

# 基于精确特征分布匹配和多域信息融合的跨域行人重识别

杨真真<sup>1,2</sup> 邵静<sup>1</sup> 杨永鹏<sup>1,3</sup> 吴心怡<sup>2</sup>

(1. 南京邮电大学宽带无线通信与传感网技术教育部重点实验室, 江苏南京 210003; 2. 南京邮电大学理学院, 江苏南京 210023; 3. 南京信息职业技术学院网络与通信学院, 江苏南京 210023)

**摘要:** 无监督域自适应行人重识别(Unsupervised Domain Adaptation for person Re-identification, UDA-ReID)任务致力于将知识从已标记的源域数据转移到目标域。和传统的单源域自适应相比,将多源域的知识迁移到目标域是一项更具挑战性的任务。由于领域上的差距,多数数据集的简单组合只能产生有限的改进。针对此问题,提出了一种基于精确特征分布匹配和多域信息融合的多源域对比学习(exact feature distribution Matching and multi-domain information Fusion based Multi-domain Contrastive Learning, MFML)方法。该方法首先采用具有混合记忆的自步对比学习提取不同域数据的特征,并对提取到的特征进行构图,然后通过两层残差图卷积网络进行多域特征融合。其次,为了增强交叉分布特征、产生更丰富的信息,通过基于排序算法的精确直方图匹配来实现精确特征分布匹配,以获得更多样化的特征增强。实验表明,与目前先进的无监督域自适应行人重识别方法相比,所提出的MFML方法在广泛使用的行人重识别数据集Market1501、MSMT17和Duke上都取得了优越的性能。

**关键词:** 行人重识别; 无监督域自适应; 精确特征分布匹配; 多域信息融合; 图卷积网络

**中图分类号:** TN911.73 **文献标识码:** A **DOI:** 10.16798/j.issn.1003-0530.2023.06.015

**引用格式:** 杨真真,邵静,杨永鹏,等. 基于精确特征分布匹配和多域信息融合的跨域行人重识别[J]. 信号处理, 2023,39(6): 1099-1107. DOI: 10.16798/j.issn.1003-0530.2023.06.015.

**Reference format:** YANG Zhenzhen, SHAO Jing, YANG Yongpeng, et al. Exact feature distribution matching and multi-domain information fusion for person re-identification[J]. Journal of Signal Processing, 2023, 39(6): 1099-1107. DOI: 10.16798/j.issn.1003-0530.2023.06.015.

## Exact Feature Distribution Matching and Multi-domain Information Fusion for Person Re-identification

YANG Zhenzhen<sup>1,2</sup> SHAO Jing<sup>1</sup> YANG Yongpeng<sup>1,3</sup> WU Xinyi<sup>2</sup>

(1. Key Laboratory of Ministry of Education in Broadband Wireless Communication and Sensor Network Technology, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu 210003, China; 2. College of Science, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu 210023, China; 3. School of Network and Communication, Nanjing Vocational College of Information Technology, Nanjing, Jiangsu 210023, China)

**Abstract:** The task of unsupervised domain adaptation for person re-identification (UDA-ReID) is to transfer knowledge from the marked source domain to the unmarked target domain. Compared with the traditional single-source domain adaptation, it is a more challenging task to transfer multi-source domain knowledge to the target domain. Due to the gap between

收稿日期: 2022-11-09; 修回日期: 2022-12-22

基金项目: 国家自然科学基金(61501251, 62171232); 南京邮电大学宽带无线通信与传感网技术教育部重点实验室开放研究基金(JZNY202113); 南京邮电大学科研项目(NY220207)

different domains, the simple combination of multiple datasets can only be improved limited. To solve this problem, a multi-domain contrastive learning method based on exact feature distribution matching and multi-domain information fusion (MFMCL) was proposed. This method firstly extracted the features of data in different domains based on the self-paced contrastive learning with hybrid memory, and constructed the knowledge graph based on the extracted features, and then fused the multi-domain features through the two-layer residual graph convolution network. In addition, in order to enhance the cross distribution features between different domains and generate more abundant image style information, the exact feature distribution matching was realized by accurate histogram matching based on sorting algorithm, so as to obtain more diversified feature enhancement. Compared with the advanced UDA-ReID methods, the experimental results show that our proposed MFMCL method achieves the best performance on the widely used ReID datasets Market1501, MSMT17 and Duke.

**Key words:** person re-identification; unsupervised domain adaptation; exact feature distribution matching; multi-domain information fusion; graph convolution network

## 1 引言

随着大规模数据集的出现,以及特征提取和度量学习方法的改进,行人重识别(person Re-identification, ReID)受到了很多关注<sup>[1]</sup>。无监督域自适应行人重识别(Unsupervised Domain Adaptation for person Re-identification, UDA-ReID)<sup>[2]</sup>旨在将源域上训练的模型自适应到未身份注释的目标域。UDA-ReID任务提出了许多新的方法<sup>[3-9]</sup>,这些方法主要分为三类:域特征对齐、图像风格迁移和基于集群聚类的方法。其中,基于集群聚类生成伪标签的自步对比学习(Self-paced Contrastive Learning, SpCL)<sup>[8]</sup>方法表现出卓越的性能。

虽然 SpCL 采用对比学习训练网络并通过混合记忆存储器在线微调伪标签,在正确识别无标注数据上取得了先进的效果,但是该模型仅使用单一源域的有限数据进行训练,在实际应用中对目标域的积极影响往往是有限的,且其他标记的数据集没有得到充分的利用。为了充分利用包含许多变化和环境因素的训练数据并且进一步提高网络对未标记数据的识别性能,本文着重于研究如何将仅单一源域训练的对比学习模型应用到多源域 UDA-ReID 任务<sup>[10-11]</sup>上,并提出多源域对比学习模型。然而,由于多个源域与目标域之间具有领域差异,仅仅将多个数据集简单组合作为训练数据可能会导致模型过拟合,甚至阻碍该模型的泛化能力。

