

# 基于局部实例匹配无监督式学习的行人重识别

吴海丽<sup>1</sup>, 张月琴<sup>2</sup>, 庞俊奇<sup>3</sup>

(1. 太原学院计算机科学与技术系, 太原 030001; 2. 太原理工大学信息与计算机学院, 太原 030002; 3. 中北大学仪器与电子学院, 太原 030051)

**摘要:** 无监督域适应(Unsupervised domain adaptation, UDA)方法通过全局特征分布匹配实现源域到目标域的知识迁移,但忽略了细粒度的局部实例信息。本文提出了一种基于双层域自适应(Two-tiered domain adaptation, TTDA)的无监督行人重识别方法,使用全尺寸网络(Omni-scale network, OSNet)作为骨干网络,在端到端深度学习框架中联合执行源域和目标域之间的全局特征分布匹配和局部实例匹配,从源域和目标域之间不同行人ID的关联中挖掘可迁移的有用知识,并通过知识选择机制提高了跨域适应性。在多个大型公开数据集上的实验结果表明,与其他先进方法相比,所提方法在源域到目标域的无监督行人重识别的平均精度均值(mean Average precision, mAP)和top- $k$ 命中率均取得显著提升。

**关键词:** 行人重识别;深度学习;无监督域;全局特征;局部实例匹配

中图分类号: TP391.41

文献标志码: A

## Unsupervised Learning Pedestrian Re-identification Based on Localized Instance Matching

WU Haili<sup>1</sup>, ZHANG Yueqin<sup>2</sup>, PANG Junqi<sup>3</sup>

(1. Department of Computer Science and Technology, Taiyuan University, Taiyuan 030001, China; 2. School of Information and Computer Science, Taiyuan University of Technology, Taiyuan 030002, China; 3. College of Instruments and Electronics, North University of China, Taiyuan 030051, China)

**Abstract:** Unsupervised domain adaptation (UDA) methods leverage global feature distribution matching to realize knowledge transfer from source domain to target domain, while ignoring fine-grained local instance information. An unsupervised person re-identification method based on two-tiered domain adaptation (TTDA) is proposed, in which the omni-scale network (OSNet) is selected as the backbone network, and global feature distribution matching and localized instance matching between source and target domains are performed jointly in an end-to-end deep learning framework. And in order to effectively mine transferable useful knowledge from associations of different pedestrian IDs between source and target domains, the cross-domain adaptability is improved with a knowledge selection mechanism. Experimental results on multiple large-scale public datasets show that compared with other state-of-the-art methods, the proposed method achieves significant improvements in terms of mean average precision (mAP) and top- $k$  hit rate for unsupervised cross-domain person re-identification tasks.

**Key words:** person re-identification; deep learning; unsupervised domain; global feature; localized instance matching

## 引 言

行人重识别旨在不同相机捕捉到的图库中检索特定行人的图像,在公共安全、智能交通和智慧城市等领域有着广阔应用前景。由于行人图像易受到照明变化、各种身体姿态和不同相机角度的影响,行人重识别任务极具挑战性<sup>[1]</sup>。随着深度学习技术的发展,行人重识别的方法已经从传统特征工程方法演化为利用深度神经网络,特别是卷积神经网络(Convolutional neural network, CNN)的端到端特征表示学习<sup>[2]</sup>。当前,监督式CNN方法已经比较成熟<sup>[3-5]</sup>,但应用到未知数据集会出现严重的性能下降。在现实监控场景下,针对特定场景采集并标注大量训练数据并不总是可行。因此,无监督域适应(Unsupervised domain adaptation, UDA)方法成为了研究热点<sup>[6]</sup>。

UDA方法主要分为基于图像合成,基于特征分布和基于聚类的适应方法。基于图像合成的UDA方法通常基于对抗生成网络(Generative adversarial networks, GAN),将源域的标注图像转换为无标注的目标域图像风格,并尝试保留行人ID信息,由此使用源ID标签在合成图库上进行监督式学习。SPGAN保留了图像转换前后的自相似性,以及转换图像与目标图像之间的域相异性<sup>[7]</sup>。PTGAN在风格迁移过程中通过语义分割施加人体区域的一致性约束<sup>[8]</sup>。CRGAN通过将源域的行人ID转移到多种不同的目标域,实现了在未注释目标域中的监督式学习<sup>[9]</sup>。但图像合成方法仅基于全局图像(像素)分布对齐,未考虑细粒度实例级知识。基于特征的UDA方法通常基于特征分布匹配,假定模型判别能力与全局特征分布统计信息相关。Yan等<sup>[10]</sup>提出了加权最大均值差异(Maximum mean discrepancy, MMD)模型,通过引入类特定的附加权重来利用源域和目标域上的类先验概率。Li等<sup>[11]</sup>引入流形正则化理念,通过源域和目标域分布重要性加权和邻接结构保留,实现流形对齐和平衡分布自适应。但特征自适应方法以全局特征表示分布对齐为基础,同样忽略了局部实例匹配。近期,一些基于聚类的UDA方法取得了较好成绩。PUL通过在目标域上迭代运行聚类算法获得伪标签,并利用生成的标签进行模型微调<sup>[12]</sup>。SSG分别向全局和局部特征分配伪标签<sup>[13]</sup>。PAST渐进地在训练集中引入可靠样本<sup>[14]</sup>。ACT利用非对称的协同教学框架,降低聚类算法生成的噪声标签<sup>[15]</sup>。MMT通过同步平均教学,利用软标签对伪标签进行在线优化,以消除聚类过程中的伪标签噪声问题<sup>[16]</sup>。但跨域行人重识别属于高度无约束场景,源域和目标域的行人ID缺少直接关联。尽管上述基于聚类的方法在跨域自适应中考虑到局部实例知识,但未实现同时从全局分布和局部实例两个层面发掘源域到目标域的可迁移知识,限制了从源域标注训练数据到目标域未标注训练数据的特征匹配精准度的进一步提高。

