

# 弱监督场景下的行人重识别研究综述\*

祁磊, 于沛泽, 高阳



(计算机软件新技术国家重点实验室(南京大学),江苏 南京 210023)

通讯作者: 高阳, E-mail: gaoy@nju.edu.cn

**摘要:** 近年来随着智能监控领域的不断发展,行人重识别问题逐渐受到学术界和工业界的广泛关注,其主要研究将不同摄像头下相同身份的行人图像进行关联.当前大部分研究工作关注在有监督场景下,即给定的训练数据都存在标记信息,然而考虑到数据标注工作的高成本,这在现实应用中往往是难以拓展的.本文关注于弱监督场景下的行人重识别算法,包括无监督场景和半监督场景,并且对当前先进的方法进行了分类和描述.对于无监督场景的行人重识别算法,本文根据其技术类型划分为五类,分别为基于伪标记的方法、基于图像生成的方法、基于实例分类的方法、基于领域自适应的方法和其他方法;对于半监督场景的行人重识别方法,本文根据其场景类型划分为四类,分别为少量的人有标记的场景、每一个人有少量标记的场景、基于 tracklet 的学习的场景和摄像头内有标记但摄像头间无标记的场景.最后,本文对当前行人重识别的相关数据集进行了整理,并对现有的弱监督方法的实验结果进行分析与总结.

**关键词:** 行人重识别;半监督学习;无监督学习;深度学习;人工智能  
**中图法分类号:** TP311

中文引用格式: 祁磊,于沛泽,高阳.弱监督场景下的行人重识别研究综述.软件学报,2020. <http://www.jos.org.cn/1000-9825/6083.htm>

英文引用格式: Qi L, Yu PZ, Gao Y. Research on Weak-Supervised Person Re-Identification. Ruan Jian Xue Bao/Journal of Software, 2020 (in Chinese). <http://www.jos.org.cn/1000-9825/6083.htm>

## Research on Weak-Supervised Person Re-Identification

QI Lei, YU Pei-Ze, GAO Yang

(State Key Laboratory for Novel Software Technology (Nanjing University), Nanjing 210023, China)

**Abstract:** Recently, with the development of the intelligent surveillance, person re-identification (Re-ID) has attracted lots of attention in the academic and industrial communities, which aims to associate person images of the same identity under different non-overlapping cameras. Most of the current research works focus on the supervised case where all given training samples have label information. Considering the high cost of data labeling, these methods designed for the supervised setting have poor generalization in practical applications. In this paper, we focus on person re-identification algorithms under the weakly supervised case including the unsupervised case and the semi-supervised case and classify and describe several state-of-the-art methods. In the unsupervised setting, these methods are divided into five categories from different technology perspectives, which include the methods based on pseudo-label, image generation, instance classification, domain adaptation, and others. In the semi-supervised setting, we divide these methods into four categories according to the case discrepancy, which are the case where a small number of persons are labeled, the case where there are few labeled images for each person, the case based on tracklet learning, and the case where there are the intra-camera labels but no inter-camera label information. Finally, we summarize several benchmark person re-identification datasets and analyze some experimental results of these weak-supervised person re-identification algorithms.

**Key words:** person re-identification; semi-supervised learning; unsupervised learning; deep learning; artificial intelligence

近年来,随着社会安防意识增强和科学技术的进步,城市中监控摄像头的数量越来越多.这些监控系统往往部署在写字楼、校园、商场、大大小小的街道和社区等各种各样的场所,在安防领域起到了重要的作用.例如,当某地发生偷窃事件时,公安机关或安保部门可以通过监控记录来获取偷窃者出现的时间及行踪轨迹等重要信息.不过对于监控中记录的视频数据,当前大多数情况都是通过人工的方式进行查看、分析,工作量非常大且效率非常低.换句话说,通过人工的方式对多个摄像头中的内容进行关联是一项非常耗时的工作.

随着人工智能技术的发展,许多曾经需要人来执行的任务都可以通过人工智能技术来实现,甚至在某些任务上机器比人完成得更加精确.例如在大规模图像数据集 ImageNet<sup>[1]</sup>上,机器对于图像的分类任务远远比人更精确,并且识别速度比人更快速.计算机视觉技术能够帮助我们有效地利用现有的大规模图像或视频数据,并进行分析和理解.对于监控视频数据而言,行人是其主要的目标对象之一.目前行人检测技术、行人跟踪技术和行人重识别技术(person re-identification,简称 Re-ID)已经在学术界和工业界受到了广泛的关注.相对于行人检测技术和行人跟踪技术,行人重识别技术起步较晚,近些年才逐渐得以关注.传统的检测和跟踪技术只关注在一个视频中的行人目标,而行人重识别技术则关注多个视频中的行人之间的关联性,即旨在将多个不同的摄像头下的同一个行人目标进行关联.如图 1 所示,实现一个完整的行人重识别系统,应当包括行人检测<sup>[2-4]</sup>、行人跟踪<sup>[5]</sup>和行人重识别<sup>[6,7]</sup>技术三个模块.

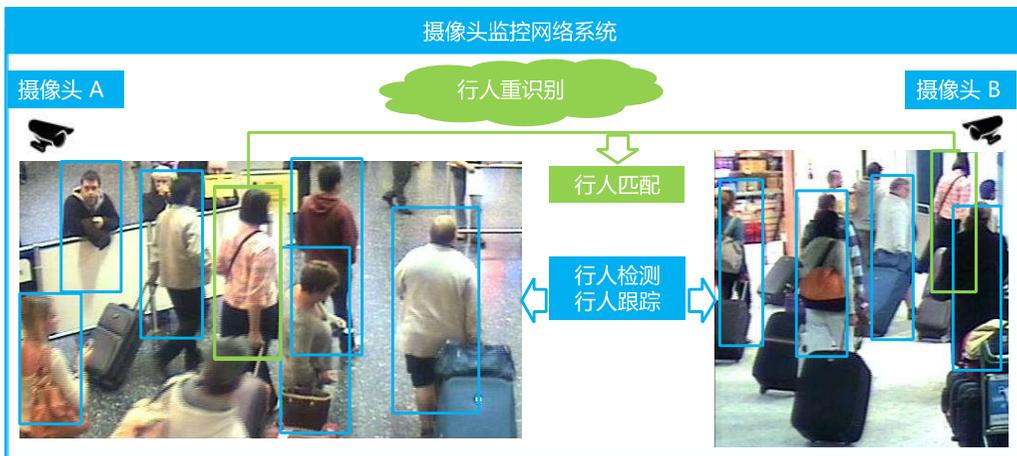


Fig.1 Schematic diagram of a person re-identification system

图 1 行人重识别系统示意图

## 1 相关背景

从技术层面来讲,行人重识别是用某个查询图像(query image)在一个大的图像数据库(gallery set)中检索和匹配相关图像的任务,也可以看作是一种只针对行人图像的图像检索(image retrieval)任务,如图 2 所示.其目标是希望获得具有判别性的特征来区分相同身份和不同身份的行人图像.因此,在行人重识别问题中,绝大部分工作都是关注在怎样获取具有判别性的特征上.由于行人图像来自多个不同的摄像头,因此该问题的研究在现实应用中面临着许多挑战,包括不同摄像头下图像的光照条件、分辨率、视角以及行人姿态等各方面差异.

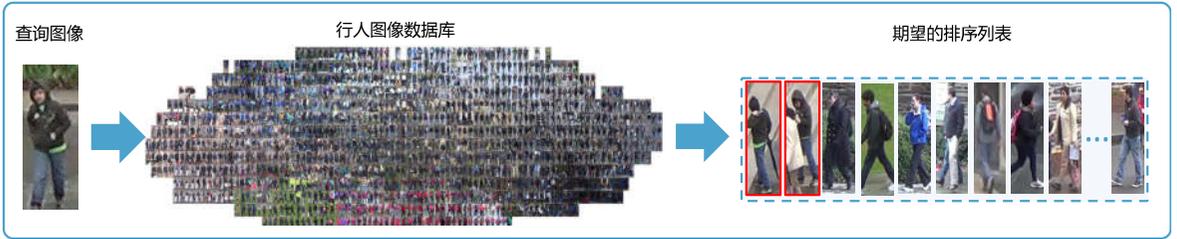


Fig.2 Schematic diagram of person retrieval

图2 行人检索示意图

当前该领域的大部分工作都关注在有监督场景下的行人重识别问题.然而在现实中,行人重识别的数据标注工作往往需要花费大量的人力和财力,特别是对跨摄像头间的行人数据进行关联的这一步骤.并且在当前深度学习时代,大部分方法都是依赖大规模的有标记数据来训练一个深度模型.而数据标注的高成本使得有监督的方法难以扩展到现实应用中,这也是阻碍行人重识别技术能够真正落地的一大因素.另一方面,在现实中我们能够轻松获得大量无标记的行人数据.因此在行人重识别问题的研究中,如何使用少标记的大规模图像数据来训练得到鲁棒的模型具有重大的研究价值和意义.

目前大部分行人重识别领域的工作主要集中在有监督场景下相关算法的研究.早些年,一些研究者主要致力于提取鲁棒的特征来强化行人特征的判别性<sup>[8-12]</sup>,也有一些研究者主要关注在学习方法上,例如设计更好的度量方法以使其更容易地识别相同的人并区分不同的人<sup>[13-16]</sup>,或者通过学习公共的子空间或字典来消除不同摄像头视角之间的差异<sup>[17-19]</sup>.近些年深度学习技术不断发展,特别是其在机器视觉应用领域取得了巨大成功,新提出的行人重识别方法基本上都是基于深度学习的.其中,一些研究工作使用注意力机制的方式来提高行人重识别模型的泛化能力<sup>[20-28]</sup>,也有一些研究工作通过设计损失函数来提升行人重识别模型的性能<sup>[29-31]</sup>.最近也出现了一些基于局部的学习方法<sup>[32-35]</sup>,该类方法虽然简单,但是可以获得更具有判别性的特征,在行人重识别任务上取得了较优的性能.虽然在有监督场景下行人重识别问题已经有了突破性的进展,但是有监督场景下的学习不利于行人重识别模型很好地泛化到其他场景下,因此考虑在深度学习需要大量的有标记数据参与训练的背景下,研究弱监督场景下的少标记学习在行人重识别任务中具有重大的意义与价值.

考虑到计算机视觉任务的相关应用在现实场景的落地需求,少标记学习在学术界和工业界渐渐受到关注.基于行人重识别任务,本文将少标记学习问题分为无监督的场景和半监督的场景,其更具体的场景分类如图3所示.

以下主要对当前存在的弱监督场景下的行人重识别方法进行总结、分类和对其性能进行分析.

## 2 无监督场景下的行人重识别问题

在深度学习时代之前,绝大部分无监督的方法主要借助于传统的领域自适应方法来学习共享的模型参数<sup>[36]</sup>、公共的子空间<sup>[37]</sup>或字典<sup>[38]</sup>.这些方法沿用了传统的领域自适应方法的数据设定,即在训练中可以使用有标记的源域数据(source domain)和无标记的目标域数据(target domain)来进行模型参数的学习.除了该设定之外,也有一些方法只使用无标记的数据.例如一些研究者使用无监督的方式训练并学习一个字典<sup>[39,40]</sup>,也有一些研究者通过无监督的方式学习一些具有判别性的特征<sup>[41,42]</sup>.