随着图卷积网络(Graph Convolutional Network, GCN)<sup>[12]</sup>的广泛应用,在多标签图像分类模型<sup>[13]</sup>中, XU 等人提出在目标域和不同源域的原型上构造一个知识图,然后利用图卷积网络 GCN 模型对知识图

中的每个节点进行信息传递并输出分类概率。该方法通过各个域的原型之间的信息传递来实现域的融合,可以有效缓解不同域之间的领域差异,从而灵活地辅助目标域预测。另外, Bai 等人<sup>[14]</sup>利用 GCN 提出了一种无监督多源域自适应(Unsupervised Multi-Source Domain Adaptation, UMSDA)方法,并通过实验表明了基于 GCN 的多域信息融合(Multi-Domain Information Fusion, MDIF)对减小不同行人数据之间领域差异的有效性。受上述启发,为了解决多源域 UDA-ReID 任务中存在的问题,本文在多源域对比学习模型中引入多域信息融合方法,在不同域的域代理节点之间以及不同域中所有实例节点之间分别构建一个知识图,并分别将其在两层 GCN 网络中传播,从而实现不同域行人特征之间的信息融合。相比于 UMSDA 中所使用的同步平均教学(Mutual Mean-Teaching, MMT)<sup>[7]</sup>框架,本文所采用的基于混合记忆存储器的对比学习框架能够为网络生成更为可靠的行人伪标签,且不会额外增加网络模型参数。因此,本文将多源域对比学习与 GCN 相结合,能够使网络取得更好的识别性能。

此外,研究者指出可以将域适应转化为一个特征分布匹配问题,通过交叉分布的特征匹配来进一步提升域泛化能力<sup>[15-16]</sup>。在高斯特征分布的假设下,传统的特征分布匹配方法通常与特征的均值和标准差相匹配。然而,真实数据的特征分布通常比高斯分布复杂得多,高斯分布除了均值和标准差外,还有高阶统计量,使用高阶特征统计量可以更准确地表示样式信息,实现特征增强。为了在不增加计算量的情况下充分利用高阶统计量, Zhang 等人首次提出了精确特征分布匹配(Exact Feature

Distribution Matching, EFDM)<sup>[17]</sup>,并通过精确匹配图像高阶特征的经验累积分布函数来实现特征增强。基于Zhang等人<sup>[17]</sup>的研究启发,本文的多源域对比学习模型还同时考虑了在特征提取时通过应用EFDM方法生成语义信息保留和风格转移的特征增强,从而进一步改善跨域性能。

综上所述,针对多源域UDA-ReID任务中存在的源域和目标域之间的特征分布差异以及多源数据集训练时存在的域差异问题,本文对SpCL框架进行了改进,提出了一种基于精确特征分布匹配和多域信息融合的多源域对比学习(exact feature distribution Matching and multi-domain information Fusion based Multi-domain Contrastive Learning, MFMCL)框架,以进一步提升跨域行人重识别的识别性能。创新点如下:

(1)提出了一种基于精确特征分布匹配和多域信息融合的多源域对比学习框架。该框架通过嵌入精确特征分布匹配模块增强不同域之间的交叉特征,接着利用多域信息融合模块融合多个源域的特征,减少域距离,同时,使用三个混合特征存储器存储不同域的聚类信息,最后通过多源域对比损失训练网络。

(2)针对源域和目标域之间的特征分布差异问题,所提出的多源域对比学习框架在特征提取的低层引入精确特征分布匹配模块,利用高阶统计数据高效隐式地进行特征分布匹配,实现了在保留源域语义信息的同时增强源域和目标域之间的交叉分布特征。

(3)针对多源数据集训练时存在的域差异问题,所提出的多源域对比学习框架在池化层后嵌入了多域信息融合模块,该模块使域信息能够在域代理节点和编码特征之间传播,从而实现域融合,减少域距离。

## 2 相关工作

### 2.1 无监督域自适应行人重识别

UDA-ReID的目的是将在有标记的源域数据集上训练的模型调整到目标域数据集,而无须进行任何进一步的注释。其中,基于聚类的方法表现出卓越的性能,例如同步平均教学(Mutual Mean-Teaching, MMT)<sup>[7]</sup>和自步对比学习(SpCL)<sup>[8]</sup>框架。该聚类方法使用标记源域数据的特征预训练、对目标域数据

的基于聚类的伪标签预测,并使用伪标签的特征表示学习和微调。MMT利用更为鲁棒的软标签对伪标签进行在线优化,从而有效解决聚类算法中的伪标签噪声问题。SpCL将所有源域类标签、目标域聚类伪标签、非聚类实例存储到混合记忆存储器,并设计了一个统一的框架来动态更新和判别混合记忆存储器里不同的入口,能够跨域为联合学习判别特征表示提供有监督信号。由于聚类算法的不完善以及源数据与目标数据的域差距,使分配的伪标签包含不正确的标签,导致误导特征学习,影响域自适应能力。群感知标签转移(Group-aware Label Transfer, GLT)算法<sup>[9]</sup>被提出来解决这类问题,该算法将在线标签细化和群感知策略相结合,从而更好地在线纠正噪声伪标签,使预测更准确。另一方面,基于内存的多源元学习(Memory-based Multi-Source Meta Learning, M<sup>3</sup>L)框架<sup>[10]</sup>引入元学习策略来模拟领域泛化的训练测试过程,同时克服了不稳定的元优化。然而,M<sup>3</sup>L等方法使用一个专家模型为不同的源域学习一个共同的特征空间,这可能会忽略单个域的区别信息。XU等人通过研究多个专家模型的混合模型,提出了基于自适应聚合的模拟嵌入(Mimic Embedding via adapTive Aggregation, META)<sup>[11]</sup>方法,并利用域特定归一化层和全局分支来分别解决模型的可伸缩性和领域不变性,以改善多源域泛化问题。

