

doi: 10.11835/j.issn.1000-582X.2026.02.007

引用格式:董能, 谢明鸿, 张亚飞, 等. 知识引导和细粒度信息增强的无监督域自适应行人再识别[J]. 重庆大学学报, 2026,49(2): 81-91.



# 知识引导和细粒度信息增强的无监督域自适应行人再识别

董 能<sup>a,b</sup>, 谢明鸿<sup>a</sup>, 张亚飞<sup>a,b</sup>, 李 凡<sup>a,b</sup>, 李华锋<sup>a,b</sup>, 谭婷婷<sup>a</sup>

(昆明理工大学 a. 信息工程与自动化学院; b. 云南省人工智能重点实验室 昆明 650500)

**摘要:**无监督域自适应旨在将源域知识迁移到目标域数据集上,是行人再识别中一个非常重要的任务。现实情况下,采集到的数据往往具有视频帧信息。因此,目标域数据集中单一视角下的行人标签极易获取。然而,已有的方法忽略了这些信息,限制了识别性能的提升。为此,笔者提出知识引导和细粒度信息增强的无监督域自适应行人再识别方法,开发了目标域单视角下行人样本已知的新范式,充分挖掘了目标域中有用的信息。同时,将源域知识作为引导,辅助模型提取目标域行人图像的判别性特征。与已有的知识迁移策略相比,这种知识引导的方式能更加有效地缓解域偏移。进一步,将行人的局部信息嵌入到全局特征中,增强了特征的细粒度信息。在 2 个公开的数据集上进行实验,充分证明了提出方法的有效性和优越性。

**关键词:**行人再识别;无监督域自适应;知识引导;细粒度信息增强

中图分类号:TP311

文献标志码:A

文章编号:1000-582X(2026)02-081-11

## Knowledge guidance and fine-grained information enhancement for unsupervised domain adaptation person re-identification

DONG Neng<sup>a,b</sup>, XIE Minghong<sup>a</sup>, ZHANG Yafei<sup>a,b</sup>, LI Fan<sup>a,b</sup>, LI Huafeng<sup>a,b</sup>, TAN Tingting<sup>a</sup>

(a. Faculty of Information Engineering and Automation; b. Key Laboratory of Artificial Intelligence in Yunnan Province, Kunming University of Science and Technology, Kunming 650500, P. R. China)

**Abstract:** Unsupervised domain adaptation (UDA) aims to transfer knowledge learned from a labeled source domain to an unlabeled target domain, playing a very important role in person re-identification. In real-world applications, video-based pedestrian data are often available, making it feasible to obtain single-camera-view labels in the target domain. However, existing UDA methods typically ignore this readily accessible information, thereby limiting performance improvements. To address this issue, we propose a knowledge-guided and fine-grained information enhancement framework for UDA person re-identification. A novel paradigm is introduced that leverages single-view labeled pedestrian samples in the target domain to fully exploit intra-domain

收稿日期:2021-10-18

基金项目:国家自然科学基金(61966021,61562053);大学生创新创业训练计划项目(202010674098)。

Supported by National Natural Science Foundation of China (61966021, 61562053) and College Students' Innovative Entrepreneurial Training Plan Program (202010674098).

作者简介:董能(1996—),男,硕士研究生,主要从事机器学习,计算机视觉方向研究。

通信作者:谢明鸿,(E-mail)minghongxie@163.com。

information. Meanwhile, source-domain knowledge is used as guidance to assist the model to extract more discriminative target-domain pedestrian representations, effectively mitigating domain shift compared with conventional knowledge-transfer strategies. Furthermore, local pedestrian cues are integrated into global features to strengthen fine-grained feature expression. Experiments conducted on two publicly datasets fully demonstrate the effectiveness and superiority of the proposed method.

**Keywords:** person re-identification; unsupervised domain adaptation; knowledge guidance; fine-grained information enhancement

行人再识别旨在从非重叠的相机视角下检索出感兴趣的行人图像,是一项重要的智能监控任务。近几年来,得益于深度学习<sup>[1-2]</sup>和大数据的兴起,行人再识别受到了国内外学者的广泛关注,并取得了一系列重大研究进展<sup>[3-5]</sup>。然而,现有的行人再识别算法大多是基于有监督学习,这类算法依赖于大量的标签样本,而现实场景中标签的获取十分困难。此外,不同数据集采集的行人图像具有不同成像风格和光照条件,这些因素会导致数据集之间的特征分布存在域差异。因此,如果直接将有监督算法训练好的模型部署到其他无标签数据集上,识别性能会急剧下降。

无监督域自适应能有效克服上述问题<sup>[6]</sup>,这类算法利用有标签的数据作为源域,无标签的数据作为目标域对模型进行训练。测试时,提取无标签数据集中待测试的行人特征进行匹配以计算识别率。现有的无监督域自适应行人再识别算法大致可以分为3类:基于自训练的无监督域自适应行人再识别<sup>[7-8]</sup>、基于图像风格迁移的无监督域自适应行人再识别<sup>[9-10]</sup>以及基于域不变特征提取的无监督域自适应行人再识别<sup>[11-12]</sup>。基于自训练的方法首先在有标签的源域上训练一个模型,用这个训练好的模型提取目标域行人的图像特征。接下来,利用聚类算法对其进行伪标签预测,最后微调模型使网络参数适用于目标域。基于图像风格迁移的无监督域自适应行人再识别方法利用循环一致性生成对抗网络<sup>[13]</sup>,将源域有标签的行人图像迁移到目标域风格下,再利用迁移后的源域图像训练识别模型,缓解风格差异导致的域偏移问题。考虑到不同数据集间的域差异会导致特征空间的分布不同,基于域不变特征提取的无监督域自适应行人再识别着重探索如何有效对齐源域和目标域之间的特征分布。上述方法尽管有效,但由于缺乏目标域样本的标签信息,无监督域自适应行人再识别算法往往识别率较低。已有的知识迁移策略一定程度上克服了域偏移问题。然而,当域差异较大时,这类方法的性能会受到限制。此外,直接在全局特征层面进行分布对齐,难以充分挖掘行人的细粒度信息。