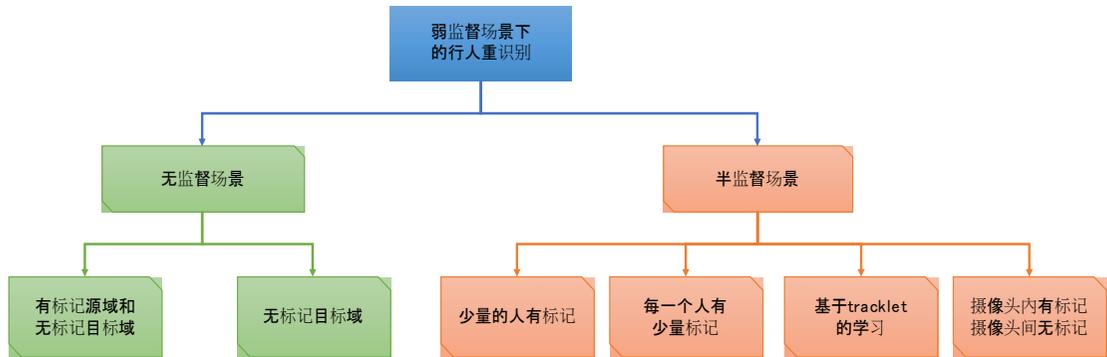


Fig.3 Classification of scene under weak-supervised person re-identification

图3 弱监督行人重识别下的场景分类

随着深度学习技术的广泛应用,近年来也出现了一些基于深度学习的无监督行人重识别方法.本文将这些方法划分为五类:基于伪标记的方法、基于图像生成的方法、基于实例分类的方法、基于领域自适应的方法和一些其他方法.在大部分无监督深度行人重识别方法中,一般会使用有标记的源域样本和无标记的目标域样本进行训练模型,其中基于伪标记的方法和基于实例分类的方法一般使用有标记的源域的数据进行模型预训练,然后使用无标记的目标域数据进行无监督学习;基于图像生成的方法一般通过将源域中的图像转化成目标域风格的图像,然后再使用这些图像来训练模型;基于领域自适应的方法旨在减少领域间数据分布的差异,通过特征的层级来对齐源域和目标域的数据分布,以将源域中的判别性信息迁移到无标记的目标域中.

## 2.1 基于伪标记的方法

伪标记方法在无监督学习中有着广泛的应用,其主要思想是为无标记的数据产生高质量的伪标记来训练和更新神经网络,如图4所示.Yu等人<sup>[43]</sup>提出了一种基于软的多标记学习的方法来解决无监督行人重识别问题,该方法通过借助有标记的辅助数据集来生成代理标签.具体地,通过在辅助数据集上为每一个类别产生一个代理(可以将其视为聚类中心),然后针对每个无标记的样本计算它们与这些代理的相似性并生成一个相似性向量(即软的多标记),进而判断两个无标记样本的相似性.例如,如果两个无标记数据生成的软的多标记相似,则它们大概率属于相同的行人.另外根据无标记数据原始特征的相似性,该方法挖掘了难的负样本对(即属于不同的人但外表比较相似的样本对).该方法最终由以下三种损失同时优化:(1)软的多标记学习旨在将无标记数据中潜在相同的人尽可能拉近,难的负样本对尽可能推远;(2)软标记分布的一致性学习旨在将不同视角下的软标记分布尽可能变成相同的分布;(3)代理样本的学习旨在找出合适的代理以对有标记的辅助数据集的每一个类别(即每一个人)进行表示.

Yang等人<sup>[44]</sup>提出基于分块的判别性特征学习方法,该方法由以下两个模块构成:(1)基于块的特征学习,该模块基于分块网络将得到的相似图像块拉近,不相似的图像块推远;(2)图像层级的块的特征学习,该模块将原始图像通过随机图像转化的方法<sup>[45]</sup>对其风格进行一定程度的转化,然后得到对应的正样本对,而对于负样本的选择则采用循环排序的方法挖掘困难的负样本(即对于一个查询样本得到的排序结果,将这些结果中的图像依次作为查询图像重新得到一个排序列表,然后根据列表中图像的重叠度来判断该查询图像和排序列表中的图像是否属于一个人),最后基于生成的正负样本对进行块的三元组损失的计算和优化.Wang等人<sup>[46]</sup>在有标记的源域上引入属性语义信息和身份判别信息分别训练两个不同的分支网络,并使用一个自编码网络将身份信息从身份分支迁移到属性分支.而对无标记的目标域数据,采用在有标记的源域上训练好的属性分支来生成属性的伪标记信息,再使用这些带有伪标记信息的目标域数据来更新网络.在测试阶段中,作者使用属性分支的特征作为最终的特征.Lv等人<sup>[47]</sup>提出使用摄像头的时空信息来提升生成正确的正样本对的概率,然后基于这些融合时空信息的正负样本对进行模型的训练和更新.在框架中,作者提出了一种基于贝叶斯推断

的融合模型,该模型能够有效地将时空信息融入到样本的相似性度量中.

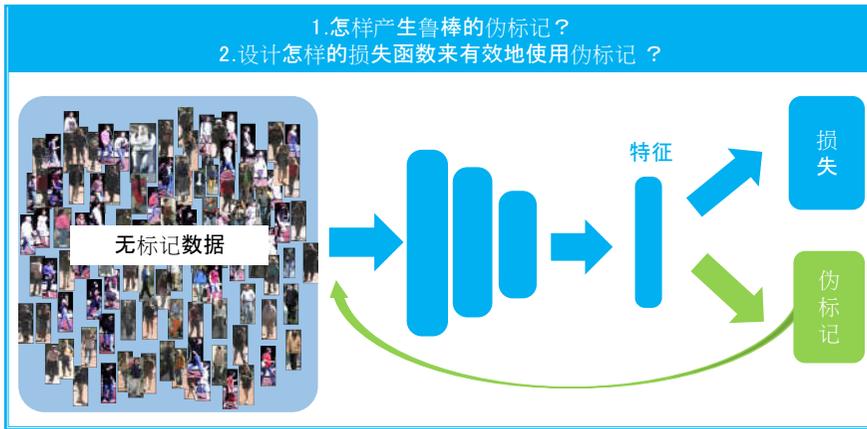


Fig.4 Schematic diagram of the methods based on pseudo-label

图 4 基于伪标记方法的学习框架

近年来许多工作也将传统的聚类方法引入到无监督的行人重识别任务中,用来产生伪标记的信息.Fu 等人<sup>[48]</sup>利用 DBSCAN 聚类算法<sup>[49]</sup>,基于在源域上预训练的模型提取特征来对无标记的数据进行聚类,然后基于聚类的结果构造三元组并使用三元组损失(triplet loss)<sup>[29]</sup>来进行训练.在每一轮训练之后,利用得到的神经网络再次提取特征并进行聚类,重新得到更新的标记信息进行训练.聚类和网络的训练是迭代的过程,这样能够不断地获得更优的标记信息和更鲁棒的特征表示.Zhang 等人<sup>[50]</sup>提出了一种自训练的渐进式增强框架,主要分为保守训练和提升训练两个步骤,并在训练过程中使用 HDBSCAN 聚类方法<sup>[51]</sup>产生伪标记.保守训练过程中使用传统的三元组损失和基于排序的三元组损失联合训练网络,提升训练过程中使用交叉熵损失进一步提升网络的泛化性能.在整个训练过程中,保守训练和提升训练也是基于迭代的方式来优化网络.Lin 等人<sup>[52]</sup>提出了一种自底向上的聚类策略来不断地融合相似的样本,该方法起初将每一个样本视为一个类,并将每一类的特征存储在一个空间中,在训练过程中不断地更新融合不同的类,并且更新每一个类的新的表示特征.作者还在文中提出了一种多样性的归一化方法以避免每一类中的图像数量差别过大.

Tang 等人<sup>[53]</sup>利用神经网络的最后两层(在残差网络中即全局平均池化层和全连接层)的输出分别作为特征并通过 DBSCAN<sup>[49]</sup>进行聚类,然后基于聚类的结果产生标记信息,在网络的最后两层上都采用三元组损失函数同时进行网络的训练,并且也使用交替学习的方式更新网络以及获得新的标记信息.Yang 等人<sup>[54]</sup>考虑到通过聚类产生伪标记的方法往往会带来许多噪声信息,因此提出了一种能够在聚类之后对样本进行过滤的方法.作者认为在学习过程中所有的样本都是重要的,因此该方法根据 DBSCAN 聚类<sup>[49]</sup>将样本划分为正常样本(在聚类过程被划分到某个类别)和异常样本(在聚类过程中没有被划分到某个类别),而在一些其他基于聚类的方法中一般会忽略这些异常样本.同时该方法使用主模型和协作模型两种模型来相互促进学习,其中主模型利用正常样本和异常样本来训练,而协作模型只使用正常样本来训练.特别地,对于用来训练主模型的异常样本,需要通过协作模型来选择其中一些置信度高的伪标记样本进行训练,而对于用来训练协作模型的正样本,需要使用主模型对已分配标记的样本再次进行过滤,以确保选择的样本都是纯净的,即保证这些样本的伪标记具有较好的可靠性.整个学习过程通过迭代的方式,不断地提升两种网络的性能.Ding 等人<sup>[55]</sup>提出一种基于分散度的聚类方法来对无标记的样本进行聚类,该聚类方法不仅仅考虑了类别间的差异性信息,而且也考虑到了类别内的紧凑程度.相比于其他的聚类方法,该方法能够更广地考虑到多个样本间的关系,并且能够有效处理不平衡的数据分布所带来的问题.

目前在无监督行人重识别领域中,为无标记样本数据生成伪标记的方法已经成为主流的技术路线.该类

方法具有思路简单清晰、性能良好的优点,特别是一些基于聚类的伪标记生成方法,可以展现出与有监督学习方法相接近的性能.然而,该类方法在伪标记生成的准确度以及如何有效利用生成的伪标记等方面仍然存在进一步提升的空间.

## 2.2 基于图像生成的方法

近些年生成对抗网络已经取得了很大的进展.在无监督行人重识别领域,一些研究者基于该技术从图像层级角度来解决领域迁移的问题,如图 5 所示.Huang 等人<sup>[56]</sup>考虑到不同领域图像背景的差异较大,且现有的图像分割方法并不能很好地将行人图像的前景和背景分开,因此提出了 SBSGAN 通过产生软掩模的方法来移除图像的背景区域,该方法能够有效地抑制图像分割方法带来的错误.考虑到当前许多基于 GAN 的方法只能产生单一风格的图像,Chen 等人<sup>[57]</sup>提出了一种对偶条件图像生成器以生成不同风格的行人图像数据,该方法能够将一张图像迁移到多个风格下.Liu 等人<sup>[58]</sup>认为领域间的差异信息由多种因素造成,如光照、分辨率、摄像头视角等,因此作者采用分而治之的方法将风格迁移网络分成多个子网络,分别针对不同的领域差异因素进行迁移,如光照迁移网络、分辨率迁移网络和视角迁移网络等.每个子网络首先进行预训练,最终通过一个选择网络来产生每个子网络的权重信息,并融合所有子网络中的信息生成最终的风格迁移图像.