### 2.2 特征分布匹配

直方图匹配(Histogram Matching, HM)<sup>[15]</sup>是一种经典的特征匹配方法,该方法除了匹配一阶和二阶特征统计量外,还引入高阶统计量来进行匹配,但同时也极大的增加了计算开销。HM在图像像素空间的输出差虽然很小,但这种微小的差异会在深度模型的特征空间中被放大,导致特征分布匹配存在明显的差异。精确直方图匹配(Exact histogram matching, EHM)<sup>[18]</sup>提出通过排序匹配来精确匹配图像像素的直方图。该方法随机或局部平均地区分等效像素值,并应用元素级变换,获得更准确的直方图匹配。其中的排序匹配是基于快速排序的策略<sup>[19]</sup>,被认为是速度最快的排序算法。根据自适应实例规范化(Adaptive Instance Normalization, AdaIN)特征分布匹配技术<sup>[20]</sup>,混合风格(Mix Style, MixStyle)<sup>[16]</sup>方法提出随机选择多个不同域的实例,并在底部CNN层的实例级特征之间采用概率凸组合的方式

匹配高斯统计量,增强了交叉分布特征。然而,真实数据集的特征分布往往比高斯分布复杂得多,此时AdaIN匹配的特征分布精度降低。

### 2.3 图卷积网络

知识图谱描述了实体及其相互关系,并组织在一个图中。在多标签图像分类任务中,WANG等人<sup>[13]</sup>提出利用知识图谱来显示其语义关系。图卷积网络(Graph Convolutional Network, GCN)<sup>[12]</sup>被设计用于直接计算图结构的数据,建模内部结构关系,从而在知识图谱上传播信息。目前图卷积网络已经被广泛应用于各种计算机视觉任务中,例如监督行人重识别<sup>[21-22]</sup>、动作识别<sup>[23]</sup>等。在多源域自适应领域,最近研究表明图卷积网络在减少领域差异方面表现出了显著的效果<sup>[13]</sup>。无监督多源域自适应(Unsupervised Multi-Source Domain Adaptation, UMSDA)方法<sup>[14]</sup>提出通过图卷积网络来减少行人数据间的领域距离。学会组合的多源领域自适应(Learning to Combine for Multi-Source Domain Adaptation, LtCMSDA)框架<sup>[24]</sup>提出在不同域的原型上构造一个知识图谱,以实现信息在语义相邻表示之间的传播。在此基础上学习一个图模型,在相关原型的指导下预测查询样本,以协助模型对目标域进行预测。

## 3 本文方法

### 3.1 整体架构

本文基于自步对比学习模型,构建了一个基于精确特征分布匹配和多域信息融合的多源域对比学习框架,该框架通过聚类生成不同级别的伪标签来训练网络,同时引入EFDM以增强行人特征,并引

入GCN以实现多域信息融合,最后利用混合记忆存储器实现多源域的自步对比学习。其整体架构如图1所示。

从图1可以看出,源域样本集 $S_1$ 、 $S_2$ 和目标域样本集 $T$ 分别经过特征提取器和广义均值池化(generalized-mean pooling, GeM)层后得到特征矩阵 $\tilde{S}_1 \in \mathbb{R}^{N \times C}$ 、 $\tilde{S}_2 \in \mathbb{R}^{N \times C}$ 和 $\tilde{T} \in \mathbb{R}^{N \times C}$ 。网络框架中的特征提取器为截取的Resnet50<sup>[25]</sup>网络结构池化前的五个阶段,分别记为L0、L1、L2、L3和L4,且除L0外其他四部分都包含了残差块,同时分别在L1和L2层后嵌入精确特征分布匹配模块EFDM(即EFDM1和EFDM2),其中EFDM模块用于产生更多样化的行人特征信息;此外,为了进一步减少领域差异,该网络框架在GeM层后嵌入了多域信息融合模块,该模块由带有残差连接的两个图卷积网络GCN1和GCN2以及激活函数(Rectified Linear Unit, ReLU)构成,其中GCN1和GCN2的具体结构遵循图卷积网络<sup>[12]</sup>模型并且引入域代理结点。信号经多域信息融合模块处理后得到的输出矩阵分别为 $\hat{F}(\tilde{S}_1)$ 、 $\hat{F}(\tilde{S}_2)$ 和 $\hat{F}(\tilde{T})$ ,并输入到域特定归一化层DSBN\_S1、DSBN\_S2和DSBN\_T得到源域和目标域样本的特征编码向量 $f^{S_1}$ 、 $f^{S_2}$ 和 $f^T$ 。一方面,得到的 $f^{S_1}$ 、 $f^{S_2}$ 和 $f^T$ 分别进入混合记忆存储器HM\_S1、HM\_S2和HM\_T中,分别训练得到对应的源域身份标签 $w^{S_1}$ 、 $w^{S_2}$ 和所有未聚类目标域实例特征 $v^T$ 和目标域聚类伪标签 $c^T$ ,同时,将上述所得结果与 $f^{S_1}$ 、 $f^{S_2}$ 和 $f^T$ 一起进入多源域对比损失 $L_{HM}$ 以实现迭代优化。之后利用每一轮迭代中获得的特征编码向量 $f^{S_1}$ 、 $f^{S_2}$ 和 $f^T$ 在线微调聚类结果。

