中国科学: 信息科学 2021 年 第 52 卷 第 12 期: 1979–2015
SCIENTIA SINICA Informationis
《中国科学》杂志社
SCIENCE CHINA PRESS
评述

智能视频监控关键技术: 行人再识别研究综述

赵才荣[1*], 齐鼎[1], 窦曙光[1], 涂远鹏[1], 孙添力[1], 柏松[2], 蒋忻洋[3], 白翔[2*], 苗夺谦[1*]

1. 同济大学电子与信息工程学院, 上海 201804
2. 华中科技大学电子信息与通信学院, 武汉 430074
3. 微软亚洲研究院 (上海), 上海 200232
* 通信作者. E-mail: zhaocairong@tongji.edu.cn, xbai@hust.edu.cn, dqmiao@tongji.edu.cn

收稿日期: 2021–06–23; 修回日期: 2021–09–14; 接受日期: 2021–10–20; 网络出版日期: 2021–12–17

国家自然科学基金 (批准号: 62076184, 61673299, 61976160, 62076182)、上海科技创新行动计划 (批准号: 20511100700)、上海市级科技重大专项 —— 人工智能基础理论与关键核心技术 (批准号: 2021SHZDZX0100) 和中央高校基本科研业务费专项资助项目

摘要 行人再识别 (person re-identification, ReID) 旨在解决跨摄像头跨场景下目标行人的关联与匹配, 作为智能视频监控系统的关键环节, 对维护社会公共秩序具有重大作用. 为了深入了解行人再识别研究现状和加速推进国内行人再识别相关研究及技术落地, 本文对该领域国家自然科学基金申报数量、资助力度以及地理分布情况进行统计, 并针对近年来发表在国际顶级会议和期刊上的行人再识别研究进行全面梳理. 具体地, 首先阐述一个标准行人再识别算法流程, 并总结其中 3 个关键技术: 表征学习、度量学习和重排序优化. 随后, 列举了实际开放场景中面临的主要难点与挑战, 并据此概括了 7 种开放行人再识别任务: 遮挡、无监督、半监督、跨模态、场景行人搜索、对抗鲁棒和快速检索. 此外, 本文整理了标准行人再识别和开放行人再识别的代表性数据集, 并且对一些代表性行人再识别算法进行比较. 最后本文对行人再识别的未来发展趋势进行展望.

关键词 行人再识别, 智能视频分析, 深度学习, 表征学习, 度量学习

1 引言

随着社会和经济的快速发展, 城市公共安全问题受到越来越多的关注. 视频监控作为保障城市安全的重要技术手段, 被广泛应用于街道、学校、商场等人流密集的公共场所. 城市视频监控网络每时每刻都在获取视频数据, 目前的视频监控技术主要以 “人工分析” 为主, 结合简单的智能化方法来处理分析视频数据, 这导致了诸如 “视频在、找不到”, “找得到、找太久”, “有服务、不可靠” 等视频监控技术应用的瓶颈. 因此, 如何实现智能视频监控, 尤其是对行人数据的智能处理、可疑行为的自动研判, 是新时代公共安全领域的迫切需要.

引用格式: 赵才荣, 齐鼎, 窦曙光, 等. 智能视频监控关键技术: 行人再识别研究综述. 中国科学: 信息科学, 2021, 52: 1979–2015, doi: 10.1360/SSI-2021-0211
Zhao C R, Qi D, Dou S G, et al. Key technology for intelligent video surveillance: a review of person re-identification (in Chinese). Sci Sin Inform, 2021, 52: 1979–2015, doi: 10.1360/SSI-2021-0211

© 2021《中国科学》杂志社

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

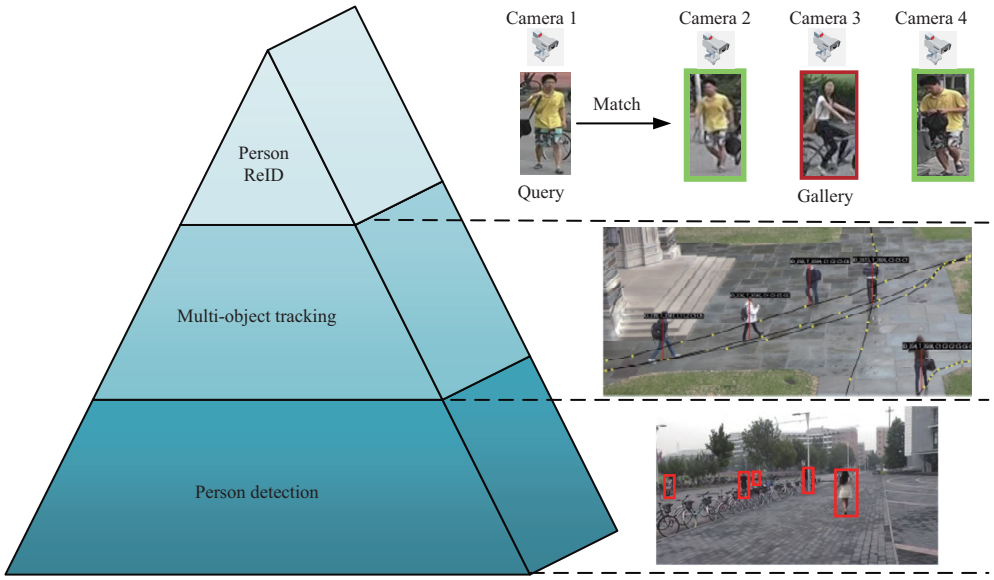

图 1 (网络版彩图) 智能视频监控

Figure 1 (Color online) Intelligent video surveillance

如图 1 所示, 智能视频监控利用模式识别和计算机视觉技术对海量监控数据进行处理和分析, 在不需要人力干预的前提下, 能够实现对监控目标的自动检测、跟踪以及识别. 行人再识别 (person re-identification, ReID) 作为智能视频监控体系中的关键一环, 旨在解决跨摄像头跨场景下目标行人的关联与匹配. 具体而言, 图 1 中摄像机 1 捕获的目标行人作为待查询对象, 摄像机 2~4 拍摄的行人图像组成候选图库, 将待查询对象与候选图库中的行人逐个匹配, 并在候选图库中找到与待查询行人相同类别的行人, 从而实现跨摄像头跨场景下目标行人关联与匹配.

2 研究背景

早期的行人再识别研究可以追溯到多摄像机追踪领域 [1]. 2005 年, 阿姆斯特丹大学 (University of Amsterdam) 的 Zajdel 等 [2] 首次提出 "行人再识别 (person re-identification)" 概念. 2007 年, Gary 等 [3] 发布了第一个用于行人再识别研究的数据集 VIPeR. 随后的十几年, 行人再识别作为一个独立的计算机视觉任务, 受到越来越多国内外学者和研究机构的关注.

国际上, 著名的行人再识别研究机构有美国德州大学圣安东尼奥分校 (University of Texas at San Antonio)、卡耐基梅隆大学 (Carnegie Mellon University)、英国伦敦大学玛丽女王学院 (Queen Mary University of London)、澳大利亚悉尼科技大学 (University of Technology Sydney)、以色列理工学院 (Technion-Israel Institute of Technology)、新加坡南洋理工大学 (Nanyang Technological University)、新加坡国立大学 (National University of Singapore) 和韩国首尔大学 (Seoul National University) 等. 近年来, 国内对于行人再识别研究的热度不断增长, 图 2 展示了近 10 年来相关国家自然科学基金的立项情况[1)], 可以看出, 项目总数和资助金额都处于稳步上升的趋势, 并且在 2018 年, 该领域的项目数量

1) 国家自然科学基金统计数据来源: http://kd.nsfc.gov.cn/baseQuery/supportQuery. http://www.letpub.com.cn/index.php?page=grant#opennewwindow.