现实情况中,采集的数据往往具有视频帧信息。因此,单一相机视角下的行人标签极易获取,这也意味着目标域数据集中单个视角下的行人样本是已知的。如果能充分利用这些样本,并与无监督域自适应行人再识别相结合,就能弥补现有算法无法利用目标域监督信息的缺陷。此外,一旦识别模型能自适应学习到适用于目标域数据集的参数,就能克服知识迁移策略的弊端,有效缓解域偏移对识别性能造成的不利影响。进一步,局部特征往往包含行人的细粒度信息。如果能够将行人的局部特征充分有效整合到全局特征中,可以提升行人特征的判别性。为此,笔者提出知识引导和细粒度信息增强的无监督域自适应行人再识别算法。

如图1所示,研究提出的算法在源域数据集上训练一个识别模型,赋予网络提取具有判别性行人特征的能力。为充分挖掘目标域中的监督信息,研究开发了目标域单视角下行人样本已知的新范式。在此基础上,利用源域知识引导模型对目标域单视角下已知的行人样本进行表征,学习适用于目标域数据集的网络参数,弱化域偏移对识别性能带来的不利影响。与已有的知识迁移策略不同,利用源域知识引导模型对目标域进行学习有助于网络优化,提取具有鲁棒性的行人特征。进一步,笔者提出的算法提取了行人图像的局部细粒度特征和全局特征,并将这些特征作为节点,以特征间相似度作为边构建图关系。接下来,将行人的局部判别性信息嵌入到全局特征中,克服现有方法难以提取行人细粒度特征的问题,提升模型的表征能力。

论文的主要贡献和提出方法的优势可以总结如下:

1)开发了目标域单视角下行人样本已知的新范式,并将该范式引入到无监督域自适应行人再识别方法中,弥补了现有算法的不足。

2)提出的算法将源域知识作为引导,辅助模型对目标域数据进行表征。与现有方法不同,提出的算法赋予了网络自主学习的能力,有效弱化了域偏移对识别性能带来的不利影响。进一步,将局部判别性信息嵌入到全局特征中,克服已有算法难以提取细粒度行人特征的问题。

3)在2个公开的大规模数据集 Market1501<sup>[14]</sup>和 DukeMTMC<sup>[15]</sup>上进行了大量的实验,充分证明提出方法的有效性和优越性。

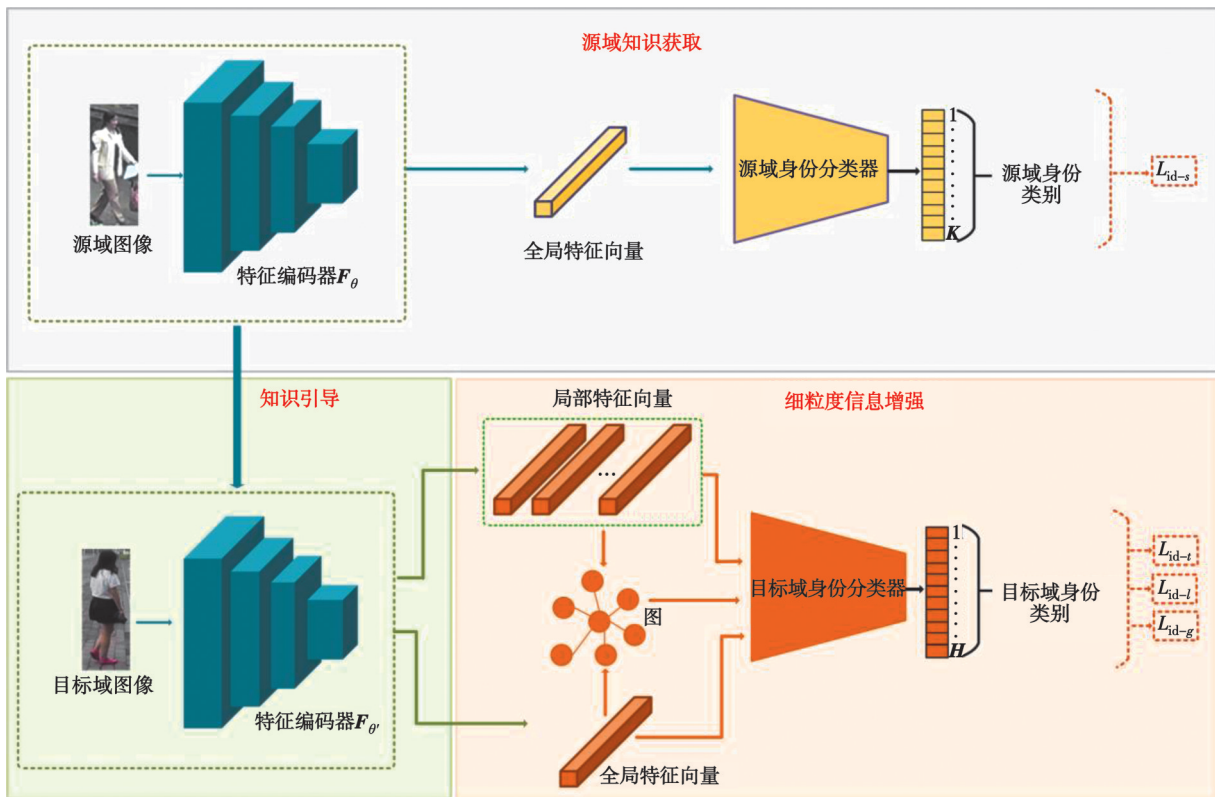


图1 知识引导和细粒度信息增强的无监督域自适应行人再识别

Fig. 1 Knowledge guidance and fine-grained information enhancement for unsupervised domain adaptation person re-identification