为提高跨域跨类别行人重识别准确度,本文提出了以全尺寸网络(Omni-scale network, OSNet)为骨干网络的双层域适应架构(Two-tiered domain adaption, TTDA),通过源域和目标域之间的全局特征分布匹配和局部实例匹配,挖掘可迁移的有用信息,并提出知识选择机制,提高跨域适应性,以充分挖掘源ID与目标ID之间的逐对底层关联。所提方法创新点包括:(1)使用基于OSNet的双层架构,在训练过程中渐进学习从每个小批中学习到的源知识,解决了灾难性遗忘问题。(2)在知识迁移中结合了全局特征分布和局部实例知识,并通过知识选择机制改善了检测模型在未知数据集上的泛化能力。

## 1 基于OSNet和无监督双层域自适应的行人重识别框架

### 1.1 模型架构

在无监督行人重识别中,源域和目标域分别表示已标注和无标注的行人重识别数据集。给定包含 $K^s$ 个行人包围框图像 $I_i^s$ 的源域数据集 $D^s = \{I_i^s, y_i^s\}_{i=1}^{K^s}$ ,每个包围框注释了ID标签 $y_i^s \in Y =$

$\{1, \dots, K_{\text{id}}^s\}$ , 即源域中共计  $K_{\text{id}}^s$  个不同的行人 ID。令  $D^t = \{I_i^t\}_{i=1}^{K^t}$  表示从目标域随机采样的  $K^t$  个未注释训练数据。利用  $D^t$  进行模型域自适应, 通过从源域获得的判别信息的知识迁移, 学习对未标注目标域的行人 ID 进行分类的最优特征。

所提基于 TTDA 的无监督重识别模型架构如图 1 所示。使用两个相同结构的 OSNet 分别作为目标识别网络和记忆网络, 通过全局分布匹配和局部实例匹配模块, 在源域中学习可泛化到目标域的目标识别模型, 通过构建知识记忆网络建立可靠的跨域跨类关联。利用特征归一化加速模型训练, 通过知识选择机制改善域自适应的鲁棒性。

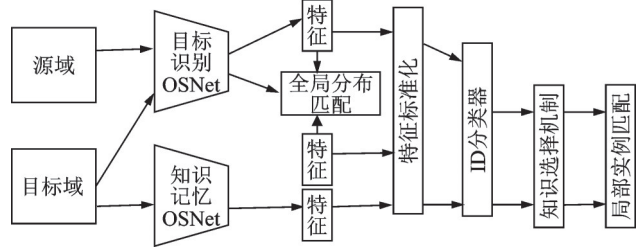


图 1 提出的域自适应双层无监督模型架构

Fig.1 The proposed two-tiered unsupervised domain adaptation model architecture

## 1.2 骨干网络

使用轻量级的 OSNet 作为骨干网络, 网络结构如表 1 所示, 其中逐层堆叠了多个全尺寸瓶颈 block<sup>[17]</sup>。OSNet 旨在学习多尺寸特征, 利用逐深度可分离卷积降低计算量, 并通过多尺寸残差块学习判别特征。

### 1.2.1 残差 Block

OSNet 以残差瓶颈为基础构建, 但配置了逐深度卷积层。深度卷积结构如图 2(a) 所示, 瓶颈 block 结构如图 2(b) 所示, 图中 AG 为聚合门, 实现动态尺度融合。

为降低计算量, 采用深度可分离卷积。利用核  $w \in \mathbb{R}^{k \times k \times c \times c'}$  将卷积层  $\text{ReLU}(w * x)$  分为 2 个单独网络层  $\text{ReLU}((u \circ v) * x)$ ,  $u \in \mathbb{R}^{k \times k \times 1 \times c'}$  为逐深度核,  $v \in \mathbb{R}^{1 \times 1 \times c \times c'}$  为逐点核,  $*$  为卷积,  $k$  为核大小,  $c$  为输入通道宽度,  $c'$  为输出通道宽度。给定输入张量  $x \in \mathbb{R}^{h \times w \times c}$ ,  $h$  为高度,  $w$  为宽度, 则计算成本从  $h \cdot w \cdot k^2 \cdot c \cdot c'$  降低至  $h \cdot w \cdot (k^2 + c) \cdot c'$ 。先进行逐点卷积 ( $1 \times 1$  普通卷积), 其后进行逐深度卷积, 有效改善对多尺寸特征的学习, 表示为  $\text{ReLU}((u \circ v) * x)$ , 将该网络层称为大 Lite  $3 \times 3$ 。

给定输入  $x$ , 残差模块旨在通过映射函数  $F$  学习残差  $\tilde{x}$ , 即

$$y = x + \tilde{x} \quad (1)$$

式中  $\tilde{x} = F(x)$ ,  $F$  表示学习单尺寸特征的 Lite  $3 \times 3$  卷积层。  $1 \times 1$  卷积层用于控制特征通道, 并不参与空间信息聚合。

