

基于深度学习的行人再识别技术研究综述

魏文钰¹, 杨文忠^{1*}, 马国祥², 黄 梅¹

(1. 新疆大学 信息科学与工程学院, 乌鲁木齐 830046; 2. 新疆大学 软件学院, 乌鲁木齐 830046)

(* 通信作者电子邮箱 ywz_xy@163.com)

摘 要: 行人再识别(Re-id)作为智能视频监控技术之一,其目的是在不同的摄像机视图中检索出指定身份的行人,因此该项技术对维护社会治安稳定具有重大研究意义。针对传统的手工特征方法难以应对行人Re-id任务中复杂的摄像机环境的问题,大量基于深度学习的行人Re-id方法被提出,极大地推动了行人Re-id技术的发展。为了深入了解基于深度学习的行人Re-id技术,整理和分析了大量相关文献,首先从图像、视频、跨模态这3个方面展开综述性介绍,将图像行人Re-id技术分为有监督和无监督两大类并分别进行概括;然后列举了部分相关数据集,并对近年来在图像和视频数据集上的一些算法进行性能的比较与分析;最后总结了行人Re-id技术的发展难点,并深入讨论了该技术未来可能的研究方向。

关键词: 行人再识别;深度学习;特征学习;度量学习;卷积神经网络

中图分类号: TP181; TP391.41 **文献标志码:** A

Survey of person re-identification technology based on deep learning

WEI Wenyu¹, YANG Wenzhong^{1*}, MA Guoxiang², HUANG Mei¹

(1. College of Information Science and Engineering, Xinjiang University, Urumqi Xinjiang 830046, China;

2. School of Software, Xinjiang University, Urumqi Xinjiang 830046, China)

Abstract: As one of intelligent video surveillance technologies, person Re-identification (Re-id) has great research significance for maintaining social order and stability, and it aims to retrieve the specific person in different camera views. For traditional hand-crafted feature methods are difficult to address the complex camera environment problem in person Re-id task, a large number of deep learning-based person Re-id methods were proposed, so as to promote the development of person Re-id technology greatly. In order to deeply understand the person Re-id technology based on deep learning, a large number of related literature were collated and analyzed. First, a comprehensive introduction was given from three aspects: image, video and cross-modality. The image-based person Re-id technology was divided into two categories: supervised and unsupervised, and the two categories were generalized respectively. Then, some related datasets were listed, and the performance of some algorithms in recent years on image and video datasets were compared and analyzed. At last, the development difficulties of person Re-id technology were summarized, and the possible future research directions of this technology were discussed.

Key words: person Re-identification (Re-id); deep learning; feature learning; metric learning; Convolutional Neural Network (CNN)

0 引言

随着“平安城市”建设的大力推进,人们在社会治安管理方面投入了大量的人力、物力,以预防、控制、打击破坏社会稳定的违法犯罪行为的发生。其中,智能监控系统作为社会治安管理系统中不可缺少的一部分,为维护社会的安全稳定发挥了巨大作用。“平安城市”的建设要求城市具备强大可靠的智能监控系统,这需要先进科学技术的支撑,以确保在追踪不法犯罪分子的过程中能提供明确可靠的线索。

一般来讲,一个完备的视频监控系统应该具备3个部分^[1]:行人检测、行人追踪、行人检索,其中,行人检索即为行

人再识别(person Re-identification, Re-id)。如图1所示,行人再识别^[1]是一种利用计算机相关技术来进行跨摄像机视图中特定行人身份匹配的技术,其中浅灰色框匹配图片为同一身份的正确匹配,深灰色框匹配图片为不同身份的错误匹配。行人再识别属于图像检索的子问题,但其与图像检索不同的是,行人再识别检索的目标是特定行人,它是为了在社会治安管理方面搜索、追踪特定行人而兴起的一项图像检索技术。

对于跨摄像机的行人检索技术来说,其存在着许多与其他计算机视觉技术不同的难点与挑战(如图2所示),包括:1)低分辨率:受摄像机等成像设备的限制,实际场景中得到的行人图像大多是分辨率较低的模糊图像;2)跨摄像机问题:由

收稿日期: 2020-01-15; **修回日期:** 2020-04-02; **录用日期:** 2020-04-04。 **基金项目:** 国家自然科学基金重点项目(U1435215); 国家自然科学基金资助项目(U1603115); 国家重点研发计划项目(2017YFC0820702-3); 新疆维吾尔自治区自然科学基金资助项目(2017D01C042)。

作者简介: 魏文钰(1995—),女,新疆哈密人,硕士研究生,主要研究方向:计算机视觉、行人再识别、信息安全; 杨文忠(1971—),男,河南南阳人,副教授,博士,CCF会员,主要研究方向:网络舆情、情报分析、信息安全、无线传感器网络; 马国祥(1993—),男,新疆伊宁人,硕士研究生,主要研究方向:计算机视觉、机器学习、信息安全; 黄梅(1996—),女,新疆和硕人,硕士研究生,主要研究方向:计算机视觉、遥感图像处理。

于复杂多变的摄像机环境,不同的摄像机提供的行人图像会存在巨大差异,这些差异包括光照不同、行人姿态不同、背景不同、成像风格不同等;3)遮挡问题:在实际视频监控中,目标行人难免会受到车辆、栏杆、其他行人等遮挡物的遮挡,导致目标行人图像不完整;4)检测不准确:经过目标检测方法处理过后得到的行人再识别数据集中的行人图像,会存在行人检测不准确而导致的行人图像部分缺失、错位等问题;5)标注困难:在实际待检索的目标域上难以获取有标注的数据进行模型训练。针对这些挑战与困难,研究者们将行人再识别技术从传统的手工设计的方法,发展到现在的卷积神经网络、生成对抗网络等深度学习方法,已经较为成熟了,且目前的算法在常用的公开数据集^[2-3]上的首位准确率(Rank-1)已经达到90%以上。但为了达到更高的再识别准确率,使行人再识别技术可以更好地运用在实际的复杂场景中,对于该技术的发展还有待进一步研究。

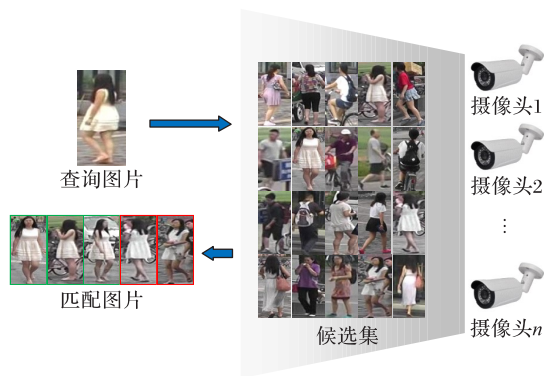


图1 行人再识别任务示意图

Fig. 1 Schematic diagram of person re-identification task



图2 行人再识别任务中的挑战

Fig. 2 Challenges in person re-identification task

根据数据源的不同,本文将现有的行人再识别方法分为图像行人再识别、视频行人再识别、跨模态行人再识别,并基于这3类基于深度学习的行人再识别技术进行详细介绍,大致描绘一个该技术从发展初期到现今的发展历程,介绍各类方法的主要思想并进行简单总结。此外,本文进行了相关数据集介绍和一些优秀方法的性能比较,最后概览全文得出总结与展望。

1 图像行人再识别

以卷积神经网络(Convolutional Neural Network, CNN)为代表的深度学习方法在计算机视觉研究领域大放异彩,包括

人脸识别^[4-5]、物体检测^[6-7]、目标跟踪^[8-9]等方面,都取得了不错的效果。在2014年之后,大量文献采用深度学习的方法来解决行人再识别问题,不仅实现了同一行人不变特征的鲁棒表示,还实现了端到端的行人再识别系统,这些方法不再使用复杂的度量算法,仅仅使用简单的欧氏距离或余弦距离就可以得到较高的匹配准确率。这些深度学习方法的性能在很大程度上超过了早期基于手工特征的行人再识别方法,很快占领了该研究领域的半壁江山,此后的研究将继续在该技术上进一步进行探索。

基于深度学习的行人再识别技术自发展以来,对用于图像方法的研究要多于用于视频的方法,同时一些用于图像的方法也可以扩展到视频行人再识别数据集,以完成视频序列的行人再识别任务。本文将图像行人再识别方法根据数据源是否有标签分为有监督方法和无监督方法两大类,下面将就这两大类展开详细介绍。

1.1 有监督方法

在机器学习中,有监督方法通常被认为是使用有标签的训练数据对模型进行监督训练,该类方法在行人再识别研究领域即为使用有ID标签的行人图片来训练行人再识别模型;随后用训练好的模型从行人图片中提取出行人不变特征,再对这些特征进行距离度量计算,从而判断出哪些行人图片属于同一身份,哪些行人图片属于不同身份。现有的有监督图像行人再识别方法可以分为基于特征学习的方法和基于度量学习的方法。基于特征学习的方法着重学习行人的不变特征表示,基于度量学习的方法着重利用度量损失函数学习行人之间的相似度,但是这些方法都共有一个目的,即对不同摄像机下的同一行人进行正确的同类匹配。

1.1.1 基于特征学习的方法

行人图像中既包含全局信息也包含局部信息,若从抽取的特征类型进行分类,特征学习可以分为全局特征学习和局部特征学习。以往的特征学习方法^[10-13]只考虑从行人图像的全局信息中抽取行人不变特征,但这对姿态变化、视角变化、行人检测不准确等问题较为敏感,从而影响检索精度。现在大多数性能好的行人再识别方法^[14-18]以抽取行人局部特征并结合行人全局特征的方法来获取行人判别信息,以克服各种跨摄像机检索的困难。根据局部特征区域的生成方式,可以将局部特征学习方法分为以下三类^[16]:1)根据先验知识^[19-21]如姿态估计、关键点定位等来定位局部区域;2)使用注意力机制的方法^[14,22-24]提取感兴趣的局部区域;3)将深度特征映射分割为预定义条状^[16-19,25-26]。这三类局部特征学习的方法将帮助模型提取出有效鲁棒的行人特征表示。