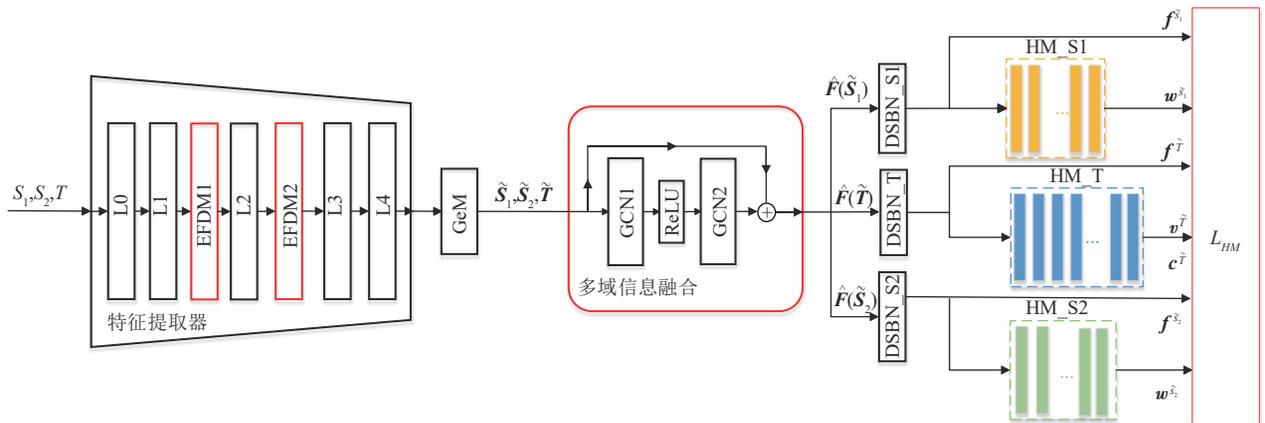


图1 提出的整体框架图

Fig. 1 The proposed overall framework

### 3.2 精确特征分布匹配

尽管 ReID 网络在源域提取和鉴别特征上表现良好,但往往不能很好地应用到目标域中,从而在跨域识别时仍表现出较低的识别精度。为了解决这个问题,本文在特征提取器模块中引入 2 个 EFDM 模块,分别嵌入到特征提取器的 L1 和 L2 层之后。嵌入的两个 EFDM(EFDM1, EFDM2)模块通过在行人特征空间中应用精确直方图匹配方法,实现对高阶统计量进行高效、隐式地匹配。以 EFDM1 为例,其具体处理流程如下所示:

首先,对于输入的源域数据集  $S$ , 目标域数据集  $T$  的每一个实例,特征提取器由 L1 层输出到 EFDM1 的源域特征向量为  $s$  (包含  $s_1, s_2$ ), 目标域行人特征向量为  $t$ 。EFDM1 模块首先将  $s$  和  $t$  按升序排序,排序后的向量分别记为  $s' = \{s'_j, j \in [1, n]\}$  和  $t' = \{t'_j, j \in [1, n]\}$ , 其索引分别存储在向量  $\tau$  和  $\eta$  中,具体表示如下:

$$\begin{aligned} \tau &= (\tau_1, \tau_2, \dots, \tau_n) \\ \eta &= (\eta_1, \eta_2, \dots, \eta_n) \end{aligned} \quad (1)$$

其中  $s'_{\tau_1}$  对应  $s$  中元素的最小值,  $s'_{\tau_n}$  对应  $s$  中元素的最大值,  $\tau_1, \tau_n$  为最小值  $s'_{\tau_1}$  与最大值  $s'_{\tau_n}$  在向量  $s$  的索引值,  $t'_{\eta_1}$  对应  $t$  中元素的最小值,  $t'_{\eta_n}$  对应  $t$  中元素的最大值,  $\eta_1, \eta_n$  为最小值  $t'_{\eta_1}$  与最大值  $t'_{\eta_n}$  在向量  $t$  的索引值。其次,为了在深度模型中实现梯度反向传播,将排序后的两个向量  $s'$  和  $t'$  经过公式(2)得到匹配的输出向量为  $o_{\tau_j}$ 。

$$o_{\tau_j} = s'_{\tau_j} + t'_{\eta_j} - \langle s'_{\tau_j} \rangle \quad (2)$$

其中,  $\langle \cdot \rangle$  表示停止梯度操作。另外,为了生成增强更多混合信息风格的不同特征,本文使用更加准确的方法对两个向量  $s'$  和  $t'$  进行处理,从而得到更精确的特征分布混合,其公式如下所示:

$$o_{\tau_j} = s'_{\tau_j} + (1 - \lambda)t'_{\eta_j} - (1 - \lambda)\langle s'_{\tau_j} \rangle \quad (3)$$

其中  $\lambda \sim \text{Beta}(\alpha, \alpha)$ ,  $\alpha \in (0, \infty)$  为超参数。通过式(3)可以得到两个源域特征并与目标域特征输入到特征提取器的 L2 层。同理, EFDM2 的处理流程与 EFDM1 类似。

### 3.3 多域信息融合

本文为了达到降低域差异的目的,还引入了多域信息融合模块。该模块包括带有残差连接的两层 GCN, 其中第一层 GCN(GCN1)负责融合全局域

表示,第二层 GCN(GCN2)确保所有实例都能够接收来自其他域的信息。其具体流程如下所示:

第一,将源域数据集  $S$ , 目标域数据集  $T$  经过特征提取器和广义均值池化处理后,并输入到多域信息融合层的特征矩阵表示分别记为  $\tilde{S}_1, \tilde{S}_2$  和  $\tilde{T}$ 。

第二,为了避免直接融合实例的特征容易产生与身份相关的噪声问题,本文针对不同的域数据集引入了可以表征全局域信息的域代理节点。域代理节点为对同一域的实例特征的加权组合,即  $\tilde{S}_1, \tilde{S}_2$  和  $\tilde{T}$  对应的域代理节点表示如下式所示:

$$\bar{s}_1 = \sum_{n=1}^N w_{1n} \cdot \tilde{S}_{1n}, \bar{s}_2 = \sum_{n=1}^N w_{2n} \cdot \tilde{S}_{2n}, \bar{t} = \sum_{n=1}^N w_n \cdot \tilde{T}_n \quad (4)$$