为实现多尺寸特征学习, 对残差函数  $F$  进行扩展。对于  $F^t$ ,  $t > 1$  为特征尺度, 堆栈  $t$  个 Lite  $3 \times 3$  层, 由此将感受野扩展为  $(2t + 1) \times (2t + 1)$ 。则要学习的残差  $\tilde{x}$  为

$$\tilde{x} = \sum_{t=1}^T F^t(x) \quad (2)$$

考虑计算成本, 设  $T = 4$ , 即最大感受野为  $9 \times 9$ 。

表 1 OSNet 网络结构

Table 1 Network structure of OSNet

阶段	结构	输出
Conv1	$7 \times 7$ 卷积	$128 \times 64, 64$
	$3 \times 3$ 最大池化	
Conv2	bottleneck $\times 2$	$64 \times 32, 256$
	transition	
	$1 \times 1$ 卷积	
Conv 3	$2 \times 2$ 平均池化	$64 \times 32, 256$
	bottleneck $\times 2$	
	transition	
Conv 4	$1 \times 1$ 卷积	$32 \times 16, 384$
	$2 \times 2$ 平均池化	
	bottleneck $\times 2$	
Conv5	$1 \times 1$ 卷积	$32 \times 16, 384$
	$2 \times 2$ 平均池化	
	bottleneck $\times 2$	
GAP	$1 \times 1$ Conv	$16 \times 8, 512$
	全局平均池化	
FC	全连接	$1 \times 1, 512$

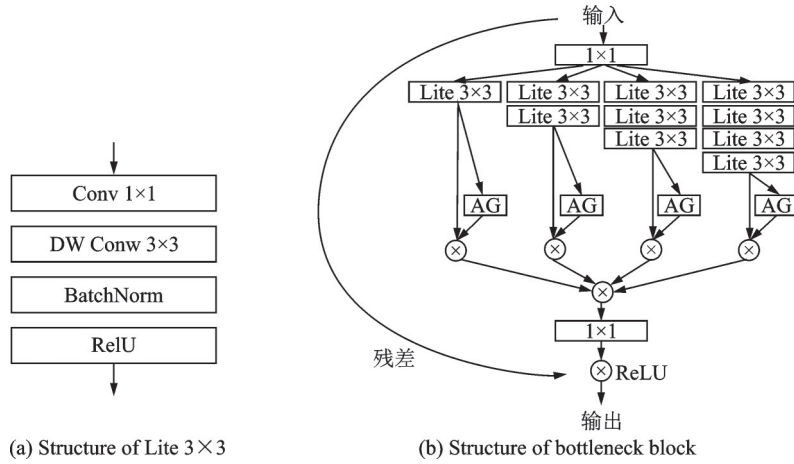


图2 残差瓶颈结构

Fig.2 Residual bottleneck structure

1.2.2 动态统一聚合门

每个卷积流提供了一个特定尺寸的特征。为学习多尺寸特征,对不同卷积流的输出进行动态合并,根据输入图像向不同尺寸分配不同权重。令  $x^t$  表示  $F^t(x)$ ,多尺寸残差  $\tilde{x}$  可表示为

$$\tilde{x} = \sum_{t=1}^T G(x^t) \odot x^t \tag{3}$$

式中:  $x^t \triangleq F^t(x)$ ;  $G(x^t)$  为数据条件向量;  $\odot$  为 Hadamard 积;  $G$  表示包含一个非参数全局池化层和一个多层感知器(Multi-layer perceptron, MLP)的迷你网络。

类似于 CNN 中卷积过滤器的参数共享, OSNet 中同一个残差块内的所有特征流共享 AG, 优点在于: (1) 参数数量与流数量  $T$  无关, 由此提高模型稳定性; (2) AG 的统一有利于执行梯度反向传播。令  $L$  为可微损失函数, 则梯度定义为

$$\frac{\partial L}{\partial G} = \frac{\partial L}{\partial \tilde{x}} \frac{\partial \tilde{x}}{\partial G} = \frac{\partial}{\partial} \left( \sum_{t=1}^T x^t \right) \tag{4}$$

式中  $\frac{\partial L}{\partial \tilde{x}} \frac{\partial \tilde{x}}{\partial G}$  意味着通过 AG 的共享, 将所有流的监督信号聚集在一起以引导  $G$  的学习。

OSNet 使用了逐通道向量门限, 对于第  $t$  个流, AG 的输出  $G(x^t)$  为向量而非标量, 由此实现了更好的特征通道调整。其次, 基于输入数据动态计算权重, 通过自适应特征和尺寸融合, 提高了特征提取的有效性和可扩展性。

1.2.3 目标网络损失函数

给定注释源训练数据  $D^s$ , 利用判别损失函数  $L_{\text{target}}$  进行模型训练, 即

$$L_{\text{target}} = L_{\text{ce}} + \lambda_{\text{tri}} L_{\text{tri}} \tag{5}$$

式中:  $L_{\text{ce}}$  为 softmax 交叉熵损失;  $L_{\text{tri}}$  为三元组损失;  $\lambda_{\text{tri}}$  为权重参数。根据经验, 设  $\lambda_{\text{tri}} = 0.3$ 。

1.3 全局特征分布匹配

全局分布匹配(Global distribution matching, GDM)旨在调整源域和目标域之间的整体统计信息。考虑到重识别图像数据有着复杂的视觉混淆性和多样性, 对视觉数据的统计信息使用非参数化特征。具体来说, 为进行分布调整, 采用最大均值差异(Maximum mean discrepancy, MMD)<sup>[10]</sup>测量源域和目

