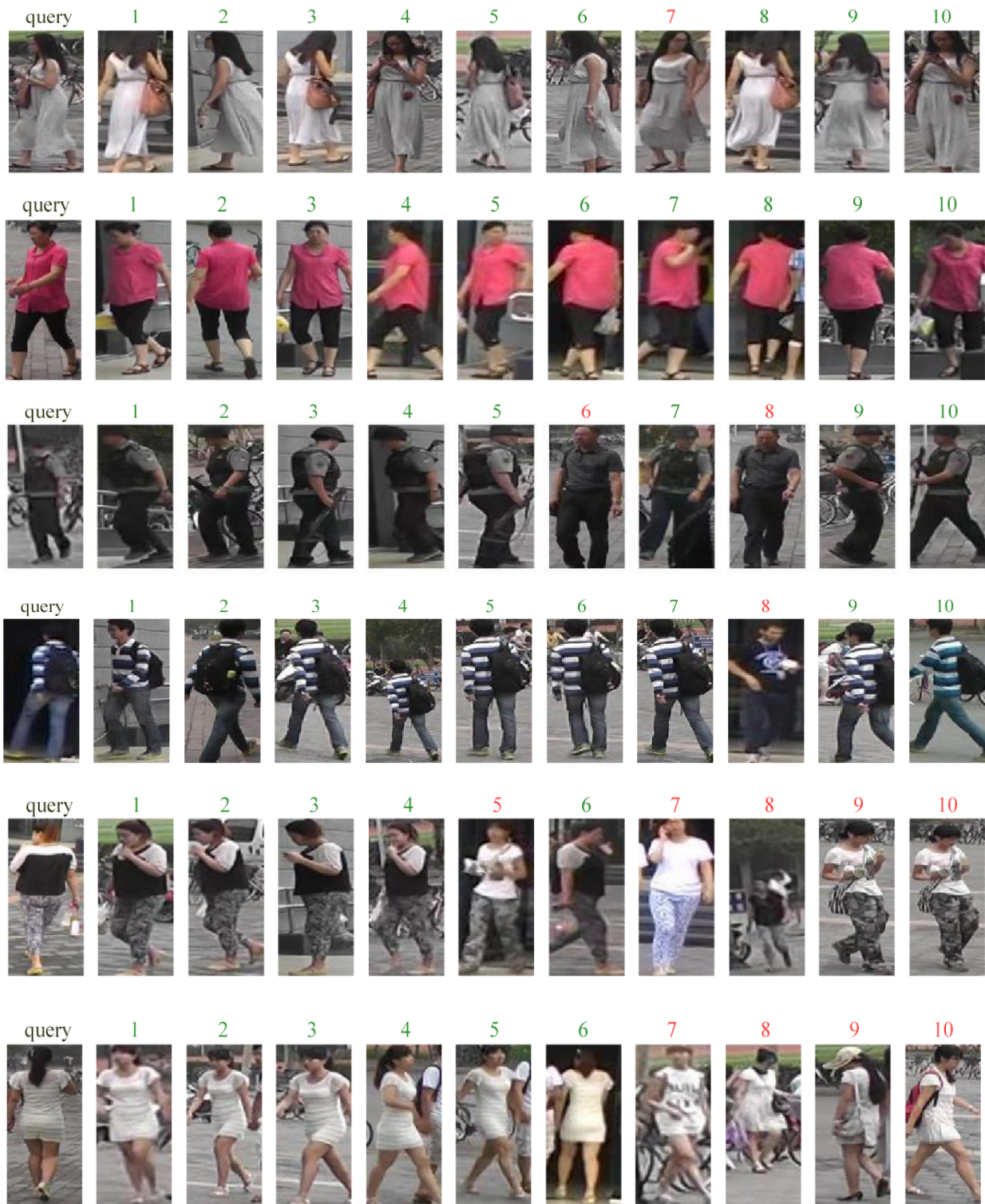


图 11 本文方法检索结果的可视化

(3) 热力图可视化

为了更有效地说明本文提出方法的有效性,我们使用经过训练的模型,通过 GradCAM<sup>[39]</sup>技术,对 Market1501 数据集中的部分图像进行了热力图可视化分析,