## 1 相关研究

### 1.1 有监督行人再识别

目前大多数行人再识别工作聚焦于有监督的训练,利用深度神经网络提取行人的全局特征。具体来说,Zheng等<sup>[14]</sup>设计了一个分类模型,利用行人的身份标签训练网络以表征行人。为获得更具判别性的全局特征,Chen等<sup>[16]</sup>在分类模型的基础上新增一个验证网络,判断2张图像是否属于同一行人。考虑到正负样本之间的相似性关系,Cheng等<sup>[17]</sup>提出利用三元组损失拉近正样本对以及推远负样本对之间的距离,进一步提高模型的表征能力。与此同时,Qian等<sup>[18]</sup>设计了一个多尺度深度表征学习模型以提取行人的特征,获取不同尺度的具有判别性的行人信息。由于模型的性能很大程度上取决于样本的数量,Zheng等<sup>[19]</sup>首次在行人再识别领域尝试利用生成对抗技术<sup>[20]</sup>生成行人样本,再将这些生成的样本送进识别模型参与训练,从而增强监督信息。

早期的行人再识别主要关注全局特征。为获取行人的细粒度信息,Sun等<sup>[21]</sup>提出利用水平切块的方式提取行人的局部特征,并将这些局部特征进行拼接用于测试。由于全局特征能够一定程度上保证行人的宏观特性,Zhao等<sup>[22]</sup>提出一个多阶段特征分解和选择性树状结构的融合框架,同时捕获行人身体的全局和局部特征。考虑到人体姿态多变的问题,Zheng等<sup>[23]</sup>引入人体姿态的先验知识,利用14个姿态关键点提取行人的局



部特征。目前,基于有监督的行人再识别方法已经取得令人惊叹的效果。然而,这类方法依赖大量的有标签样本,而样本标签的获取需要耗费大量的人力物力。另外,由于不同数据集间的域偏移问题,直接将这类算法训练好的模型部署到其他数据集上,识别性能会急剧下降。

### 1.2 无监督域自适应行人再识别

基于无监督域自适应的行人再识别能够有效解决上述问题。已有的算法大致可以分为基于自训练、基于图像风格迁移以及基于域不变特征提取的方法。

基于自训练的无监督域自适应行人再识别首先在源域上训练一个识别模型,随后预测目标域行人图像的伪标签,最后对模型进行微调以适应目标域。Fan等<sup>[24]</sup>利用预训练好的模型提取特征,再用K均值聚类算法预测行人标签。为提高伪标签的准确性,Yu等<sup>[12]</sup>提出多级软标签的概念,充分挖掘样本之间的标签信息。考虑到不同行人之间的差异性以及同一身份行人之间的相似性,Lin等<sup>[25]</sup>提出一种自下而上的聚类算法,并使用联合学习的策略获取行人图像之间的相关性。然而,现实情况中成对的样本是稀缺的。在这样的情况下预测伪标签会引入大量噪声,限制识别率的提升。

得益于循环一致性生成对抗网络在图像转换任务上取得的巨大成就,基于图像风格迁移的无监督域自适应行人再识别算法被提出。Deng等<sup>[26]</sup>提出相似性保留的生成对抗网络,将源域的行人图像迁移到目标域风格下。为保留行人的身份信息,Wei等<sup>[27]</sup>引入语义分割技术,并设计身份一致性损失来保证行人外观的不变性。然而,这类算法需要大量的时间成本和硬件消耗,且迁移过程中容易导致行人信息的丢失。

近年来,基于域不变特征提取的无监督域自适应行人再识别受到广泛关注。Liu等<sup>[28]</sup>设计了一个自适应迁移网络以弱化源域和目标域之间的域偏移,提升了跨域行人再识别的性能。考虑到行人局部细粒度信息的强判别性,Yang等<sup>[29]</sup>提出基于分块的判别性特征提取方法来提升模型的表征能力。Zhong等<sup>[30]</sup>引入一个记忆模块存储目标域行人图像的特征,进一步提高识别性能。然而,由于缺少目标域样本的标签,这类算法往往难以训练,且提取到的特征缺乏判别性,因此识别率较低。

### 1.3 元学习行人再识别

元学习,又被称为学会学习,是继增强学习之后机器学习领域的一个重要研究分支。该方法的提出起源于人类终身学习的思想,即利用已有的经验知识来对新事物进行学习和认知。对于机器学习而言,就是要通过已学得模型参数指导网络自适应的进行优化以适用于新任务。近年来,有关元学习的研究呈现出百家争鸣的状态。Santoro等<sup>[31]</sup>设计了一个记忆模块并将其嵌入到网络中存储学习到的知识。接下来,通过外部记忆获取的方式将这些知识应用到后期的新任务上,使模型能够更加准确的对目标物进行预测。众所周知,深度学习是通过梯度更新的方式对模型进行优化。一旦神经网络具有梯度预测能力,在面临新任务时,模型就能够准确找到最优的学习方向,实现快速学习。基于这个思想,Andrychowicz等<sup>[32]</sup>构建了一个具有梯度预测能力的通用型神经网络。