标域之间的特征相异性,即

$$L_{\text{MMD}^2} = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \phi(\mathbf{f}_{s,i}) - \frac{1}{n_t} \sum_{i=1}^{n_t} \phi(\mathbf{f}_{t,i}) \right\|_{\text{H}}^2 \quad (6)$$

式中: $\mathbf{f}_s \in \mathbf{R}^{n_s \times d}$ 和 $\mathbf{f}_t \in \mathbf{R}^{n_t \times d}$ 分别为每个小批中 $n_s$ 个源域图像和 $n_t$ 个目标域图像的特征向量; $d$ 为特征维度。利用映射函数 $\phi(\cdot)$ 将特征样本映射到再生核 Hilbert 空间 (Reproducing kernel hilbert space, RKHS)<sup>[18]</sup>。利用核映射函数 $k(\cdot)$ ,将 GDM 损失 $L_2$ 计算为

$$L_{\text{GDM}} = \frac{1}{n_s^2} \sum_{i=1}^{n_s} \sum_{i'=1}^{n_s} k(\mathbf{f}_{s,i}, \mathbf{f}_{s,i'}) + \frac{1}{n_t^2} \sum_{j=1}^{n_t} \sum_{j'=1}^{n_t} k(\mathbf{f}_{t,j}, \mathbf{f}_{t,j'}) - \frac{2}{n_s n_t} \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} k(\mathbf{f}_{s,i}, \mathbf{f}_{t,j}) \quad (7)$$

#### 1.4 局部实例匹配

通过局部实例匹配(Local instance matching, LIM),实施实例级的细粒度判别学习,进一步改善跨域行人重识别的自适应性。LIM旨在训练过程中逐步发现与个体目标样本相对应的可迁移源信息。通过目标样本与视觉相似的源数据之间的关联,指导跨域知识迁移。通过在不同域的不同身份ID中发掘出LIM关联,将目标行人图像分入源身份ID。具体来说,给定未注释目标行人图像样本 $I^t$ ,预测其在源域的分类标签空间中的类概率向量为

$$p(I^t) = \{p(1|I^t), p(2|I^t), \dots, p(K_{\text{id}}^s|I^t)\} \quad (8)$$

该向量表示目标行人图像与所有源类别的视觉相似程度,其中编码了UDA中所需的跨域迁移知识。

##### 1.4.1 源域知识发掘

将使用小批训练样本的深度学习考虑为随机学习过程,每个小批有着不同的特征分布。深度模型需要在整个训练过程中不停重复适应新的分布,提高了跨域跨类自适应难度。为此,所提框架强制模型始终输出固定分布的特征表征。具体来说,对CNN输出的特征表示进行标准化。在来自源域和目标域的每批特征向量上执行每维度归一化,将标准化特征 $\hat{\mathbf{f}}$ 计算为

$$\hat{\mathbf{f}} = \frac{\mathbf{f} - E[\mathbf{f}]}{\sqrt{V[\mathbf{f}] + \epsilon}} \quad (9)$$

式中: $E[\mathbf{f}]$ 和 $V[\mathbf{f}]$ 分别表示每批特征值的每维度期望值和方差; $\epsilon$ 表示一个非常小的正数。

##### 1.4.2 知识记忆网络

考虑跨域行人重识别任务特点,为解决深度学习模型的灾难性遗忘问题<sup>[19]</sup>,进一步改善知识质量,所提框架在训练过程中渐进式地记忆从每个小批中学习到的源信息。

采用与目标模型相同架构建立一个知识记忆网络 $\theta^{\text{mem}}$ ,并用其获得行人身份ID的后验概率知识。在每次迭代 $\tau$ ,通过指数滑动平均(Exponential moving average, EMA)让记忆网络 $\theta^{\text{mem}}$ 随目标识别网络 $\theta^{\text{tar}}$ 进行更新,即

$$\theta_{\tau}^{\text{mem}} = \alpha \theta_{\tau-1}^{\text{mem}} + (1 - \alpha) \theta_{\tau}^{\text{tar}} \quad (10)$$

式中 $\alpha$ 为平滑系数。由此, $\theta^{\text{mem}}$ 中吸收并记忆了从每个小批推导出的判别信息,从而实现更强的知识提取性能。在小批训练中,利用 $\theta_{\tau}^{\text{mem}}$ 得到每个未注释目标样本在源域类空间中的后验概率向量。

##### 1.4.3 知识选择机制

源知识迁移旨在提高目标识别模型的泛化性。为获取源域和目标域之间的底层可迁移知识,设计

知识选择机制,以增强记忆网络中与目标域样本相关的知识,提高在个体样本上的自适应性和可靠性。

在无监督跨域重识别任务中,由于源域与目标域之间有着完全不同的行人ID,并非所有目标行人图像均能与某些源行人ID高度相关。将源域知识表示为向量概率,利用最大似然(Maximum likelihood, ML)估计知识的可迁移性和可靠性为

$$\text{ML}(\mathbf{I}^t) = \max(\{p(1|\mathbf{I}^t), p(2|\mathbf{I}^t), \dots, p(K_{\text{id}}^s|\mathbf{I}^t)\}) \quad (11)$$

然后,采用知识选择阈值策略,选择 $\text{ML}(\mathbf{I}^t)$ 超过预定义阈值 $u$ 的目标样本 $\tilde{\mathbf{I}}^t$ 。跨域背景下, $\text{ML}(\mathbf{I}^t)$ 值通常不会很高,选择适中的阈值有助于在源样本与目标样本之间建立鲁棒关联。根据实证经验,设 $u = 0.3$ 。