图 12 呈现了相应的结果. 图 12(a)为原始图像,图 12(b)为基于基线模型生成的热力图,图 12(c)为基于本文提出的方法进行模型训练后生成的热力图. 通过观察图 12 可以看出,与基线模型相比,本文提出的方法可以更有效地提取

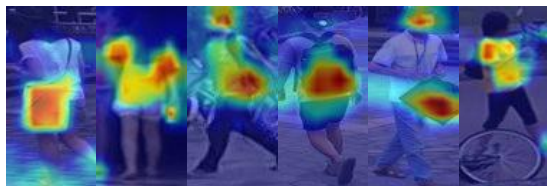
具有较强判别能力的特征,例如背包、头部、行人穿着等,进一步验证了本文方法在图像特征提取方面的优越性.



(a) 原始图像



(b) 基于基线模型生成的热力图



(c) 基于本文方法训练后生成的热力图

图12 热力图可视化结果

## 5 总结

本文提出了一种基于伪标签正则化损失的无监督行人重识别方法. 该方法首先提出了一种聚类引导的注意力分数,根据伪标签与聚类中心的语义相关程度来估计伪标签的置信度,以此来识别噪声标签并给正确标签分配更多的权重. 此外设计了一种新的伪标签正则化损失函数,利用伪标签进行软样本挖掘,构建正负样本对并为每个正负样本对计算一个连续的权重分数,充分利用伪标签的判别能力. 利用伪标签的置信度和样本相似度从两个角度对伪标签噪声进行约束,从而有效减少标签噪声对训练过程的影响. 多个数据集的实验结果验证了该方法的有效性,并通过消融实验验证了消除伪标签噪声的必要性和本文消除伪标签噪声的合理性. 但该方法在区分难样本、噪声样本和离群值方面仍然有改进空间,未来将探索在现有算法的基础上提升无监督行人重识别模型的准确率.

## 参考文献

- [1] 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.  
KUANG C, CHEN Y. Multi-granularity feature fusion network for person re-identification[J]. Acta Electronica Sinica, 2021, 49(8): 1541-1550. (in Chinese)
- [2] GU X Q, CHANG H, MA B P, et al. Clothes-changing person re-identification with RGB modality only[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 1050-1059.
- [3] 孙锐, 张磊, 余益衡, 等. 基于局部异构聚合图卷积网络的跨模态行人重识别[J]. 电子学报, 2023, 51(4): 810-825.  
SUN R, ZHANG L, YU Y H, et al. Cross-modality person re-identification based on locally heterogeneous polymerization graph convolutional network[J]. Acta Electronica Sinica, 2023, 51(4): 810-825. (in Chinese)
- [4] 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, 14(4): 83.
- [5] FU Y, WEI Y C, WANG G S, et al. Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2019: 6111-6120.
- [6] ZHANG X Y, CAO J W, SHEN C H, et al. Self-training with progressive augmentation for unsupervised cross-domain person re-identification[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2019: 8221-8230.
- [7] LLOYD S. Least squares quantization in PCM[J]. IEEE Transactions on Information Theory, 1982, 28(2): 129-137.
- [8] ESTER M, KRIEGLER H P, SANDER J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[C]//Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. New York: ACM, 1996: 226-231.
- [9] GE Y X, CHEN D P, LI H S. Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification[EB/OL]. (2020-01-30) [2023-08-30]. <http://arxiv.org/abs/2001.01526>.
- [10] 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: ACM, 2020: 11309-11321.
- [11] 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.
- [12] HU Z, ZHU C, HE G. Hard-sample guided hybrid contrast learning for unsupervised person re-identification