文献[19]提出一种基于姿态的深度卷积模型(Pose-driven Deep Convolutional model, PDC)结构,其由两部分组成:特征嵌入子网(Feature Embedding sub-Net, FEN)和特征加权子网(Feature Weighting sub-Net, FWN)。FEN通过姿态估计算法^[27]得到14个人体关键点进而得到人体6个身体局部区域,然后使用FWN对不同局部区域加权融合得到全局与局部特征。文献[20]将人体结构信息融入到CNN框架中,使用文献[28]的算法来定位人体关键点,得到7个人体局部区域,最后使用特征融合方法合并不同区域的特征向量。文献[21]为了获得更加精确的人体局部特征,利用Deeper Cut方法^[29]来估计定位对姿态变化、摄像机视角变化具有鲁棒性的14个人体关键点,然后根据这些关键点得到3个人体局部区域(如图3所示),在

训练阶段,将这3个部分区域和全局图像输入网络中,提取出4个特征描述符,并将它们串联起来,最后得到全局-局部对齐描述符(Global-Local-Alignment Descriptor, GLAD)。总的来说,利用先验知识来定位人体局部区域虽然可以学习到具有语义信息的局部特征,但是会引入额外的误差,影响提取到的局部特征的质量,从而影响行人检索算法的性能。

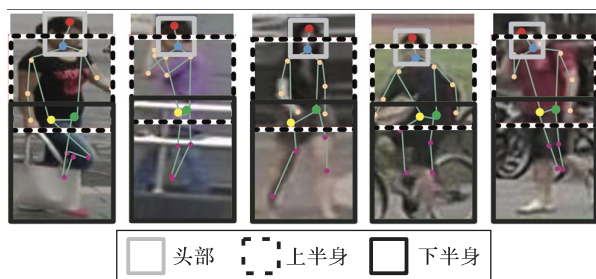


图3 关键点定位示意图^[21]

Fig. 3 Schematic diagram of located keypoints^[21]

注意力机制模型在一些具有挑战性的序列数据识别和建模任务上表现出了优异的性能,包括标题生成^[30]、隐式情感分析^[31]和动作识别^[32]等。简而言之,它与人类处理视觉信息的方式类似,基于注意力机制的算法倾向于选择性地集中于部分信息,同时忽略其他感知信息,因此该方法在行人再识别中可用于自适应地定位行人图像的辨别性局部区域。文献[22]提出了一个端到端的比较性注意力网络(Comparative Attention Network, CAN),该网络以长短时记忆(Long Short Term Memory, LSTM)网络为基本架构,融入了时间信息,可以通过多次“观察”比较行人外观后,学习图像中哪些局部区域与待识别的行人相关,并自动集成不同局部区域的信息,以确定一对图像是否属于同一个人。针对细粒度的行人分析任务,文献[23]提出了一种基于多方向注意力模块的网络架构HP-Net(HydraPlus-Net),该网络能够从低层到语义层捕获多个注意力特征,并加入了注意力特征的多尺度选择,以此丰富了行人图像的最终特征表示。深度学习模型由于有大量的参数而导致算法效率不高,文献[24]提出了轻量级(参数少)的注意力网络架构HA-CNN(Harmonious Attention CNN),该网络能同时学习任意行人图像中的硬区域级注意力和软像素级注意力,以得到行人不变性特征表示。文献[14]提出了在卷积神经网络中将非局部运算与二阶统计相结合的观点,设计了一种二阶非局部注意(Second-Order Non-local Attention, SONA)模型,通过非局部操作对特征映射中的位置相关性进行建模,可以将卷积操作捕获的局部信息集成到模型中,为行人再识别任务提供了良好的判别特征。注意力机制在很多行人再识别方法中都非常有效,但是使用该方法抽取的局部特征区域缺乏语义信息,不利于理解和解释。

图像特征分割是基于局部特征学习的行人再识别技术中常用的方法,该方法一般将经过深度网络的特征映射横向切分为几个不重叠的预定义的条状或块状区域,分别进行局部特征学习。文献[25]提出了一种基于局部特征学习的卷积基线网络(Part-based Convolutional Baseline, PCB)。如图4所示,PCB以整幅图像作为基线网络ResNet-50输入,将所得的特征映射在卷积层上均匀划分为 p 个部分,经过Softmax多分类损失函数训练得到 p 个分类器。该文献还提出细化局部分区的池化方法(Refined Part Pooling, RPP),目的是为了重新

分配每个局部区域内的离群点,加强每个局部区域内部的一致性。图像特征分割方法一般假设不同的图像中行人是对齐的,但这种假设显然不符合数据集中的实际情况,基于此问题,文献[26]提出AlignedReID方法执行局部区域对齐。首先共同学习全局特征和局部特征,然后在学习局部特征的网络分支中,用动态规划求最小路径来对齐局部特征。为了缓解行人图像错位而引起的离群值问题,文献[16]使用空间金字塔池化网络(Horizontal Pyramid Matching, HPM),在不同金字塔尺度进行全局平均池化和全局最大池化的融合操作,将最后的特征图划分为多个水平条,然后独立地对不同尺度的水平条进行分类操作。还有一些文献^[17-18]也采用类似的金字塔池化方法抽取图像的局部特征,都取得了不错的效果。图像特征分割作为局部特征学习的方法之一,在行人再识别的深度学习方法中取得了很好的行人检索效果。但是该类方法增加了训练模型的难度,降低了训练效率。同时,很多方法没有考虑到每个局部区域之间的渐进线索^[17],从而忽略了空间上下文信息,这将严重影响错位行人图像的匹配精度。

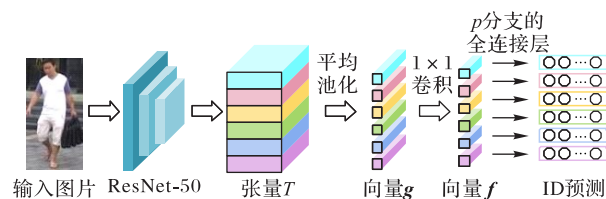


图4 PCB模型示意图^[25]

Fig. 4 Schematic diagram of PCB model^[25]

单一的全局特征学习或局部特征学习都会丢失代表行人不变特性的部分信息,因此,作为行人再识别深度学习方法中的主流方法,基于特征学习的方法常以局部特征作为对全局特征的补充,两者结合共同表示行人判别特征,这在一定程度上大大提升了行人特征的鲁棒性。但是,由于行人所在的复杂环境,常使得摄像机得到的行人图像存在遮挡问题,而这类问题会因为引入过多噪声,严重影响特征学习方法的行人检索性能。此外,实际场景中不准确的行人检测算法也会影响特征学习方法所提取特征的鲁棒性。因此,特征学习的方法需要考虑更多复杂的实际监控场景,以便提取出更能应对多种复杂环境的、更为鲁棒的行人不变特征表示。

1.1.2 基于度量学习的方法

基于深度学习的度量学习方法不同于早期的距离度量方法^[33-34]需要设计复杂的度量算法来衡量特征之间的相似性,度量学习以损失函数的形式藏匿于网络结构中,测试阶段中查询集和候选集之间的相似度量任务仅仅需要简单的余弦距离或欧氏距离来完成。度量学习在传统的度量距离(欧氏距离、余弦距离、马氏距离等)的基础上,设计出在深度网络的可导的度量损失函数,经过大量样本训练来减小该损失进而优化其参数,从而达到使同类样本距离变小、异类样本距离变大的目的。

常见的度量学习损失函数有对比损失^[35]、三元组损失^[36-38]、四元组损失^[39]、中心损失^[40]等,其中,最为常用的是三元组损失。对比损失一般使用在Siamese网络中,文献[38]为了解决一些网络难以区分正样本和难分负样本的问题,提出了一种使用匹配门函数的Siamese网络来增强局部特征的细节信息,该网络就是通过对比损失函数进行优化的。文献[36]第一次将三元组损失函数模型使用在图像检索任务中,由于行人再识别是一种特殊的图像检索任务,随后就有一些

工作将三元组损失使用在行人再识别任务中。为了提取能克服姿态变化、视角变化的行人特征表示,文献[38]使用了一种多通道 CNN 模型,学习行人的全局特征和局部特征,同时使用改进的三元组损失函数来进一步增强所学特征的辨别能力。由于三元组损失函数在度量学习的应用研究中,产生了很多相关变体^[41-43],所以文献[38]就对这些变体进行性能对比,然后提出自己改进的三元组损失,即难分样本采样损失(Triplet Loss with batch Hard mining, TriHard Loss),该方法采用难分正样本和难分负样本以及固定样本组成的三元组进行三元组损失函数的计算。实验证明,难分样本采样损失的性能要优于其他三元组损失的性能。文献[39]认为,传统的三元组损失在测试集上泛化性能较差,原因是类内间距依然很大,因此在三元组损失的基础上提出了四元组损失。该方法以四张图片作为输入,引入一种在线边缘难分样本挖掘方法,自适应地选择难分样本进行网络训练。分类模型中单一的 Softmax 损失仅仅只会将不同身份的人进行归类,却没有考虑视角变化、空间错位所引起的类内差异,因此,文献[40]将多种损失函数结合起来,使用在一个分类模型中,来学习行人的联合深度特征。这些损失函数包括 Softmax 损失、中心损失、类内中心损失,其中,中心损失将同一行人的特征进行集中,旨在减少类内差距;类内中心损失则使不同类间的距离最大化,旨在进一步扩大类间距离。

度量学习的最终目的是缩小类内距离、增大类间距离,并在训练阶段学习行人之间的相似度量,因此,度量学习的训练目标与测试方式是一致的。度量学习侧重于学习相似度量,使用的是标签的弱监督信息;特征学习则侧重于学习行人不变特征,使用的是标签的强监督信息,因此度量学习可用于大规模数据集的模型训练。此外,度量学习方法需要在网络训练前随机选择样本组成多元组,这一过程无疑会增加计算量,降低模型训练效率。近几年一些行人再识别深度学习模型已经将两种方法融合在一起,联合优化两种学习任务,以获取超越单一模型框架的性能。