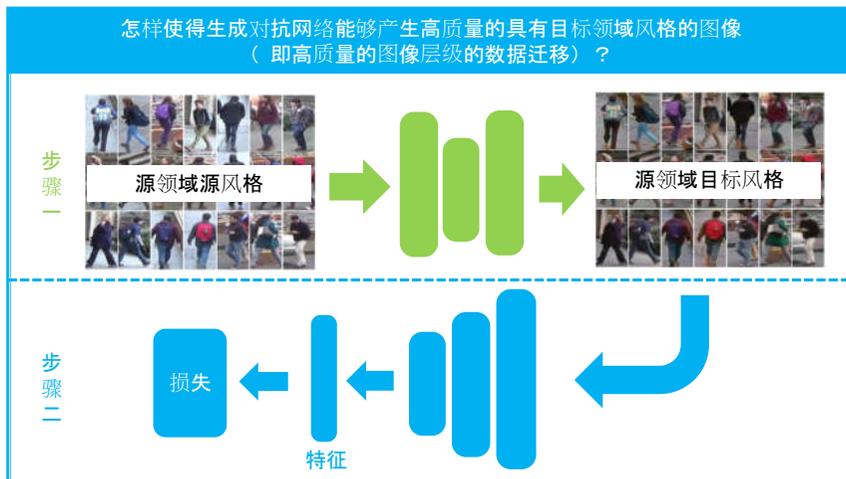


Fig.5 Schematic diagram of the methods based on image generation

图 5 基于图像生成方法的学习框架

Zhong 等人<sup>[59]</sup>利用 StarGAN<sup>[60]</sup>对目标域中不同摄像头风格下的图像进行转化,训练过程中的正样本对来自于同一个摄像头风格下.结合原始目标域图像、源域图像和这些转化的图像一起生成三元组来训练更新神经网络.特别地,对于一个三元组,如果一个 anchor 样本可以容易地在有标记的源域中得到它对应的正样本,同样也能从目标域中得到它的负样本,这样的三元组能够减小源域和目标域之间的差异.Bak 等人<sup>[61]</sup>认为剧烈变化的光照条件是跨领域行人重识别问题的一个巨大挑战,然而对于当前单一数据集来说,光照情况相对比较单一.因此作者提出了一个合成的行人重识别数据集(Synthetic Person Re-Identification,简称 SyRI),其包含在 140 种不同的光照条件下的 100 个虚拟的行人.该方法首先训练一个 140 类的光照推断模型,用来推断一个目标域与哪种光照情况下的源域数据接近,然后利用 CycleGAN<sup>[62]</sup>将该源域的数据转化成目标域风格的图像,再用来训练特征提取网络.Deng 等人<sup>[63]</sup>在 CycleGAN 的基础上引入两个重要的跨域行人重识别特性来保证图像风格迁移的质量.其一,每张风格迁移后的图像应该与转换前的图像保持身份信息的一致性;其二,任何图像从源域迁移到目标域后都应为目标域中所有图像的身份信息不同.为了在原始的 CycleGAN 上引入这两个特性,作者利用对比损失<sup>[64]</sup>的子网络来约束原始的 CycleGAN 的训练.Wei 等人<sup>[65]</sup>提出了 PTGAN(Person Transfer GAN)来对图像进行从源域到目标域的迁移,该方法在 CycleGAN<sup>[62]</sup>的基础上引入行人前景分割图像

来保证行人区域迁移前后的一致性。

这类方法的思想是从图像层面进行风格迁移,其很大程度上依赖于生成对抗网络所生成图像的质量.与其他场景的不同点在于,从监控摄像头获取的行人图像往往质量较低并且存在一些噪声,导致风格转换后图像的质量并不高.因此该方法在无监督场景下的性能提升并不是很理想,需要进一步研究更加适合于行人场景的生成对抗网络来解决该类问题。

### 2.3 基于实例分类的方法

在传统的图像分类问题中,无监督学习已经取得很大的研究进展.受非参数化实例分类的方法<sup>[66]</sup>的启发,近期也有一些研究者将其引入到无监督行人重识别任务中来.非参数化实例分类的方法考虑到在分类任务中外表相似的类别与外表相似性较小的类别相比有一个更大的预测概率值,这说明这些相似的类别在特征学习网络中存在潜在的相关性,因此该方法将所有独立的样本当作一个单独的类别来训练网络.假设我们有 $n$ 张图像  $x_1, x_2, \dots, x_n$ , 它们的特征分别为  $v_1, v_2, \dots, v_n$ , 对于一个图像  $x$ , 其对应的特征为  $v$ , 属于第  $i$  个样本(类别)的概率为

$$P(i|v) = \frac{\exp(v_i^T v / \tau)}{\sum_{j=1}^n \exp(v_j^T v / \tau)} \tag{1}$$

其中每个样本的特征  $v_j$  被存储在内存银行(memory bank)  $M$  中.特别地,  $M$  在每一个 epoch 之后会被更新.  $\tau$  是一个超参数,用来调节特征向量在单位球体上的集中程度<sup>[67]</sup>.基于公式(1),对于样本  $x$ , 其对应的特征为  $v$ , 我们得到基于实例的损失函数为

$$L(x) = - \sum_j^n r_j \log P(j|v) \tag{2}$$

其中  $r_j \in \{0, 1\}^n$  表示图像  $x$  对应的指示值,即如果  $x$  和  $x_j$  被判定为邻近的样本则为 1, 否则为 0.

该类方法旨在关注如何得到更好的邻近关系  $r$  来学习模型,如图 6 所示.Zhong 等人<sup>[68]</sup>沿用了非参数化实例分类的框架,并且对于每一个独立的样本(即每一个类),通过对抗生成网络 StarGAN<sup>[60]</sup>生成一些其他摄像头风格的图像来增加每一个类的样本数量,即类似于一种数据增广的方式,并且在训练过程中考虑拉近一些邻近样本间的距离来强化类别之间的关联.对于邻近样本的选择,Zhong 等人<sup>[69]</sup>进一步提出了一种基于图的预测方式来判别两个样本是否是真实的邻近样本,该方法主要考虑了所有样本间的关系进一步确保选择的真正同类样本的正确性.Ding 等人<sup>[70]</sup>通过设置一个距离阈值来选择每一个实例的邻近样本,并且考虑到每个实例的邻近样本的不均衡性会导致偏向于学习某些样本,在损失函数中融入了一种平衡机制来抑制该问题。

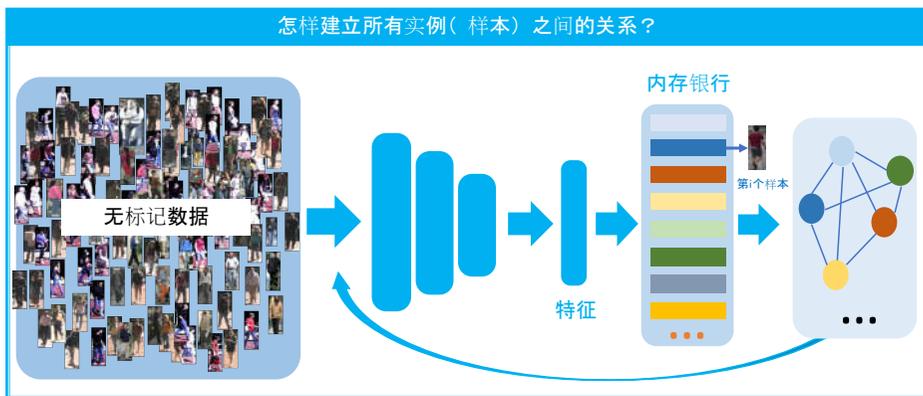


Fig.6 Schematic diagram of the methods based on instance classification

图 6 基于实例分类方法的学习框架

基于实例分类的方法虽然在性能方面展现出了优越性,然而其对于样本之间的关联问题仍然需要进一步

步研究,即考虑如何采用有效的算法更加精确地进行样本关联度匹配.

#### 2.4 基于领域自适应的方法

在深度学习的无监督行人重识别方法中,许多研究工作沿用了传统的领域自适应的架构,即考虑消除或减少领域间的差异来将判别性的信息从源域迁移到目标域中,如图7所示.Lin 等人<sup>[71]</sup>提出了一种多任务中间层的特征对齐方法(Multi-task Mid-level Feature Alignment,简称 MMFA)来解决无监督跨域行人重识别问题.该方法联合身份学习和属性学习一起来训练更新网络,对无标记的目标域采用基于源域训练的模型生成的伪属性标记来进行模型的训练,并通过 MMD(Maximum Mean Discrepancy)<sup>[72]</sup>的方法减少源域和目标域之间的差异.考虑到跨域行人重识别问题中数据分布的差异不仅仅存在于领域之间,也存在于同一领域下的不同摄像头之间(不同摄像头下光照、分辨率、背景和视角等方面也存在差异),Delorme 等人<sup>[73]</sup>和 Qi 等人<sup>[74]</sup>都针对性地提出了基于摄像头的对抗网络来解决在跨域行人重识别任务中的数据分布差异问题.其中,Delorme 等人<sup>[73]</sup>在源域和目标域中所有的摄像头之间做等价的对抗,并且对无标记的目标域中的数据采用标记平滑的方式<sup>[32]</sup>分配到源域类别中进行训练.Qi 等人<sup>[74]</sup>提出了源域和目标域中摄像头交互式的对抗,并在理论上证明了该对抗方式能够将源域和目标域所有的摄像头下的数据映射到同一空间中,另外还利用时序信息从无标记的目标域中挖掘一些判别性的信息来训练更新网络.同时作者在文中也提到,对于在行人重识别任务中使用的这种传统领域自适应框架,挖掘无标记的判别性信息是非常重要的,这是因为单一地减少数据分布差异可能会破坏目标域中的原始样本间的关系,因此引入目标域中的一些信息能够一定程度地保证这种信息不会被破坏.

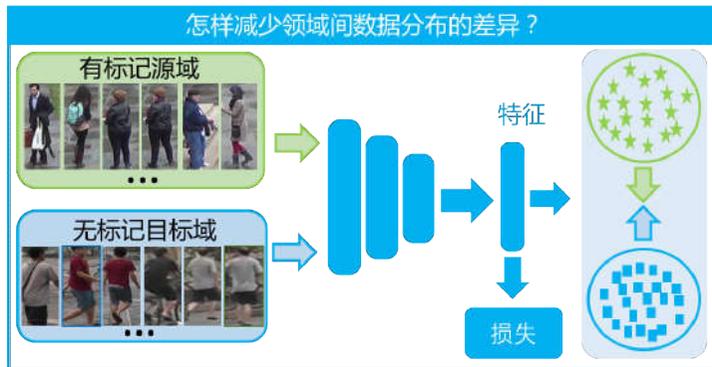


Fig.7 Schematic diagram of the methods based on domain adaptation

图7 基于领域自适应方法的学习框架

由于从数据分布的视角来解决无监督行人重识别问题是一种间接的处理方法,因此该方法与基于伪标记的方法和基于实例分类的方法相比,在性能方面稍有些不足.但是与基于图像生成的方法相比,该类方法的性能更好.因此,这说明了在行人重识别问题中,从特征层级的迁移效果要比从图像层级的迁移效果更好.

#### 2.5 其他方法

除上述的几大类方法之外,还有少量从其他角度设计的方法.Wu 等人<sup>[75]</sup>观察到摄像头内样本的相似性分布和摄像头间样本的相似性分布不一致,提出了摄像头一致性的学习方式以使得摄像头内的数据分布和摄像头间的样本相似性分布趋于一致,并且在学习过程中保持摄像头内样本间的相似性分布与其在预训练上的模型一致.也有少部分的研究者关注在基于领域泛化的行人重识别任务上,在该任务的训练过程中,只存在有标记的源域样本,对于目标域没有任何可用的数据.Kumar 等人<sup>[76]</sup>探索了只简单地结合多个源域来训练一个模型的方式,在目标域上也具有良好的泛化性能.Jia 等人<sup>[77]</sup>考虑到领域间的差异性主要是由不同领域间的风格

信息的差异引起的,受风格迁移学习的启发,作者提出在神经网络的低层使用实例归一化来减少不同领域的风格的影响,同时在高层使用特征归一化进一步地减少领域间风格信息的影响.Song 等人<sup>[78]</sup>提出领域不变性的映射网络来解决行人重识别任务在未见领域上的泛化问题,该方法专注于在一张行人图像和身份分类器的权重之间学习一种映射.具体地,对于每一个来自于候选集中的图像,可以生成一个分类器的权重向量.为了获得领域间的不变性,作者使用了元学习(meta-learning)中的插曲训练机制(episodic training)来更新网络的参数.在测试过程中,对于一张来自查询集合的图像和一张来自候选集合的图像,利用查询分支中提取的特征向量和候选分支中提取的权重向量进行点乘的值作为这两张图像的相似性.