- [C]//2021 7th IEEE International Conference on Network Intelligence and Digital Content (IC-NIDC). Piscataway: IEEE, 2021: 91-95.
- [13] XU M Y, GUO H Y, JIA Y H, et al. Pseudo label rectification with joint camera shift adaptation and outlier progressive recycling for unsupervised person re-identification[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(3): 3395-3406.
- [14] LI P N, WU K Y, HUANG S Z Q, et al. Pseudo labels refinement with intra-camera similarity for unsupervised person re-identification[EB/OL]. (2023-04-25)[2023-08-30]. <http://arxiv.org/abs/2304.12634>.
- [15] YIN J H, ZHANG X Y, MA Z Y, et al. A real-time memory updating strategy for unsupervised person re-identification[J]. IEEE Transactions on Image Processing, 2023, 32: 2309-2321.
- [16] YE M, YUEN P C. PurifyNet: A robust person re-identification model with noisy labels[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 2655-2666.
- [17] YE M, LI H, DU B, et al. Collaborative refining for person re-identification with label noise[J]. IEEE Transactions on Image Processing, 2022, 31: 379-391.
- [18] SUN H, LI M K, LI C G. Hybrid contrastive learning with cluster ensemble for unsupervised person re-identification[C]//Asian Conference on Pattern Recognition. Cham: Springer, 2022: 532-546.
- [19] ZHENG K C, LAN C L, ZENG W J, et al. Exploiting sample uncertainty for domain adaptive person re-identification[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(4): 3538-3546.
- [20] 彭锦佳, 王辉兵. 基于异构卷积神经网络集成的无监督行人重识别方法[J]. 电子学报, 2023, 51(10): 2902-2914. PENG J J, WANG H B. An unsupervised person re-identification method based on heterogeneous convolutional neural networks ensemble[J]. Acta Electronica Sinica, 2023, 51(10): 2902-2914. (in Chinese)
- [21] DAI Z Z, WANG G Y, YUAN W H, et al. Cluster contrast for unsupervised person re-identification[C]//Asian Conference on Computer Vision. Cham: Springer, 2023: 319-337.
- [22] ZHAO Y Y, ZHONG Z, YANG F X, et al. Learning to generalize unseen domains via memory-based multi-source meta-learning for person re-identification[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 6273-6282.
- [23] 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.
- [24] WANG X S, HUA Y, KODIROV E, et al. Deep metric learning by online soft mining and class-aware attention [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 5361-5368.
- [25] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 770-778.
- [26] 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, 2015: 1116-1124.
- [27] 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. Amsterdam: Springer, 2016: 17-35.
- [28] 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.
- [29] ZHANG Z Z, LAN C L, ZENG W J, et al. Relation-aware global attention for person re-identification[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 3183-3192.
- [30] LIN Y T, DONG X Y, ZHENG L, 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.
- [31] 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.
- [32] 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.
- [33] 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.

- [34] 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.
- [35] 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, 2019: 8320-8329.
- [36] 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, 2019: 598-607.
- [37] 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.
- [38] LI M X, ZHU X T, GONG S G. Unsupervised noisy tracklet person re-identification[EB/OL]. (2021-01-16) [2023-08-30]. <http://arxiv.org/abs/2101.06391>.
- [39] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization[C]//2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 618-626.



**钱宇华** 男,1976年出生于山西晋城. 博士,山西大学教授,博士生导师. 国家优秀青年基金获得者,全球高被引科学家. 主要研究方向为粗糙集理论、粒度计算与人工智能.

E-mail: jinchengqyh@126.com



**阮秋琦** 男,1944年出生于河北玉田. 博士,北京交通大学教授,博士生导师. 北京交通大学学位委员会副主席,国务院学位委员会学科评议组成员,中国电子学会会士、理事、国际交流委员会副主任. 主要研究方向为数字图像处理、模式识别.

E-mail: qqraun@bjtu.edu.cn

## 作者简介



**贾洁茹** 女,1990年出生于山西运城. 博士,山西大学讲师. 国家自然科学基金青年项目获得者. 主要研究方向为计算机视觉、模式识别. 中国电子学会会员编号:E190036481M.

E-mail: jierujia@sxu.edu.cn



**张硕蕊** 女,1999年出生于山西运城. 山西大学计算机与信息技术学院硕士. 主要研究方向为图像处理、计算机视觉.

E-mail: 202122407069@email.sxu.edu.cn