最后,将选定知识从记忆网络迁移到目标模型,完成源域知识到目标域的跨域自适应。为此,利用KL散度(Kullback-Leibler divergence),记忆模型与目标模型对选定目标样本的预测结果 $\tilde{\mathbf{I}}^t$ 进行配准,得到局部实例匹配损失 $L_{\text{lim}}$ 为

$$L_{\text{lim}} = \sum_{j=1}^{K_{\text{id}}^s} p(j|\tilde{\mathbf{I}}^t, \theta^{\text{mem}}) \lg \frac{p(j|\tilde{\mathbf{I}}^t, \theta^{\text{mem}})}{p(j|\tilde{\mathbf{I}}^t, \theta^{\text{tar}})} \quad (12)$$

## 1.5 整体损失函数

所提跨域重识别模型的整体损失函数定义为目标模型的重识别损失 $L_{\text{target}}$ ,全局分布匹配损失 $L_{\text{GDM}}$ 和局部实例匹配损失 $L_{\text{lim}}$ 的加和为

$$L_{\text{total}} = L_{\text{target}} + L_{\text{GDM}} + L_{\text{lim}} \quad (13)$$

利用随机梯度下降,基于整体损失函数 $L_{\text{total}}$ 对整个框架进行端到端训练。

## 2 实验与分析

### 2.1 数据集

实验中使用了Market-1501、DukeMTMC-reID和MSMT17三个主流公开数据集,评估所提TTDA框架的性能。

Market-1501(Market)<sup>[20]</sup>是使用5个高分辨率相机和1个低分辨率相机在清华大学的超市前采集行人图像得到的大型行人重识别数据集,并使用行人探测器DPM自动检测行人包围框。该数据集包含1 501个不同行人ID的共计32 668张图像,图像分辨率为128 像素×64 像素。

DukeMTMC-reID(Duke)<sup>[21]</sup>是使用8个静态高清相机在美国杜克大学采集到的行人重识别数据集,包含1 404个不同行人ID的36 411张图像。

MSMT17(MSMT)<sup>[8]</sup>是使用15个相机在大学校园采集的大规模行人重识别数据集,包含4 101不同行人ID的共计126 441张图像。该数据集包含大量不同场景和显著照明变化,因此识别难度较大。表2给出了实验数据集的统计详情,图3给出了实验数据集的图片示例。

表2 行人重识别数据集

Table 2 Person re-identification dataset

数据集	相机	行人ID数	训练ID数	测试ID数	总图像数
Market	6	1 501	751	750	32 688
Duke	8	1 404	702	702	36 411
MSMT	15	4 101	1 041	3 060	126 411

2.2 模型实施和评估指标

实验平台操作系统为 Ubuntu 16.04 LTS,使用 Intel Xeon E5-2640-v4 2.4 处理器,2×16 GB RAM,GeForce RTX 2080 Ti GPU,显存为 11 GB。使用 PyTorch 实施所有实验。通过随机梯度下降(Stochastic gradient descent, SGD)训练目标识别网络,动量设为 0.9。权重衰减为 0.000 5,批大小为 64(32 个源样本与 32 个目标域样本),代数为 60。所有输入图像缩放为 256 像素×128 像素。在训练中针对目标识别网络和知识记忆网络分别应用包括随机裁剪、随机翻转和颜色抖动的数据增广处理。测试中,使用欧氏距离作为行人重识别的匹配指标。

使用行人重识别中主流的累积匹配特征(Cumulative matching characteristics, CMC)曲线和平均精度均值(Mean average precision, mAP)作为性能评估指标<sup>[22]</sup>。基于查询目标与要查询的目标图像之间的相似性排序,计算出在图库中检索到前  $k$  个图像(top- $k$ )包含正确查询结果的概率  $Acc_k$ 。若检索出的前  $k$  个图库样本中包含查询目标 ID,则  $Acc_k = 1$ ;否则,  $Acc_k = 0$ 。通过每张查询图像的  $Acc_k$  的加和除以查询图像总数,得到 CMC 曲线,表示为 Rank- $k$  命中率。例如 Rank-1 准确率表示匹配列表中第 1 个目标的正确匹配概率。mAP 同时考虑查询的平均精度和精度-召回率曲线,反映查询结果中包含正确 ID 的所有图像在检索队列中的靠前程度,以更全面地分析重识别算法的性能。

下面给出的实验结果均是一次实验的结果。本文实验中采用的 3 个数据集均是大规模行人重识别公开数据集,训练和测试数据样本量足够大,因此一次实验的结果已经能够充分说明模型性能<sup>[23]</sup>。

2.3 消融实验

首先,通过消融实验,分析所提 TTDA 框架中每个组件的作用,并比较不同设定下的性能变化情况。

2.3.1 骨干网络

表 3 给出了所提框架在使用不同的骨干网络<sup>[24]</sup>时,以 Duke 数据集作为源数据集,将 Market 数据集作为目标数据集时的性能结果。从表中可发现,Alex-Net 和 VGG-19 作为骨干网络时的性能较差,说明这两种网络架构不能很好地提取出不同行人之间的细微外观差异。ResNeXt 结合