### 3 半监督场景下的行人重识别问题

近年来,一些研究者也开始关注如何利用较少的标记信息来训练一个较优的模型.特别地,不同于无监督学习的定义,半监督学习在行人重识别中的有许多不同的设定.对于现有的方法,本文将其划分成如下几个场景.

#### (1) 少量的人有标记.

Liu 等人<sup>[79]</sup>提出利用半监督的对偶字典学习来解决少标记的行人重识别问题.该方法利用少量的标记数据来学习在不同摄像头之间的特征关系,而大量的未标记数据用来获得鲁棒的稀疏表示.Wu 等人<sup>[80]</sup>假定了只有少量的行人标记样本的情况,通过在其他有标记的数据集上训练好的多个不同模型来迁移信息,这些模型可以被视为多个教师模型.该方法使用教师学生网络训练机制来进行网络的训练,通过使用大量的无标记数据训练更新学生网络,并通过少量有标记样本来判断每个在源域上训练的模型的权重(即对于不同的样本,不同教师模型贡献程度不同).Xin 等人<sup>[81]</sup>使用少量的有标记数据训练模型,然后通过多视角聚类方法对无标记的数据进行聚类,再联合有标记数据和带有伪标记的无标记数据更新网络,并继续基于新的网络再次执行聚类算法,整个过程是迭代交替的.

#### (2) 每一个人有少量标记.

Wu 等人<sup>[82]</sup>提出了一种渐进式的学习方法来解决该问题,首先根据每个人的少量标记数据训练初始模型,然后对大量无标记数据中置信度较高的数据分配伪标记,其余置信度较低的数据暂不分配标记信息.在训练过程中,结合有标记和伪标记的数据一起使用传统的交叉熵损失更新训练网络.对于暂未分配标记的数据,接下来采用实例分类的方法,将每一个独立的样本当作一个类别并在网络中进行训练,然后基于训练好的网络再次重新分配伪标记.该方法的学习过程也是采用迭代更新的方式.然而该场景存在一定的局限性,其很难拓展到实际应用中.这是因为在该设定下通常很难获得整个数据集上所有行人的数量,除非对整个数据集进行标记,这将耗费大量的人工成本,与半监督设定的初衷相悖.因此采用该设定的研究工作较少.

#### (3) 基于 tracklet 的学习.

Li 等人<sup>[83]</sup>把行人重识别任务划分为摄像头内的学习和摄像头间的学习.对于摄像头内的学习,作者假设摄像头内经过跟踪算法已得到若干 tracklet,并提出了一种基于时序的稀疏采样方法以获得摄像头内不重复的 tracklet(即这些 tracklet 尽可能属于不同的人),然后使用交叉熵损失来学习这些有标记的样本.对于摄像头间的学习,作者提出一种损失函数以使得相近的跨摄像头的 tracklet 尽可能相似.在此基础上,Li 等人<sup>[84]</sup>进一步对摄像头内的学习做出改进,不再需要选择一些不重复的 tracklet,即可以使用全部的 tracklet,同时提出了一种软分类学习的方式来自动探索摄像头内 tracklet 的关系.在这两个研究工作的实验过程中,作者给定了图像数据集中摄像头内的标记信息.Wu 等人<sup>[75]</sup>沿用了文献<sup>[83]</sup>中的 tracklet 的选择方式,该方法主要提出了基于图的关联方式来建立跨摄像头 tracklet 之间的关联.特别地,基于视频的方法致力于使用给定的 tracklet 来探索时序信息以便于将其融入到特征表示中.然而,上述半监督方法主要是使用 tracklet 信息作为部分的标记信息来执行学习任务.

#### (4) 摄像头内有标记,摄像头间无标记.

受基于无监督的 tracklet 的学习的启发,Qi 等人<sup>[85]</sup>定义了一种新的半监督行人重识别设定,并且分别从数

据分布的角度和伪标记学习的角度提出了两种不同的解决方案<sup>[85,86]</sup>.在该设定下,每一个摄像头内均给定标记信息,而摄像头间是没有标记信息的.由于在行人重识别问题中,标记摄像头间的信息需要花费大量的成本,而摄像头内的标记信息可以借助于跟踪算法和少量的人工标记即可完成,因此这种半监督行人重识别的设定在实际应用中较大的意义.在文献<sup>[85]</sup>中,作者考虑到不同摄像头间数据分布的差异性(由背景、光照、视角等因素带来的影响),提出了一种基于摄像头对齐的对抗学习网络以将不同摄像头的的数据映射到同一空间中.在文献<sup>[86]</sup>中,作者通过在跨摄像头间生成渐进式的软标记来探索跨摄像头间样本之间的关系.同时,Zhu 等人<sup>[87]</sup>提出了类似的问题,即对某一个摄像头中的数据分别在其他每个摄像头下找到最相似的人.然而这样的方式存在一个问题,即当一个人在某个摄像头中没有出现的时候,该方法会强制性选择一个错误的样本进行关联.

## 4 数据集和评价标准

为了评估行人重识别的相关算法,我们往往需要在一些公开数据集上进行实验,并通过统一的评价标准来评估所提出方法的性能.在该部分,我们对行人重识别的相关数据集和评价标准进行了总结.

### 4.1 数据集

近年来行人重识别问题在科研中的关注逐渐得到提高,因此也出现了越来越多的更大规模的数据集.这些数据集主要分为两类,即基于图像的数据集和基于视频的数据集.我们将选取部分常用的数据集分别进行介绍.

#### (1) 图像数据集.

常用的图像数据集主要包括 Market1501<sup>[88]</sup>、DukeMTMC-reID<sup>[32]</sup>、MSMT17<sup>[65]</sup>和 CUHK03<sup>[89]</sup>,这些数据集既可以用于无监督任务,也可以用于半监督任务.其基本信息概括在表 1 中.

**Table 1** Information of some image-based person re-identification datasets

表 1 部分行人重识别图像数据集信息

数据集	ID 数量	训练集 ID 数量	训练集图像数量	测试集 ID 数量	测试集图像数量	摄像头数量
Market1501 <sup>[88]</sup>	1501	751	12936	750	16384	6
DukeMTMC-ReID <sup>[32]</sup>	1404	702	16522	702	17661	8
MSMT17 <sup>[65]</sup>	4101	1041	32621	3060	82161	15
CUHK03-NP <sup>[93]</sup> (labeled)	1467	767	7368	700	5328	10 (5 pairs)
CUHK03-NP <sup>[93]</sup> (detected)	1467	767	7365	700	5332	10 (5 pairs)

Market-1501<sup>[88]</sup>是在大学校园内一个超市前面采集的,由 6 个摄像头拍摄得到.该数据集包含 1501 个行人的 32668 张图像(标注框),且每个行人都至少在两个摄像头中出现.训练集包括 751 个行人的 12936 张图像,查询集包括 750 个行人的 3368 张图像,测试集包括 750 个行人的 16384 张图像.该数据集采用 DPM 检测器<sup>[90]</sup>来检测行人标注框,而非采用人工裁剪的方式来获得,这样更加贴近现实应用,即可以考虑到行人检测的标注框会存在偏移与不对齐的情况.DukeMTMC-reID<sup>[32]</sup>是基于多摄像头多目标行人跟踪数据集 DukeMTMC<sup>[91]</sup>构建而成的,由 8 个摄像头拍摄得到.该数据集包含 1404 个行人的 36411 张图像(标注框),且该 1404 个行人中每个行人都至少在两个摄像头中出现.训练集包含 702 个行人的 16522 张图像,查询集包括 702 个行人的 2228 张图像,测试集包括 702 个行人的 17661 张图像.该数据集通过人工剪裁的方式来获得行人标注框.MSMT17<sup>[65]</sup>是从校园中部署的 15 个摄像头中拍摄得到的,包含 4101 个行人的 126441 张图像.训练集包含 1041 个行人的 32621 张图像(其中包括验证集的 2373 张图像),查询集包含 3060 个行人的 11659 张图像,测试集包含 3060 个行人的 82161 张图像.该数据集采用 Faster RCNN 检测器<sup>[92]</sup>来检测行人标注框.CUHK03<sup>[89]</sup>由 5 组摄像头拍摄得到,每组摄像头包含两个摄像头,且采用人工裁剪(labeled)和 DPM 检测器<sup>[90]</sup>(detected)两种方

式来检测行人标注框.该数据集存在两种测试协议,本文仅介绍新的一种协议,以下称为 CUHK03-NP<sup>[93]</sup>.在 CUHK03-NP(labeled)中,训练集包含 767 个行人的 7368 张图像,查询集包含 700 个行人的 1400 张图像,测试集包含 700 个行人的 5328 张图像;在 CUHK03-NP(detected)中,训练集包含 767 个行人的 7365 张图像,查询集包含 700 个行人的 1400 张图像,测试集包含 700 个行人的 5332 张图像.图 8 展示了部分数据集的实例图像,其中左图来自 Market-1501<sup>[88]</sup>,右图来自 DukeMTMC-reID<sup>[32]</sup>,上下两行分别代表不同摄像头下的行人图像.



Fig.8 Examples from the image-based datasets

图 8 部分图像数据集的示例图像

(2) 视频数据集.

常用的视频数据集除了较早出现的 PRID2011<sup>[94]</sup>和 iLIDS-VID<sup>[95]</sup>以外,主要包括 MARS<sup>[96]</sup>·DukeMTMC-SI-Tracklet<sup>[84]</sup>和 DukeMTMC-VideoReID<sup>[97]</sup>,这些数据集常用于半监督任务.其基本信息概括在表 2 中.

Table 2 Information of some video-based person re-identification datasets

表 2 部分行人重识别视频数据集信息

数据集	ID 数量	训练集 ID 数量	测试集 ID 数量	图像数量	tracklet 数量	摄像头数量
MARS <sup>[96]</sup>	1261	625	636	71191003	20478	6
DukeMTMC-SI-Tracklet <sup>[84]</sup>	1788	702	1086	833984	12647	8
DukeMTMC-VideoReID <sup>[97]</sup>	1812	702	702(and 408 distractors)	815420	4832	8

MARS<sup>[96]</sup>是在大学校园中的 6 个摄像头采集得到的,包含 1261 个行人的 20478 个 tracklet 共计 1191003 张图片,分别将 626 和 635 个行人作为训练集和测试集,其所有的轨迹片段都是由 DPM 检测器<sup>[90]</sup>和 GMMCP 跟踪器<sup>[98]</sup>自动生成的.DukeMTMC-SI-Tracklet<sup>[84]</sup>和 DukeMTMC-VideoReID<sup>[97]</sup>均来自来自 DukeMTMC<sup>[91]</sup>,由 8 个摄像头进行拍摄,人工裁剪得到标注框.DukeMTMC-SI-Tracklet 由 1788 个行人的 19,135 个 tracklet 共计 833984 张图片组成,并分别将 702 个和 1086 个行人作为训练集和测试集;DukeMTMC-VideoReID 由 1812 个行人的 4832 个 tracklet 共计 815420 张图片组成,并分别将 702 个、702 个和 408 个行人作为训练集、测试集和干扰项.图 9 展示了部分数据集的实例图像,均来自 MARS<sup>[96]</sup>,其中上下两行分别代表不同的 tracklet.