1.2 无监督方法

在实际的应用场景中,待检索的目标域多为无标签的数据,且标注工作异常困难。但是现有的深度学习行人再识别方法需要大量数据来进行有监督训练才可以达到更好的效果。因此,为了解决目标域数据集标签难以获取的问题,一些无监督的行人再识别方法被提出。现有的无监督行人再识别方法可以根据是否跨域划分为两类(如图 5 所示):1)单域无监督行人再识别方法:利用聚类^[44-45]或行人轨迹片段^[46-47]的方法来进行目标域上的无监督模型训练;2)跨域无监督行人再识别方法:利用有标签的源域数据来辅助训练用于目标域的深度学习模型。这些无监督行人再识别方法由于缺乏真实有效的监督信息,因此在算法性能上和有监督方法还存在一定差距。

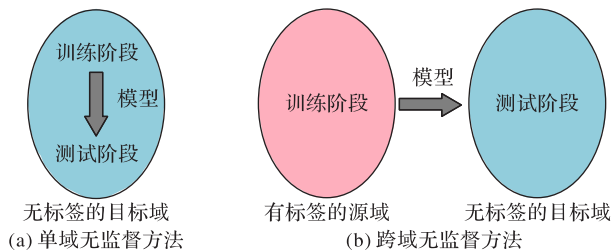


图 5 无监督方法示意图

Fig. 5 Schematic diagram of unsupervised methods

1.2.1 单域无监督方法

单域无监督行人再识别方法大多不借助源域的辅助信息,仅仅在目标域上利用处理后的无标签数据来训练行人再识别模型,基于此,有些工作采用聚类的方法^[44-45]进行无标签数据的处理。文献[44]提出了无监督的非对称度量学习方法 CAMEL(Clustering-based Asymmetric MEtric Learning),该方法旨在通过联合学习非对称度和最优聚类,来学习每个摄像机图像的特征,然后将不同摄像机的特征映射到一个共享空间中进行特征对齐,从而可以获得更好的跨图像匹配性能。文献[45]提出了一种自底向上聚类(Bottom-Up Clustering, BUC)的方法将 CNN 提取的特征进行聚类,最大化身份的多样性,同时保持每个身份中的相似性,在此过程中,利用身份多样性和相似性来学习判别特征。为了减少跨域方法中对源域的依赖,还有一些方法^[46-48]利用行人轨迹片段来实现无监督行人再识别。文献[46-47]提出的 TAUDL(Tracklet Association Unsupervised Deep Learning)和 UTAL(Unsupervised Tracklet Association Learning)可以从行人图像中自动生成的行人轨迹片段中提取行人判别信息,即用跨摄像机的行人轨迹片段关联学习,找到最有可能是同一个人的跨摄像机视图轨迹片段。文献[48]对之前行人轨迹片段的方法^[46-47]进行扩展,提出无监督图关联(Unsupervised Graph Association, UGA)框架,其中包括单摄像机特征学习和跨摄像机特征学习。在单摄像机学习中应用多分支分类器来对每个摄像机内图像各自进行分类,在跨摄像机学习中建立一个跨摄像机图来关联行人轨迹片段,从而学习行人的跨视图不变特性。

由于单域无监督方法的模型训练和测试在同一数据域中,没有涉及到其他数据域,因此这类方法的系统实施过程会较为容易;但是,本就缺乏监督信息的无监督方法,在没有借助其他辅助信息的情况下,将更难达到有监督方法的性能水平。就此而言,如何缩小单域无监督方法和有监督方法的性能差距,将会是研究者的一个极大挑战。

1.2.2 跨域无监督方法

目前大多数的无监督行人再识别的研究重点放在跨域无监督行人再识别,此类方法旨在如何以无监督的方式将预先训练好的模型从源域有效地传递到目标域^[49]。一般情况下,直接将在源域中训练的模型应用到目标域,由于两个域之间存在一定差异,会导致效果不理想,而无监督的跨域方法会减小这种域差异,使得源域的模型也可以有效地用于无标签的目标域。文献[50]提出的 PUL(Progressive Unsupervised Learning)方法,使用预训练的模型从未标记的目标域训练集中提取图像特征,然后使用 K-means 聚类算法对这些特征聚类,再从这些类中选择可靠的样本对原始模型进行微调,随后使用这个新模型来提取特征,并开始另一次迭代训练。如此迭代学习可以使得模型最终从不可靠的聚类结果中提取出可靠的特征信息。文献[51]提出的 ARN(Adaptation and Re-identification Network)框架利用数据集之间的信息来学习域不变特性,通过域不变特性来进行跨域行人匹配。为了提高 CNN 模型在源域上提取判别特征的能力,文献[52]提出一种聚类和动态采样(Clustering and Dynamic Sampling, CDS)方法,利用行人属性来加强源域模型的训练,然后对目标域样本

进行迭代聚类,动态选择目标域中的信息样本对源域模型进行微调(Fine-tuning)。

还有一些工作采用了无监督域适应的方法来处理跨域行人再识别任务。近年来,无监督域适应(Unsupervised Domain Adaptation, UDA)^[53-54]是机器学习领域所研究的一个热点,该类问题的任务是在有标签的源域数据集上训练的模型应用在另一个任务相关却特征分布不同的目标域数据集^[55]。然而,现有的无监督域自适应方法大多基于跨域类标签相同的假设,而不同的Re-id数据集的人身份是完全不同的。因此,许多无监督域适应方法^[53-54]不能直接用于行人再识别任务。一些行人再识别的域适应方法^[55-58]采用生成对抗网络(Generative Adversarial Network, GAN)^[59]来对不同数据集之间的样本进行图像迁移,即将源域图像转换为目标域图像的样式风格并保持人物身份不变,然后使用转换后的图像来训练模型。

生成对抗网络^[59]是一种强大的生成式模型(如图6所示),其中蕴含了博弈思想:生成器以随机噪声为输入生成可以以假乱真的虚假样本来干扰鉴别器的判断,判别模型以真实样本和生成样本为输入来判断该样本是真还是假,这两者在各自任务的驱使下相互博弈,不断进行迭代优化,从而达到最优的模型性能,以生成高质量的虚假样本。文献[55]提出了数据集之间的行人迁移模型 PTGAN (Person Transfer Generative Adversarial Network),使用了与CycleGAN^[60]相似的网络结构,可以有效减小数据集之间的域差异。文献[55]为了解决领域自适应问题只考虑了域间的摄像机差异,而文献[56]不仅考虑了域之间的摄像机差异性,还考虑了域内不同摄像机间的差异性,其使用基于CycleGAN^[60]所提出的CamStyle(Camera Style)方法来训练两两相机之间的转换模型,同时采用了标签平滑正则化(Label Smooth Regularization, LSR),减小有噪声的生成样本所带来的过拟合风险。这些使用CycleGAN^[60]的跨域行人再识别方法在迁移过程中确保了相机风格改变、行人身份不变。然而,CamStyle方法^[56]使用多达28个CycleGAN^[60]模型来训练多个摄像机对之间的样式转换,因此该方法复杂度高,训练模型困难。基于此,文献[57]提出了基于StarGAN^[61]的多摄像机迁移GAN(multi-Camera Transfer GAN, CTGAN),该方法只需要训练一个GAN模型,就可以将源数据集中的图像转换为目标数据集中每个摄像机的样式,并使用SCDA(Selective Convolutional Descriptor Aggregation)方法^[62]去除背景噪声,保留有用的深度特征。在以往基于GAN的跨域方法中,域转移后得到的图像一般只具有一种或者几种预定义的样式,文献[58]为了改进这一缺陷,提出了一种实例引导上下文生成图像的方法,通过设计一个成对的条件GAN即CRGAN(Context Rendering GAN),实现了用一个源域的行人图像来生成具有目标域中 N 中背景的 N 张图像,用这种方法得到的新数据集不仅有标签信息(来自源域的行人ID标签),而且具有目标域中丰富的样式信息,以此可以对无标签的目标域实现有监督Re-id模型训练。GAN技术在行人再识别领域除了可以用于跨域无监督方法之外,还可以用于单纯的数据增强^[63]、学习行人不变特性^[64]等。

除了使用GAN方法,文献[65]提出了一种联合属性信息和身份信息的深度学习框架TJ-AIDL(Transferable Joint

Attribute-Identity Deep Learning),可以同时学习全局的身份信息与局部的属性信息,该模型可以被转移到任何无标签的目标域中进行行人再识别的无监督域适应任务。文献[66]也同样利用属性信息来丰富行人特征,以促进实现行人再识别的无监督域适应任务。

与单域无监督方法相比,跨域无监督方法的研究涉及广泛,且由于GAN技术的兴起,越来越多的工作更倾向于研究跨域无监督方法。在两种无监督方法中,由于跨域方法借助了源域的辅助信息,大多数跨域方法可以取得更好的无监督算法性能,但是跨域方法过于依赖源域和目标域之间的相似性,这降低了跨域无监督模型的鲁棒性。

虽然有监督的行人再识别技术在最近的研究中已经逐渐趋向于稳定,但是无监督的行人再识别技术由于缺乏有效的监督信息,其性能远不如监督学习的方法。为了减少对标注数据的依赖,使搭建好的行人再识别模型高效地部署在实际应用场景中,无监督技术还有待进一步的研究和发展。

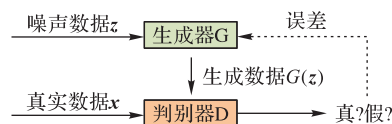


图6 生成对抗网络结构示意图

Fig. 6 Schematic diagram of generative adversarial network structure

2 视频行人再识别

近年来,随着数据量的增加,基于视频的行人再识别研究越来越多。基于图像的行人再识别方法只能从单个图像中获取有限的信息,并且在解决单个图像中遮挡、姿态变化和摄像机视角等问题上效果有限。相比之下,基于视频的行人的再识别比单个图像包含更多的信息。由于图像序列通常包含丰富的时间信息,因此更适合在复杂的环境和较大的几何变化下对行人进行再识别,而且基于视频序列的方法更符合实际监控系统中的行人再识别任务的需求,这样会避免一些监控视频的预处理过程。