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

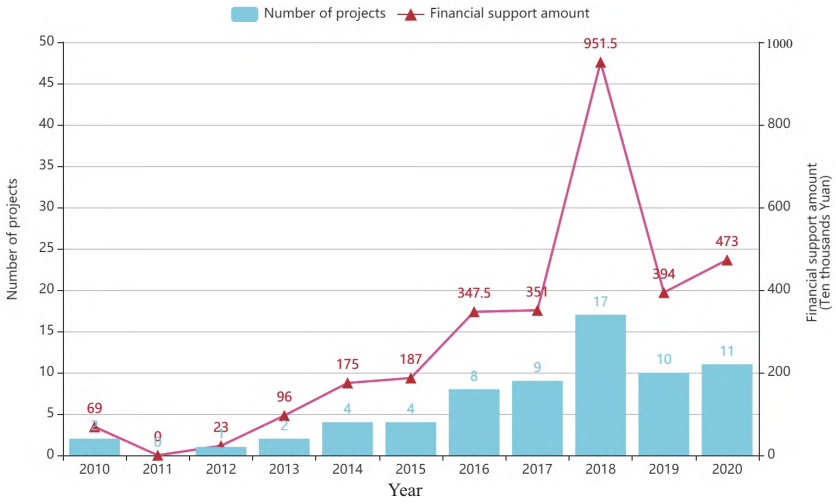

图 2 (网络版彩图) 行人再识别基金项目数量与资助金额

Figure 2 (Color online) Number of projects and financial support amount of National Natural Science Foundation of China for person ReID research

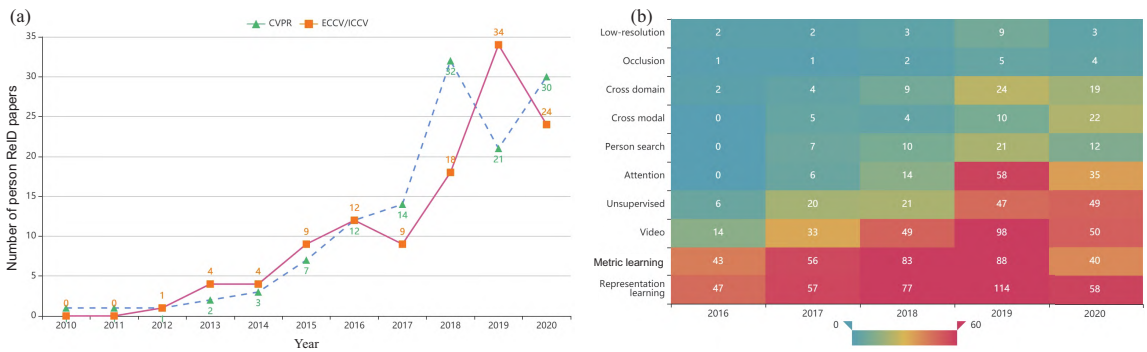

图 3 (网络版彩图) (a) 行人再识别论文在顶级会议中的数量; (b) 近年行人再识别关键词频

Figure 3 (Color online) (a) Number of top conference papers in recent years; (b) key word frequency of person ReID in recent years

和资助金额达到峰值, 这是因为当年由国家自然科学基金和企业共同资助的多个联合项目立项. 近年来, 国内行人再识别研究活跃的高校和研究机构主要包括中山大学智能科学与系统实验室、北京大学多媒体学习研究组、中国科学院自动化研究所模式识别国家重点实验室、中国科学院自动化研究所生物识别与安全研究中心、香港中文大学多媒体实验室、同济大学视觉与智能学习实验室、华中科技大学视觉与深度学习研究组、武汉大学国家多媒体软件工程技术研究中心、清华大学信息处理研究所、腾讯优图、微软亚洲研究院、旷世科技、阿里巴巴、京东 AI 研究院、商汤科技、依图科技、云从科技等.

行人再识别的众多科研成果也相继发表在计算机视觉领域顶级会议和顶级刊物上. 图 3(a) 展示了近 10 年来计算机视觉三大顶级会议 CVPR, ICCV, ECCV 收录的行人再识别论文数量变化情况. 在 TPAMI, TIP, TMM, TOMM, TCSVT 等国际顶级刊物上, 也发表了大量优秀的行人再识别研究成果. 随着深度学习的快速发展, 其在多个重要的计算机视觉任务上取得了重大突破. 因此, 基于深度学习的行人再识别研究也成为了近年来的主流. 图 3(b) 展示了近年来基于深度学习的行人再识别研究

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

热点及变化趋势[2)]: 表征学习和度量学习持续得到研究人员的关注, 近年来一直保持着较高的研究热度; 基于视频、无监督、注意力机制的行人再识别研究热度显著上升; 此外, 跨模态、无监督、遮挡问题也开始得到研究人员更多关注. 具体地, 2016~2017 年, 大量研究集中在行人的特征表达和相似性度量上, 这一时期提出的数据集规模较小, 涉及的开放性问题也较少, 因此只能供研究者在相对理想环境下进行标准行人再识别研究. 2018~2020 年, 注意力研究升温, 标准行人再识别性能有了进一步提升. 此外, 更多的基于开放问题的行人再识别研究被提出. 针对遮挡、跨模态、无监督、低分辨率等问题的研究成为另一大趋势. 在这一阶段, 规模更大、考虑问题全面、场景拟真的数据集被提出也是推动这一研究趋势的重要原因.

3 标准行人再识别

标准行人再识别是指在相对理想环境下, 仅考虑少量光照、姿态、视角变化因素, 更多关注于行人自身的特征表达和相似性度量. 图 4 是一个标准行人再识别算法流程: 首先从监控系统中获取原始视频数据; 随后对视频抽帧, 得到一系列场景图片; 然后, 使用行人检测算法将场景中的行人检测并裁剪出来; 接着, 通过表征学习的方法提取行人的鲁棒特征表示; 再利用度量学习方法计算行人之间的相似度得分, 并从高到低进行排序. 根据排序结果完成目标行人的再识别; 此外, 还可以在原始排序基础上, 通过重排序算法进一步优化性能. 本节将着重阐述标准行人再识别的 3 个关键技术: 表征学习、度量学习和重排序优化.

3.1 表征学习

表征学习方法旨在根据行人外观, 学习行人的一般特征表示, 并要求对光照、姿态、视角等变化具有一定抗干扰能力. 传统的表征学习方法主要关注于设计手工特征, 包括底层的颜色、纹理和高层属性、语义等特征, 这些手工设计特征的过程往往十分烦琐并且难以应对复杂的背景变化. 随着深度学习的发展, 利用深度网络从大量数据中学习网络参数, 并自动提取强判别力行人特征的方法逐渐成为主流. 本小节将详细介绍基于深度学习的行人表征方法, 并根据数据类型不同, 分别从图像特征学习、视频特征学习两方面进行阐述.

3.1.1 图像特征学习

在图像特征学习中, 根据学习特征类型, 行人特征表示可以分为全局特征表示和局部特征表示 [4] 两种. 以往的一些方法关注于学习行人图像全局特征 [5~8], 即从全局信息中提取行人不变的特征, 但这对图像的成像质量要求较高, 尤其是对图像中行人外观的变化: 姿态变化、视角变化、遮挡等问题较为敏感, 从而影响识别精度. 结合卷积神经网络的局部相关性, 而后大多的行人再识别方法大都以结合行人局部特征和全局特征的方法来获取有效判别信息, 以克服上述困难. 与单一全局特征学习相比, 局部特征辅助全局特征学习和表示在行人再识别领域得到更为广泛的研究和应用. 如图 5 所示, 目前基于局部特征的行人再识别主要分为基于水平分块对齐、基于姿态估计、基于注意力机制.

• **基于水平分块对齐.** 基于水平分块的方法通过将一张完整的行人图像均分成几个固定大小的条纹块, 进而提取各个部件块的有效特征. Sun 等 [9] 提出水平部件分割网络 (part convolutional baseline, PCB) 将行人图片均匀划分为 6 个部件并提取特征, 模型分块计算分类损失并联合优化训练. 考虑到同一部件在不同图像中可能因为行人没有对齐的原因而具有不同的语义, 进一步提出细化局部分区的

2) 图 3(b) 涉及关键词: 表征学习、度量学习、视频、无监督、注意力、行人搜索、跨模态、跨域、遮挡、低分辨率.

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

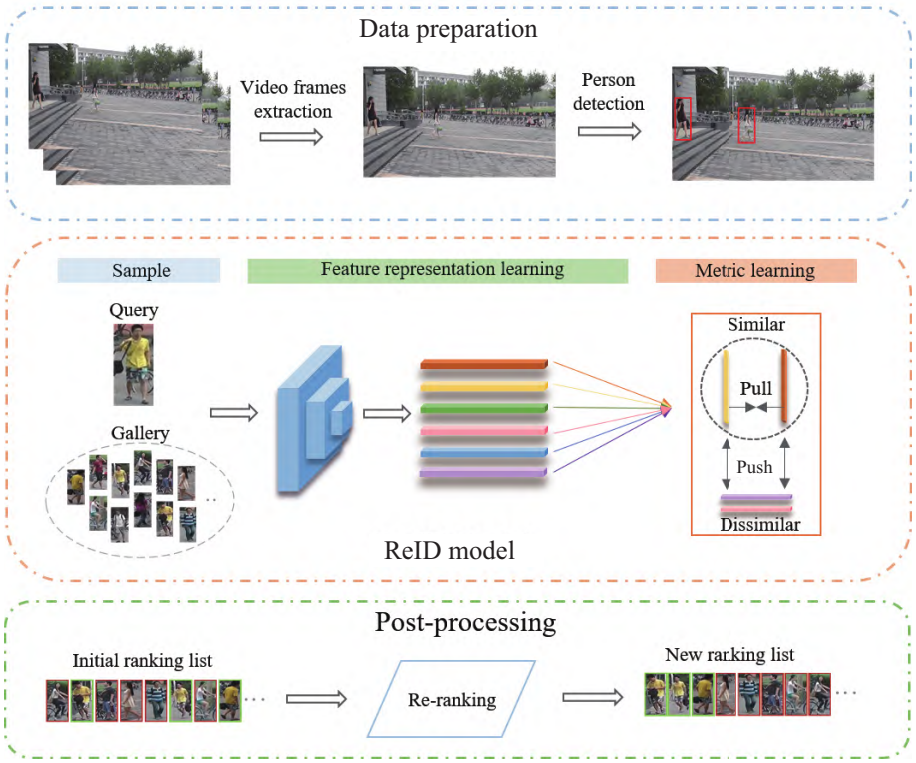

图 4 (网络版彩图) 标准行人再识别算法流程

Figure 4 (Color online) Standard person re-identification algorithm flow

图 5 (网络版彩图) 3 种基于局部特征的表征学习形式. (a) 基于水平分块对齐; (b) 基于姿态估计; (c) 基于注意力机制

Figure 5 (Color online) Three different types of feature representation learning. (a) Part-level alignment-based; (b) pose estimation-based; (c) attention mechanism-based