Fig.9 Examples from the video-based datasets

图9 部分视频数据集的示例图像

## 4.2 评价标准

对于行人重识别算法的性能,通常使用累积匹配特性(Cumulative Match Characteristic, CMC)曲线和平均精度均值(mean Average Precision, mAP)来进行评估。

CMC 曲线能够综合反映分类器的性能,可以表示匹配目标出现在大小为  $k$  的候选列表中的概率.直观上,CMC 曲线可以通过 Rank- $k$  准确率的形式给出,即目标正确匹配出现在匹配列表前  $k$  位的概率.在行人重识别问题中,通常关注  $k = \{1, 5, 10, 20\}$  时的性能,即匹配目标的  $k = \{1, 5, 10, 20\}$  准确率.例如 Rank-1 准确率表示正确匹配出现在匹配列表第 1 位的概率,即查找 1 次即可返回正确匹配的概率.通常最后的 Rank- $k$  准确率是指对所有检索目标进行查询后取结果的平均值。

然而当测试集中存在多个正确匹配时, Rank- $k$  准确率不能完整地算法进行评估.Zheng 等人<sup>[88]</sup>考虑到行人重识别的目标应将所有的正确匹配都检索出来,即在考虑查准率的同时,应当同时考虑查全率,因此建议采用 mAP 来将算法的检索召回能力考虑进去.具体地,mAP 的计算过程需遍历所有检索目标,对于每个检索目标分别计算 AP(Average Precision)并取平均,而 AP 的计算过程即为求 PR(Precision-Recall)曲线下的面积的过程,即考虑了目标在某些阈值下的查准率和查全率.因此在后续工作中,通常将 mAP 与 Rank- $k$  准确率结合在一起作为行人重识别问题的评价指标,这样能够达到对算法性能进行全面评价的目标。

## 5 现有方法的性能及分析

在该部分,我们对现有弱监督场景下行人重识别算法的实验结果进行总结,并给出了分析和比较。

### 5.1 无监督方法

对于现有的无监督方法,我们总结了当前基于深度学习的方法在 3 个大规模数据集上的实验结果,即 Market1501<sup>[88]</sup>、DukeMTMC-reID<sup>[32]</sup>和 MSMT17<sup>[65]</sup>.其中包括基于伪标记的方法,如 TJ-AIDL<sup>[46]</sup>、TFusion-uns<sup>[47]</sup>、DC<sup>[55]</sup>、HCR<sup>[99]</sup>、BUC<sup>[52]</sup>、PAUL<sup>[44]</sup>、MAR<sup>[43]</sup>、PCB-R-PAST<sup>[50]</sup>、SSG<sup>[48]</sup>、ISSDA<sup>[53]</sup>和 ACT<sup>[54]</sup>;基于图像生成的方法,如 HHL<sup>[59]</sup>、SyRI<sup>[61]</sup>、PTGAN<sup>[65]</sup>、SPGAN<sup>[63]</sup>、ATNet<sup>[58]</sup>、DA-2S<sup>[56]</sup>和 CR-GAN<sup>[57]</sup>;基于实例分类的方法,如 ECN<sup>[68]</sup>、AE<sup>[70]</sup>和 LAIM<sup>[69]</sup>;基于领域自适应的方法,如 MMFA<sup>[71]</sup>、CAT<sup>[73]</sup>和 UCDA<sup>[74]</sup>.实验结果总结在表 3、表 4 和表 5 中.表 3 展示了现有无监督的方法在 Market-1501 数据集上的结果,其中 \* 表示没有使用有标记的源域数据进行模型的预训练(即直接使用 ImageNet 预训练的模型),† 表示使用除 DukeMTMC-reID、MSMT17 和 CUHK03 之外的行人重识别数据集进行模型的预训练,- 表示没有对应的实验结果,DukeMTMC-reID/MSMT17/CUHK03 表示分别使用这 3 个数据集作为源域的实验结果;表 4 展示了现有无监督的方法在 DukeMTMC-reID 数据集上的结果,其中 \* 表示没有使用有标记的源域数据进行模型的预训练(即直接使用 ImageNet 预训练的模型),- 表示没有对应的实验结果,Market-1501/MSMT17/CUHK03 表示分别使用这 3 个数据集作为源域的实验结果;表 5 展示了现有无监督的方法在 MSMT17 数据集上的结果,其中 \* 表示没有使用有标记的源域数据进行模型的预训练(即直接使用 ImageNet 预训练的模型),- 表示没有对应的实验结果,Market-1501/DukeMTMC-reID/CUHK03 表示分别使用这 3 个数据集作为源域的实验结果。

对于基于伪标记的方法,TJ-AIDL<sup>[46]</sup>是通过结合属性学习并对目标域中的数据生成伪属性的方法来进行学习的,TFusion-uns<sup>[47]</sup>利用时序信息产生更可靠的伪标记信息,PAUL<sup>[44]</sup>和 MAR<sup>[43]</sup>以有标记的源域数据为基准生成无标记目标域的伪标记信息,DC<sup>[55]</sup>、HCR<sup>[99]</sup>、BUC<sup>[52]</sup>、PCB-R-PAST<sup>[50]</sup>、SSG<sup>[48]</sup>、ISSDA<sup>[53]</sup>和 ACT<sup>[54]</sup>都是基于聚类的算法,其中 ACT 对聚类后的结果进行了进一步处理以便于找出确定性的伪标记信息和非确定性的伪标记信息,因此相对于其他方法,该方法能够得到相对更好的结果.另外从表 3 和表 4 中可以发现,基于聚类的伪标记方法相对于其他伪标记的方法有更好的性能。

对于基于图像生成的方法,当前 CR-GAN<sup>[57]</sup>的性能最优.不同于其他从领域层级或摄像头层级的风格迁

移,如 PTGAN<sup>[65]</sup>、SPGAN<sup>[63]</sup>和 ATNet<sup>[58]</sup>等,CR-GAN 是基于图像层级的风格转化,即根据一张特定图像的风格对一张目标的图像进行图像风格的迁移.因此该方法相对于其他方法有更好的性能.此外在所有这些基于图像生成的方法中,HHL<sup>[59]</sup>只使用了图像生成网络对目标域内的不同摄像头间的图像进行转化,以产生不同摄像头风格的正样本对.

**Table 3** Results of existing unsupervised methods on Market-1501

**表 3** 现有无监督的方法在 Market-1501 数据集上的结果

DukeMTMC-reID/MSMT17/CUHK03->Market-1501						
方法	mAP	Rank-1	Rank-5	Rank-10	Rank-20	发表
TJ-AIDL <sup>[46]</sup>	26.5/-/-	58.2/-/-	74.8/-/-	81.1/-/-	86.5/-/-	CVPR2018
†TFusion-uns <sup>[47]</sup>	—	60.8	74.4	79.3	—	CVPR2018
*DC <sup>[55]</sup>	41.3	69.2	83.0	87.8	—	arxiv 2019
*HCR <sup>[99]</sup>	55.3	79.5	91.7	<b>95.7</b>	—	arxiv 2019
*BUC <sup>[52]</sup>	38.3	66.2	79.6	84.5	—	AAAI2019
PAUL <sup>[44]</sup>	36.8/40.1/-	66.7/68.5/-	-/82.4/-	-/87.4/-	—	CVPR2019
MAR <sup>[43]</sup>	-/40.0/-	-/67.7/-	-/81.9/-	-/87.3/-	—	CVPR2019
PCB-R-PAST <sup>[50]</sup>	54.6/-/57.3	78.4/-/79.5	—	—	—	ICCV2019
SSG <sup>[48]</sup>	58.3/-/-	80.0/-/-	90/-/-	92.4/-/-	—	ICCV2019
ISSDA <sup>[53]</sup>	63.1/-/-	81.3/-/-	<b>92.4/-/-</b>	95.2/-/-	—	CVPRW2019
ACT <sup>[54]</sup>	60.6/-/ <b>64.1</b>	80.5/-/ <b>81.2</b>	—	—	—	AAAI 2020
HHL <sup>[59]</sup>	31.4/-/29.8	62.2/-/56.8	78.8/-/74.7	84.0/-/ 81.4	<b>88.3/-/86.3</b>	ECCV2018
*SyRI <sup>[61]</sup>	—	65.7	—	—	—	ECCV2018
PTGAN <sup>[65]</sup>	—	—	—	—	—	CVPR2018
SPGAN <sup>[63]</sup>	22.8/-/-	51.5/-/-	70.1/-/-	76.8/-/-	82.4/-/-	CVPR2018
SPGAN+LMP <sup>[63]</sup>	26.9/-/-	58.1/-/-	76.0/-/-	82.7/-/-	87.9/-/-	CVPR2018
ATNet <sup>[58]</sup>	25.6/-/-	55.7/-/-	73.2/-/-	79.4/-/-	84.5/-/-	CVPR2019
DA-2S <sup>[56]</sup>	27.3/-/28.5	58.5/-/57.6	—	—	—	ICCV2019
CR-GAN <sup>[57]</sup>	29.6/-/30.4	59.6/-/58.5	-/75.8	/-/ 81.9	—	ICCV2019
CR-GAN+LMP <sup>[57]</sup>	<b>33.2/-/-</b>	<b>64.5/-/-</b>	<b>79.8/-/-</b>	<b>85.0/-/-</b>	—	ICCV2019
ECN <sup>[68]</sup>	43/-/-	75.1/-/-	87.6/-/-	91.6/-/-	94.5/-/-	CVRP2019
*AE <sup>[70]</sup>	54.0	77.5	89.8	93.4	—	arxiv 2019
AE <sup>[70]</sup>	58.0/-/-	81.6/-/-	91.9/-/-	94.6/-/-	—	arxiv 2019
LAIM <sup>[69]</sup>	<b>63.8/-/-</b>	<b>84.1/-/-</b>	<b>92.8/-/-</b>	<b>95.4/-/-</b>	<b>96.9/-/-</b>	arxiv 2019
MMFA <sup>[71]</sup>	27.4/-/-	56.7/-/-	—	—	—	BMVC2018
CAT+LSRO+LMP <sup>[73]</sup>	27.8/-/-	57.8/-/-	—	—	—	arxiv 2019
UCDA <sup>[74]</sup>	<b>34.5/-/-</b>	<b>64.3/-/-</b>	—	—	—	ICCV2019
CASCL <sup>[75]</sup>	35.6/35.5/-	64.7/65.4/-	80.2/80.6/-	85.6/86.2/-	—	ICCV2019

对于基于实例分类的方法,当前方法基本都是关注在如何建立样本间的关系上,其中 LAIM<sup>[69]</sup>引入了图的关系来增强找出相同实例的可靠性,因此该方法目前在此类型的方法中具有较好的表现.

对于基于领域自适应的方法,MMFA<sup>[71]</sup>使用传统的 MMD 方法来减少领域间差异.在行人重识别问题中,数据分布的差异不仅存在于领域间,也存在于相同领域的不同摄像头间,而 CAT<sup>[73]</sup>和 UCDA<sup>[74]</sup>都考虑到了这一方面,提出了基于摄像头感知的领域对抗学习.对比于 CAT, UCDA 提出了一种跨领域等视角的对抗学习方

法来减少所有摄像头视角层级的数据分布的差异,并且利用了时序信息在无标记的目标域中挖掘判别性信息,以保证在减少数据分布差异的过程中目标域数据内部结构的不变性.因此在该类方法中,UCDA 目前具有最好的性能.