考虑到人类的注意力具有自我学习和提升的能力,Vinyals等<sup>[33]</sup>提出利用以往的任务训练一个注意力网络优化模型,使其在面对新任务时能够直接关注到最重要的部分。只要基准模型具有足够的泛化性,就能对新任务进行很好的表征。为此,Finn等<sup>[34]</sup>提出多任务同时启动的思想,通过获取不同任务学习的合成梯度方向来进行网络参数更新。不同于分类任务,行人再识别面临的一个主要问题是同一行人的样本数量较少,这也是导致识别性能不高的一个重要原因。为此,Zhao等<sup>[35]</sup>充分利用未知域下的行人样本并提出基于记忆模块的多源域元学习框架训练一个具有泛化性的模型。可见,将元学习的思想引入到行人再识别中,研究如何有效赋予识别模型自主学习的能力能进一步推动行人再识别技术的发展。

基于以上讨论,为了缓解完全有监督行人再识别可扩展性弱以及无监督域自适应行人再识别性能低的问题,研究开发了单视角下行人样本已知的新范式,并将这种范式引入到无监督域自适应行人再识别中,提出知识引导和细粒度信息增强的无监督域自适应行人再识别。该方法将源域的知识作为引导信息,辅助网络对目标域行人已知的样本进行学习,有效缓解域偏移对识别性能带来的不利影响。更进一步,将行人的局部判别性信息嵌入到全局特征中,克服已有算法难以提取行人细粒度特征的问题。

## 2 方 法

如图 1 所示,研究提出的算法在源域数据集上训练一个识别模型,获取知识,赋予网络提取判别性行人特征的能力。考虑到目标域单个相机视角下行人样本是已知的,将模型在源域学到的知识作为引导,辅助模型对目标域已知的行人样本进行表征,使模型参数适用于目标域数据集。进一步,细粒度信息有助于特征判别性的提升。因此,设计了一个图卷积神经网络,将行人的局部信息嵌入到全局特征中,增强特征的细粒度信息,提升模型的表征能力。

### 2.1 源域知识获取

为获取源域知识,在源域数据集上训练一个特征提取器  $F$  和一个源域身份分类器  $W_s$ ,赋予网络提取判别性行人特征的能力。给定源域的行人图像  $X_s = \{x_s^i\} (i=1, 2, \dots, N_s)$ , 对应的身份标签为  $Y_s = \{y_s^k\} (k=1, 2, \dots, K)$ 。其中  $N_s$  表示源域行人图像的数量,  $K$  表示源域行人身份的个数。

首先,利用特征提取器提取源域行人图像的特征,再将其送进分类器得到分类结果。为了赋予特征提取器提取判别特征的能力,使用交叉熵损失优化特征提取器及源域身份分类器。损失函数如公式(1)所示

$$L_{\text{id-s}}(F, W_s) = -\frac{1}{N_s} \sum_{i=1}^{N_s} I_{[y_s^k=k]} \log p(W_s(F_\theta(x_s^i))), \quad (1)$$

式中:  $F_\theta$  表示在源域上学习到的特征提取器,其参数为  $\theta$ 。  $I_{[y_s^k=k]}$  是一个指示函数。通常情况下,当  $y_s^k = k$  时,  $I_{[y_s^k=k]} = 1$ ; 否则  $I_{[y_s^k=k]} = 0$ 。然而,这种取值只考虑了正确标签位置的损失,忽略了其他标签位置,容易导致过拟合。为克服这一问题,采用标签平滑正则化<sup>[36]</sup>对指示函数进行改进,即

$$I_{[y_s^k=k]} = \begin{cases} 1 - \varepsilon \frac{K-1}{K}, & y_s^k = k, \\ \frac{\varepsilon}{K}, & \text{otherwise,} \end{cases} \quad (2)$$

这里的  $\varepsilon$  是一个常数,取 0.1。

### 2.2 知识引导

现实情况中采集的行人图像具有视频帧信息,因此一个相机视角下的行人样本往往是已知的。如果能够充分利用这些数据,并将模型在源域数据集上学到的知识作为引导信息,指导模型进行优化,能有效缓解源域与目标域之间的域偏移对识别性能造成的不利影响。

提出的知识引导机制如图 2 所示,选取目标域数据集中一个相机视角下的行人图像。由于含有帧信息,可以对不同行人的图像进行标签分配。例如,目标域某个相机视角下共有  $H$  个行人的  $N_t$  张图像,可以将这些行人图像的标签从 1 分配到  $H$ 。令目标域一个相机视角下的行人图像为  $X_t = \{x_t^j\} (j=1, 2, \dots, N_t)$ , 其对应的身份标签为  $Y_t = \{y_t^h\} (h=1, 2, \dots, H)$ 。

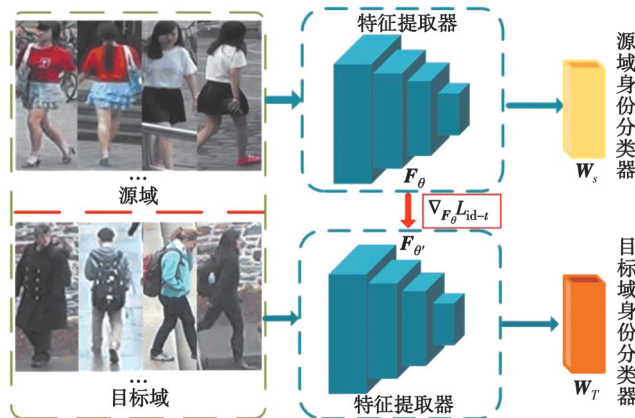


图 2 知识引导机制

Fig. 2 Knowledge guidance mechanism

在源域知识获取阶段,特征提取器在源域数据集上学到一定的知识,其参数为 $\theta$ 。利用这些知识引导模型对目标域图像进行表征,将这个网络参数应用到目标域数据单视角已知的行人样本上求取梯度 $\nabla_{F_\theta} L_{\text{id}-t}(F_\theta)$ 。其中

$$L_{\text{id}-t}(F_\theta, W_T) = -\frac{1}{N_T} \sum_{j=1}^{N_T} I_{[y_t^j=h]} \log p(W_T(F_\theta(x_t^j))), \quad (3)$$