池化方法 RPP (refined part pooling), 目的是重新分配每个局部区域内的离群点, 加强每个局部部件内部语义的一致性, 以获得多尺度的局部特征, 弥补全局特征丢失的信息. Fu 等 [10] 在 PCB 网络的基础上, 使用空间金字塔池化网络 (horizontal pyramid matching, HPM), 在不同金字塔尺度下进行特征的

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

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优化特征表示. Zheng 等 [11] 同样提出了一个水平分块由粗到细的金字塔模型, 通过不同尺度的水平切分, 同时整合全局和局部特征之间的渐进线索, 在特征表示和匹配上实现了更佳的性能. Bai 等 [12] 通过引入 LSTM 模块顺序建模行人的局部信息, 以此学习身体各部件间的上下文依赖关系, 从而克服了序列级别行人表示不一致问题. Zhao 等 [13] 通过定义重构矩阵引入中间层特征, 采用非精确增广拉格朗日乘子 (inexact augmented Lagrange multiplier, IALM) 算法求解, 然后结合低层特征和判别传递特征描述行人图像的外观, 为了融合局部和全局特征, 还设计了联合传递约束来求解最优函数.

• **基于姿态估计.** 基于姿态估计的方法是将人体姿态信息应用到行人再识别问题上. 首先利用姿态估计器提取图像中行人若干个骨骼关键点信息, 再结合 CNN 模型提取对应区域的行人局部特征, 实现特征表示. 但是, 仅提取关键特征点等局部特征, 有可能丢失一些重要的细节, 例如行人背包、雨伞等. 因此, 基于姿态评估的方法需要同时使用全局特征. Zheng 等 [14] 提出了一个姿态嵌入不变性的行人描述符 (pose invariant embedding, PIE), 实现描述特征与标准行人姿态对齐, 解决因姿态估计错误和信息丢失造成的识别性能下降问题. Zhao 等 [15] 提出 Spindle Net, 利用人体区域信息和姿态信息引导学习 7 个不同语义的身体部件, 并与全局特征融合进一步提升表征性能. 为了获得更加精确的人体局部特征, Wei 等 [16] 结合关键特征点信息构建 4 个局部特征提取子网, 分别提取人体的头部、上半身体、下半身体 3 个粗粒度部件特征和全局特征, 最后得到全局 – 局部对齐描述符 (global-local-alignment descriptor, GLAD) 进行联合表示. Xu 等 [17] 提出了利用姿态信息引导网络 (pose-guided part attention, PPA) 学习行人身体区域特征, 自适应忽略不需要的背景信息, 在一定程度上解决图像部分遮挡的问题. Su 等 [18] 提出一种姿态驱动的深度卷积模型 (pose-driven deep convolutional, PDC) 来改进深度网络的特征提取和匹配模型. 模型有效地利用行人局部语义部件的线索, 以缓解姿态变化造成的特征学习的误差, 从而学到稳健的特征表示. 为了匹配全身和局部的特征, 进一步设计了一个姿态驱动的特征权重子网络来学习自适应的特征融合.

• **基于注意力机制.** 基于注意力机制的局部特征表示方法是指使网络自发地寻找并学习图像中的感兴趣区域信息, 即在神经网络的设计过程中结合人类视觉注意力机制的特点, 强化感兴趣区域的局部细节特征, 使网络更多地聚焦于图像中行人所在区域的特征, 削弱目标以外背景信息的响应. Liu 等 [19] 提出了一个端到端的比较注意力网络 (comparative attention network, CAN), 可以选择性地关注图像中行人某一部分的信息, 再通过行人不同部分特征的整合和比对进行行人再识别. Li 等 [20] 提出了一种新颖的注意力模型 (harmonious attention CNN, HA-CNN), 将基于像素的注意力机制与基于区域的注意力机制联合优化进行行人特征表示, 解决图像中行人区域的特征学习和对齐问题. Xu 等 [17] 提出了注意力感知组成网络 (attention-aware compositional network, AACN), 利用行人姿态注意力信息识别并排除图像中非行人部分信息的干扰, 同时根据每个姿态关键点的可见性评分预测行人各个关键点的可见性, 使网络聚焦于行人可见区域的特征. Song 等 [21] 提出了基于掩码 (mask) 引导的对比注意模型 (mask-guided contrastive attention model, MGCAM) 来分别学习行人区域和背景区域的特征并将其区分开. Chen 等 [22] 提出了基于注意力机制的网络 (attentive but diverse network, ABD-Net), 利用注意力机制强化网络特征的通道聚合和行人位置的感知, 使得网络获取行人区域特征.

综上所述, 由于单一粒度的特征表示存在稳定性和鲁棒性较差的问题, 并且网络深度的加深会导致图像特征信息的逐步丢失, 设计多粒度的深度特征融合表示方法成为提升行人表征能力的主流方法. 此外, 在神经网络的设计过程中结合人类视觉注意力机制的特点, 可以强化关键的局部细节特征, 通过设计有效的注意力机制模块, 使网络更多聚焦于图像中行人所在区域的特征, 削弱目标以外背景信息的响应.

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

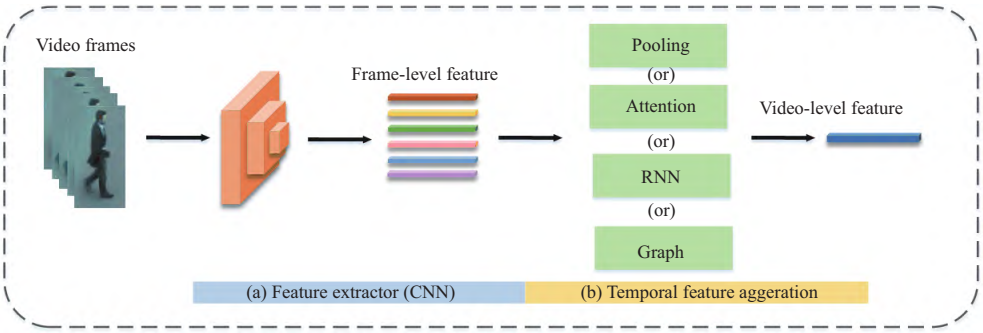

图 6 (网络版彩图) 视频特征学习形式

Figure 6 (Color online) Video feature learning format

3.1.2 视频特征学习

视频序列相比单帧图像包含更多丰富的时空信息, 并且在实际应用中视频才是实际监控的原始数据形式. 如图 6 所示, 视频行人再识别方法的一般形式为先使用特征提取器从视频帧中提取帧组的特征, 再通过特征聚合得到视频组特征描述符. 与提取图像特征相比, 视频时空特征关键在于如何聚合序列帧的特征. 根据特征聚合方法的不同, 可分为基于池化、循环神经网络、注意力机制和图等方法.

• **基于池化.** 基于池化的方法一般通过 CNN 得到视频帧的多组特征, 并使用最大池化 [23]、时序池化 [24]、平均池化甚至强化学习 [25] 等方法来聚合这些特征. Chung 等 [24] 提出一种双流卷积神经网络, 其中空间网络以 RGB 图像帧作为输入, 时序网络以光流 (optical flow) 作为输入. Liu 等 [26] 提出质量感知网络, 首先使用全卷积网络生成中间特征, 再通过质量生成模块和特征生成模块产生每个图像的质量分数和图像级的表征, 最后通过集合池化 (set pooling) 聚合生成图像组级的表征. 与之类似, Song 等 [27] 提出基于区域的质量评估网络将表征输入到基于区域的质量预测器, 然后通过集合聚合单元对所有图像的分数和特征进行聚合. 与之前方法不同, Zhao 等 [28] 提出基于属性驱动的方法用于特征分解, 通过属性识别的置信度对子功能进行重新加权, 然后在时间维度上进行时序聚合作为最终表示. Zhang 等 [25] 提出一种基于可解释强化学习的方法, 先训练一个 CNN 网络作为图像级特征提取器, 再通过一个智能体聚合时序级的特征.