早期的基于视频的行人再识别技术使用一些手工设计特征的方法^[67-69],虽然这些方法提取的特征可解释性强,但却忽略了行人视频序列中的大量信息,使得行人检索精度不高。最近几年,基于深度学习的视频行人再识别方法^[70-76]被提出。文献[70]提出了一种用于深度视频的基本方法RCN(Recurrent Convolutional Network),其利用CNN从连续的视频帧中提取特征,然后利用循环神经网络(Recurrent Neural Network, RNN)将时间序列中的上下文信息合并到每个视频帧中,并采用最大池化或平均池化操作获得最终的序列特征表示。文献[71]提出一种时间注意力机制模型TAM(Temporal Attention Model)来提取表示视频中特定行人的判别帧。此外,还采用RNN模型SRM(Spatial Recurrent Model)从六个方向结合上下文信息,增强行人特征映射中每个行人的位置表示。文献[72]提出了ASTPN(Attentive Spatial-Temporal Pooling Network)架构,其中引入了时间建模中的共享注意力矩阵,实现了帧选择过程中的查询集与候选集视频序列之间的信息交换,在这种情况下,判别帧可以根据其注意

力权重来进行行人特征对齐。文献[73]提出了同时学习视频内距离和视频间距离的方法,使视频的特征表达更加紧凑,并借此区分不同行人的视频序列。在实际监控视频中,常有噪声视频片段的干扰,导致行人再识别算法性能显著下降,针对该问题,文献[74]提出了一种新的基于视频的自适应加权(Self-Paced Weighting, SPW)身份识别方法,该方法首先使用一种离群点检测的方法来评估视频子序列的噪声程度,然后采用加权距离的方法对两个不同身份的视频序列进行距离测量。在最新的工作中,文献[75]为了解决跨摄像机检索任务中的视图偏差问题,提出了一种深度非对称度量学习(Deep Asymmetric Metric learning, DAM)方法,该方法将所提出的非对称距离度量学习损失函数嵌入到一个双流卷积网络中,用于联合学习特定视图和特定特征的转换,以解决基于视频的行人再识别中的视图偏差问题。由于该模型框架中的运动特征依赖于光流,而光流被广泛应用于运动建模,并且需要考虑计算的成本,因此对于在线视频处理,该方法仍然不够有效。为了解决基于视频的行人再识别任务中标注成本高等问题,文献[76]首次提出并研究了行人再识别中的弱监督问题,将弱监督的行人再识别问题转化为一个多实例多标签(Multi-Instance Multi-Label, MIML)学习问题,开发了一种跨视图多标签多实例(Cross-View Multi-Instance Multi-Label, CV-MIML)学习方法,该方法能够利用视频片段内的类似实例来进行内部对齐,并且通过将分布原型嵌入到 MIML 中来捕获不同相机视图之间的潜在匹配实例,最后将该方法嵌入到深度神经网络中,形成端到端的深度跨视图多标签多实例学习模型。

与基于单帧图像的人像识别方法相比,基于视频的行人再识别方法具有更多的时空信息、运动信息和更多的外观线索,可以在某种程度上提高匹配性能,因而受到关注。现有的许多基于图像的方法也可以扩展到基于视频的方法中,但由于基于视频的方法的输入是行人图像序列,这将大大增加计算量,影响了计算效率。因此,在此后的行人再识别研究中,在致力于提高检索精度的同时,也需要考虑计算成本,使一个更快更高效的行人再识别方法能更好地应用在实际视频监控系统中,以维护社会安全稳定。

3 跨模态行人再识别

现有的大多数行人再识别方法的处理数据都集中在可见光相机采集的同一类型数据上,并将行人再识别任务作为一个单模态数据匹配问题。然而,在实际生活应用中,仅用可见光相机采集的单模态行人数据会出现画质模糊、分辨率低、光照变化大等问题,这将影响可见光相机捕捉的行人图像的质量,导致实际应用中检索精度不高。为了克服实际情况中复杂多变的环境对行人检索精度的影响,引入了多模态行人数据(如图7所示),包括常见的可见光相机采集的RGB图像、低/高分辨率图像^[77-78]、红外(Infra-Red, IR)图像^[79-80]、深度图像^[81]、文本描述^[82]、素描图像^[83]。文献[77]针对跨分辨率的行人图像匹配问题提出了 SING (Super-resolution and Identity joiNt learninG) 方法,通过增强低分辨率图像中的有效局部信息,从而解决了跨分辨率行人匹配中的信息量差异问题。文献[78]提出对比中心损失(Contrastive Center Loss, CCL)方法,使深度模型在不受分辨率差异影响的情况下,从不同分

率的图像中学习行人特征。针对红外图像数据和可见光图像数据的跨模态行人再识别,文献[79]提出端到端对齐生成对抗网络 AlignGAN,其可以联合利用像素对齐和特征对齐,缓解跨模态数据的变化,学习跨模态行人图像的不变特征。文献[80]提出一种双级差异减小方法 D²RL (Dual-level Discrepancy Reduction Learning),将不同模态的图像红外图像和 RGB 图像转换为统一模态以减小模态差异,再通过减小特征差异来处理外观上的不同。对于行人深度图像,文献[81]利用深度图像和 RGB 图像之间的内在联系,采用一个深度跨模态网络将特征表示进行跨模态转移,以便从两种模态的数据中提取出相似的行人特征。针对文本描述和行人图像的匹配的问题,文献[82]提出一种基于身份感知的两阶段深度学习框架。该框架在第一阶段有效地筛选出容易出现的错误匹配,作为第二阶段网络训练的起点,在第二阶段使用联合注意力机制的 CNN-LSTM 结构共同优化匹配结果。考虑到在实际情况下,图像数据并不是随时都可以获得,文献[83]提出了素描行人再识别模型,并创建了素描行人再识别数据集。对于该模型,文献中提出了一种深度对抗学习架构,用来共同学习行人判别特征和域不变特征。

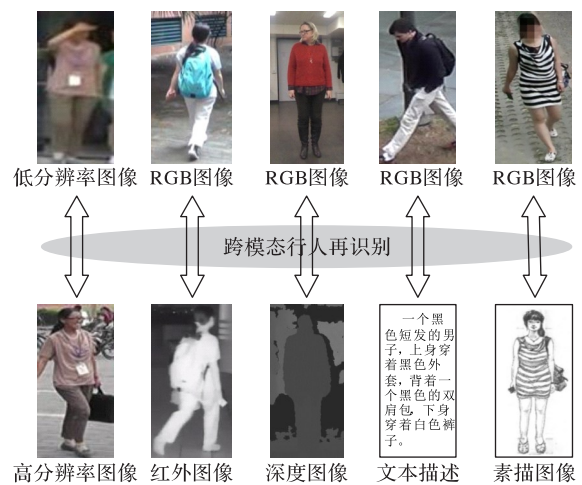


图7 跨模态行人再识别

Fig. 7 Cross-modal person re-identification

行人再识别技术作为当今社会维护公共生命财产安全的重要科技手段之一,应当涉及到监控系统中各种模态数据,并形成一套完善的多模态行人再识别体系,以更好地保障社会的安全稳定。跨模态行人再识别涉及到两种及两种以上类型的行人数据,这给行人匹配带来了一定的难度,但同时也丰富了行人的模态信息,更加贴合实际监控情况。相较于单模态行人再识别,跨模态行人再识别的研究起步较晚,技术发展不够成熟,因此如何将多模态的行人信息整合起来进行行人匹配,将是以后基于深度学习行人再识别技术的一个研究重点和难点。

4 常用数据集

深度学习技术的研究需要依赖大量数据来训练模型,数据样本越多、采集环境越贴近实际情况,越有利于开发鲁棒性的行人再识别模型。自从行人再识别任务被提出以来,许多用于该研究领域的数据集也相继被公布,这些数据集的规

模、模态、检测方法也不尽相同。根据数据类型不同,现有的行人再识别数据集可以分为单帧图片数据集、视频序列数据集、多模态数据集。对于这三类中一些常见的数据集,本文做了简单的对比总结,具体见表 1 所示。

表 1 行人再识别数据集
Tab. 1 Datasets of person re-identification

数据集名称	模态	发布年份	ID 数	相机数		图片数(序列数)	检测器
				室内	室外		
VIPeR ^[84]	RGB 图像	2007	632	0	2	1 264	手工
GRID ^[85]	RGB 图像	2009	250	8	0	1 275	手工
CUHK01 ^[86]	RGB 图像	2012	971	0	2	3 884	手工
CUHK03 ^[10,87]	RGB 图像	2014	1 467	10	0	14 096	手工+DPM
Market-1501 ^[2]	RGB 图像	2015	1 501	0	6	32 668	DPM
DukeMTMC-reID ^[3]	RGB 图像	2017	1 812	0	8	36 411	手工
MSMT17 ^[60]	RGB 图像	2018	4 101	3	12	126 441	Faster RCNN
PRID-2011 ^[89]	RGB 序列	2011	200	0	2	400	手工
iLIDS-VID ^[90]	RGB 序列	2014	300	2	0	600	手工
MARS ^[91]	RGB 序列	2016	1 261	0	6	20 715	DPM+GMMCP
LVreID ^[92]	RGB 序列	2018	3 772	3	12	14 943	Faster RCNN
CAVIAR ^[93]	低/高分辨率	2011	72	2	0	610	手工
SYSU-MM01 ^[94]	红外、RGB 图像	2017	491	2	4	287 628	手工
RegDB ^[95]	红外、RGB 图像	2017	412	0	2	8 340	手工
BIWI RGBD-ID ^[96]	深度、RGB 序列	2014	50	—	—	106	—
RobotPKURGBD-ID ^[97]	深度、RGB 序列	2017	90	—	—	180	—
SketchRe-ID ^[83]	素描、RGB 图像	2018	200	0	2	600	手工