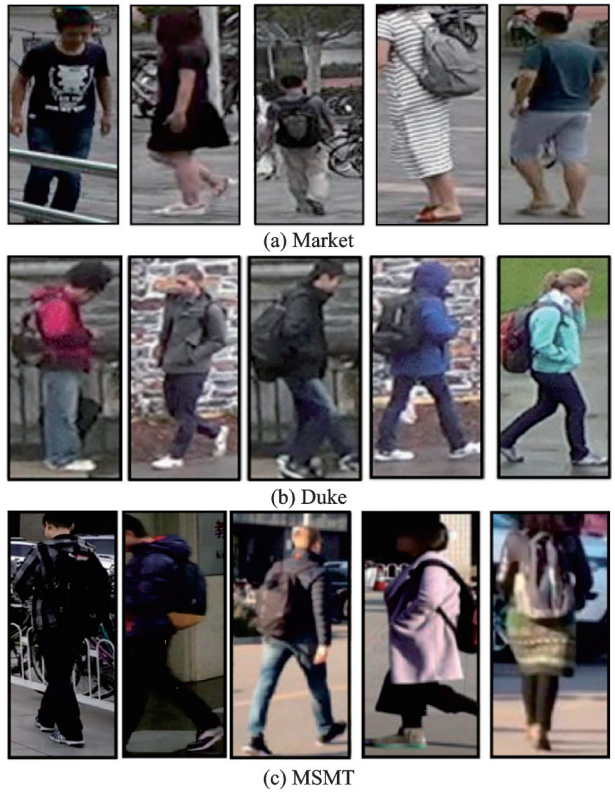


图 3 实验数据集图像示例

Fig.3 Image examples of the experimental datasets

表 3 不同骨干网络性能比较

Table 3 Performance comparison of different backbone networks %

网络	Duke → Market			
	Rank-1	Rank-5	Rank-10	mAP
Alex-Net	45.8	50.5	58.6	20.5
VGG-19	50.2	58.1	66.3	26.7
ResNet-50	71.5	76.8	80.5	33.6
ResNeXt	76.2	83.5	88.6	45.5
OSNet	88.7	97.6	98.4	76.2

了inception与resnet的优点,并利用分组卷积降低了错误率。但ResNext中所有卷积流使用相同深度。与之相比,OSNet融合不同尺寸的特征,并通过逐通道门限聚合和共享机制增强了对判别特征的提取能力,取得了最优性能。

### 2.3.2 域自适应分析

所提TTDA框架中结合了全局特征分布匹配GDM和局部实例匹配LIM模块。图4给出了使用OSNet作为骨干网络,从源数据集MSMT到目标数据集Market的无监督行人重识别中,使用不同框架配置时的top- $k$ 和mAP性能对比。其中,模型1表示仅使用单个OSNet做无监督分类,模型2表示仅使用TTDA框架中的GDM模块,模型3表示仅使用TTDA中的LIM模块,模型4表示OSNet+TTDA,即所提完整方法。从图4中可发现,仅使用原始OSNet时,无法适应不同数据集之间的跨域差异,模型性能较差。仅使用GDM或LIM模块时,性能较原始模型显著提升,证明不同数据集的不同行人ID中提取出的特征存在底层关联。而同时使用GDM和LIM模块时,性能进一步提升,证明将全局特征分布与局部实例信息相结合,能够有效改善跨域跨类行人重识别性能。此外,对于所提完整模型,同样将Market数据集作为目标数据集,以MSMT为源数据集时的性能优于以Duke为源数据集,证明所提方法能够充分提取源数据集中有用的可迁移知识。这是因为MSMT数据集相比Duke数据集有着更大规模的样本和更多场景,源数据集中的训练样本的多样化有助于提升所提方法在目标数据集中的识别性能。

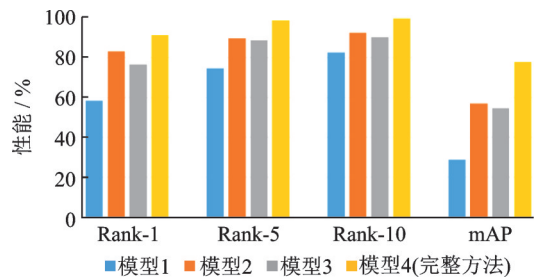


图4 模块作用分析

Fig.4 Function analysis of different modules

图5给出了在仅使用GDM模块和使用完整模型时,从MSMT到Market的跨域无监督行人重识别的可视化结果。图中,左侧为查询图像,右侧分别为不使用LIM模块和完整模型的top-1至top-10的检索结果。蓝色框表示查询图像与检索图像为相同行人ID,红色框表示查询图像与检索图像为不同行人ID。从图5中可发现,结合实例信息和知识选择机制显著提高了所提模型的top- $k$ 命中率。

### 2.3.3 知识选择机制

本文框架中,使用阈值 $u$ 确定源域ID空间中的最大似然(式(11)),以控制从源域到目标域的知识迁移程度。表4给出了以Duke为源域,Market为目标域时,不同 $u$ 数值对所提框架的top- $k$ 命中率和mAP的性能影响。其中, $u=0$ 表示对所有源知识进行相等程度的迁移,即不使用知识选择机制。 $u=1$ 则表示完全不考虑局部实例知识的适应性。从表4中可发现, $u$ 取值范围在 $[0.1, 0.4]$ 之间时所提模型能够获得最优性能,证明在跨域自适应过程中,某些无用的源知识可能会造成模型的泛化性能下降,所提知识选择机制有助于充分利用适合目标域的源域知识。