对比这几大类方法,基于图像生成的方法和基于领域自适应的方法相较于基于伪标记的方法和基于实例分类的方法表现性能相对较弱.主要原因可能是基于图像生成或领域自适应的方法可以看作是从数据分布的层级来解决无标记的学习问题,其中基于图像生成的方法旨在缩小源域和目标域图像分布间的差异,而基于领域自适应的方法是从特征表示的层级来缩小小领域间的差异.这些方法属于隐式地解决无标记的问题,而基于伪标记或实例分类的方法是显式地通过对无标记的数据直接产生伪标记或者是建立这些样本间的关联来进行学习.因此,基于伪标记和实例分类的方法相比较于其他方法能够更加直接地解决无标记行人重识别问题.

**Table 4** Results of existing unsupervised methods on DukeMTMC-reID

**表 4** 现有无监督的方法在 DukeMTMC-reID 数据集上的结果

Market-1501/MSMT17/CUHK03->DukeMTMC-reID						
方法	mAP	Rank-1	Rank-5	Rank-10	Rank-20	发表
TJ-AIDL <sup>[46]</sup>	23.0/-/	44.3/-/	59.6/-/	65/-/	<b>70.0/-/</b>	CVPR2018
*DC <sup>[55]</sup>	30.0	51.5	64.6	70.1	—	arxiv 2019
*HCR <sup>[99]</sup>	46.8	66.0	78.6	83.4	—	arxiv 2019
*BUC <sup>[52]</sup>	27.5	47.4	62.6	68.4	—	AAAI2019
PAUL <sup>[44]</sup>	35.7/53.2/-	56.1/72.0/-	<b>-/82.7/-</b>	<b>-/86.0/-</b>	—	CVPR2019
MAR <sup>[43]</sup>	<b>-/48.0/-</b>	<b>-/67.1/-</b>	<b>-/79.8/-</b>	<b>-/84.2/-</b>	—	CVPR2019
PCB-R-PAST <sup>[50]</sup>	54.3/-/51.8	72.4/-/69.9	—	—	—	ICCV2019
SSG <sup>[48]</sup>	53.4/-/	<b>73.0/-/</b>	80.6/-/	83.2/-/	—	ICCV2019
ISSDA <sup>[53]</sup>	54.1/-/	72.8/-/	82.9/-/	85.9/-/	—	CVPRW2019
ACT <sup>[54]</sup>	<b>54.5/-/35.4</b>	72.4/-/52.8	—	—	—	AAAI 2020
HHL <sup>[59]</sup>	27.2/-/23.4	46.9/-/42.7	61/-/57.5	66.7/-/64.2	71.9/-/69.1	ECCV2018
SPGAN <sup>[63]</sup>	22.3/-/	41.4/-/	56.6/-/	63.0/-/	69.6/-/	CVPR2018
SPGAN+LMP <sup>[63]</sup>	26.4/-/	46.9/-/	62.6/-/	68.5/-/	<b>74.0/-/</b>	CVPR2018
ATNet <sup>[58]</sup>	24.9/-/	45.1/-/	59.5/-/	64.2/-/	70.1/-/	CVPR2019
DA-2S <sup>[56]</sup>	30.8/-/27.8	53.5/-/47.7	—	—	—	ICCV2019
CR-GAN <sup>[57]</sup>	30.0/-/26.9	52.2/-/46.5	<b>-/61.6</b>	<b>-/67.0</b>	—	ICCV2019
CR-GAN+LMP <sup>[57]</sup>	<b>33.3/-/</b>	<b>56.0/-/</b>	<b>70.5/-/</b>	<b>74.6/-/</b>	—	ICCV2019
ECN <sup>[68]</sup>	40.4/-/	63.3/-/	75.8/-/	80.4/-/	84.2/-/	CVRP2019
*AE <sup>[70]</sup>	39.0	63.2	75.4	79.4	—	arxiv 2019
AE <sup>[70]</sup>	46.7/-/	67.9/-/	79.2/-/	83.6/-/	—	arxiv 2019
LAIM <sup>[69]</sup>	<b>54.4/-/</b>	<b>74.0/-/</b>	<b>83.7/-/</b>	<b>87.4/-/</b>	<b>90.0/-/</b>	arxiv 2019
MMFA <sup>[71]</sup>	24.7/-/	45.3/-/	—	—	—	BMVC2018
CAT+LSRO+LMP <sup>[73]</sup>	28.7/-/	50.9/-/	—	—	—	arxiv 2019
UCDA <sup>[74]</sup>	<b>36.7/-/</b>	<b>55.4/-/</b>	—	—	—	ICCV2019
CASCL <sup>[75]</sup>	30.5/37.8/-	51.5/59.3/-	66.7/73.2/-	71.7/77.8/-	—	ICCV2019

5.2 半监督方法

在这部分,我们总结了当前基于深度学习的半监督行人重识别算法在 3 个大规模图像数据集 Market1501<sup>[88]</sup>、DukeMTMC-reID<sup>[32]</sup>和 MSMT17<sup>[65]</sup>以及 3 个大规模视频数据集 MARS<sup>[96]</sup>、DukeMTMC-VideoReID<sup>[84]</sup>和 DukeMTMC-SI-Tracklet<sup>[97]</sup>上的实验结果.特别地,当前在行人重识别问题半监督的定义有很多种,该部分总结的方法包括:1)少部分人有标记的场景,如 Distilled-ReID<sup>[80]</sup>和 MVC<sup>[81]</sup>;2)每一个人有少量标记的场景,如 One-Example<sup>[82]</sup>;3)基于 tracklet 的场景,例如 TAUDL<sup>[83]</sup>、UTAL<sup>[84]</sup>、TSSL<sup>[100]</sup>、TASTR<sup>[101]</sup>和 UGA<sup>[75]</sup>;4)摄像头内有标记但摄像头间无标记的场景,如 ACAN<sup>[85]</sup>、MTML<sup>[87]</sup>和 PCSL<sup>[86]</sup>.所有方法的实验结果总结在表 6 和表 7 中.表 6 展示了现有的半监督方法在图像数据集 Market1501、DukeMTMC-ReID 和 MSMT17 上的结果,其中 - 表示没有对应的实验结果;表 7 展示了现有的半监督方法在视频数据集 MARS、DukeMTMC-VideoReID 和 DukeMTMC-SI-Tracklet 上的结果,其中 - 表示没有对应的实验结果.特别地,与大部分无监督方法会涉及到源域和目标域的数据集不同的是,半监督的方法中只有目标域的数据集.

Table 5 Results of existing unsupervised methods on MSMT17

表 5 现有无监督的方法在 MSMT17 数据集上的结果

Market-1501/DukeMTMC-reID/CUHK03->MSMT17					
方法	mAP	Rank-1	Rank-5	Rank-10	发表
SSG <sup>[48]</sup>	13.2/13.3/-	31.6/32.2/-	—	49.6/51.2/-	ICCV2019
PTGAN <sup>[65]</sup>	2.9/3.3/1.7	10.2/11.8/6.5	—	24.4/27.4/17.2	CVPR2018
ECN <sup>[68]</sup>	8.5/10.2/-	25.3/30.2/-	36.3/41.5/-	42.1/46.8/-	CVPR2019
*AE <sup>[70]</sup>	8.5	26.6	37	41.7	arxiv 2019
AE <sup>[70]</sup>	9.2/11.7/-	25.5/32.3/-	37.3/44.4/-	42.6/50.1/-	arxiv 2019
LAIM <sup>[69]</sup>	15.2/ <b>16.0</b> -	40.4/ <b>42.5</b> -	53.1/ <b>55.9</b> -	58.7/ <b>61.5</b> -	arxiv 2019

Table 6 Results of existing semi-supervised methods on image-based datasets

表 6 现有的半监督方法在图像数据集上的结果

方法	Market-1501					DukeMTMC-ReID					MSMT17					发表
	mAP	R-1	R-5	R-10	R-20	mAP	R-1	R-5	R-10	R-20	mAP	R-1	R-5	R-10	R-20	
Distilled-ReID <sup>[80]</sup>	35.4	63.7	—	—	—	36.7	57.4	—	—	—	—	—	—	—	—	CVPR2019
MVC <sup>[81]</sup>	52.6	75.2	—	—	—	37.8	57.6	—	—	—	—	—	—	—	—	PR2019
One-Example <sup>[82]</sup>	26.2	55.8	72.3	78.4	83.5	28.5	48.8	63.4	68.4	73.1	—	—	—	—	—	TIP2019
TAUDL <sup>[83]</sup>	41.2	63.7	—	—	—	43.5	61.7	—	—	—	—	—	—	—	—	ECCV2018
UTAL <sup>[84]</sup>	46.2	69.2	—	—	—	44.6	62.3	—	—	—	13.1	31.4	—	—	—	TPAMI2019
TSSL <sup>[100]</sup>	43.3	71.2	—	—	—	38.5	62.2	—	—	—	—	—	—	—	—	AAAI2020
TASTR <sup>[101]</sup>	—	—	—	—	—	54.9	74.1	<b>85.5</b>	<b>89.0</b>	<b>91.8</b>	—	—	—	—	—	arxiv2019
UGA <sup>[75]</sup>	<b>70.3</b>	<b>87.2</b>	—	—	—	<b>53.3</b>	<b>75.0</b>	—	—	—	<b>21.7</b>	<b>49.5</b>	—	—	—	ICCV2019
ACAN-GRL <sup>[85]</sup>	50.6	73.3	87.6	91.8	—	46.6	65.1	80.6	85.1	—	11.2	27.1	40.9	47.3	—	arxiv2019
ACAN-OCE <sup>[85]</sup>	47.7	72.2	86.3	90.4	—	45.1	67.6	81.2	85.2	—	12.6	33.0	48.0	54.7	—	arxiv2019
MTML <sup>[87]</sup>	65.2	85.3	—	96.2	<b>97.6</b>	50.7	<b>71.7</b>	—	86.9	<b>89.6</b>	18.6	44.1	—	63.9	<b>70.0</b>	ICCVW2019
PCSL-C <sup>[86]</sup>	<b>69.4</b>	<b>87.0</b>	<b>94.8</b>	<b>96.6</b>	—	50.0	70.6	82.4	85.7	—	<b>20.7</b>	<b>48.3</b>	<b>62.8</b>	<b>68.6</b>	—	arxiv2019
PCSL-D <sup>[86]</sup>	65.4	82.0	92.2	95.1	—	<b>53.5</b>	<b>71.7</b>	<b>84.7</b>	<b>88.2</b>	—	20.5	45.1	61.0	67.6	—	arxiv2019

对于少量的人有标记的场景, MVC<sup>[81]</sup>和 Distilled-ReID<sup>[80]</sup>的设定并不相同,因此它们并不具有可比性. 对于每一个人有少量标记的场景,在现实应用中该方法并不是可行的,这是因为如果需要知道整个数据集行人的数量,就必须对整个数据集进行标注,即这种设定不具有现实应用价值,因此针对该类场景的研究方法较少. 对于基于 tracklet 的场景,本文中我们将其归类为半监督的方法,因为 tracklet 中的每个图像默认为同一个标记. 特别地,一些方法在基于图像的行人重识别数据集中,假定每个摄像头内的人的所有图像在一个 tracklet 内,即对于这些数据集已经给定了摄像头内的标记信息. 这类方法在近年来得到了较多的关注,一些研究工作主要关注在如何在摄像头内获得不重复的 tracklet 上,例如使用时序信息来缓解该问题,除此之外如何建立跨摄像头间的联系也是该类方法需重点解决的问题. 由于这类方法采用了基于 tracklet 的标记信息,因此该类方法相对于无监督的行人重识别方法,整体来看具有更好的性能. 进一步, Qi 等人<sup>[85]</sup>基于 tracklet 的场景定义了一种新的半监督场景的学习方式,即摄像头内有标记而摄像头间无标记的场景. 相对于基于 tracklet 学习方法,该方法在基于视频的图像数据集上具有更好的性能,主要原因在于基于 tracklet 的方法通过采样的方法,并未完全使用摄像头的数据,而基于摄像头内给定标记的场景能够有效地利用所有的数据,并且摄像头内的标记并不需要大量的人工成本,因此该类方法在现实中具有重要的研究意义. 另外, ACAN<sup>[85]</sup>基于数据分布的视角来解决跨摄像头间无标记的问题,而 MTML<sup>[87]</sup>和 PCSL<sup>[86]</sup>直接采用关联的方法来建立跨摄像头样本间的关联性. 从实验结果来看,直接建立样本间的关联性相较于从数据分布的视角解决跨摄像头间无标记的问题具有更好的性能.