其中  $\bar{s}_1, \bar{s}_2$  表示  $\tilde{S}_1, \tilde{S}_2$  的域代理节点,  $\bar{t}$  表示目标域  $\tilde{T}$  的代理节点,  $N$  表示  $\tilde{S}_1, \tilde{S}_2$  和  $\tilde{T}$  的样本数量,  $w_{1n}, w_{2n}$  和  $w_n$  是可学习的权值,  $\tilde{S}_{1n}, \tilde{S}_{2n}$  和  $\tilde{T}_n$  分别为  $\tilde{S}_1, \tilde{S}_2$  和  $\tilde{T}$  的第  $n$  个实例对应的向量。

第三,将  $\tilde{S}_1, \tilde{S}_2, \tilde{T}$  和三个对应的域代理节点按行堆叠生成一个待输入到多域信息融合模块的图信号  $F^{(0)} \in \mathbb{R}^{(3N+3) \times C}$ ,  $C$  为每个实例的特征维数。

第四,获取第  $l$  个 GCN 层的邻接矩阵  $A^{(l)}$  ( $l = 1, 2$ )。  $A^{(l)} \in \mathbb{R}^{(3N+3) \times (3N+3)}$  是根据两个实例节点之间的连接关系得到的,其  $A^{(l)}$  的每个元素  $A^{(l)}_{r,c}$  由式(5)可得:

$$A^{(l)}_{r,c} = \begin{cases} 1, & r = c \text{ or node } r \text{ connect with node } c \\ 0, & \text{else} \end{cases} \quad (5)$$

其中,  $r$  和  $c$  分别表示邻接矩阵  $A^{(l)}$  的行索引和列索引。其中,为了融合全局域的信息,本文使邻接矩阵  $A^{(1)}$  和  $A^{(2)}$  的三个域代理节点  $\bar{s}_1, \bar{s}_2$  和  $\bar{t}$  之间存在连接,同时,为了确保所有实例都能够接收来自其他域的信息,本文使  $A^{(2)}$  中的  $\tilde{S}_1, \tilde{S}_2$  和  $\tilde{T}$  对应的每个实例特征都连接到其对应的域代理节点中。另外,为了增强 GCN1 和 GCN2 对特征的代表能力,本文还对邻接矩阵  $A^{(l)}$  进行归一化操作,如式(6)所示:

$$\tilde{A}^l = (D^l)^{-\frac{1}{2}} A^{(l)} (D^l)^{-\frac{1}{2}} \quad (6)$$

其中  $D^l$  为邻接矩阵  $A^{(l)}$  的度矩阵。

最后,将上述得到的图信号  $F^{(0)}$  和邻接矩阵  $A^{(l)}$  输入到包括带有残差连接的两层 GCN 的多域信息融合模块中进行处理。其中每个 GCN 层都可以写成一个如式(7)所示的非线性变换:

$$F^l = \rho(\tilde{A}^l F^{l-1} W^l) \quad (7)$$

其中,  $\rho$  为经典的非线性函数,本文采用的是带泄露

的修正线性单元(Leaky Rectified Linear Unit, LeakyReLU),  $\mathbf{W}$ 为可学习的加权矩阵。最后,将经过两层GCN处理后的结果  $\mathbf{F}^{(2)}$ 和原始图信号  $\mathbf{F}^{(0)}$ 进行残差连接处理得到最终的多域信息融合特征  $\hat{\mathbf{F}}$ 为:

$$\hat{\mathbf{F}} = \mathbf{F}^{(0)} + \mathbf{F}^{(2)} \quad (8)$$

并对其按域分解最终得到  $\tilde{\mathbf{S}}_1$ 、 $\tilde{\mathbf{S}}_2$ 和  $\tilde{\mathbf{T}}$ 对应的处理结果  $\hat{\mathbf{F}}(\tilde{\mathbf{S}}_1)$ 、 $\hat{\mathbf{F}}(\tilde{\mathbf{S}}_2)$ 和  $\hat{\mathbf{F}}(\tilde{\mathbf{T}})$ 。

通过上述多域信息融合方法,从而实现在多个域之间进行信息融合,进一步减小域间隙,提高域泛化能力。

### 3.4 对比学习损失

本文使用对比学习损失  $L_{HM}$ 作为多源域自适应网络的整体优化损失函数。对比学习损失  $L_{HM}$ 的输入分别为经过域特定归一化层DSBN\_S1、DSBN\_S2和DSBN\_T得到的源域和目标域样本的特征编码向量  $f^s$ 、 $f^{s_2}$ 和  $f^T$ ,以及通过混合记忆存储器HM\_S1、HM\_S2和HM\_T分别对上述三个特征向量进行训练后的源域身份标签  $w^s$ 、 $w^{s_2}$ 以及所有未聚类目标域实例特征  $v^T$ 和目标域聚类伪标签  $c^T$ 。其过程如下所述:

第一,在混合记忆存储器HM\_S1和HM\_S2里分别初始化源域身份标签  $w^{s_1} \in \mathbb{R}^{N \times n^1}$ ,  $w^{s_2} \in \mathbb{R}^{N \times n^2}$ ,并将其堆叠为  $w^s = w^{s_1} \cup w^{s_2}$ ,同时,在HM\_T中初始化所有未聚类目标域实例特征  $v^T \in \mathbb{R}^{N \times n^c}$ 和目标域聚类伪标签  $c^T \in \mathbb{R}^{N \times n^c}$ 。其中  $n^1$ ,  $n^2$ 为源域类别的数量,  $n^c$ 为目标域集群聚类的数量,  $n^v$ 为目标域非集群实例特征的数量。