现在大部分行人再识别技术都是基于单帧 RGB 图像进行实验研究的,因此已发布的该类数据集较多,其规模、环境复杂度也相对较大。较早发布的数据集,如:VIPeR^[84]、GRID^[85]、CUHK01^[86]等不仅人物数量少、图片数少,且大都用手工标注的方法检测行人。随着深度学习的发展,小规模数据已经无法满足深度学习模型的训练要求,于是提出了 CUHK03^[10,87]、Market-1501^[2]等大规模数据集,且采用了自动检测提取行人的方法 DPM(Deformable Part Model),降低了数据采集的难度。2018 年公布的 MSMT17^[60]数据集采用了更加先进、准确的目标检测方法 Faster RCNN^[88]对更多行人进行图像数据采集,其包含了更丰富的场景信息,更加接近实际情况,因此对行人再识别算法的鲁棒性要求也更高。

基于视频的行人再识别数据集比较常见的有 PRID-2011^[89]、iLIDS-VID^[90]、MARS^[91],这些数据集包含的人数分别是 200、300、1 261,所有视频的平均序列长度分别为 100 帧、73 帧、58 帧。其中,MARS 是包含人数、视频序列数、相机数最多的数据集,且采用了较为先进的自动检测方法 DPM+GMMCP(Generalized Maximum Multi Clique Problem),但是该数据集平均序列长度短、相机场景单一,不够接近真实的复杂场景。2018 年发布了现有规模最大的视频数据集 LVreID^[92],其包含了 3 个室内相机和 12 个室外相机,共有 3 772 人的 14 943 个视频序列,且平均序列长度为 200 帧,这意味着可以在长序列视频片段中捕捉到更丰富的行人信息。LVreID 使用更快、更精确的检测算法 Faster RCNN^[88]进行包围框检测,生成了更高质量的行人视频序列数据。

由于对多模态的研究工作较少,因此不同类型的跨模态数据集也相应较少。CAVIAR^[93]是 2011 年提出的唯一一个用于跨分辨率行人再识别研究的真实数据集,其中所涵盖的图像分辨率在 17×39~72×144 不等。后来对跨分辨率行人再识别的研究^[77-78]所使用的数据集都是从 VIPeR^[84]、CUHK03^[87]、Market-1501^[2]等现有的数据集中创建的虚拟数据集,其规模大小都与

这些现有的数据集相同。SYSU-MM01^[94]是 2017 年发布的可见光-红外跨模态行人数据集,其包含 6 个摄像头拍摄的图像,包括 2 个在户外黑暗环境下工作红外摄像机和 4 个白天工作的可见光摄像机,可见光摄像机又包含 2 个户内场景和 2 个户外场景。另一个同年发布的红外数据集 RegDB^[95]含 412 人,其中每个人有 10 张来自可见光摄像机的图像和 10 张来自红外摄像机的图像。深度图像是表示场景中每个点与摄像机之间的距离的图像,可以包含行人的外观、骨架结构等信息。BIWI RGBD-ID^[96]是由 RGB-D 摄像机采集的可用于行人再识别研究的深度图像数据集,其包含 50 个训练集序列和 56 个测试集序列,共有 50 个行人。该数据集包括 RGB 图像、行人分割图、深度图像和骨架结构信息。RobotPKURGBD-ID^[97]是使用 Microsoft Kinect SDK 与 Kinect 传感器一起收集的深度图像行人数据集,包含 180 个 90 人的视频序列。该数据集同样由 RGB 图像、行人分割图、深度图像和骨架结构信息组成。SketchRe-ID^[83]是 2018 年发布的一个行人素描数据集,其包括 200 个行人,每个行人有一张来自素描画家的素描图像和两张来自不同可见光摄像机的 RGB 图像。目前进行素描行人再识别的研究较少,因此这是现有的唯一一个用于素描行人再识别研究的数据集。行人再识别数据集随着深度学习技术的发展,其规模越来越大,所包含的摄像机数也越来越多,也更能代表实际监控系统的复杂场景,但数据集的发展不能仅仅止步于此。首先,深度学习技术日益更新,模型也日益复杂,所需要的训练数据量也更加庞大,因此,创建规模更大的数据集来辅助模型的研究和训练是深度学习技术的发展要求,也是必要条件;其次,实际监控系统的环境千变万化,数据模态也多种多样,应用场景也不尽相同,但现有的数据集并不能很好地代表生活中的复杂环境。因此,需要对辅助行人再识别技术发展的数据集有所创新,使之尽可能地包括更多摄像机场景,以便于后续能开发出一套适用于各种实际复杂场景的完备智能监控系统。

5 算法性能比较

由于跨模态方法所涉及的数据集种类较多,而且跨模态方法相对较少,因此本文仅对基于RGB图像的行人再识别方法进行了性能比较,其中包括用于单帧图像数据的方法比较和用于视频数据的方法比较。

5.1 用于单帧图像的方法

表2所示为近几年的一些基于图像的行人再识别方法比较,所涉及的测试数据集是Market-1501^[2]和DukeMTMC-reID^[3],所使用的准确度评估方法是Rank-1和mAP(mean Average Precision)。Rank-1表示所有查询图片所匹配的第一张图像的平均准确率,mAP表示所有查询图片的每个查询图片与候选集匹配的平均精度求和再取平均的值。表中前20项数据是有监督的方法性能比较,其中除了前文介绍方法外,

还增加了其他文献中的一些方法来进行性能比较,包括:IDE (ID-disciminative Embedding)^[1]、BoW (Bag-of-Words)+KISSME (Keep It Simple and Straightforward MEtric)^[2]、Pyramid-Net^[17]、PPS (Part Power Set)^[18]、AANet (Attribute Attention Network)^[98]、Auto-ReID^[99]、OSNet (Omni-Scale Network)^[100]、DG-Net (Discriminative and Generative Network)^[101]、MHN (Mixed High-order attention Network)^[102]、Multi-Scale^[103]、UMFL (Unified Multifaceted Feature Learning)-enabled ABD (Attentive But Diverse)^[104]、SCR (Spatial and Channel partition Representation network)^[105]。后15项数据是无监督的方法性能比较,其中除了前文介绍方法外,也增加了其他文献中的一些方法,包括:BoW (Bag-of-Words)^[2]、HHL (Hetero-Homogeneous Learning)^[106]、ECN (E; Exemplar-invariance. C: Camera-invariance. N: Neighborhood-invariance)^[107]。

表2 图像行人再识别方法性能比较
Tab. 2 Performance comparison of image-based person re-identification methods

类别	方法	来源	描述	Market-1501		DukeMTMC-reID	
				Rank-1	mAP	Rank-1	mAP
有监督	BoW+KISSME ^[2]	ICCV2015	手工特征	44.4	20.8	25.1	12.1
	IDE ^[1]	arXiv2016	CNN基线	72.5	46.0	—	—
	Quadruplet ^[39]	CVPR2017	度量学习,四元组损失	81.5	64.9	73.5	54.3
	PDC ^[19]	ICCV2017	局部特征,姿态估计	84.4	63.4	—	—
	GLAD ^[21]	IEEE2018 TMM2018	局部特征,姿态估计	89.9	73.9	—	—
	HA-CNN ^[24]	CVPR2018	注意力机制	91.2	75.7	80.5	63.8
	AlignedReID ^[26]	arXiv2017	局部特征对齐	91.8	79.3	—	—
	PCB ^[25]	ECCV2018	图像特征分割,基线网络	92.8	77.4	81.1	66.1
	PCB+RPP ^[25]	ECCV2018	图像特征分割	93.8	81.6	83.3	69.2
	AANet (ResNet-50) ^[98]	IEEE TIP2019	注意力机制,属性信息	93.9	82.4	86.4	71.6
	PPS ^[18]	IJCAI2019	多尺度特征,金字塔模型	94.3	85.3	88.7	78.4
	Auto-ReID ^[99]	ICCV2019	神经网络架构搜索	94.5	85.1	88.5	75.1
	OSNet ^[100]	ICCV2019	全尺度特征学习	94.8	84.8	88.6	73.5
	DG-Net ^[101]	CVPR2019	GAN,数据增强	94.8	86.0	86.6	73.4
	MHN-6(PCB) ^[102]	ICCV2019	注意力机制	95.1	85.0	89.1	77.2
	Multi-Scale ^[103]	arXiv2019	注意力机制	95.5	89.0	89.4	79.2
	MGN ^[15]	ACMMM2018	多粒度图像特征分割	95.7	86.9	88.7	78.4
	Pyramid-Net ^[17]	CVPR2019	特征分割,金字塔模型	95.7	88.2	89.0	79.0
	UMFL-enabled ABD ^[104]	arXiv2019	三元组损失,随机擦除	95.7	88.7	89.5	79.3
	SCR ^[105]	IEEE2020	局部特征,注意力机制	95.7	89.0	91.1	81.4
无监督	BoW ^[2]	ICCV2015	手工特征	35.8	14.8	17.1	8.3
	PTGAN ^[55]	ACMMM2018	CycleGAN,风格迁移	38.5	—	27.4	—
	PUL ^[50]	ACMMM2018	聚类	45.5	20.5	30.0	16.4
	CAMEL ^[44]	ICCV2017	聚类,非对称度量	54.5	26.3	—	—
	TJ-AIDL ^[65]	CVPR2018	属性信息+身份信息	58.2	26.5	44.3	23.0
	HHL ^[106]	ECCV2018	CycleGAN,图像转换	62.2	31.4	46.9	27.2
	TAUDL ^[46]	ECCV2018	tracklet关联学习	63.7	41.2	61.7	43.5
	BUC ^[45]	AAAI2019	聚类	66.2	38.2	47.4	27.5
	UTAL ^[47]	TPAMI2019	tracklet关联学习	69.2	46.2	62.3	44.6
	ARN ^[51]	CVPR2018	学习域不变特性	70.3	39.4	60.2	33.4
	CDS ^[52]	ICME2019	聚类,行人属性	71.6	39.9	67.2	42.7
	ECN ^[107]	CVPR2019	CycleGAN	75.1	43.6	63.3	40.4
	CR-GAN+TAUDL ^[58]	ICCV2019	GAN,风格迁移	77.7	54.0	68.9	48.6
	UGA ^[48]	ICCV2019	tracklet,无监督图关联	87.2	70.3	75.0	53.3