• **基于循环神经网络.** 一些研究通过循环神经网络建模视频中的运动信息. McLaughlin 等 [29,30] 提出的循环 – 卷积网络是该方向十分经典的方法. 给定一个行人的视频序列, 使用包含一个循环神经网络的卷积神经网络从每个帧中提取特征. 然后使用时空池化层将来自所有时间帧的特征组合起来, 为序列图像提供一个整体特征. Dai 等 [31] 设计了一个 Bi-LSTM 网络, 利用时序池化和残差计算部分组成的时空残差学习模块来分别学习通用特征和特定特征. 光流信息可用于编码相邻帧间的短期动作信息, 因此常被用于基于循环神经网络的方法中作为网络的输入之一 [29]. Liu 等 [32] 受 FlowNet [33] 启发, 提出一种端到端的累积运动上下文 (accumulative motion context, AMOC) 网络, 通过联合空间外观学习和从原始视频帧积累运动上下文信息学习视频特征表示. AMOC 使用双流卷积结构, 包含空间信息分支和运动信息分支. 空间信息分支从单帧图像中学习行人特征表示, 运动信息分支提取帧间的光流信息. 随后, 通过循环聚合的方式将行人表征信息和光流信息进行融合. 这种包含运动信息的特征表示不仅能够提高视频行人再识别精度, 还对一些遮挡、大位移情景更加鲁棒. Li 等 [34] 提出一种基于外观信息和运动信息的增强模型 (appearance and motion enhancement model, AMEM) 以学习更丰富的视频特征. 其中, 外观信息增强模块利用视频数据中的行人属性学习来关注行人外观的不同方

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

面, 运动信息增强模块通过预测连续帧捕捉特定身份的行走风格.

• **基于注意力机制.** 基于注意力机制的方法一般在 CNN 或者 RNN-CNN 结构的基础上构建注意力模块, 可以自动学习到哪些帧的特征或者帧上哪些区域有利于再识别. 在空间域或者时间序列上构建注意力模块是基于注意力机制最常见的方法 [35~40]. Xu 等 [36] 提出一种时空联合注意池化网络, 使得特征提取器能够感知当前输入的视频序列, 从而使来自匹配项的相互依赖能够直接影响到彼此表征的计算. Zhou 等 [41] 通过一个时间注意模型, 自动地提取出给定视频中最具判别力的帧. Chen 等 [42] 将长视频序列分成多个短视频片段, 利用一种新的时间协同注意深度神经网络来估计片段的相似度. Li 等 [40] 提出一个关系引导的空间注意模块来搜索具有区分度的区域, 和一个关系引导的时间细化模块以进一步细化帧间的特征表示. Jiang 等 [43] 提出一个通用的时域整合框架来整合帧上的语义特征和时间特征.

现有的大多数基于注意力机制的方法都是分别学习空间注意和时间注意的, 忽略了两者之间的相关性. 近期, 一些融和时域和空间域注意力的方法被提出 [37, 44~47]. Subramaniam 等 [44] 提出一种新颖的基于共分割的注意力模块, 能够在视频多个帧之间以无监督的方式激活一组共同的显著特征. Chen 等 [37] 提出联合注意时空特征聚集网络通过质量感知注意力模块评测图像的质量, 使用帧感知注意力模块衡量图像帧对于时序特征的贡献. Zhu 等 [45] 针对现有的大多数方法都是分别学习空间注意和时间注意的情况, 提出自适应时空注意网络. 该网络包含多个自适应时空融合模块, 以便在多层次特征图上探索更加精确的时空注意.

除了基于空间域和时域设计注意力模块, 自注意力机制也被用于视频行人再识别中 [48~50]. Li 等 [48] 提出全局 – 局部时序表征来利用多尺度时序信息. 短期时序信息通过平行空洞卷积建模表示行人的外观和运行, 通过时序自注意捕获长期关系. Zhang 等 [49] 提出自协作注意网络 (self-and-collaborative attention network, SCAN). SCAN 首先使用共享的卷积神经网络来获取帧级的特征, 再通过自注意力子网络和协作注意力子网络生成时序级的表征. Hou 等 [51] 提出了一种时间互补学习网络来提取连续视频帧的互补特征, 通过删除先前帧激活的部分挖掘新的互补的部分. Hou 等 [52] 提出一种双边互补网络来提取连续帧间的互补空间特征, 辅以多个平行的空间注意力模块, 以关注帧间的不同区域. 为了解决视频行人错位问题, Jiang 等 [53] 提出一种自分离网络 (self-separated network, SSN), 利用注意力机制, 在保留单一图像原始空间信息的同时, 找出不同的行人区域. 随后, SSN 被应用于一个视频序列, 出现在多个帧的相同部分被对齐, 并基于三维卷积聚合最终特征表示.

• **基于图.** 近年来, 不少学者提出与图模型相结合的视频行人再识别方法. Wu 等 [54] 利用姿势对齐连接和特征相似度连接来构建自适应的结构感知邻接图. 在邻接图上执行特征传播, 以迭代方式细化区域特征, 并且考虑相邻节点的信息以表示部分特征. Yan 等 [55] 提出多粒度超图, 在多个粒度层面对时空依赖性进行建模来追求更好的表征能力. Yang 等 [56] 提取时空图卷积网络从不同帧的时间关系和帧内的空间关系进行建模, 其空间分支提取人体的结构信息, 而时间分支挖掘相邻帧间显著特征.

综上所述, 目前主流的基于深度学习的视频行人再识别方法通常先将连续的图像帧输入到一个共享的 CNN 中提取每个图像的图像级特征, 再通过构建的时序模块得到代表视频级的最终特征, 从而使得最终的特征包含单个图像的外观特征和帧间的时序特征. 基于深度学习的方法早期使用各种池化方法来聚合多帧图像中的特征, 随着该领域的发展, 将 RNN 与 CNN 相结合成为一个经典方向. 由于 RNN-CNN 模型存在难以训练的问题 (如大量参数问题和梯度问题), 注意力机制替换 RNN-CNN 结构成为一个主流方向. 对于大规模应用场景, 基于图的方法提供了全新的思想, 仍有提升的潜力.

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

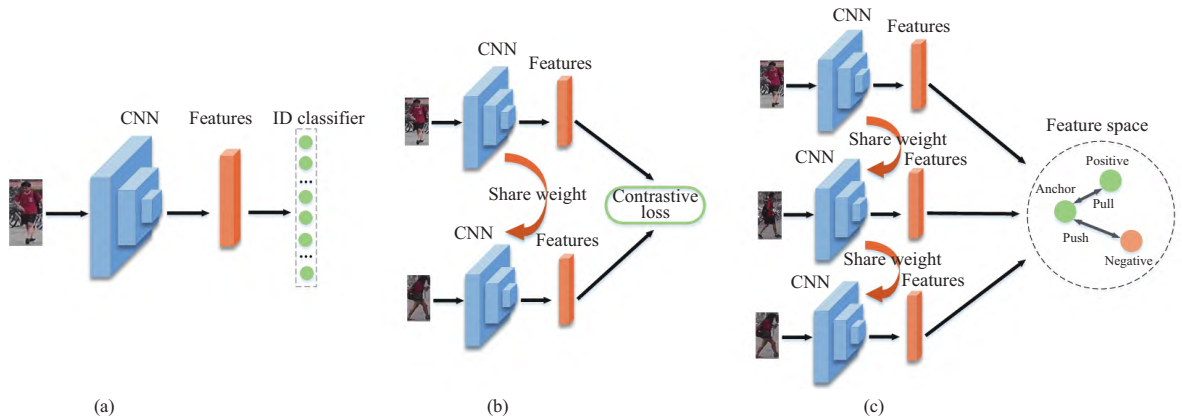

图 7 (网络版彩图) 3 种基于不同损失的行人再识别模型

Figure 7 (Color online) Person ReID based on three different losses. (a) Classification loss; (b) verification loss; (c) triplet loss

3.2 度量学习

度量学习方法一般通过网络学习出多张图片的相似度, 并根据图像标签与相似度的差异构建损失, 在行人再识别问题上, 具体表现为同一行人的不同图片间的相似度大于不同行人的不同图片. 本小节详细介绍 3 种常用度量损失函数, 即分类损失、验证损失、三元组损失, 对应分别使用单张图像、图像对、三元组输入网络构建损失.

• **分类损失.** 分类损失将行人再识别任务视为图像分类问题, 通过将网络提取的高层语义特征输入全连接层得到分类结果, 其网络结构大致如图 7(a) 所示.

根据该分类结果输入以下表达式计算交叉熵损失:

$$L_{\mathrm{id}}(i)=\frac{1}{n}\sum_{i=1}^{n}\log\left(p\left(y_i\mid x_i\right)\right), \tag{1}$$

其中 y_i 代表样本 i 的标签, x_i 为输入样本, n 代表每个批次中训练样本数量. 该类损失通过最大化样本真实类别的后验概率以聚类方式对样本的间距进行优化, 以簇的形式对样本进行聚类分割, 从而优化不同样本之间的距离. 该损失在训练过程中会自动挖掘出难分类样本, 且易于与其他损失函数结合使用从而进一步优化判别空间, 因此被广泛应用于行人再识别任务中. 近年来, 多种分类损失的变体形式 [57~59] 也逐渐被提出. 标签平滑化策略 [60] 也被广泛应用于交叉熵损失的计算中, 使得模型避免标签过拟合, 提升网络的泛化能力.