第二,在训练过程中不断更新聚类结果,并通过公式(9)~(11)将属于  $k$ 类的源域身份标签  $w_k^s$ 以及所有未聚类目标域实例特征  $v^T$ 和目标域聚类伪标签  $c^T$ 进行更新,更新过程如下式所示:

$$w_k^s = m_s w_k^s + \frac{(1 - m_s)}{|\mathbf{D}_k|} \sum_{f_i^s \in \mathbf{D}_k} f_i^s \quad (9)$$

$$v_i^T = m_t v_i^T + (1 - m_t - f_i^T) \quad (10)$$

$$c_k^T = \frac{1}{|\mathbf{I}_k|} \sum_{v_i^T \in \mathbf{I}_k} v_i^T \quad (11)$$

其中,  $f^s = f^{s_1} \cup f^{s_2}$ ,  $f_i^s$ 是  $f^s$ 第  $i$ 个特征向量,  $f = f^{s_1} \cup f^{s_2} \cup f^T$ ,  $f_i^T$ 是  $f^T$ 的第  $i$ 个特征向量,  $v_i^T$ 是  $v^T$ 的第  $i$ 个特征向量,  $c_k^T$ 是属于第  $k$ 类的目标域聚类伪标签,  $m_s \in [0, 1]$ 表示更新源域类质心的动量系数,  $m_t \in [0, 1]$ 表示更新目标域实例的动量系数;  $\mathbf{D}_k$ 表示属于源域类  $k$ 的特征集;  $\mathbf{I}_k$ 表示包含聚类  $k$ 中所有特

征向量的第  $k$ 个聚类集,  $|\cdot|$ 表示该集合中的特征数。

第三,本文采用对比损失  $L_{HM}$ 进行训练,使编码的特征向量不断接近其分配类别,其公式如下所示:

$$L_{HM} = -\log(\exp(\langle f_i, z^+ \rangle / \sigma) / (q(f_i, w_k^s) + q(f_i, c_k^T) + q(f_i, v_i^T))) \quad (12)$$

其中  $q(f_i, w_k^s) = \sum_{k=1}^{n^s} \exp(\langle f_i, w_k^s \rangle / \sigma)$ ,  $q(f_i, c_k^T) = \sum_{k=1}^{n^c} \exp(\langle f_i, c_k^T \rangle / \sigma)$ ,  $q(f_i, v_i^T) = \sum_{k=1}^{n^v} \exp(\langle f_i, v_i^T \rangle / \sigma)$ 。  $\langle \cdot, \cdot \rangle$

表示两个特征向量之间的内积,  $z^+$ 表示与  $f_i$ 对应的正类原型,  $\sigma$ 为温度参数。

## 4 实验结果与分析

### 4.1 数据集与评价指标

目前广泛使用的ReID数据集有Market1501<sup>[26]</sup>、MSMT17<sup>[27]</sup>和Duke<sup>[28]</sup>,此外还有小规模数据集CUHK03(新协议)<sup>[29]</sup>。本文在这几个数据集上验证了提出的基于精确特征分布匹配和多域信息融合的多源域对比学习网络的有效性和优越性。

本文实验主要使用两个定量评估指标:rank-n和平均均值精度(mean Average Precision, mAP)。rank-n指的是在搜索的前  $n$ 个结果中正确结果的分数,表1和表2中分别简记为R-1、R-5和R-10。在算法返回的排序列表中,通过计算前  $n$ 个图像中的正确分类的图像的数量来获得排名的准确性。图像分类任务中大多数都有多种标签,mAP则是多分类任务中平均精度的平均总和,是ReID任务中常用的定量评估指标。

### 4.2 实验细节

本文使用Resnet50作为特征提取的骨干网络,动量优化选择Adam优化算法,其权值衰减率为0.0005。模型初始学习率设置为0.00035,并在每迭代20轮之后缩小10%。实验中设置EFDM中的超参数  $\alpha = 0.1$ ,并采用0.5的概率来决定EFDM在训练阶段是否被激活。公式(12)中的温度参数  $\sigma$ 参考文献[8]设置为0.05。源域身份标签  $w^{s_1}$ ,  $w^{s_2}$ 初始化为数据集给定的原始身份标签,  $v^T$ 和  $c^T$ 都初始化为0。伪标签生成遵循基于聚类的方法,使用具有噪声的基于密度的空间聚类方法<sup>[8]</sup>进行近距离聚类。混合记忆存储器通过使用图像预训练编码器提取整个训练集进行初始化,然后动态更新公式(9)和

(10), 其中  $m_s = m_t = 0.2$ 。

### 4.3 消融实验

本小节分别以 Market1501、MSMT17 和 Duke 作为消融实验中无标注的测试数据集, 以验证在本文所提出的基于精确特征分布匹配和多域信息融合的多源域对比学习模型对多源域自适应行人重识别性能的影响。将经过调整和配置后得到的网络模型作为实验的强基线模型, 记为 M-SpCL。在特征提取网络底层嵌入 EFDM 进行图像特征增强, 然后在特征提取后采用残差 GCN 进行多域信息融合, 并分别验证其有效性, 实验结果如表 1 所示。

表 1 中, “MSMT17+CUHK03->Market1501”是指 MSMT17 和 CUHK03 为源域, Market1501 为目标域, 其余类似。从表 1 中可以看出, 引入的 M-SpCL+EFDM、M-SpCL+MDIF 都比基线模型 M-SpCL 具有更好的性能, 提出的 MFMCCL 即同时融合了两者的对比学习模型的改善效果更为显著。具体来说, 提出的 MFMCCL 与 M-SpCL 相比, 在 Market1501 数据集上 mAP 提升了 0.7%, rank-1 提升了 0.8%; 在 MSMT17 数据集上 mAP 提升了 1.9%, rank-1 提升了 1.8%; 在 Duke 数据集上 mAP 提升了 1.0%, rank-1 提升了 1.9%。可以得出结论, 首先, EFDM 在多源域对比学习模型中起着积极的作用, 它使得模型在训练时