在有监督的方法中,2016年以来日渐崛起的深度学习方法使行人再识别方法发展到了一个新的高度,直至2019年,在Market-1501数据集上的Rank-1最高已达到95.7%,2020年最新的方法^[105]在DukeMTMC-reID数据集上的Rank-1也已达到

90%以上,其中最优秀的几个方法大都采用局部特征学习的方法来学习跨摄像机行人的不变特性。然而在局部特征学习的方法中,注意力机制和图像特征分割的方法是近两年来使用较多效果较好的方法,在这些方法的模型中,也会结合使用验证损失

或分类损失,使模型的复杂度上升,使算法的性能更优。行人局部特征作为全局特征的辅助信息,可以在很大程度上提高行人检索精度,因此,在以后的行人再识别研究中,如何获取更鲁棒、更具判别能力的行人局部特征将是一个重要的研究思路。

在无监督的方法中,GAN在近两年受到极大欢迎,使用该方法可以从目标域到源域进行图像风格迁移,生成具有目标域相机风格的行人图片,从而减小不同数据域之间的差距。GAN技术的发展使得无监督行人再识别技术有了极大的突破。除此之外,行人轨迹片段(tracklet)的方法也推动了无监督行人再识别技术的发展,表2中所示的最好的方法UGA^[48]所使用的就是该类方法,在Market-1501和DukeMTMC-reID的Rank-1已分别达到87.2%和75.0%,已经可以媲美一些有监督深度学习方法。相较于有监督的行人再识别技术,无监督方法的研究难度大,但实用性更强,因此无监督行人再识别技术

还有很大的发展空间。

5.2 用于视频的方法

表3所示的是近几年基于视频序列的方法性能比较,所涉及的视频数据集是PRID-2011^[89]和iLIDS-VID^[90],所使用的准确度评估方法是Rank-1和Rank-5。Rank-1表示所有查询序列所匹配的第一段序列的平均准确率,Rank-5表示所有查询序列所匹配的前五段序列的平均准确率。表中共有13项数据,除了前文所提到的一些方法外,也增加了其他文献中的一些方法,包括:TDL(Top-push Distance Learning)^[108]、GAN+Keypoint^[109]、QAN(Quality Aware Network)^[110]、STMP(Spatial and Temporal Mutual Promotion)^[111]、CSA(Competitive Snippet-similarity Aggregation)^[112]、M3D(Multi-scale 3D deep convolutional neural network)^[113]、GLTR(Global-Local Temporal Representations)^[114]。

表3 视频行人再识别方法性能比较

单位:%

Tab. 3 Performance comparison of video-based person re-identification methods

unit:%

方法	来源	描述	PRID-2011		iLIDS-VID	
			Rank-1	Rank-5	Rank-1	Rank5
TDL ^[108]	CVPR2016	手工特征	56.7	80.0	56.3	87.6
RCN ^[70]	CVPR2016	CNN+RNN	70.0	90.0	58.0	84.0
SPPL ^[73]	IEEE TIP2018	度量学习	76.6	95.6	48.7	81.1
ASTPN ^[72]	ICCV2017	注意力机制	77.0	95.0	62.0	86.0
TAM+SRM ^[71]	CVPR2017	注意力机制,RNN	79.0	94.4	55.2	86.5
GAN+Keypoint ^[109]	自动化学报2018	GAN,姿态估计	80.2	96.0	66.3	88.4
SPW ^[74]	AAAI2018	加权距离度量学习	83.5	96.3	69.3	89.6
DAM ^[75]	PR2019	非对称度量学习	87.0	97.0	77.3	94.0
QAN ^[110]	CVPR2017	三元组损失,Softmax损失	90.3	98.2	68.0	86.8
STMP ^[111]	AAAI2019	时域信息+空域信息	92.7	98.8	84.3	96.8
CSA ^[112]	CVPR2018	相似性估计,LSTM	93.0	99.3	85.4	96.7
M3D ^[113]	AAAI2019	双流卷积网络	94.4	100.0	74.0	94.3
GLTR ^[114]	ICCV2019	时间自注意力模型	95.5	100.0	86.0	98.0

由表3可以看出,深度学习的方法远远领先于传统手工特征的方法。神经网络模型可以同时从视频序列中提取空间信息和时间信息,并以各种各样的损失函数作为度量手段,实现了高效的跨摄像机追踪方法。2019年出现的M3D^[113]和GLTR^[114]方法在PRID-2011和iLIDS-VID的Rank-5均达到了100.0%,这些突破令人惊叹,但同时也提高了技术研究难度,后续的研究将需要在规模更大、场景更复杂、更符合实际情况的数据集中寻找突破,但是更大规模的视频数据也意味着需要更多的计算量,这将对硬件设备有了更高的要求。此外,如何把视频行人再识别部署到一个完整的跨摄像机追踪系统,也将是未来行人再识别研究的一个重要部分。

6 总结与展望

行人再识别任务自提出以来一直是计算机视觉领域的研究热点,其对于维护社会安全稳定具有重大意义。该技术的发展至今,虽然已在现有的数据集上达到了很好的性能,但是行人再识别技术在很多方面还存在一些问题,对于未来在该领域的研究,本文展开以下几个方面的讨论:

1)姿态变化、视角变化。由于行人再识别任务的特殊性,不同摄像机下的行人姿态变换和相机视角变化一直以来都是该任务的研究重点。最近的方法大都采用局部特征学习的方法来解决这一问题,利用局部特征和全局特征的互补性,学习

更鲁棒的行人不变特征表示。未来研究将需要解决更多相机视角、更为复杂场景下的姿态变化、视角变化,因此如何克服这些变换学习到强大的行人不变性特征将有待进一步研究。

2)行人遮挡。遮挡类问题不仅是行人再识别研究中一大难点,也是其他计算机视觉任务(如:人脸识别、行人检测等)中的难点。被遮挡的行人由于丢失大量信息而导致提取特征不完整从而影响再识别效果。但是目前对于遮挡行人再识别的研究较少,如果此遮挡问题在行人再识别中有了更进一步的研究,那对整体行人再识别技术的发展都会有一个更大的推动。

3)不准确的行人检测框。在进行行人再识别之前,需要将行人从监控视频中检测出来,这将依赖于行人检测算法的准确度。但是在复杂的真实场景中,总会出现一些行人检测偏差,这会严重影响后续的行人检索精度。对于此类问题,文献[26]提出了局部特征对齐的方法,使相对应的局部特征进行对齐,提高匹配准确率,这种方法也可以解决行人遮挡类问题;但是,若将检测任务和再识别任务集成到一个统一的框架中,是否也是此类问题的一个解决思路,这将有待研究。

4)规模更大的数据集。深度学习依赖于大量数据进行模型训练,且一个规模大、周期长、场景多、环境复杂的数据集更符合实际情况。现有的行人再识别数据集还存在一定的局限性,为了将所研究的行人再识别技术更好地应用于实际监控

系统中,需要创建规模更大的数据集来辅助行人再识别技术研究。虽然GAN技术可以对目前的数据集进行一定程度上的扩充,但是由于GAN生成的图像质量不高,无法真正代替真实的图片数据,因此创建规模更大的真实数据集也是一个待解决的任务。

5)无监督方法。数据标注问题是实际研究中难以避免的问题,一般的有监督方法需要依赖标注好的信息来监督深度模型的训练,但数据标注工作将花费大量财力物力,因此标注工作异常困难。针对此类问题,一些无监督方法被提出,此类方法不需要目标域有标注信息,而是借助聚类、轨迹片段、无监督域适应等方法实现无标注目标域上的行人再识别技术。由于缺乏监督信息,因此无监督方法与有监督方法还存在一定差距,其在现有数据集上的准确率也有待提高。

6)跨模态行人再识别。单一的RGB模态的图像数据很容易受到光照等因素的影响,且不同相机采集的图像分辨率也不同,因此单一模态的行人图像数据将限制行人再识别精度的进一步提高。跨模态行人再识别技术由于引入多模态数据,更加符合实际生活中智能监控系统的设备多样性。但是该类技术的研究还不完善,算法性能有待提高。因此,利用深度学习技术来捕捉不同模态数据之间行人不变特征将是一个更具挑战性的问题。

7)跨摄像机追踪系统的设计与开发。行人再识别独立于单摄像机下的行人检测、行人追踪等技术,是智能监控系统研究的一个分支。将这三项技术结合起来,才能实现一个完整的跨摄像机追踪系统。未来的研究需要将这三者有效结合,实现一个完备且高效的视频监控系统。

参考文献 (References)