• **验证损失.** 基于相似样本在特征空间中仍然相似的假设, 利用图像间的成对关系将检索问题转化为验证问题, 一般用在孪生网络中, 其结构如图 7(b) 所示.

当两个输入样本相似时, 对比损失可以有效地对细节进行建模, 即对这两个样本的差异性进行度量. 该损失最早于 2014 年由 Li 等 [61] 引入行人再识别任务, 用于处理行人再识别中的光照及几何变换问题. 其基本形式如下:

$$L_{\mathrm{con}}\left(i,j\right)=\left(1-\delta_{ij}\right)\left\{\max\left(0,\rho-d_{ij}\right)\right\}^2+\delta_{ij}d_{ij}^2, \tag{2}$$

其中, δ_{ij} 代表标签是否相似, 当样本 i, j 具有相同的类别标签时, 其值为 1, 否则为 0. 而 ρ 则是人工设定的阈值参数, 其值越大, 则网络对于不相似样本对的优化力度越大, d_{ij} 表示样本 i, j 所得到的嵌入特征之间的欧氏距离.

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

在验证问题中, 二分类验证是一个常用的模型, 其通过判断输入图像对相似与否优化特征的判别力, 该损失于 2015 年被 Zheng 等 [62] 提出并应用于行人再识别任务中. 具体而言, 是对输入样本得到的特征作差的平方结果进行分类, 判断该结果的正负. 结合交叉熵的二分类验证损失公式表示如下:

$$L_{\text{veri}}(i,j) = -\delta_{ij}\log\left(p\left(\delta_{ij} \mid f_{ij}\right)\right) - \left(1-\delta_{ij}\right)\log\left(1-p\left(\delta_{ij} \mid f_{ij}\right)\right), \tag{3}$$

其中 δ_{ij} 定义与对比损失中相同, 为相似度标签, 而 $f_{ij} = (f_i - f_j)^2$, f_i, f_j 分别代表样本 i 和 j 的特征, $p\left(\delta_{ij} \mid f_{ij}\right)$ 代表 f_{ij} 被分类为 δ_{ij} 的概率. 2016 年, Wang 等 [63] 进一步将验证损失应用到单张图像匹配及交叉图像二分类的联合任务中, 意在挖掘图像之间的关联性. 2016 年, Yu 等 [64] 将长短时记忆神经网络与双分支网络结合, 以序列形式处理局部图像, 以验证损失进行监督学习, 从而增强局部特征的判别力. 进一步 Song 等 [21] 于 2018 年提出使用二元掩码协助网络消除图像的背景噪声并从中获取步态特征信息进而提升高层特征质量, 取得了不错的效果. 为了解决已有方法仅考虑局部约束相似度的问题, Chen 等 [65] 将条件随机场应用于组相似度学习, 学习到的相似度包含多张行人图像的局部及组相似度, 并进一步将该相似度应用于对比损失的计算中. 2019 年, Chen 等 [66] 则提出将注意力网络、多尺度网络与验证损失结合, 提取得到多尺度的视角不变性的行人图像特征. 2019 年, Zhou 等 [67] 则基于对比损失, 并从深层特征图出发, 提取不同特征尺度下的行人掩码, 提出使用注意力一致性的正则化策略, 使得低层特征掩码更为准确, 最终性能得到进一步改善.

• **三元组损失.** 三元组损失将行人再识别任务当作检索排序问题进行处理, 是一种被广泛应用于行人再识别领域的损失函数. 相比于验证损失, 其构造的三元组中包含锚样本 X_i, 与 X_i 类别相同的正例样本 X_i^+, 以及与 X_i 类别不同的负例样本 X_i^-, 即一个三元组包含一对正样本对及一对负样本对, 进一步引入预定义的距离阈值参数后, 其损失函数形式如下:

$$L_{\text{tri}} = \left(d_{a,p} - d_{a,n} + \alpha\right)_+, \tag{4}$$

其中, $d_{a,p}$ 为正样本对的距离, $d_{a,n}$ 为负样本对的距离, 而 α 为设置的阈值, 基于三元组损失的模型需要以 3 张图像作为网络的输入, 其网络结构图如图 7(c) 所示.

2015 年, Ding 等 [68] 首次将三元组损失引入行人再识别任务中, 并且将阈值引入三元组损失的定义中, 使得三元组损失不仅仅依赖于类内距离小于类间距离这一简单约束, 并针对随机映射产生三元组造成生成较多无用数据的情况, 提出在每次迭代过程中, 随机选取一定数量的行人身份, 根据该身份信息从数据集中选取对应类别的特定数量的行人图像作为锚样本, 进行随机三元组构造, 解决了模型优化效率过低的问题. 而基于上述形式, Cheng 等 [69] 考虑到已有的三元组损失仅使得类内距离小于类间距会导致最终学习到的聚类簇相对较大的问题, 引入另一个预先定义的阈值, 使得类内距离小于该阈值, 从而学习得到更为紧凑的类簇. Zhou 等 [70] 提出使用点到集合的距离替代点到点的距离计算相似度. 由于三元组损失使用简单的硬挖掘策略易受异常值的影响, 并且困难样本在梯度下降训练过程中易被淡化, Yu 等 [64] 进一步改进三元组损失, 通过对每组样本进行动态加权, 赋予困难程度高的样本以更大的权重进行多损失联合训练, 取得了进一步提升. 而将三元组损失与其他损失结合亦或是嵌入到各种结构中也是较为热门的做法, Wang 等 [71] 提出将三元组损失与注意力损失等多种损失结合, 以多任务方式训练网络, 以挖掘行人有效特征. Suh 等 [72] 则提出一种基于身体部位对齐的双分支网络结构, 对人体姿势提取得到部分特征图计算三元组损失, 取得了不错的性能提升. Zhao 等 [73] 提出一种困难挖掘中心 – 三元组 (hard mining center-triplet loss, HCT) 损失, 有效地实现类内和类间距离的优化. 公式具体如下:

$$L_{\text{hct}} = \left(d(c_p, f(p^i)) - d(c_p, f(l^j)) + \alpha\right)_+, \tag{5}$$

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

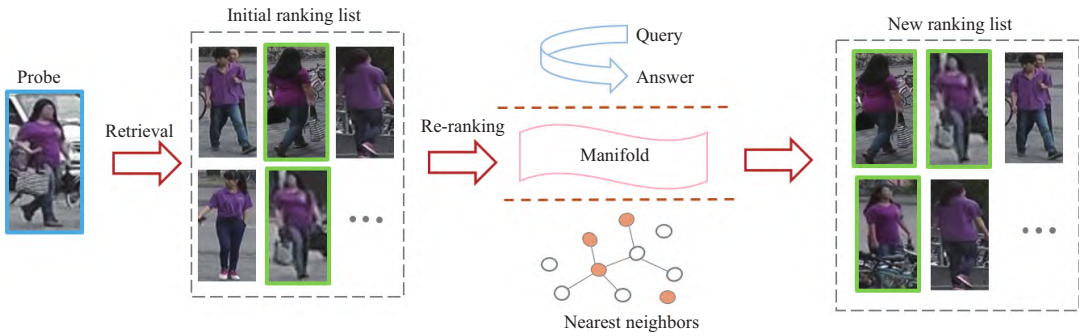

图 8 (网络版彩图) 基于重排序的图像行人再识别

Figure 8 (Color online) Illustration of person ReID based on re-ranking

其中 c_p 为第 p 类的深度特征中心, $f(p^i)$ 为第 p 类中第 i 样本的深度特征, $l \in [1, p]$. HCT 损失同时降低了计算和挖掘困难训练样本的成本, 从而提高了对鲁棒特征的学习效果.

2017 年 Chen 等 [74] 进一步提出四元组损失, 在原有三元组中引入另一负例样本, 并且两个负例样本属于不同的行人. 通过优化四元组损失, 可以获得比三元组损失更小的类内差异以及更大的类间差异, 但由于构造四元组需要耗费大量的时间及计算资源, 相比于三元组损失, 其应用范围较为有限.

基于度量学习的行人再识别方法具有较强的灵活性, 易与多种结构集成使用, 但大多数损失需要人为设定阈值, 因此其性能易受到人为因素影响.