**Table 7** Results of existing semi-supervised methods on video-based datasets

**表 7** 现有的半监督方法在视频数据集上的结果

方法	MARS					DukeMTMC-VideoReID					DukeMTMC-SI-Tracklet					发表
	mAP	R-1	R-5	R-10	R-20	mAP	R-1	R-5	R-10	R-20	mAP	R-1	R-5	R-10	R-20	
One-Example <sup>[82]</sup>	42.6	62.8	75.2	<b>80.4</b>	83.8	63.3	72.9	84.3	<b>88.3</b>	91.4	—	—	—	—	—	TIP2019
TAUDL <sup>[83]</sup>	29.1	43.8	59.9	—	72.8	—	—	—	—	—	20.8	26.1	42.0	—	57.2	ECCV2018
UTAL <sup>[84]</sup>	35.2	49.9	66.4	—	77.8	<b>72.1</b>	<b>74.5</b>	<b>88.7</b>	—	<b>96.3</b>	36.6	43.8	62.8	—	76.5	TPAMI2019
TSSL <sup>[100]</sup>	30.5	56.3	—	—	—	64.6	73.9	—	—	—	—	—	—	—	—	AAAI2020
UGA <sup>[75]</sup>	<b>39.3</b>	<b>58.1</b>	<b>73.4</b>	—	<b>81.4</b>	—	—	—	—	—	—	—	—	—	—	ICCV2019
ACAN-GRL <sup>[85]</sup>	49.1	59.2	77.1	—	86.7	—	—	—	—	—	43.0	52.0	71.0	—	82.0	arxiv2019
ACAN-OCE <sup>[85]</sup>	47.5	57.7	75.1	—	84.0	—	—	—	—	—	40.3	50.4	68.0	—	81.2	arxiv2019
PCSL-C <sup>[86]</sup>	<b>62.2</b>	<b>70.5</b>	<b>87.0</b>	—	<b>94.2</b>	—	—	—	—	—	37.3	47.3	65.4	—	75.9	arxiv2019
PCSL-D <sup>[86]</sup>	57.5	65.6	83.1	—	91.9	—	—	—	—	—	<b>43.1</b>	<b>52.3</b>	<b>72.3</b>	—	<b>84.4</b>	arxiv2019

## 6 总结

本文主要总结了弱监督场景下的行人重识别算法,包括无监督场景和半监督场景,并且对近年的方法进行了分类和描述. 对于无监督的行人重识别算法,我们根据其技术类型划分为五类,分别为基于伪标记的方法、基于图像生成的方法、基于实例分类的方法、基于领域自适应的方法和其他类型的方法. 对于半监督的行人重识别方法,根据其场景类型划分为四类,分别为少量的人有标记的场景、每一个人有少量标记的场景、基于 tracklet 学习的场景和摄像头内有标记但摄像头间无标记的场景. 最后,我们对当前行人重识别的相关数据集进行总结,并对现有的弱监督方法的实验结果进行总结与分析.

研究弱监督场景下的行人重识别问题,能够帮助行人重识别技术更好地拓展到现实应用中. 而基于弱监督场景下的行人重识别算法,其着重研究利用无标记或少量标记的数据来学习具有更好泛化性能的模式. 对该领域的探索不仅具有理论价值,还有很高的应用价值. 该领域虽然在近年来得到了一定的关注,但目前仍然不能完全达到有监督场景下的性能. 该领域仍然有一些研究问题亟待解决:

### (1) 实例间的关系评估.

基于实例分类的无监督方法在近年来得到了广泛的关注,但其主要的挑战集中在如何有效地挖掘每个样本之间的实际关系,即以一对样本而言观察它们是否属于相同的类别.如果所有样本之间的关系能够被很好地评估,那么这一类方法的性能将等价于有监督场景下的行人重识别任务的性能.

### (2) 领域泛化问题.

虽然弱监督场景下的行人重识别算法相比于传统的有监督场景更能够有利于应用到现实当中,然而这些方法仍然需要收集无标记的样本来学习.在将来,通用性的行人重识别算法也许是该领域能够真正落地的一大发展趋势,即只通过在现有的数据进行训练就能够很好地泛化到其他未见场景中.这也是实现通用人工智能技术的必要的一条路.我们首先需要解决单一任务上的通用型,才能进一步去探索在不同任务上的通用性.这一类问题结合风格迁移和元学习的相关方法或许将在未来的研究中展现出很大的前景.

## References:

- [1] Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. *IEEE Conference on Computer Vision and Pattern Recognition*, Ieee, 2009: 248-255.
- [2] 杨涛,李静,潘泉,张艳宁.基于场景模型与统计学习的鲁棒行人检测算法. *自动化学报*, 2010, 36(4):499-508.
- [3] 郭立君,刘曦,赵杰煜,史忠植.结合运动信息与表观特征的行人检测方法. *软件学报*, 2012, 23(2):299-309.
- [4] 刘威,段成伟,遇冰,柴丽颖,袁淮,赵宏.基于后验 HOG 特征的多姿态行人检测. *电子学报*, 2015, 43(2):217-224.
- [5] 高君宇,杨小汕,张天柱,徐常胜.基于深度学习的鲁棒性视觉跟踪方法. *计算机学报*, 2016, (7):1419-1434.
- [6] 杜宇宁,艾海舟.基于二次相似度函数学习的行人再识别. *计算机学报*, 2016, (8):1639-1651.
- [7] 桑海峰,王传正,吕应宇,何大阔,刘晴.基于多信息流动卷积神经网络的行人再识别. *电子学报*, 2019, 47(02):97-103.
- [8] Matsukawa T, Okabe T, Suzuki E, Sato Y. Hierarchical gaussian descriptor for person re-identification. *IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 1363-1372.
- [9] Liao SC, Hu Y, Zhu XY, Li SZ. Person re-identification by local maximal occurrence representation and metric learning. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015: 2197-2206.
- [10] Yang Y, Yang JM, Yan JJ, Liao SC, Yi D, Li SZ. Salient color names for person re-identification. *European conference on computer vision*, Springer, Cham, 2014: 536-551.
- [11] Bazzani L, Cristani M, Murino V. Symmetry-driven accumulation of local features for human characterization and re-identification. *Computer Vision and Image Understanding*, 2013, 117(2): 130-144.
- [12] Gray D, Tao H. Viewpoint invariant pedestrian recognition with an ensemble of localized features. *European conference on computer vision*, Springer, Berlin, Heidelberg, 2008: 262-275.
- [13] Paisitkriangkrai S, Shen CH, Van Den Hengel A. Learning to rank in person re-identification with metric ensembles. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 1846-1855.
- [14] Liao SC, Li SZ. Efficient psd constrained asymmetric metric learning for person re-identification. *Proceedings of the IEEE International Conference on Computer Vision*, 2015: 3685-3693.
- [15] Xiong F, Gou M, Camps O, Szaier M. Person re-identification using kernel-based metric learning methods. *European conference on computer vision*, Springer, Cham, 2014: 1-16.
- [16] Liu CX, Change Loy C, Gong SG, Wang GJ. Pop: Person re-identification post-rank optimization. *Proceedings of the IEEE International Conference on Computer Vision*, 2013: 441-448.
- [17] Karanam S, Li Y, Radke RJ. Person re-identification with discriminatively trained viewpoint invariant dictionaries. *Proceedings of the IEEE International Conference on Computer Vision*, 2015: 4516-4524.
- [18] Zhang L, Xiang T, Gong SG. Learning a discriminative null space for person re-identification. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016: 1239-1248.
- [19] Chen DP, Yuan ZJ, Hua G, Zheng NN, Wang JD. Similarity learning on an explicit polynomial kernel feature map for person re-identification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 1565-1573.

- [20] Chen BH, Deng WH, Hu JN. Mixed high-order attention network for person re-identification. Proceedings of the IEEE International Conference on Computer Vision, 2019: 371-381.
- [21] Xia BN, Gong Y, Zhang YZ, Poellabauer C. Second-order non-local attention networks for person re-identification. Proceedings of the IEEE International Conference on Computer Vision, 2019: 3760-3769.
- [22] Fang PF, Zhou JM, Roy SK, Petersson L, Harandi M. Bilinear attention networks for person retrieval. Proceedings of the IEEE International Conference on Computer Vision, 2019: 8030-8039.
- [23] Chen TL, Ding SJ, Xie JY, Yuan Y, Chen WY, Yang Y, Ren Z, Wang ZY. Abd-net: Attentive but diverse person re-identification. Proceedings of the IEEE International Conference on Computer Vision, 2019: 8351-8361.
- [24] Chen GY, Lin CZ, Ren LL, Lu JW, Zhou J. Self-critical attention learning for person re-identification. Proceedings of the IEEE International Conference on Computer Vision, 2019: 9637-9646.
- [25] Tay CP, Roy S, Yap KH. AANet: Attribute attention network for person re-identifications. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 7134-7143.
- [26] Zheng M, Karanam S, Wu ZY, Radke RJ. Re-identification with consistent attentive siamese networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 5735-5744.
- [27] Li W, Zhu XT, Gong SG. Harmonious attention network for person re-identification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 2285-2294.
- [28] Si JL, Zhang HG, Li CG, Kuen J, Kong XF, Kot AC, Wang G. Dual attention matching network for context-aware feature sequence based person re-identification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 5363-5372.
- [29] Hermans A, Beyer L, Leibe B. In defense of the triplet loss for person re-identification. arXiv preprint arXiv:1703.07737, 2017.
- [30] Zhang YY, Zhong QY, Ma L, Xie D, Pu SL. Learning incremental triplet margin for person re-identification. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33: 9243-9250.
- [31] Chen WH, Chen XT, Zhang JG, Huang KQ. Beyond triplet loss: a deep quadruplet network for person re-identification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 403-412.
- [32] Zheng ZD, Zheng L, Yang Y. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. Proceedings of the IEEE International Conference on Computer Vision, 2017: 3754-3762.
- [33] Sun YF, Zheng L, Yang Y, Tian Q, Wang SJ. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). Proceedings of the European Conference on Computer Vision (ECCV), 2018: 480-496.
- [34] Wei LH, Zhang SL, Yao HT, Gao W, Tian Q. Glad: Global-local-alignment descriptor for pedestrian retrieval. Proceedings of the 25th ACM international conference on Multimedia, ACM, 2017: 420-428.
- [35] Su C, Li JN, Zhang SL, Xing JL, Gao W, Tian Q. Pose-driven deep convolutional model for person re-identification. Proceedings of the IEEE International Conference on Computer Vision, 2017: 3960-3969.
- [36] Ma AJ, Yuen PC, Li JW. Domain transfer support vector ranking for person re-identification without target camera label information. Proceedings of the IEEE international conference on computer vision, 2013: 3567-3574.
- [37] Wang XJ, Zheng WS, Li X, Zhang JG. Cross-scenario transfer person reidentification. IEEE Transactions on Circuits and Systems for Video Technology, 2015, 26(8): 1447-1460.
- [38] Peng PX, Xiang T, Wang YW, Pontil M, Gong SG, Huang TJ, Tian YH. Unsupervised cross-dataset transfer learning for person re-identification. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 1306-1315.
- [39] Kodirov E, Xiang T, Fu ZY, Gong SG. Person re-identification by unsupervised  $l_1$  graph learning. European conference on computer vision, Springer, Cham, 2016: 178-195.
- [40] Kodirov E, Xiang T, Gong SG. Dictionary learning with iterative laplacian regularisation for unsupervised person re-identification. BMVC, 2015, 3: 8.
- [41] Zhao R, Ouyang WL, Wang XG. Unsupervised salience learning for person re-identification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013: 3586-3593.
- [42] Yang Y, Wen LY, Lyu SW, Li SZ. Unsupervised learning of multi-level descriptors for person re-identification. Thirty-First AAAI Conference on Artificial Intelligence, 2017.