## 2.4 与其他方法的比较

表5给出了分别使用Duke数据集为源数据集,Market数据集为目标数据集(Duke→Market),以及使用Market数据集为源数据集,Duke数据集为目标数据集(Market→Duke)时,所提TTDA方法与其他先进方法的性能比较。其中,SPGAN、CR-GAN和PTGAN为基于图像合成的方法,PUL、SSG和MMT为基于特征自适应和聚类的方法。从表5中可发现,基于图像合成的方法性能较差,证明通过图像风格迁移进行跨域自适应并不能取得理想结果。与当前性能领先的MMT方法相比,所提TTDA方法在Market→Duke和Duke→Market的mAP结果分别提升了7.7%和1.6%,证明了所提框架结合全局分布匹配和局部实例匹配进行选择知识迁移的有效性。





图5 在Market数据集上的可视化结果

Fig.5 Visualization results on the Market dataset

表4 阈值  $u$  对 rank-1 和 mAP 性能的影响

Table 4 Effect of threshold  $u$  on rank-1 and mAP performance

$u$ 值	Rank-1	Rank-5	Rank-10	mAP
0	83.5	95.9	96.9	71.0
0.1	87.5	96.3	97.8	74.9
0.2	87.9	96.6	98.1	75.3
0.4	87.6	96.3	98.0	75.1
0.6	83.6	96.1	97.2	73.2
0.8	81.9	95.3	96.8	71.2
1.0	79.8	94.8	96.1	68.9

表6给出了使用MSMT数据集作为源数据集,分别应用到Market数据集和Duke数据集进行无监督行人重识别时,表5中表现较好的几个模型的识别结果。与其他两个数据集相比,MSMT数据集样本数量更大,场景更具多样化。通过比较使用不同源域数据集,在目标数据集进行行人重识别的性能,能够更好地检验模型提取源域有用知识和避免灾难性遗忘的能力。从结果中可发现,CR-GAN使用MSMT作为源域时,在Market和Duke数据集上的性能均出现了下降,证明基于重建的方法在多样化场景中提取有用特征的能力会下降。SSG方法和MMT方法利用伪标签分配和噪声移除技术提高了不同域知识提取的适应性,性能比之前稍有提升。所提TTDA方法通过结合GDM和LIM模块提高跨域自适应性,并通过知识选择机制提高了对源域有用知识的提取,在使用更大的数据集作为源域时性能得到了明显改善。

表5 Duke作为源数据集时与其他先进方法的性能比较

Table 5 Performance comparison with other state-of-the-art methods taking Duke as the source dataset

方法	Duke → Market				Market → Duke			
	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10	mAP
PTGAN <sup>[8]</sup>	38.6	57.5	66.1	22.6	27.3	45.2	50.7	19.5
SPGAN <sup>[7]</sup>	57.8	74.6	82.3	26.7	47.2	62.5	68.0	26.5
CR-GAN <sup>[9]</sup>	65.5	79.8	84.5	33.2	55.9	71.0	75.8	34.9
PUL <sup>[12]</sup>	44.6	61.2	67.8	20.5	30.0	43.4	48.5	16.4
SSG <sup>[13]</sup>	80.0	90.0	92.4	58.3	73.0	80.6	83.2	53.4
MMT <sup>[16]</sup>	88.4	96.2	97.8	74.6	75.1	87.3	91.2	61.0
TTDA	88.7	97.6	98.4	76.2	81.6	91.3	93.6	68.7

表6 MSMT作为源数据集时与其他先进方法的性能比较

Table 6 Performance comparison with other state-of-the-art methods taking MSMT as the source dataset

方法	MSMT → Market				MSMT → Duke			
	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10	mAP
CR-GAN <sup>[9]</sup>	63.8	77.4	82.6	31.5	52.1	68.7	74.7	29.8
SSG <sup>[13]</sup>	81.4	90.5	92.6	59.1	74.3	81.8	83.7	54.0
MMT <sup>[16]</sup>	88.5	96.3	98.0	75.2	75.3	87.5	91.4	61.6
TTDA	91.7	98.6	99.7	88.3	84.2	93.8	95.7	72.7

### 3 结束语

本文提出了以OSNet为骨干网络,基于TTDA框架的无监督行人重识别方法。实验结果表明,OSNet能够动态聚合不同尺寸的特征,在降低计算量的同时有效提高了特征自适应性和可扩展性。所提TTDA方案同时利用全局特征分布和局部实例信息,并通过知识选择机制有效提高从源域到目标域的知识迁移质量,在大型公开数据集上的性能优于其他先进的无监督行人重识别方法。现实智能监控系统中的应用中难免会存在一些未知场景,可能会影响到所提方法的适应性。未来,可将所提方法与近期提出的“自学习”等技术相结合,进一步提高所提方法对新场景的适应能力,并结合背景减除等方法进一步提高行人重识别的匹配性能。

#### 参考文献:

- [1] 夏道勋, 郭方, 刘浩杰, 等. 开放式行人再识别研究进展综述[J]. 数据采集与处理, 2021, 36(3): 449-467.  
XIA Daoxun, GUO Fang, LIU Haojie, et al. Review on research progress of open-world person re-identification[J]. Journal of Data Acquisition and Processing, 2021, 36(3): 449-467.
- [2] YE M, SHEN J, LIN G, et al. Deep learning for person re-identification: A survey and outlook[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(6): 2872-2893.
- [3] 杨婉香, 严严, 陈思, 等. 基于多尺度生成对抗网络的遮挡行人重识别方法[J]. 软件学报, 2020, 31(7): 1943-1958.  
YANG Wanxiang, YAN Yan, CHEN Si, et al. Multi-scale generative adversarial network for person re-identification under occlusion[J]. Journal of Software, 2020, 31(7): 1943-1958.
- [4] LIN Y, ZHENG L, ZHENG Z, et al. Improving person re-identification by attribute and identity learning[J]. Pattern