保留了更多的源域匹配信息。其次, 通过采用基于 GCN 的多域信息融合方法, 使对比学习模型能够对提取到的多源域信息进行融合, 以改善模型性能。最后, 两种方法的组合得到提出的基于 MFMCCL 的多源域对比学习模型, 取得了最佳的性能。

### 4.4 对比实验

为了验证所提出的基于精确特征分布匹配和多域信息融合的多源域对比学习模型的优越性, 将其与先进的 UDA-ReID 方法 MMT<sup>[7]</sup>、SpCL<sup>[8]</sup>、GLT<sup>[9]</sup>、M<sup>3</sup>L<sup>[10]</sup>、META<sup>[11]</sup> 及 UMSDA<sup>[14]</sup> 进行比较。表 2 中给出了本文提出的方法和其他先进的 UDA-ReID 方法的对比结果。

由表 2 可以看出, 与次最佳性能的域适应方法 GLT 相比, 以 Market1501 为测试数据集时, 提出的 MFMCCL 分别在 mAP 和 rank-1 上提高了 2.4% 和 1.6%; 以 MSMT17 为测试数据集时, 提出的方法分别在 mAP 和 rank-1 上提高了 4.6% 和 3.1%; 以 Duke 为测试数据集时, 与次最佳域适应方法 SpCL 相比, 提出的方法分别在 mAP 和 rank-1 上提高了 2.3% 和 2.3%。此外, 以 MSMT17 和 CUHK03 为源域, Market1501 为目标域时, 与次最佳性能的多源域模型 UMSDA 相比, 提出 MFMCCL 在 mAP 和 rank-1 上分别高出 0.4% 和 1.6%; 以 Market1501 和 CUHK03 为源

表 1 识别结果(%)

Tab. 1 Recognition results(%)

方法	MSMT17+CUHK03->Market1501				Market1501+CUHK03->MSMT17				Market1501+CUHK03->Duke			
	mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10
M-SpCL	89.3	93.0	96.2	97.0	37.0	63.1	70.7	74.3	80.4	85.0	91.6	93.6
M-SpCL+EFDM	89.5	93.6	96.3	97.3	37.6	63.9	72.2	74.7	80.6	85.9	91.7	93.7
M-SpCL+MDIF	89.4	93.3	96.3	97.0	38.5	63.5	71.3	74.8	80.8	85.9	91.7	93.6
MFMCCL	90.0	93.8	96.7	97.5	38.9	64.9	73.0	76.3	81.4	86.9	92.2	93.7

表 2 不同 UDA-ReID 方法的识别结果(%)

Tab. 2 Recognition results of different UDA-ReID methods(%)

方法	MSMT17->Market1501				Market1501->MSMT17				Market1501->Duke			
	mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10
单源域												
MMT	86.7	90.7	95.0	96.7	34.5	57.2	66.3	70.6	77.6	81.4	87.8	91.1
SpCL	87.2	91.4	95.5	96.8	34.6	60.6	68.9	72.7	79.1	84.6	90.4	92.6
GLT	87.6	92.2	96.0	96.9	34.3	61.8	69.3	73.1	79.0	83.8	90.4	92.7
多源域												
M <sup>3</sup> L	66.8	78.5	85.9	88.6	13.8	27.7	37.1	41.9	43.3	53.9	64.7	68.9
META	67.3	80.4	87.8	91.4	25.2	48.1	57.7	62.6	59.3	69.2	78.5	82.7
UMSDA	89.6	92.2	95.6	96.6	23.6	43.5	52.2	56.8	76.7	81.9	87.8	90.6
MFMCCL	90.0	93.8	96.7	97.5	38.9	64.9	73.0	76.3	81.4	86.9	92.2	93.6

域, MSMT17为目标域时,与次最佳的多源域模型META相比,提出的方法在mAP和rank-1上分别高出13.7%和16.8%;以Market1501和CUHK03为源域, Duke为目标域时,与次最佳的多源域模型UMSDA相比,提出的方法在mAP和rank-1上分别高出4.7%和5.0%,表明本方法对于多源域任务适应性更强。总体来说,本文基于自步对比学习策略所构建的基于精确特征分布匹配和多域信息融合的多源域对比学习架构,在定性识别精度上表现出了较强的性能,远远超越了现有的方法。

## 5 结论

本文针对多源域自适应行人重识别任务,提出了一种基于精确特征分布匹配和多域信息融合的多源域对比学习框架。为了提高源域与目标域的交叉特征匹配,提出采用精确特征分布匹配方法,隐式地利用高阶统计数据进行交叉分布特征匹配,以实现更多样化的特征增强。并通过基于图卷积网络的多域信息融合方法进一步减小域距离。另外,将采用聚类算法获得的目标域数据的伪标签存入混合记忆存储器中,并计算多源域对比学习损失以优化模型。最后,在开放ReID数据集Market1501、MSMT17和Duke上进行了详细的消融实验和对比实验。实验结果表明,提出的MFML方法显著提高了多源域自适应行人重识别的性能。