- [43] Yu HX, Zheng WS, Wu AC, Guo XW, Gong SG, Lai JH. Unsupervised person re-identification by soft multilabel learning. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019: 2148-2157.
- [44] Yang QZ, Yu HX, Wu AC, Zheng WS. Patch-based discriminative feature learning for unsupervised person re-identification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019: 3633-3642..
- [45] Dosovitskiy A, Springenberg JT, Riedmiller M, Brox T. Discriminative unsupervised feature learning with convolutional neural networks. *Advances in neural information processing systems*, 2014: 766-774.
- [46] Wang JY, Zhu XT, Gong SG, Li W. Transferable joint attribute-identity deep learning for unsupervised person re-identification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 2275-2284.
- [47] Lv JM, Chen WH, Li Q, Yang C. Unsupervised cross-dataset person re-identification by transfer learning of spatial-temporal patterns. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 7948-7956.
- [48] Fu Y, Wei YC, Wang GS, Zhou YQ, Shi HH, Huang TS. Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification. *Proceedings of the IEEE International Conference on Computer Vision*, 2019: 6112-6121.
- [49] Ester M, Kriegel HP, Sander J, Xu XW. A density-based algorithm for discovering clusters in large spatial databases with noise. *Kdd*, 1996, 96(34): 226-231.
- [50] Zhang XY, Cao JW, Shen CH, You MY. Self-training with progressive augmentation for unsupervised cross-domain person re-identification. *Proceedings of the IEEE International Conference on Computer Vision*, 2019: 8222-8231.
- [51] Campello RJGB, Moulavi D, Sander J. Density-based clustering based on hierarchical density estimates. *Pacific-Asia conference on knowledge discovery and data mining*, Springer, Berlin, Heidelberg, 2013: 160-172.
- [52] Lin YT, Dong XY, Zheng L, Yan Y, Yang Y. A bottom-up clustering approach to unsupervised person re-identification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 33: 8738-8745.
- [53] Tang HT, Zhao YR, Lu HT. Unsupervised person re-identification with iterative self-supervised domain adaptation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019: 0-0.
- [54] Yang FX, Li K, Zhong Z, Luo ZM, Sun X, Cheng H, Guo XW, Huang FY, Ji RR, Li SZ. Asymmetric co-teaching for unsupervised cross domain person re-identification. *arXiv preprint arXiv:1912.01349*, 2019.
- [55] Ding GD, Khan S, Tang ZM, Zhang J, Porikli F. Towards better validity: Dispersion based clustering for unsupervised person re-identification. *arXiv preprint arXiv:1906.01308*, 2019.
- [56] Huang Y, Wu Q, Xu JS, Zhong Y. SBSGAN: Suppression of inter-domain background shift for person re-identification. *Proceedings of the IEEE International Conference on Computer Vision*, 2019: 9527-9536.
- [57] Chen YB, Zhu XT, Gong SG. Instance-guided context rendering for cross-domain person re-identification. *Proceedings of the IEEE International Conference on Computer Vision*, 2019: 232-242.
- [58] Liu JW, Zha ZJ, Chen D, Hong RC, Wang M. Adaptive transfer network for cross-domain person re-identification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019: 7202-7211.
- [59] Zhong Z, Zheng L, Li SZ, Yang Y. Generalizing a person retrieval model hetero-and homogeneously. *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018: 172-188.
- [60] Choi Y, Choi M, Kim M, Ha JW, Kim S, Choo J. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 8789-8797.
- [61] Bak S, Carr P, Lalonde JF. Domain adaptation through synthesis for unsupervised person re-identification. *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018: 189-205.
- [62] Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings of the IEEE international conference on computer vision*, 2017: 2223-2232.
- [63] Deng WJ, Zheng L, Ye QX, Kang GL, Yang Y, Jiao JB. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 994-1003.
- [64] Hadsell R, Chopra S, LeCun Y. Dimensionality reduction by learning an invariant mapping. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, IEEE, 2006, 2: 1735-1742.

- [65] Wei LH, Zhang SL, Gao W, Tian Q. Person transfer gan to bridge domain gap for person re-identification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 79-88.
- [66] Wu ZR, Xiong YJ, Yu SX, Lin DH. Unsupervised feature learning via non-parametric instance discrimination. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 3733-3742.
- [67] Wang F, Xiang X, Cheng J, Yuille AL. Normface: L2 hypersphere embedding for face verification. Proceedings of the 25th ACM international conference on Multimedia, ACM, 2017: 1041-1049.
- [68] Zhong Z, Zheng L, Luo ZM, Li SZ, Yang Y. Invariance matters: Exemplar memory for domain adaptive person re-identification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 598-607..
- [69] Zhong Z, Zheng L, Luo ZM, Li SZ, Yang Y. Learning to adapt invariance in memory for person re-identification. arXiv preprint arXiv:1908.00485, 2019.
- [70] Ding YH, Fan HH, Xu ML, Yang Y. Adaptive exploration for unsupervised person re-identification. arXiv preprint arXiv:1907.04194, 2019.
- [71] Lin S, Li HL, Li CT, Kot AC. Multi-task mid-level feature alignment network for unsupervised cross-dataset person re-identification. arXiv preprint arXiv:1807.01440, 2018.
- [72] Gretton A, Borgwardt KM, Rasch MJ, Schölkopf B, Smola A. A kernel two-sample test. Journal of Machine Learning Research, 2012, 13(Mar): 723-773.
- [73] Delorme G, Xu YH, Lathuilière S, Horaud R, Alameda-Pineda X. CANU-ReID: A Conditional Adversarial Network for Unsupervised person Re-Identification. arXiv preprint arXiv:1904.01308, 2019.
- [74] Qi L, Wang L, Huo J, Zhou LP, Shi YH, Gao Y. A novel unsupervised camera-aware domain adaptation framework for person re-identification. International Conference on Computer Vision, 2019.
- [75] Wu AC, Zheng WS, Lai JH. Unsupervised person re-identification by camera-aware similarity consistency learning. Proceedings of the IEEE International Conference on Computer Vision, 2019: 6922-6931.
- [76] Kumar D, Siva P, Marchwica P, Wong A. Fairest of them all: Establishing a strong baseline for cross-domain person re-id. arXiv preprint arXiv:1907.12016, 2019.
- [77] Jia JR, Ruan QQ, Hospedales TM. Frustratingly easy person re-identification: Generalizing person re-id in practice. arXiv preprint arXiv:1905.03422, 2019.
- [78] Song JF, Yang YX, Song YZ, Xiang T, Hospedales TM. Generalizable person re-identification by domain-invariant mapping network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 719-728.
- [79] Liu X, Song ML, Tao DC, Zhou XC, Chen C, Bu JJ. Semi-supervised coupled dictionary learning for person re-identification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 3550-3557..
- [80] Wu AC, Zheng WS, Guo XW, Lai JH. Distilled person re-identification: Towards a more scalable system. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 1187-1196..
- [81] Xin XM, Wang JJ, Xie RJ, Zhou SP, Huang WL, Zheng NN. Semi-supervised person re-identification using multi-view clustering. Pattern Recognition, 2019, 88: 285-297.
- [82] Wu Y, Lin YT, Dong XY, Yan Y, Bian W, Yang Y. Progressive learning for person re-identification with one example. IEEE Transactions on Image Processing, 2019, 28(6): 2872-2881.
- [83] Li MX, Zhu XT, Gong SG. Unsupervised person re-identification by deep learning tracklet association. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 737-753.
- [84] Li MX, Zhu XT, Gong SG. Unsupervised tracklet person re-identification. IEEE transactions on pattern analysis and machine intelligence, 2019.
- [85] Qi L, Wang L, Huo J, Shi YH, Gao Y. Adversarial camera alignment network for unsupervised cross-camera person re-identification. arXiv preprint arXiv:1908.00862, 2019.
- [86] Qi L, Wang L, Huo J, Shi YH, Gao Y. Progressive cross-camera soft-label learning for semi-supervised person re-identification. arXiv preprint arXiv:1908.05669, 2019.
- [87] Zhu XP, Zhu XT, Li MX, Murino V, Gong SG. Intra-camera supervised person re-identification: A new benchmark. Proceedings of the IEEE International Conference on Computer Vision Workshops, 2019: 0-0.

- [88] Zheng L, Shen LY, Tian L, Wang SJ, Wang JD, Tian Q. Scalable person re-identification: A benchmark. *Proceedings of the IEEE international conference on computer vision*, 2015: 1116-1124.
- [89] Li W, Zhao R, Xiao T, Wang XG. Deepreid: Deep filter pairing neural network for person re-identification. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014: 152-159.
- [90] Felzenszwalb PF, Girshick RB, McAllester D, Ramanan D. Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence*, 2009, 32(9): 1627-1645.
- [91] Ristani E, Solera F, Zou R, Cucchiara R, Tomasi C. Performance measures and a data set for multi-target, multi-camera tracking. *European Conference on Computer Vision*. Springer, Cham, 2016: 17-35.
- [92] Ren SQ, He KM, Girshick R, Sun J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 2015: 91-99.
- [93] Zhong Z, Zheng L, Cao DL, Li SZ. Re-ranking person re-identification with k-reciprocal encoding. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 1318-1327.
- [94] Hirzer M, Belezni C, Roth PM, Bischof H. Person re-identification by descriptive and discriminative classification. *Scandinavian conference on Image analysis*, Springer, Berlin, Heidelberg, 2011: 91-102.
- [95] Wang TQ, Gong SG, Zhu XT, Wang SJ. Person re-identification by video ranking. *European conference on computer vision*. Springer, Cham, 2014: 688-703.
- [96] Zheng L, Bie Z, Sun YF, Wang JD, Su C, Wang SJ, Tian Q. Mars: A video benchmark for large-scale person re-identification. *European Conference on Computer Vision*, Springer, Cham, 2016: 868-884.
- [97] Wu Y, Lin YT, Dong XY, Yan Y, Ouyang WL, Yang Y. Exploit the unknown gradually: One-shot video-based person re-identification by stepwise learning. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 5177-5186.
- [98] Dehghan A, Modiri Assari S, Shah M. Gmmcp tracker: Globally optimal generalized maximum multi clique problem for multiple object tracking. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 4091-4099.
- [99] Zeng KW. Hierarchical clustering-guided re-id with triplet loss. *arXiv preprint arXiv:1910.12278*, 2019.
- [100] Wu GL, Zhu XT, Gong SG. Tracklet self-supervised learning for unsupervised person re-identification. *AAAI Conference on Artificial Intelligence*, 2019.
- [101] Xie QK, Zhou WG, Qi GJ, Tian Q, Li HQ. Progressive unsupervised person re-identification by tracklet association with spatio-temporal regularization. *arXiv preprint arXiv:1910.11560*, 2019.