3.3 重排序优化

重排序通常作为行人再识别任务的一种后处理手段, 如图 8 所示, 通过发掘候选集的上下文信息和相似度信息对检索结果进行优化, 从而进一步提升检索精度 [75]. 早期的一些研究 [76,77] 通过向用户问询得到查询图像的负样本, 以此优化排序. Wang 等 [77] 设计一种基于人工验证增量学习的混合人机识别模型, 该模型可从人类反馈中累积学习, 从而提升排序准度. 此类方法由于增加了人机交互量, 会增大用户负担和影响算法运行效率, 因此不利于大规模数据集的应用, 后续的重排序研究大多是自动且无监督的. 一些学者 [78,79] 关注于如何对候选样本集构建流形结构. Loy 等 [78] 使用亲密度矩阵构建候选样本集的流形近似, 并对查询信息进行传播, 实现了检索结果的重排序. Bai 等 [79] 提出一种基于流形的亲和力学习算法, 用于优化排序结果. 还有一些方法 [80~83] 通过挖掘候选集的局部近邻上下文信息, 进一步提升了重排序算法的速度和精度. Bai 等 [82] 提出一种稀疏上下文激活方法, 将邻域集编码成一个向量, 并通过广义 Jaccard 距离计算样本间相似度. 在此基础上, Zhong 等 [81] 引入一种 k 互近邻编码策略来挖掘近邻信息. Leng 等 [80] 提出一种双向排名算法, 从内容和上下文两个角度计算查询图像和图库之间的相似性, 从而得到更准确的排序结果. Ye 等 [83] 将全局和局部特征的共同最近邻作为新的查询对象, 通过聚合全局和局部特征的新排序, 优化初始排序.

现有的重排序行人再识别可分为 3 类: 基于用户反馈、基于候选集流形结构和基于近邻信息挖掘. 尽管重排序算法是一种提升检索精度的有效手段, 但是其往往涉及复杂的流形分析以及大量的参数计算. 这些不仅制约行人再识别的速度, 还一定程度上影响模型的稳定性.

4 开放行人再识别

在开放监控场景中, 存在大量光照、视角、姿态、遮挡、跨域、对抗样本攻击等不利因素以及海量

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

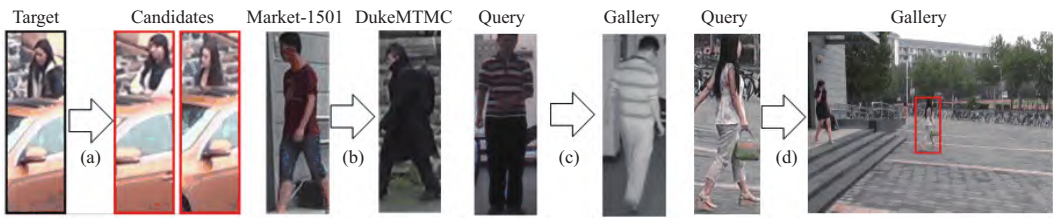

图 9 (网络版彩图) 开放行人再识别面临挑战

Figure 9 (Color online) Challenges for open person ReID. (a) Occlusion; (b) cross-domain; (c) cross-modal; (d) person search

数据带来的检索速度问题. 标准行人再识别难以有效解决上述问题. 为了使行人再识别技术能够更好应对这些实际应用中的困难, 国内外很多研究者开展众多的研究工作, 取得了一定进展. 如图 9 所示, 本节针对不同问题, 将行人再识别研究划分为遮挡行人再识别、无监督行人再识别、半监督行人再识别、跨模态行人再识别、场景行人搜索、对抗鲁棒行人再识别和快速检索 – 哈希行人再识别. 下面将对这些内容一一阐述.

4.1 遮挡行人再识别

遮挡行人再识别问题主要研究的是在行人部分信息被遮挡的情况下, 如何获取有效的特征表示以及合理度量遮挡行人特征间的相似性 [84].

• **非遮挡特征表示学习.** 在遮挡行人的特征表示方面, 主要研究如何获取遮挡行人有效的特征描述, 减少遮挡噪声的干扰. Zhuo 等 [85] 提出了一种基于行人身体注意力的深度特征学习框架. 通过遮挡模拟器自动为一般行人图像生成人工遮挡, 然后设计多任务损失训练网络, 强化其对遮挡行人图像特征的学习和行人身份的识别. Miao 等 [86] 提出了一种利用姿态信息引导网络提取人体有效区域特征的新方法 —— 姿态引导特征对齐. 此外, 该工作也为遮挡行人再识别的研究提供了当前最大规模的遮挡行人数据集 Occluded-DukeMTMC. Wang 等 [87] 在利用姿态信息的基础上, 有效挖掘了图像局部特征间的高阶关联信息, 以此强化局部特征对齐. He 等 [88] 结合全卷积网络和金字塔池化提取遮挡行人图像的空间金字塔特征, 提出了有效的金字塔重建法 (foreground-aware pyramid reconstruction, FPR), 可以精确地计算遮挡行人间的相似性分数. Yoo 等 [89] 提出了一个姿态引导的行人局部可见性预测匹配网络, 通过姿态估计注意力定位遮挡区域和可见区域, 强化可见区域特征的细节描述, 从而获取更具鉴别力的遮挡行人特征表示. Zhao 等 [90] 提出一种新颖的增量式生成遮挡对抗抑制 (incremental generative occlusion adversarial suppression, IGOAS) 网络. IGOAS 首先通过增量式生成遮挡块产生从易到难的遮挡数据, 再通过对抗抑制分支抑制生成的遮挡区域. Li 等 [91] 提出一种部分感知 Transformer 网络, 包括像素级和部分级 Transformer 编码器 – 解码器结构. 在像素级 Transformer 编码器中, 采用自注意机制来获取完整的图像上下文信息, 在部分级 Transformer 解码器中生成行人部位感知掩码.

• **有效对齐和度量.** 考虑到不同图像遮挡区域的差异性, 主要研究如何在表征基础上, 进一步实现特征之间有效的对齐和度量. Zheng 等 [92] 提出了联合局部 – 局部特征匹配与全局 – 局部特征匹配优化的框架实现遮挡行人之间的有效度量. He 等 [93] 提出了一种无需显式对齐的深度空间特征重构方法 (deep spatial feature reconstruction, DSR), 利用字典学习模型的重构误差计算遮挡图像不同空间特征图之间的相似性, 实现鲁棒表征. Sun 等 [94] 提出了一种基于半监督学习的可见性感知的局部模型 VPM, 通过预测行人不同部位的可见性分数判断对应部位是否被遮挡, 实现不同遮挡行人的区域特征自适应对齐和度量. Chen 等 [95] 认为基于先验知识的模型由于存在领域偏差, 往往不能起到很好的效

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

果, 为此提出一种遮挡增强方法, 能够为已有数据生成多样化和精确标记的遮挡, 并在生成的遮挡数据上设计一种遮挡感知掩码网络, 通过注意力机制获取未遮挡的行人区域.

目前对遮挡行人再识别的研究工作仍处在探索阶段, 现有的方法在遮挡行人数据集上的性能与实际应用仍有一定差距. 因此, 遮挡场景下的行人再识别问题仍需系统性研究, 为实际场景中行人身份识别问题提供解决方案.

4.2 无监督行人再识别

基于深度学习的行人再识别方法已在许多大规模数据集 [60, 96, 97] 上取得了较好的性能, 但相机以及拍摄场景的差异性和不同数据集之间的数据分布不同, 导致在一个场景数据集下训练的模型很难直接应用到另一场景中. 为加速行人再识别技术的实际应用, 近年来, 越来越多的研究聚焦于无监督行人再识别研究. 现有的主流无监督行人再识别方法可以大致分为两类: 基于无监督域适应的行人再识别方法以及目标域数据结构挖掘的行人再识别方法.

• **无监督域适应.** 无监督域适应包含源数据集和未标注的目标数据集, 通常也称为源域和目标域. 由于行人再识别中的源数据集和目标数据集采集自不同监控系统, 包含完全不同的类别标签 [98], 因此, 一般的无监督域适应方法无法直接应用在行人再识别场景中.

近几年, 许多基于对抗生成网络 (generative adversarial networks, GAN) [99] 的图像翻译方法 [100~102] (图像翻译是指将图片内容从一个域转换到另一个域) 被提出. Zhu 等 [100] 提出了 CycleGAN 模型, 使用循环一致损失来训练不成对数据, 将图片内容从源域迁移到目标域, 而不需要源域与目标域图像内容匹配. Choi 等 [102] 在此基础上, 提出了 StarGAN 模型, 仅使用单个生成器即可实现多个域之间的图像迁移.

为解决行人再识别的无监督域适应问题, 一些方法 [103, 104] 使用对抗生成网络在图像层面上降低源域和目标域的域间差异. Wei 等 [103] 提出了一种行人迁移模型, 在 CycleGAN [100] 模型基础上新增行人身份损失函数, 以保证迁移后的图片与目标域图片有相似的风格且在迁移前后行人的外观和身份信息保持不变. 同时, 该文还公开了一个更大型更有挑战的行人再识别数据集 (MSMT17), 旨在促进行人再识别方法的研究. Deng 等 [104] 提出了类似的基于 CycleGAN 网络的相似性保留对抗生成网络模型, 实现图像整体风格从源域到目标域的迁移, 使用新生成的图像进行行人再识别网络的有监督学习. Zhai 等 [105] 提出一种基于领域自适应的多专家头脑风暴网络 (multiple expert brainstorming network, MEB-Net), 采用互学习策略, 在源域内预先训练多个不同架构的网络作为具有特定特征和知识的专家模型, 然后通过专家模型之间的相互学习来完成自适应. Wu 等 [106] 提出一种轨迹自监督学习 (tracklet self-supervised learning, TSSL) 方法, 通过轨迹相干学习、轨迹邻域紧致性学习和轨迹聚类结构学习 3 种自监督方式优化行人的特征嵌入空间.