- Recognition, 2019, 95(1): 151-161.
- [5] 徐胜军, 刘求缘, 史亚, 等. 基于多样化局部注意力网络的行人重识别[J]. 电子与信息学报, 2022, 44(1): 211-220.  
XU Shengjun, LIU Qiuyuan, SHI Ya, et al. Person re-identification based on diversified local attention network[J]. Journal of Electronics & Information Technology, 2022, 44(1): 211-220.
- [6] MADADI Y, SEYDI V, NASROLLAHI K, et al. Deep visual unsupervised domain adaptation for classification tasks: A survey[J]. IET Image Processing, 2020, 14(14): 3283-3299.
- [7] DENG W, ZHENG L, YE Q, et al. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2018. Salt Lake City, UT, USA: IEEE Press, 2018: 994-1003.
- [8] WEI L, ZHANG S, GAO W, et al. Person transfer GAN to bridge domain gap for person re-identification[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE Press, 2018: 79-88.
- [9] CHEN Y, ZHU X, GONG S. Instance-guided context rendering for cross-domain person re-identification[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision 2019. Seoul, Korea: IEEE Press, 2019: 232-242.
- [10] YAN H, DING Y, LI P, et al. Mind the class weight bias: Weighted maximum mean discrepancy for unsupervised domain adaptation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017. Honolulu, Hawaii, USA: IEEE Press, 2017: 2272-2281.
- [11] LI Y, CHENG L, PENG Y, et al. Manifold alignment and distribution adaptation for unsupervised domain adaptation[C]// Proceedings of 2019 IEEE International Conference on Multimedia and Expo (ICME). Shanghai, China: IEEE Press, 2019: 688-693.
- [12] FAN H, ZHENG L, YAN C, et al. Unsupervised person re-identification: Clustering and fine-tuning[J]. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2018, 14(4): 1-18.
- [13] FU Y, WEI Y, WANG G, et al. Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision 2019. Seoul, Korea: IEEE Press, 2019: 6112-6121.
- [14] ZHANG X, CAO J, SHEN C, et al. Self-training with progressive augmentation for unsupervised cross-domain person re-identification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision 2019. Seoul, Korea: IEEE Press, 2019: 8222-8231.
- [15] YANG F, LI K, ZHONG Z, et al. Asymmetric co-teaching for unsupervised cross-domain person re-identification[C]// Proceedings of the AAAI Conference on Artificial Intelligence 2020. New York, USA: IEEE Press, 2020: 12597-12604.
- [16] GE Y, CHEN D, LI H. Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification[C]//Proceedings of International Conference on Learning Representations 2019. New Orleans, LA, USA: IEEE Press, 2019: 1-15.
- [17] ZHOU K, YANG Y, CAVALLARO A, et al. Omni-scale feature learning for person re-identification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision 2019. Seoul, Korea (South): IEEE Press, 2019: 3702-3712.
- [18] 朱虎飞, 丁子豪, 杨永亮, 等. 再生核 Hilbert 空间中两阶段稀疏表示目标跟踪算法[J]. 控制理论与应用, 2022, 39(4): 730-740.  
ZHU Hufei, DING Zihao, YANG Yongliang, et al. Two-stage sparse representation objective tracking algorithm in reproducing kernel Hilbert space[J]. Control Theory & Applications, 2022, 39(4): 730-740.
- [19] ZHANG B, GUO Y, LI Y, et al. Memory recall: A simple neural network training framework against catastrophic forgetting [J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 33(5): 2010-2022.
- [20] ZHENG L, SHEN L, TIAN L, et al. Scalable person re-identification: A benchmark[C]//Proceedings of the IEEE International Conference on Computer Vision 2015. Santiago, Chile: IEEE Press, 2015: 1116-1124.
- [21] ZHENG Z, ZHENG L, YANG Y. Unlabeled samples generated by Gan improve the person re-identification baseline in vitro [C]//Proceedings of the IEEE International Conference on Computer Vision 2017. Venice, Italy: IEEE Press, 2017: 3754-3762.
- [22] 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 Cheng, CHEN Ying. Multi-granularity feature fusion network for person re-identification[J]. Acta Electronica Sinica, 2021, 49(8): 1541-1550.

- [23] 唐轲, 郎丛妍. 基于噪声标签自适应的行人再识别方法[J]. 数据采集与处理, 2021, 36(1): 103-11.  
TANG Ke, LANG Congyan. Noise label based self-adaptive person re-identification[J]. *Journal of Data Acquisition and Processing*, 2021, 36(1): 103-11.
- [24] BIANCO S, CADENE R, CELONA L, et al. Benchmark analysis of representative deep neural network architectures[J]. *IEEE Access*, 2018, 6(1): 64270-64277.

作者简介:



吴海丽(1984-), 通信作者, 女, 讲师, 研究方向: 人工智能、大数据、图像识别, E-mail: cattleman01@163.com。



张月琴(1963-), 女, 教授, 硕士生导师, 研究方向: 数据挖掘、智能信息处理。



庞俊奇(1984-), 男, 博士, 高级工程师, 研究方向: 测试、计量、仪器、光学。

(编辑: 张黄群)