## 参考文献

- [1] YE Mang, SHEN Jianbing, LIN Gaojie, et al. Deep learning for person re-identification: A survey and outlook[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(6): 2872-2893.
- [2] LIN X T, REN P Z, YEH C, et al. Unsupervised person re-identification: a systematic survey of challenges and solutions [EB/OL]. 2021: arXiv: 2109.06057 [cs.CV]. <https://arxiv.org/abs/2109.06057>.
- [3] WANG Mei, DENG Weihong. Deep visual domain adaptation: A survey[J]. *Neurocomputing*, 2018, 312: 135-153.
- [4] MEKHAZNI D, BHUIYAN A, EKLADIOUS G, et al. Unsupervised domain adaptation in the dissimilarity space for person re-identification [C]//*Computer Vision-ECCV 2020*, 2020: 159-174.
- [5] ZOU Y, YANG X D, YU Z D, et al. Joint disentangling and adaptation for cross-domain person re-identification [C]. *Proceedings of the European Conference on Computer Vision (ECCV)*. Glasgow, UK. Springer, 2020: 87-104.
- [6] CHEN Hao, WANG Yaohui, 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)*. Nashville, TN, USA. IEEE, 2021: 2004-2013.
- [7] GE Y, CHEN D, LI H. Mutual mean-teaching: pseudo label refinery for unsupervised domain adaptation on person re-identification [C]. *Proceedings of the International Conference on Learning Representation (ICLR)*. Addis Ababa, Ethiopia. ICLR: 2020: 1-15.
- [8] GE Y, ZHU F, CHEN D, et al. Self-paced contrastive learning with hybrid memory for domain adaptive object re-id [J]. *Proceedings of the Neural Information Processing Systems*, 2020: 11309-11321.
- [9] ZHENG Kecheng, LIU Wu, HE Lingxiao, et al. Group-aware label transfer for domain adaptive person re-identification [C]//2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville, TN, USA. IEEE, 2021: 5306-5315.
- [10] ZHAO Yuyang, ZHONG Zhun, YANG Fengxiang, 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)*. Nashville, TN, USA. IEEE, 2021: 6273-6282.
- [11] XU B Q, LIANG J, HE L X, et al. Mimic embedding via adaptive aggregation: learning generalizable person re-identification [EB/OL]. 2022: arXiv: 2112.08684v3 [cs.CV]. <https://arxiv.org/abs/2112.08684v3>.
- [12] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks [EB/OL]. 2017: arXiv: 1609.02907 [cs.LG]. <https://arxiv.org/abs/1609.02907>.
- [13] XU Minghao, WANG Hang, NI Bingbing. Graphical modeling for multi-source domain adaptation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022: 1.
- [14] BAI Zechen, WANG Zhigang, WANG Jian, et al. Unsupervised multi-source domain adaptation for person re-identification [C]//2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville, TN, USA. IEEE, 2021: 12909-12918.
- [15] KALISCHEK N, WEGNER J D, SCHINDLER K. In the light of feature distributions: Moment matching for Neural Style Transfer [C]//2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville, TN, USA. IEEE, 2021: 9377-9386.
- [16] ZHOU K Y, YANG Y X, QIAO Y, et al. Domain gener-

- alization with mixstyle [C]. International Conference on Learning Representations (ICLR). Venice, Italy. ICLR, 2021: 1-15.
- [17] ZHANG Yabin, LI Minghan, LI Ruihuang, et al. Exact feature distribution matching for arbitrary style transfer and domain generalization [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA. IEEE, 2022: 8025-8035.
- [18] COLTUC D, BOLON P, CHASSERY J M. Exact histogram specification [J]. IEEE Transactions on Image Processing, 2006, 15(5): 1143-1152.
- [19] ROLLAND J P, VO V, BLOSS B, et al. Fast algorithms for histogram matching: Application to texture synthesis [J]. Journal of Electronic Imaging, 2000, 9: 39-45.
- [20] HUANG Xun, BELONGIE S. Arbitrary style transfer in real-time with adaptive instance normalization [C]//2017 IEEE International Conference on Computer Vision. Venice, Italy. IEEE, 2017: 1510-1519.
- [21] YAN Yichao, ZHANG Qiang, NI Bingbing, et al. Learning context graph for person search [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA. IEEE, 2019: 2153-2162.
- [22] JIANG Bo, WANG Xixi, LUO Bin. pH-GCN: Person re-identification with part-based hierarchical graph convolutional network [EB/OL]. 2019: arXiv: 1907.08822 [cs.CV]. <https://arxiv.org/abs/1907.08822>.
- [23] 辛华磊, 丁英强, 高猛, 等. 基于多分区时空图卷积网络的骨骼动作识别 [J]. 信号处理, 2022, 38(2): 241-249.
- XIN Hualei, DING Yingqiang, GAO Meng, et al. Multi-partitioned spatiotemporal graph convolutional network for skeletal action recognition [J]. Journal of Signal Processing, 2022, 38(2): 241-249. (in Chinese)
- [24] WANG H, XU M H, NI B B, et al. Learning to combine: knowledge aggregation for multi-source domain adaptation [C]. Proceedings of European Conference on Computer Vision (ECCV). Glasgow, UK. Springer, 2020: 1-19.
- [25] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. IEEE, 2016: 770-778.
- [26] ZHENG Liang, SHEN Liyue, TIAN Lu, et al. Scalable person re-identification: A benchmark [C]//2015 IEEE International Conference on Computer Vision. Santiago, Chile. IEEE, 2015: 1116-1124.
- [27] WEI Longhui, ZHANG Shiliang, GAO Wen, et al. Person transfer GAN to bridge domain gap for person re-identification [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. IEEE, 2018: 79-88.
- [28] ZHENG Zhedong, ZHENG Liang, YANG Yi. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro [C]//2017 IEEE International Conference on Computer Vision. Venice, Italy. IEEE, 2017: 3774-3782.
- [29] ZHONG Zhun, ZHENG Liang, CAO Donglin, et al. re-ranking person re-identification with k-reciprocal encoding [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. IEEE, 2017: 3652-3661.

#### 作者简介



杨真真 女, 1984年生, 山东临沂人。南京邮电大学副教授, 博士。研究方向为深度学习、计算机视觉。  
E-mail: yangzz@njupt.edu.cn



邵静 女, 1995年生, 江苏徐州人。南京邮电大学研究生。研究方向为深度学习、计算机视觉。  
E-mail: 1481726137@qq.com



杨永鹏 男, 1984年生, 山东高密人。南京信息职业技术学院讲师, 南京邮电大学博士研究生。研究方向为深度学习、计算机视觉。  
E-mail: yangyp@njcit.cn



吴心怡 女, 1999年生, 江苏无锡人。南京邮电大学研究生。研究方向为深度学习、计算机视觉。  
E-mail: 1323746365@qq.com