- [1] ZHENG L, YANG Y, HAUPTMANN A G. Person re-identification: past, present and future[EB/OL]. [2020-03-20]. <https://arxiv.org/pdf/1610.02984.pdf>.
- [2] ZHENG L, SHEN L, TIAN L, et al. Scalable person re-identification: a benchmark[C]// Proceedings of the 2015 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2015: 1116-1124.
- [3] RISTANI E, SOLERA F, ZOU R S, et al. Performance measures and a data set for multi-target, multi-camera tracking[EB/OL]. [2020-03-20]. <https://arxiv.org/pdf/1609.01775.pdf>.
- [4] 龚锐,丁胜,章超华,等. 基于深度学习的轻量级和多姿态人脸识别方法[J]. 计算机应用, 2020, 40(3): 704-709. (GONG R, DING S, ZHANG C H, et al. Lightweight and multi-pose face recognition method based on deep learning[J]. Journal of Computer Applications, 2020, 40(3): 704-709.)
- [5] PENG C, WANG N, LI J, et al. DLFace: deep local descriptor for cross-modality face recognition[J]. Pattern Recognition, 2019, 90: 161-171.
- [6] 董静,耿达,郭迎港,等. 室内环境下基于R-CNN的光照自适应物体检测[J]. 计算机工程与应用, 2019, 55(2):168-173, 252. (DONG J, GENG D, GUO Y G, et al. Illumination adaptive object detection based on R-CNN under indoor environment[J]. Computer Engineering and Applications, 2019, 55(2):168-173, 252.)
- [7] GHIASI G, LIN T Y, LE Q V. NAS-FPN: learning scalable feature pyramid architecture for object detection[C]// Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 7029-7038.
- [8] 王俊岭,王硕豪. 基于孪生网络的深度学习目标跟踪算法[J]. 计算机工程与设计, 2019, 40(10):3014-3019. (WANG J L, WANG S H. Deep learning siamese networks for object tracking[J]. Computer Engineering and Design, 2019, 40(10): 3014-3019.)
- [9] SUN S, AKHTAR N, SONG H, et al. Deep affinity network for multiple object tracking[EB/OL]. [2020-03-20]. <https://arxiv.org/pdf/1810.11780.pdf>.
- [10] LI W, ZHAO R, XIAO T, et al. DeepReID: deep filter pairing neural network for person re-identification[C]// Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014: 152-159.
- [11] WANG F, ZUO W, LIN L, et al. Joint learning of single-image and cross-image representations for person re-identification[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 1288-1296.
- [12] ZHENG L, ZHANG H, SUN S, et al. Person re-identification in the wild[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 3346-3355.
- [13] XIAO T, LI H, OUYANG W, et al. Learning deep feature representations with domain guided dropout for person re-identification[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 1249-1258.
- [14] XIA B, GONG Y, ZHANG Y, et al. Second-order non-local attention networks for person re-identification[EB/OL]. [2019-12-10]. <https://arxiv.org/pdf/1909.00295.pdf>.
- [15] WANG G, YUAN Y, CHEN X, et al. Learning discriminative features with multiple granularities for person re-identification[C]// Proceedings of the 26th ACM International Conference on Multimedia. New York: ACM, 2018: 274-282.
- [16] FU Y, WEI Y, ZHOU Y, et al. Horizontal pyramid matching for person re-identification[C]// Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2019: 8295-8302.
- [17] ZHENG F, DENG C, SUN X, et al. Pyramidal person re-identification via multi-loss dynamic training[C]// Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 8506-8514.
- [18] SHEN Y, JI R, HONG X, et al. A part power set model for scale-free person retrieval[C]// Proceedings of the 28th International Joint Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2019: 3397-3403.
- [19] SU C, LI J, ZHANG S, et al. Pose-driven deep convolutional model for person re-identification[C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 3980-3989.
- [20] ZHAO H, TIAN M, SUN S, et al. Spindle net: person re-identification with human body region guided feature decomposition and fusion[C]// Proceedings of the 2017 IEEE

- Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 907-915.
- [21] WEI L, ZHANG S, YAO H, et al. GLAD: global-local-alignment descriptor for scalable person re-identification[J]. IEEE Transactions on Multimedia, 2019, 21(4): 986-999.
- [22] LIU H, FENG J, QI M, et al. End-to-end comparative attention networks for person re-identification[J]. IEEE Transactions on Image Processing, 2017, 26(7): 3492-3506.
- [23] LIU X, ZHAO H, TIAN M, et al. HydraPlus-Net: attentive deep features for pedestrian analysis[C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 350-359.
- [24] LI W, ZHU X, GONG S. Harmonious attention network for person re-identification[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 2285-2294.
- [25] SUN Y, ZHENG L, YANG Y, et al. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline)[C]// Proceedings of the 2018 European Conference on Computer Vision, LNCS 11208. Cham: Springer, 2018: 501-518.
- [26] ZHANG X, LUO H, FAN X, et al. AlignedReID: surpassing human-level performance in person re-identification[EB/OL]. [2019-12-10]. <https://arxiv.org/pdf/1711.08184.pdf>.
- [27] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 3431-3440.
- [28] WEI S E, RAMAKRISHNA V, KANADE T, et al. Convolutional pose machines[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 4724-4732.
- [29] INSAFUTDINOV E, PISHCHULIN L, ANDRES B, et al. DeeperCut: a deeper, stronger, and faster multi-person pose estimation model[C]// Proceedings of the 2016 European Conference on Computer Vision, LNCS 9910. Cham: Springer, 2016: 34-50.
- [30] XU K, BA J, KIROS R, et al. Show, attend and tell: Neural image caption generation with visual attention[EB/OL]. [2019-12-10]. <https://arxiv.org/pdf/1502.03044.pdf>.
- [31] ZHANG Z, ZOU Y, GAN C. Textual sentiment analysis via three different attention convolutional neural networks and cross-modality consistent regression[J]. Neurocomputing, 2018, 275: 1407-1415.
- [32] 胡正平,刁鹏成,张瑞雪,等. 基于注意力机制的时间分组深度网络行为识别算法[J]. 模式识别与人工智能, 2019, 32(10):892-900. (HU Z P, DIAO P C, ZHANG R X, et al. Temporal group deep network action recognition algorithm based on attention mechanism[J]. Pattern Recognition and Artificial Intelligence, 2019, 32(10):892-900.)
- [33] LIAO S, HU Y, ZHU X, et al. Person re-identification by local maximal occurrence representation and metric learning[C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 2197-2206.
- [34] KÖSTINGER M, HIRZER M, WOLHLHART P, et al. Large scale metric learning from equivalence constraints[C]// Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2012: 2288-2295.
- [35] VARIOR R R, HALOI M, WANG G. Gated siamese convolutional neural network architecture for human re-identification[C]// Proceedings of the 2016 European Conference on Computer Vision, LNCS 9912. Cham: Springer, 2016: 791-808.
- [36] WANG J, SONG Y, LEUNG T, et al. Learning fine-grained image similarity with deep ranking[C]// Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014: 1386-1393.
- [37] CHENG D, GONG Y, ZHOU S, et al. Person re-identification by multi-channel parts-based CNN with improved triplet loss function [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 1335-1344.
- [38] HERMANS A, BEYER L, LEIBE B. In defense of the triplet loss for person re-identification[EB/OL]. [2019-12-10]. <https://arxiv.org/pdf/1703.07737.pdf>.
- [39] CHEN W, CHEN X, ZHANG J, et al. Beyond triplet loss: a deep quadruplet network for person re-identification[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 1320-1329.
- [40] ZHONG W, ZHANG T, JIANG L F, et al. Discriminative representation learning for person re-identification via multi-loss training[J]. Journal of Visual Communication and Image Representation, 2019, 62: 267-278.
- [41] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: a unified embedding for face recognition and clustering[C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 815-823.
- [42] GENG M, WANG Y, XIANG T, et al. Deep transfer learning for person re-identification[EB/OL]. [2019-12-10]. <https://arxiv.org/pdf/1611.05244.pdf>.
- [43] SONG H O, XIANG Y, JEGELKA S, et al. Deep metric learning via lifted structured feature embedding[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 4004-4012.
- [44] YU H, WU A, ZHENG W. Cross-view asymmetric metric learning for unsupervised person re-identification[C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 994-1002.
- [45] LIN Y, DONG X, ZHENG L, et al. A bottom-up clustering approach to unsupervised person re-identification[C]// Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2019: 8738-8745.
- [46] LI M, ZHU X, GONG S. Unsupervised person re-identification by deep learning tracklet association[C]// Proceedings of the 2018 European Conference on Computer Vision, LNCS 11208. Cham: Springer, 2018: 772-788.
- [47] LI M, ZHU X, GONG S. Unsupervised tracklet person re-identification[EB/OL]. [2019-12-10]. <https://arxiv.org/pdf/>

- 1903.00535.pdf.
- [48] WU J, LIU H, YANG Y, et al. Unsupervised graph association for person re-identification[C]// Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 8320-8329.
- [49] ZHANG X, CAO J, SHEN C, et al. Self-training with progressive augmentation for unsupervised cross-domain person re-identification[C]// Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 8221-8230.
- [50] 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, 2018, 14(4): No. 83.
- [51] LI Y J, YANG F E, LIU Y C, et al. Adaptation and re-identification network: an unsupervised deep transfer learning approach to person re-identification[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE, 2018: 285-291.
- [52] WU J, LIAO S, LEI Z, et al. Clustering and dynamic sampling based unsupervised domain adaptation for person re-identification [C]// Proceedings of the 2019 IEEE International Conference on Multimedia and Expo. Piscataway: IEEE, 2019: 886-891.
- [53] LONG M, ZHU H, WANG J, et al. Deep transfer learning with joint adaptation networks[C]// Proceedings of the 34th International Conference on Machine Learning. New York: JMLR. org, 2017: 2208-2217.
- [54] CHEN C, CHEN Z, JIANG B, et al. Joint domain alignment and discriminative feature learning for unsupervised deep domain adaptation[C]// Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2019: 3296-3303.
- [55] WEI L, ZHANG S, GAO W, et al. Person transfer GAN to bridge domain gap for person re-identification[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 79-88.
- [56] ZHONG Z, ZHENG L, ZHENG Z, et al. Camera style adaptation for person re-identification[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 5157-5166.
- [57] ZHOU S, KE M, LUO P. Multi-camera transfer GAN for person re-identification[J]. Journal of Visual Communication and Image Representation, 2019, 59: 393-400.
- [58] CHEN Y, ZHU X, GONG S. Instance-guided context rendering for cross-domain person re-identification[C]// Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 232-242.
- [59] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]// Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2014:2672-2680.
- [60] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 2242-2251.
- [61] CHOI Y, CHOI M, KIM M, et al. StarGAN: unified generative adversarial networks for multi-domain image-to-image translation [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 8789-8797.
- [62] WEI X, LUO J, WU J, et al. Selective convolutional descriptor aggregation for fine-grained image retrieval[J]. IEEE Transactions on Image Processing, 2017, 26(6): 2868-2881.
- [63] ZHENG Z, ZHENG L, YANG Y. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro[C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 3774-3782.
- [64] ZHANG C, WU L, WANG Y. Crossing generative adversarial networks for cross-view person re-identification[J]. Neurocomputing, 2019, 340: 259-269.
- [65] WANG J, ZHU X, GONG S, et al. Transferable joint attribute-identity deep learning for unsupervised person re-identification[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 2275-2284.
- [66] ZHU X, MORERIO P, MURINO V. Unsupervised domain-adaptive person re-identification based on attributes[C]// Proceedings of the 2019 IEEE International Conference on Image Processing. Piscataway: IEEE, 2019: 4110-4114.
- [67] GHEISSARI N, SEBASTIAN T B, HARTLEY R. Person reidentification using spatiotemporal appearance[C]// Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2006: 1528-1535.
- [68] TRUONG CONG D N, ACHARD C, KHOUDOUR L, et al. Video sequences association for people re-identification across multiple non-overlapping cameras[C]// Proceedings of the 15th International Conference on Image Analysis and Processing, LNCS 5716. Berlin: Springer, 2009: 179-189.
- [69] BEGAGKAR-GALA A, SHAH S K. Multiple person re-identification using part based spatio-temporal color appearance model[C]// Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops. Piscataway: IEEE, 2011: 1721-1728.
- [70] MCLAUGHLIN N, MARTINEZ DEL RINCON J, MILLER P. Recurrent convolutional network for video-based person re-identification[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 1325-1334.
- [71] ZHOU Z, HUANG Y, WANG W, et al. See the forest for the trees: Joint spatial and temporal recurrent neural networks for video-based person re-identification[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 6776-6785.
- [72] XU S, CHENG Y, GU K, et al. Jointly attentive spatial-temporal pooling networks for video-based person re-identification[C]//

- Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 4743-4752.
- [73] ZHU X, JING X, YOU X, et al. Video-based person re-identification by simultaneously learning intra-video and inter-video distance metrics[J]. IEEE Transactions on Image Processing, 2018, 27(11): 5683-5695.
- [74] HUANG W, LIANG C, YU Y, et al. Video-based person re-identification via self paced weighting[C]// Proceedings of the 32nd AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2018:2273-2280.
- [75] MENG J, WU A, ZHENG W. Deep asymmetric video-based person re-identification[J]. Pattern Recognition, 2019, 93: 430-441.
- [76] MENG J, WU S, ZHENG W. Weakly supervised person re-identification[C]// Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 760-769.
- [77] JIAO J, ZHENG W S, WU A, et al. Deep low-resolution person re-identification[C]// Proceedings of the 32nd AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2018: 6967-6974.
- [78] ZHUANG Z, AI H, CHEN L, et al. Cross-resolution person re-identification with deep antithetical learning[C]// Proceedings of the 14th Asian Conference on Computer Vision, LNCS 11363. Cham: Springer, 2018: 233-248.
- [79] WANG G, ZHANG T, CHENG J, et al. RGB-infrared cross-modality person re-identification via joint pixel and feature alignment[C]// Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 3622-3631.
- [80] WANG Z, WANG Z, ZHENG Y Q, et al. Learning to reduce dual-level discrepancy for infrared-visible person re-identification[C]// Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 618-626.
- [81] HAFNER F, BHUIYAN A, KOOIJ J F P, et al. A cross-modal distillation network for person re-identification in RGB-depth[EB/OL]. [2019-12-10]. <https://arxiv.org/pdf/1810.11641.pdf>.
- [82] LI S, XIAO T, LI H, et al. Identity-aware textual-visual matching with latent co-attention[C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 1908-1917.
- [83] PANG L, WANG Y, SONG Y Z, et al. Cross-domain adversarial feature learning for sketch re-identification[C]// Proceedings of the 26th ACM International Conference on Multimedia. New York: ACM, 2018: 609-617.
- [84] GRAY D, TAO H. Viewpoint invariant pedestrian recognition with an ensemble of localized features[C]// Proceedings of the 2008 European Conference on Computer Vision, LNCS 5302. Berlin: Springer, 2008: 262-275.
- [85] LOY C C, XIANG T, GONG S. Multi-camera activity correlation analysis[C]// Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2009: 1988-1995.
- [86] LI W, ZHAO R, WANG X. Human reidentification with transferred metric learning[C]// Proceedings of the 11th Asian Conference on Computer Vision, LNCS 7724. Berlin: Springer, 2012: 31-44.
- [87] ZHONG Z, ZHENG L, CAO D, et al. Re-ranking person re-identification with k-reciprocal encoding[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 3652-3661.
- [88] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]// Proceedings of the 28th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2015: 91-99.
- [89] HIRZER M, BELEZNAI C, ROTH P M, et al. Person re-identification by descriptive and discriminative classification[C]// Proceedings of the 2011 Scandinavian Conference on Image Analysis, LNCS 6688. Berlin: Springer, 2011: 91-102.
- [90] WANG T, GONG S, ZHU X, et al. Person re-identification by video ranking[C]// Proceedings of the 2014 European Conference on Computer Vision, LNCS 8692. Cham: Springer, 2014: 688-703.
- [91] ZHENG L, BIE Z, SUN Y, et al. MARS: a video benchmark for large-scale person re-identification[C]// Proceedings of the 2016 European Conference on Computer Vision, LNCS 9910. Cham: Springer, 2016: 868-884.
- [92] LI J, ZHANG S, WANG J, et al. LVreID: person re-identification with long sequence videos[EB/OL]. [2019-12-10]. <https://arxiv.org/pdf/1712.07286v1.pdf>.
- [93] CHENG D S, CRISTANI M, STOPPA M, et al. Custom pictorial structures for re-identification[C]// Proceedings of the 22nd British Machine Vision Conference. Durham: BMVA Press, 2011: No. 686.
- [94] WU A, ZHENG W, YU H, et al. RGB-infrared cross-modality person re-identification[C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 5390-5399.
- [95] NGUYEN D T, HONG H G, KIM K W, et al. Person recognition system based on a combination of body images from visible light and thermal cameras[J]. Sensors, 2017, 17(3): No. 605.
- [96] MUNARO M, FOSSATI A, BASSO A, et al. One-shot person re-identification with a consumer depth camera[M]// GONG S, CRISTANI M, YAN S, et al. Person Re-Identification, Advances in Computer Vision and Pattern Recognition. London: Springer, 2014: 161-181.
- [97] LIU H, HU L, MA L. Online RGB-D person re-identification based on metric model update[J]. CAAI Transactions on Intelligence Technology, 2017, 2(1): 48-55.
- [98] TAY C P, ROY S, YAP K H. AANet: attribute attention network for person re-identifications[C]// Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 7127-7136.
- [99] QUAN R, DONG X, WU Y, et al. Auto-ReID: searching for a part-aware ConvNet for person re-identification[EB/OL]. [2019-12-10]. <https://arxiv.org/pdf/1903.09776.pdf>.

- [100] ZHOU K, YANG Y, CAVALLARO A, et al. Omni-scale feature learning for person re-identification[EB/OL]. [2020-03-20]. <https://arxiv.org/pdf/1905.00953.pdf>.
- [101] ZHENG Z, YANG X, YU Z, et al. Joint discriminative and generative learning for person re-identification[C]// Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 2133-2142.
- [102] CHEN B, DENG W, HU J. Mixed high-order attention network for person re-identification[C]// Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 371-381.
- [103] WU D, WANG C, WU Y, et al. Attention deep model with multi-scale deep supervision for person re-identification[EB/OL]. [2020-03-20]. <https://arxiv.org/pdf/1911.10335.pdf>.
- [104] YAN C, PANG G, BAI X, et al. Unified multifaceted feature learning for person re-identification[EB/OL]. [2020-03-20]. <https://arxiv.org/pdf/1911.08651.pdf>.
- [105] CHEN H, LAGADEC B, BREMOND F. Learning discriminative and generalizable representations by spatial-channel partition for person re-identification[C]// Proceedings of the 2020 IEEE Winter Conference on Applications of Computer Vision. Piscataway: IEEE, 2020: 2472-2481.
- [106] ZHONG Z, ZHENG L, LI S, et al. Generalizing a person retrieval model hetero-and homogeneously[C]// Proceedings of the 2018 European Conference on Computer Vision, LNCS 11217. Cham: Springer, 2018: 176-192.
- [107] ZHONG Z, ZHENG L, LOU Z, et al. Invariance matters: exemplar memory for domain adaptive person re-identification [C]// Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 598-607.
- [108] YOU J, WU A, LI X, et al. Top-push video-based person re-identification[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 1345-1353.
- [109] 刘一敏, 蒋建国, 齐美彬, 等. 融合成对抗网络和姿态估计的视频行人再识别方法[J]. 自动化学报, 2020, 46(3): 576-584. (LIU Y M, JIANG J G, QI M B, et al. Video-based person re-identification method based on GAN and pose estimation[J]. Acta Automatica Sinica, 2020, 46(3): 576-584.)
- [110] LIU Y, YAN J, OUYANG W. Quality aware network for set to set recognition[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 4694-4703.
- [111] LIU Y, YUAN Z, ZHOU W, et al. Spatial and temporal mutual promotion for video-based person re-identification[C]// Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2019: 8786-8793.
- [112] CHEN D, LI H, XIAO T, et al. Video person re-identification with competitive snippet-similarity aggregation and co-attentive snippet embedding[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 1169-1178.
- [113] LI J, ZHANG S, HUANG T. Multi-scale 3D convolution network for video based person re-identification[C]// Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2019: 8618-8625.
- [114] LI J, ZHANG S, WANG J, et al. Global-local temporal representations for video person re-identification[C]// Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 3957-3966.

This work is partially supported by the Key Project of National Natural Science Foundation of China (U1435215), the National Natural Science Foundation of China (U1603115), the National Key Research and Development Program of China (2017YFC0820702-3), the Natural Science Foundation of Xinjiang Uygur Autonomous Region (2017D01C042).

WEI Wenyu, born in 1995, M. S. candidate. Her research interests include computer vision, person re-identification, information security.

YANG Wenzhong, born in 1971, Ph. D., associate professor. His research interests include internet public opinion, intelligence analysis, information security, wireless sensor network.

MA Guoxiang, born in 1993, M. S. candidate. His research interests include computer vision, machine learning, information security.

HUANG Mei, born in 1996, M. S. candidate. Her research interests include computer vision, remote sensing image processing.