式中: $F_\theta$ 表示目标域上学习得到的特征提取器,其参数为 $\theta$ 。利用梯度更新的方式来学习该参数,即

$$F_{\theta'} = F_\theta - \alpha \nabla_{F_\theta} L_{\text{id}-t}(F_\theta), \quad (4)$$

式中: $\alpha$ 表示学习率。

通过上述方式,就能够赋予网络自主学习的能力,使模型参数适用于目标域行人图像,弱化域偏移对识别性能带来的不利影响。

### 2.3 细粒度信息增强

为了提取目标域图像中更具判别性的细粒度特征,利用特征提取器分别提取目标域行人图像的局部和全局特征。接下来,通过图卷积网络将行人的局部信息嵌入到全局特征中,增强行人特征中的细粒度信息。

首先,将特征提取器提取的行人特征图经过全局平均池化得到全局特征向量。同步地,将该特征图水平分为 $N(N=6)$ 块,同样利用全局平均池化得到 $N$ 个局部特征向量,并利用身份损失对其进行约束

$$L_{\text{id}-l}(F_\theta, W_T) = -\frac{1}{N_T} \frac{1}{m} \sum_{m=1}^6 \sum_{j=1}^{N_T} I_{[y_t^j=h]} \log p(W_T(f_m)), \quad (5)$$

接下来,利用局部和全局特征向量构建图 $G=(V, E_n)$ 。其中, $V=\{v_1, v_2, \dots, v_n, \dots, v_N, v\}$ 是节点, $\{v_1, v_2, \dots, v_n, \dots, v_N\}$ 表示局部特征, $v$ 是全局特征。 $E_n$ 是边,表示全局特征与各局部特征之间的相似性关系。本文利用高斯核函数来定义这种关系,即

$$E_n = e^{-\frac{\|v-v_n\|^2}{2\sigma^2}}, \quad (6)$$

式中 $\sigma$ 为常数,设为1。构建好图后,通过图卷积网络将目标域行人图像的局部细粒度信息整合进全局特征,得到一个新的全局特征 $f_g$ 。利用身份损失约束该特征,有

$$L_{\text{id}-g}(F_\theta, G, W_T) = -\frac{1}{N_T} \sum_{j=1}^{N_T} I_{[y_t^j=h]} \log p(W_T(f_g)), \quad (7)$$

这样,就能够克服已有算法难以提取细粒度特征的缺陷,进一步提升识别性能。

### 2.4 总损失

综上所述,提出算法的整体损失函数为

$$L_{\text{total}} = L_{\text{id}-s} + L_{\text{id}-t} + L_{\text{id}-l} + \lambda L_{\text{id}-g}, \quad (8)$$

式中, $\lambda$ 用于权衡损失项 $L_{\text{id}-g}$ 所占的相对权重。

## 3 实验

### 3.1 数据集

为证明提出方法的有效性,在2个公开的大规模数据集Market1501和DukeMTMC上进行实验。其中,Market1501数据集包含来自6个不同相机视角下的1501个行人,共有32668张图像,这些图像采集于夏天。根据文献[14]中的协议,将751个行人的12936张图像作为训练集,余下的作为测试集。与Market1501不同,DukeMTMC的36411张行人图像是用8个相机在冬天拍摄的,共有1404个行人。其中,702个行人的16522张图像作为训练集,余下的作为测试集。在无监督域自适应行人再识别设置中,分别将其中一个数据集作为源域,另一个作为目标域,且在源域训练集上进行训练并在目标域测试集上进行测试。

### 3.2 实验细节

研究提出的算法采用在ImageNet<sup>[37]</sup>上预训练好的ResNet-50<sup>[38]</sup>作为特征提取器,将其最后一个卷积层的

步长以及下采样层的步长统一设为 1<sup>[39]</sup>。特征提取器之后加入 7 个全局平均池化层用于得到行人的局部和全局特征向量。源域和目标域的身份分类器是全连接层,其维度等于行人的身份数量。另外,设计了一个图卷积网络增强行人的细粒度信息。在训练过程中,对所有的行人图像进行随机翻转、随机填充以及随机擦除等预处理操作,并将图像大小统一缩放到 384×192。批大小设置为 16,学习率为 0.003 5,训练迭代 110 次。测试时,利用特征提取器提取目标域测试集中行人图像的局部和全局特征,再将其送进图卷积网络以将行人的局部信息嵌入到全局特征中。接下来,用全局特征进行相似性度量得到测试结果。算法采用平均预测精度(mAP)和 Rank 值作为评价指标。

3.3 方法比较

为证明提出方法的有效性,将提出的算法与其他先进的无监督域自适应行人再识别方法进行比较。这些方法大致可以分为 3 类,基于自训练的方法:CAMEL<sup>[40]</sup>、DECAMEL<sup>[41]</sup>、PUL<sup>[24]</sup>;基于图像风格迁移的方法:Camstyle<sup>[9]</sup>、SPGAN<sup>[26]</sup>、PTGAN<sup>[27]</sup>、SBSGAN<sup>[42]</sup>、ATNet<sup>[28]</sup>、CSGLP<sup>[10]</sup>;基于域不变特征提取的方法:TJ-AIDL<sup>[43]</sup>、CFSM<sup>[44]</sup>、SSAE<sup>[45]</sup>、CaNE<sup>[46]</sup>。比较结果如表 1 所示。