部分研究在特征层面实现无监督域适应 [107~110], 即将不同分布的源域和目标域图像映射到一个特征空间中, 使源域和目标域数据的分布差异尽可能小. Wang 等 [107] 提出了一种联合学习目标域图像语义属性和身份判别特征表示的迁移模型, 可同时学习行人图像的全局特征表示和局部属性信息. Peng 等 [108] 提出了基于非对称多任务字典学习方法来学习目标域的判别特征表示. Yang 等 [110] 提出了基于局部块的渐进适应网络, 通过渐进学习方式使源域和目标域的全局特征以及局部特征对齐. Jiang 等 [111] 提出一种端到端自监督智能体学习 (self-supervised agent learning, SAL) 算法, 设计了包括源域监督学习、目标域相似度一致性学习和跨域自监督学习 3 种学习机制来学习领域不变但具有辨别力的特征.

此外, 还有一些方法在目标域内部进行细粒度的风格迁移, 并结合源域信息, 提高模型在目标域

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

上的泛化能力. Zhong 等 [112] 提出了一种异构同质学习法, 该方法使用 StarGAN 模型 [102] 实现目标域不同相机之间的图像风格迁移, 并保证身份信息在迁移前后不发生变化. 基于相机不变性和域连通性两个约束条件, 使用源域图像、目标域图像以及迁移后图像构建三元组训练神经网络模型, 实现目标域判别模型的学习.

• **目标域数据结构挖掘.** 为了解决无监督行人再识别问题, 部分工作聚焦于无标注目标域数据集的内部数据结构挖掘. 大多数无监督行人再识别方法会使用源域数据训练一个深度神经网络模型作为初始行人特征提取器. 随后, 在目标域上学习度量 [113, 114] 或使用无监督聚类的方法获取行人图像伪标签并调整深度模型 [115, 116]. Yu 等 [113] 提出了一种基于聚类的非对称度量学习模型, 采用非对称度量进行不同视域的不同映射学习, 将各个视域的图片映射到共享的空间中进行聚类, 联合优化聚类以及各个映射. Fan 等 [115] 提出了一种基于 K-means 聚类算法和行人再识别 IDE 网络微调交替迭代训练的渐进学习方法, 使用聚类结果作为标签信息训练分类网络, 同时从网络中学习判别特征用于无监督聚类. Fu 等 [117] 将目标数据集无标注图像送入源数据集预训练好的网络中, 提取上半身、下半身以及全身的特征向量集, 在 3 个集合中分别使用基于密度的无监督聚类算法 [118] 获得聚类分组, 并给每个组分配伪标签, 再使用伪标签优化特征提取网络, 聚类和优化阶段不停迭代直至网络收敛. Zhang 等 [119] 通过渐进增强学习获取目标数据集上的局部结构和全局数据分布. 由于不同摄像机拍摄同一个行人存在外观上的显著差异, Xuan 等 [120] 将样本相似度计算分解为 “相机内” 和 “相机间” 两个阶段, 以逐步寻找可靠的伪标签. Yang 等 [121] 提出一种动态对称交叉熵损失来抵抗无监督聚类过程中产生的噪声伪标签, 并且设计了一个相机感知元学习算法以适应相机移动.

Yu 等 [122] 提出了一种基于软多标签学习 (soft multilabel learning) 的目标域标签估计方法, 通过引入参考代理学习, 标记目标域图像基于辅助参考域图像的相似性作为行人软标签. Zhong 等 [123] 引入了目标域的 3 种不变性约束来挖掘隐藏标签信息. Yang 等 [124] 提出了一种基于图像块判别特征提取的无监督学习方法. Huang 等 [125] 提出了一种 EANet 模型, 利用源域和目标域额外的人体解析任务辅助行人再识别训练, 有效实现了行人图像的特征定位与对齐, 提高了模型在目标域上的泛化能力. 这些使用辅助任务监督学习的方法需要满足两个前提: 辅助任务的标签可以自动获取、辅助任务本身对行人再识别性能有促进作用.

无监督行人再识别任务近几年受到了越来越多的关注, 也有越来越多的相关方法被提出. 目前主流的方法分为两类: 基于无监督域适应相关研究方法, 将从源域学习到的知识迁移至目标域中; 利用图模型、聚类等方法对原始提取的特征进一步标注, 挖掘目标域数据内部的相似性结构信息. 但无监督行人再识别方法与单域行人再识别方法的性能仍存在较大差异, 如何提高模型在未知域上的检索性能, 仍旧是行人再识别领域面临的一大挑战.

4.3 半监督行人再识别

无监督学习要求训练样本完全无标记, 而半监督学习允许对少量的训练样本进行标记. 近年来, 有很多工作关注于半监督设置下的行人再识别研究. 针对行人再识别中存在严重的未标记数据不平衡问题, Li 等 [126] 提出了一种半监督区域度量学习方法, 不再从未标记的数据中寻找匹配的图像对 (正样本), 而是提出利用交叉行人得分分布对齐的标签传播估计正邻居. 随后, 利用正邻居集生成多个正区域, 学习区域到点的判别度量. 基于未标记数据是开放集的假设, Chang 等 [127] 提出过渡性半监督度量学习框架, 设计了一种基于图的过渡性困难样本挖掘方法, 用于深入挖掘无标签数据中的困难三元组, 以及一种基于程度的关系置信评分方法, 进一步减少错误的三元组. 此外, 还对特征一致性损失进行研究, 并采用课程学习策略来改进半监督行人表征学习. Ding 等 [128] 通过挖掘特征空间中未标记的

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

训练样本和已标记的训练样本之间的关系实现半监督学习, 并成功为无标签的样本分配伪标签. Zheng 等 [60] 使用 DCGAN 生成无标签数据, 并提出离群值标签平滑正则 (label smoothing regularization for outliers, LSRO) 方法, 为无标签图像分配统一的标签分布. Liu 等 [129] 设计了一个传导式中心点投影 (transductive centroid projection, TCP) 模块, 用来替代 CNN 的最后一个全连接层. TCP 模块的目的是将未标记的数据作为标记数据, 同时减少类内冲突的概率. Xin 等 [130, 131] 提出了一种多视图自适应聚类方法, 将标签传播到未标记的数据中, 然后通过自适应的方式使排名损失和识别损失最小化以实现对系统的微调.

一些研究人员关注于单样本学习 (one-shot) 设置下的视频行人再识别问题, 即只有一个行人轨迹 (tracklet) 是拥有完整身份标签的, 目的是利用这个有标签轨迹和其他无标签轨迹来共同学习一个行人再识别模型. Ye 等 [132] 提出了一种动态图匹配 (dynamic graph matching, DGM) 方法, 通过迭代更新图匹配和标签估计学习更好的特征. Liu 等 [133] 在图库中使用 k 倒数最近邻 (k-nearest neighbors, KNN) 更新分类器, 并在查询集中使用 KNN 进行负样本挖掘来改进最近邻. Wu 等 [134, 135] 认为初始迭代阶段估计的伪标签并不可靠, 因此采用渐进伪标签选择策略, 这使得整个训练过程中伪标签子集是逐渐丰富且可靠的.

4.4 跨模态行人再识别

在实际监控系统中, 采集数据的设备来源往往是不同的, 这就会导致同一个目标出现在不同的模态场景中. 这种模态异构会极大降低标准行人再识别模型的识别精度. 因此, 跨模态行人再识别被提出, 其研究的关键是将不同模态下的行人特征映射到相同的特征空间中进行度量. 如图 10 所示, 本小节将从以下 3 类模态异构进行阐述: 图像 – 视频行人再识别、文本 – 图像行人再识别和可见光 – 红外行人再识别.

4.4.1 图像 – 视频行人再识别

在图像 – 视频行人再识别中, 查询对象是单个图像, 候选图库是视频. 任务的一个应用实例是根据犯罪嫌疑人的图片快速查找和追踪大量城市监控录像中的嫌疑人. 在图像 – 视频行人再识别任务中, 图像与视频包含的信息不对等, 图像级特征与视频级特征之间的差异成为此任务的核心挑战.