表 1 与一些最先进的无监督域自适应方法进行比较  
Table 1 Comparison of the proposed method with some state-of-the-art unsupervised domain adaptation methods

方法	Market1501→DukeMTMC			DukeMTMC→Market1501		
	Rank-1	Rank-5	mAP	Rank-1	Rank-5	mAP
CAMEL(ICCV, 17)	40.30	57.60	19.80	54.50	73.10	26.30
PUL(TOMCCAP, 18)	30.00	43.40	16.40	45.50	60.70	20.50
DECAMEL(TPAMI, 20)	—	—	—	60.24	—	32.44
Camstyle(CVPR, 18)	48.40	62.50	25.10	58.50	<b>78.20</b>	27.40
SPGAN(CVPR, 18)	46.40	62.30	26.20	57.70	75.80	26.70
PTGAN(CVPR, 19)	27.40	43.60	13.50	38.60	57.30	15.70
SBSGAN(ICCV, 19)	53.50	—	30.80	58.50	—	27.30
ATNet(CVPR, 19)	45.10	59.50	24.90	55.70	73.20	25.60
CSGLP(TIFS, 20)	47.80	62.30	27.10	61.20	77.50	31.50
TJ-AIDL(CVPR, 19)	44.30	59.60	23.00	58.20	74.80	26.50
CFSM(AAAI, 19)	49.80	—	27.30	61.20	—	28.30
SSAE(PR, 20)	50.20	—	28.10	60.70	—	26.60
CaNE(WACV, 20)	—	—	—	57.20	73.00	27.40
提出的方法	<b>58.45</b>	<b>71.05</b>	<b>39.92</b>	<b>61.68</b>	74.25	<b>37.49</b>

注:加黑数据表示最优的性能。

从表 1 可以看出,研究提出的算法在无监督域自适应行人再识别任务上取得了最优性能,充分证明了算法的有效性和优越性。具体来说,与 DECAMEL 相比,提出算法的 mAP 在 DukeMTMC→Market1501 任务上,mAP 和 Rank-1 分别提高了 5.05% 和 1.44%。与基于图像风格迁移的无监督域自适应行人再识别方法不同,笔者提出的算法不需要预先迁移行人图像的风格,也能够有效地学习适用于目标域行人图像的网络参数。例如,在 DukeMTMC→Market1501 和 DukeMTMC→Market1501 任务上,提出算法的性能都优于目前最先进的 SBSGAN。其中,在 DukeMTMC→Market1501 任务上,mAP 从 27.30% 提升到了 37.49%,Rank-1 精度提升了 3.18%。在 Market1501→DukeMTMC 任务上,mAP 和 Rank-1 也分别提高了 9.12% 和 4.95%。此外,相比较于基于域不变特征提取的方法 TJ-AIDL、CFSM、SSAE 以及 CaNE 而言,笔者提出算法的性能更加优越。在 DukeMTMC→Market1501 任务上,mAP 超出了最优的 SSAE10.89%,Rank-1 超出其 0.98%。在 Market1501→DukeMTMC 任务上也远远优于其他基于域不变特征提取的方法。



笔者提出的方法之所以能取得良好性能,主要原因在于:充分挖掘了目标域中可用的监督信息,弥补了现有算法的不足;利用模型在源域数据集上学到的先验知识引导模型对目标域样本进行表征,有效缓解域偏移对识别性能造成的不利影响;细粒度信息的增强有助于模型提取更具判别性的行人特征。

### 3.4 消融实验

为赋予网络提取具有判别性行人特征的能力,本文在源域数据集上训练识别模型。考虑到目标域单相机视角下的行人标签极易获取,开发了单视角样本已知的新范式。同时,将源域知识作为引导信息,指导模型对目标域已知的行人样本进行表征以缓解域偏移问题。进一步,将行人的局部信息嵌入到全局特征中,增强行人特征中的细粒度信息,提高识别性能。为证明每一块内容的作用,在这一部分进行了消融实验,实验结果如表2所示。

表2 消融实验  
Table 2 Ablation study

模块	Market1501→DukeMTMC			DukeMTMC→Market1501		
	Rank-1	Rank-5	mAP	Rank-1	Rank-5	mAP
源域知识获取	31.05	46.27	16.45	48.90	66.30	21.86
源域知识获取+知识引导	47.48	62.52	28.69	57.93	72.98	30.91
源域知识获取+知识引导+细粒度信息增强	<b>58.45</b>	<b>71.05</b>	<b>39.92</b>	<b>61.68</b>	<b>74.25</b>	<b>37.49</b>

注:加黑数据表示最优的性能。

由表2可以看出,仅使用源域知识获取,在Market1501→DukeMTMC任务上mAP仅为16.45%,DukeMTMC→Market1501任务上的mAP为21.86%。当利用模型在源域数据集上学到的知识对目标域单视角下已知的行人样本进行表征学习后,在Market1501→DukeMTMC任务上的mAP提升了12.24%,DukeMTMC→Market1501任务上的mAP提升到了30.91%。这充分说明了这种方式能够赋予模型自主学习的能力,有效缓解了源域和目标域之间的域偏移。进一步,当局部信息嵌入到全局特征中,识别性能进一步提升。比如在Market1501→DukeMTMC任务上,mAP提升了11.23%,Rank-1从47.48%提升到58.45%。同样,在DukeMTMC→Market1501任务上也有显著提升。可见,提出的算法可以增强行人特征中的细粒度信息,提高模型的表征能力。

### 3.5 参数分析

在提出的知识迁移和细粒度信息增强的行人再识别算法中, $\lambda$ 用于权衡损失项 $L_{id-g}$ 所占的相对权重,取值为1。为证明取值最优,分别在Market1501和DukeMTMC数据集上做了参数分析的实验,结果如下图3所示。“M”代表Market1501,“D”表示DukeMTMC。

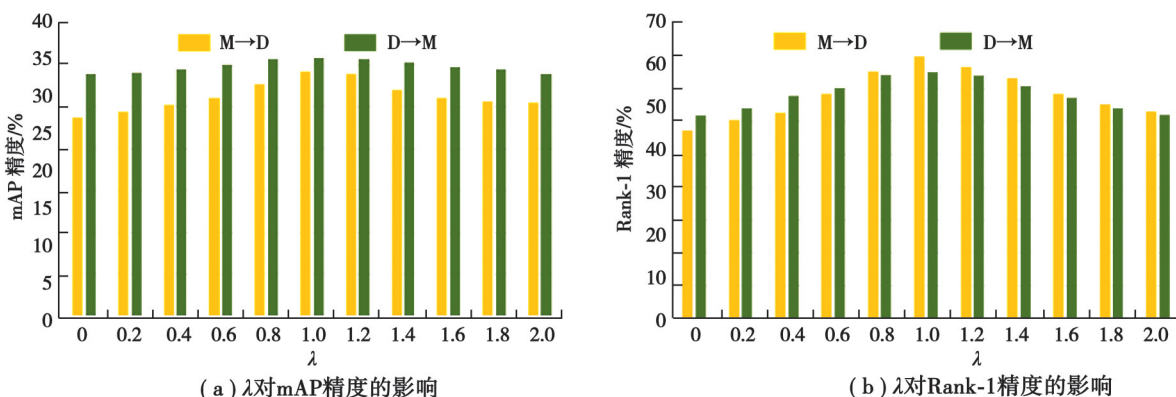


图3  $\lambda$ 对性能的影响分析

Fig. 3 Parametric analysis



从图3可以看出,当 $\lambda$ 从0增长至1时,在2个任务上的mAP精度和Rank-1值都呈稳定提升状态。当 $\lambda = 1$ 时精度最高,当 $\lambda > 1$ 时,2个任务上的性能都开始下降,可见, $\lambda$ 的最优取值为1。

## 4 结 论

研究针对无监督域自适应行人再识别任务,提出知识引导和细粒度信息增强的无监督域自适应行人再识别算法。在该方法中,开发了目标域单视角下行人样本已知的新范式,并将这种范式与无监督域自适应行人再识别相结合,充分挖掘了目标域中可用的监督信息。考虑到知识迁移固有的缺陷,提出将源域知识作为引导信息。利用信息指导模型对目标域数据进行表征,赋予网络自主学习的能力,弱化域偏移对识别性能带来的不利影响。进一步,将行人的局部信息嵌入到全局特征中,使网络具有提取细粒度判别性特征的能力。最后,进行了大量的实验验证,并与其他一些先进的无监督域自适应行人再识别算法进行比较,体现了提出方法的有效性和优越性。

## 参考文献

- [1] 刘智,冯欣,张杰.基于深度卷积神经网络和深度视频的人体行为识别[J].重庆大学学报,2017,40(11):99-106.  
Liu Z, Feng X, Zhang J. Action recognition based on deep convolution neural network and depth sequences[J]. Journal of Chongqing University, 2017, 40(11): 99-106. (in Chinese).
- [2] 许晓泓,何霆,王华珍,等.结合Transformer模型与深度神经网络的数据到文本生成方法[J].重庆大学学报,2020,43(7):91-100.  
Xu X H, He T, Wang H Z, et al. Research on data-to-text generation based on transformer model and deep neural network[J]. Journal of Chongqing University, 2020, 43(7): 91-100. (in Chinese).
- [3] Zheng M, Karanam S, Wu Z Y, et al. Re-identification with consistent attentive siamese networks[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019: 5735-5744.
- [4] 雷大江,滕君,王明达,等.基于卡方核的正则化线性判别行人再识别算法[J].重庆大学学报,2018,41(9):66-76.  
Lei D J, Teng J, Wang M D, et al. Chi square kernel regularized linear discriminant analysis for person re-identification[J]. Journal of Chongqing University, 2018, 41(9): 66-76. (in Chinese).
- [5] Wu A C, Zheng W S, Guo X W, et al. Distilled person re-identification: towards a more scalable system[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019: 1187-1196.
- [6] 张林华,方正云,李仕林,等.双重对抗无监督域自适应绝缘子检测算法[J].重庆大学学报,2021,44(3):122-131.  
Zhang L H, Fang Z Y, Li S L, et al. Unsupervised domain adaptation insulator detection algorithm based on dual adversarial[J]. Journal of Chongqing University, 2021, 44(3): 122-131. (in Chinese).
- [7] Yang F, Li K, Zhong Z, et al. Asymmetric co-teaching for unsupervised cross-domain person re-identification[J]. Proceedings of AAAI Conference on Artificial Intelligence, 2020, 34(7): 12597-12604.
- [8] Li H F, Yan S L, Yu Z T, et al. Attribute-identity embedding and self-supervised learning for scalable person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(10): 3472-3485.
- [9] Zhong Z, Zheng L, Zheng Z D, et al. Camstyle: a novel data augmentation method for person re-identification[J]. IEEE Transactions on Image Processing, 2018, 28(3): 1176-1190.
- [10] Ren C X, Liang B H, Ge P F, et al. Domain adaptive person re-identification via camera style generation and label propagation [J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 1290-1302.
- [11] Qi L, Wang L, Huo J, et al. A novel unsupervised camera-aware domain adaptation framework for person re-identification[C]//IEEE International Conference on Computer Vision (ICCV). Seoul, South Korea: IEEE, 2019: 8079-8088.
- [12] Yu H Y, Zheng W S, Wu A C, et al. Unsupervised person re-identification by soft multi-label learning[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019: 2148-2157.
- [13] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017: 2223-2232.
- [14] Zheng L, Shen L, Lu T, et al. Scalable person re-identification: A benchmark[C]//IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015: 1116-1124.

- [15] 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 (ECCV). Amsterdam, The Netherlands: Springer, 2016: 17-35.
- [16] Chen H R, Wang Y W, Shi Y M, et al. Deep transfer learning for person re-identification[C]//IEEE 4th International Conference on Multimedia Big Data (BigMM). Xi'an, China: IEEE, 2018: 1-5.
- [17] Cheng D, Gong Y H, Zhou S P, et al. Person re-identification by multi-channel parts-based cnn with improved triplet loss function[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016: 1335-1344.
- [18] Qian X L, Fu Y W, Xiang T, et al. Multi-scale deep learning architectures for person re-identification[C]//IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017: 5399-5408.
- [19] Zheng Z D, Zheng L, Yang Y. Unlabeled samples generated by gan improve the person re-identification baseline in vitro[C]//IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017: 3754-3762.
- [20] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//International Conference on Neural Information Processing Systems (NeurIPS). Montreal, Canada: NeurIPS Foundation, 2014: 2672-2680.
- [21] Sun Y F, Zheng L, Yang Y, et al. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline)[C]//European Conference On Computer Vision (ECCV). Munich, Germany: Springer, 2018: 480-496.
- [22] Zhao H Y, Tian M Q, Sun S Y, et al. Spindle net: person re-identification with human body region guided feature decomposition and fusion[C]//IEEE Conference on Computer Vision and Pattern (CVPR). Honolulu, USA: IEEE, 2017: 1077-1085.
- [23] Zheng L, Huang Y J, Lu H C, et al. Pose-invariant embedding for deep person re-identification[J]. IEEE Transactions on Image Processing, 2019, 28(9): 4500-4509.
- [24] 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): 1-18.
- [25] 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.
- [26] Deng W J, Zheng L, Ye Q X, et al. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, USA: IEEE, 2018: 994-1003.
- [27] Wei L H, Zhang S L, Gao W, et al. Person transfer gan to bridge domain gap for person re-identification[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, USA: IEEE, 2018: 79-88.
- [28] Liu J W, Zha Z J, Chen D, et al. Adaptive transfer network for cross-domain person re-identification[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019: 7202-7211.
- [29] Yang Q Z, Yu H X, Wu A C, et al. Patch-based discriminative feature learning for unsupervised person re-identification[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019: 3633-3642.
- [30] Zhong Z, Zheng L, Luo Z M, et al. Invariance matters: Exemplar memory for domain adaptive person re-identification[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019: 598-607.
- [31] Santoro A, Bartunov S, Botvinick M, et al. One-shot learning with memory-augmented neural networks[J]. Proceedings of Machine Learning Research, 2016, 48(1): 1842-1850.
- [32] Andrychowicz M, Denil M, Gomez S, et al. Learning to learn by gradient descent by gradient descent[C]//International Conference on Neural Information Processing Systems (NeurIPS). Barcelona, Spain: NeurIPS Foundation, 2016: 3981-3989.
- [33] Vinyals O, Blundell C, Lillicrap T, et al. Matching networks for one shot learning[C]//International Conference on Neural Information Processing Systems (NeurIPS). Barcelona, Spain: NeurIPS Foundation, 2016: 3630-3638.
- [34] Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks[J]. Proceedings of the 34th International Conference on Machine Learning, 2017, 70(1): 1126-1135.

- [35] 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]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE, 2021: 6277-6286.
- [36] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016: 2818-2826.
- [37] Deng J, Dong W, Socher R, et al. ImageNet: A large-scale hierarchical image database[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Miami, USA: IEEE, 2009: 248-255.
- [38] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016: 770-778.
- [39] Luo H, Gu Y Z, Liao X Y, et al. Bag of tricks and a strong baseline for deep person re-identification[C]//IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, USA: IEEE, 2019: 1487-1495.
- [40] Yu H X, Wu A C, Zheng W S. Cross-view asymmetric metric learning for unsupervised person re-identification[C]//IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017: 994-1002.
- [41] Yu H Y, Wu A C, Zheng W S. Unsupervised person re-identification by deep asymmetric metric embedding[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(4): 956-973.
- [42] Huang Y, Wu Q, Xu J S, et al. Sbsgan: suppression of inter-domain background shift for person re-identification[C]//IEEE International Conference on Computer Vision (ICCV). Seoul, South Korea: IEEE, 2019: 9527-9536.
- [43] Wang J Y, Zhu X T, Gong S G, et al. Transferable joint attribute-identity deep learning for unsupervised person re-identification [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City: IEEE, 2018: 2275-2284.
- [44] Chang X B, Yang Y X, Xiang T, et al. Disjoint label space transfer learning with common factorised space[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 3288-3295.
- [45] Li H F, Kuang Z Y, Yu Z T, et al. Structure alignment of attributes and visual features for cross-dataset person re-identification [J]. Pattern Recognition, 2020, 106: 107414.
- [46] Yuan Y, Chen W Y, Chen T L, et al. Calibrated domain-invariant learning for highly generalizable large scale re-identification [C]//IEEE Winter Conference on Applications of Computer Vision (WACV). Colorado, USA: IEEE, 2020: 3589-3598.

(编辑 侯 湘)