为了实现图像与视频的匹配, Zhu 等 [136] 提出了一种联合特征投影矩阵和异构字典对学习的方法, 将图像和视频的异构特征转化为相同维数的编码系数. Wang 等 [137] 提出一种端到端的深度学习方法来匹配图像和视频, 让网络聚焦于可用帧而忽略视频中其他无用的帧. Zhang 等 [138] 提出了由特征表征子网和相似子网两部分组成的时间记忆相似性学习神经网络. 该网络使用 LSTM 网络中对视频序列的时间信息进行编码, 再将图像和视频序列的特征向量送入相似子网络中进行距离度量学习. 为了缓解图像与视频匹配的难度, Gu 等 [50] 提出了一种新的时域知识传播方法, 将视频表征网络学习到的时域知识传播到图像表征网络中. 通过反向传播, 可以传递时域知识来增强图像特征, 缓解信息不对称问题.

4.4.2 文本 – 图像行人再识别

文本 – 图像行人再识别通常给定一个目标行人的自然语言描述, 要求从候选图库中检索该行人. 传统的图像行人再识别方法在一些特定问题中可能难以建树. 例如在刑侦场景中, 犯罪嫌疑人的图像可能难以获取, 但是目击者的自然语言描述相对更容易获得 [139, 140]. 文本 – 图像行人再识别方法由于其查询限制少, 在智能视频监控系统中有着广阔的应用前景.

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

图 10 (网络版彩图) 3 种常见的跨模态任务

Figure 10 (Color online) Three common cross-modal tasks

Li 等 [141] 首先提出文本 – 图像行人再识别任务, 为此, 建立了一个包含行人自然语言描述的文本 – 图像行人再识别数据集 CUHK-PEDES, 并提出一种基于门控神经注意机制的递归神经网络来学习文本 – 图像对之间共同的语义特征. Chen 等 [142] 将自然语言描述作为监督信息来训练网络, 并且提出全局和局部两种互补的文本 – 图像关联方案, 以指导网络学习不同粒度的视觉特征. Aggarwal 等 [143] 从文本描述中挖掘属性表示, 并提出一种分层模型来建立两种不同层级的文本 – 图像特征空间. Jing 等 [144] 将文本 – 图像行人再识别看作是跨模态 – 跨域任务, 并提出一种跨模态 – 跨域对齐网络 MAN (moment alignment network), 将域对齐、模态对齐和示例对齐进行联合建模, 以互补的方式减少域差异和语义差距. Jing 等 [140] 则认为行人的视觉特征和自然语言描述间存在细粒度级相关性. 为此, 提出一种姿态引导的多粒度注意力网络使得行人不同身体部位的视觉特征和相应文本描述对齐. Wu 等 [145] 设计了两种颜色推理辅助任务来构建细粒度的跨模态关联, 包括文本引导的图像着色和图像引导的文本填空. Li 等 [146] 基于 Transformer 架构, 将行人按照语义划分, 以匹配具有视觉 – 语言共同注意的区域. Zhu 等 [147] 提出一种深度环境 – 行人分离模型来提取和匹配行人特征, 包括环境 – 行人分离模块 (surroundings-person separation module, SPSM), 环境 – 行人融合模块 (surroundings-person fusion module, SPFM), 信号去噪模块 (signal denoising module, SDM)、行人描述模块 (person describing module, PDM) 和显著注意力模块 (salient attention module, SAM). 此外, 该工作在 MSMT17 数据集基础上进行自然语言标注, 构建了文本 – 图像行人再识别数据集 RSTPReid.

自然语言描述相较于行人图像更容易获取, 因此, 文本 – 图像行人再识别有着巨大潜力. 然而, 文本和图像两种模态的差异较大, 直接衡量两种异质模态特征间的相似性是十分困难的. 现有的文本 – 图像行人再识别方法大多采用 “双塔模型”, 即使用两分支网络, 一支提取行人文本特征, 另一支提取行人图像特征, 随后将两种模态特征作细粒度级别对齐并映射到同一个语义空间, 最后进行文本 – 图像匹配.

4.4.3 可见光 – 红外行人再识别

可见光 – 红外行人再识别主要研究 RGB 行人图像和红外行人图像之间的跨模态度量问题. 第一个数据集于 2017 年提出 [148]. 可见光 – 红外行人再识别可以突破光线不足情况下传统 RGB 图像的局限, 是 RGB 行人再识别应用的补充方法.

• **提取共享特征.** 可见光 – 红外行人再识别的一个思路是提取不同模态图像中的共享特征. Wu 等 [148] 提出将不同模态的图像统一变换为双通道, 即灰度通道和红外通道叠加, 并用 0 填充缺失的

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

通道, 进而输入网络进行特征提取. Wu 等 [149] 提出了提取共享特征的思路, 利用一个双输入的网络分别提取 RGB 图片和红外图片的特征, 并通过一种跨模态的焦点损失函数进行相似性学习. 类似地, Ye 等 [150] 同样使用双输入网络, 并利用一种双向的排序损失学习跨模态的相似性, 同时辅以身份损失 (identity loss) 进一步提升性能. Dai 等 [151] 尝试利用生成对抗网络来解决可见光 – 红外行人再识别问题, 为此分别设计一个学习图像表示的生成器和一个模态分类的鉴别器, 以从不同模态中学习判别性特征. 此外, 将分类损失和跨模态三元组损失结合, 极大降低了类间的模糊性, 并尽可能提高了实例间的跨模态相似性. Wei 等 [152] 提出一种注意力提升机制, 将特征分解为注意和非注意部分, 以改善模态内的判别, 并提出一种共注意力学习机制, 通过共享特征学习弥合模态间的鸿沟.

• **RGB 和红外特征结合.** 可见光 – 红外行人再识别的另一个思路是将不同模态图像中的特征结合起来再进行相似性度量. Lu 等 [153] 关注特异性特征, 提出了共享与特异特征的变换算法, 在给定某一模态的查询图片时, 先找到一些简单的置信度高的跨模态样本 (假设其中大部分是正确的匹配), 再利用这些跨模态样本的特异特征去搜索更难的跨模态样本. Wang 等 [154] 使用 CycleGAN [100] 生成 RGB 图像对应的红外图像, 又使用一个判别器来对抗训练生成的红外图像真实性. 该文仅实现了 RGB 到红外图像的单向转换, 没有实现双向转换, 而 Wang 等 [155] 实现了 RGB 图像和红外图像的相互转换. 他们将模态的差异性和行人外观的差异性分别处理, 首先使用一个子网络对输入图像进行模态转换, 组成 RGB 和红外结合的图像, 再使用另一个子网络进行特征提取和相似性度量. 由于可见光和红外两种模态间缺乏成对的标签, 很多工作尝试建立全局集合级别 (set-level) 对齐来减小模态差异, 但却忽略了实例级别 (instance-level) 不对齐造成的影响. Wang 等 [156] 提出一种生成跨模态图像对的方法, 通过编码器从两种模态图像中区分出模态不变特征和模态独有特征, 再利用解码器生成跨模态成对图像, 在此基础上综合考虑了集合级别和实例级别对齐. 由于 RGB 图像与红外图像颜色信息间的不平衡, RGB 图像特征容易过度拟合服装颜色信息, 这不利于模态的对齐. 为此, Zhao 等 [157] 提出通过不相关颜色一致性学习模块 (color-irrelevant consistency learning, CICL) 学习不相关颜色特征, 通过身份感知模态自适应模块 (identity-aware modality adaptation, IAMA) 对齐身份级别特征分布.

不同于以往直接从可见光和红外两种原始模态中提取特征的思路, Li 等 [158] 引入一种介于可见光和红外之间的 X 模态作为辅助, 构建 X-RGB-IR 跨模态学习框架. 其中, X 模态以自监督方式通过轻量网络生成, 并设计一种模态间隙约束来指导 3 种模态间的信息交换. 受神经架构搜索 (neural architecture search, NAS) 启发, Chen 等 [159] 提出一种神经特征搜索 (neural feature search, NFS) 范式, 以实现特征选择过程的自动化. NFS 结合了双层特征搜索空间和可微分的搜索策略, 以共同选择粗粒度通道和细粒度空间像素中的身份相关线索. 此外, 通过一个跨模态对比优化机制指导 NFS 搜索那些能够最小化模态差异同时最大化类间距离的特征.

可见光 – 红外行人再识别数据集标注成本高昂, 大大限制了实际应用性, 一些学者关注于无监督跨模态行人再识别的研究. 由于模态内的识别比跨模态识别容易得多, 并且可以为跨模态识别提供共享信息, 一种同质 – 异构两阶段方法 [160] 被提出, 第 1 阶段采用无监督方式学习模态内的特征表示, 并生成可见光和红外图像集的伪标签, 第 2 阶段通过异构学习方式, 利用两个伪标签图像集来学习模态不变和视图不变特征表示. 此外, 还提出一种基于自模态搜索和循环模态搜索的跨模态重排序方法, 极大提高了无监督跨模态行人再识别的性能.

在实际监控场景中, 由于光照因素影响, 可见光摄像机获取的行人外观会出现模糊不清的情况, 而红外设备可以突破这种限制, 即使在黑暗环境中仍然能拍摄到较为清晰的行人图像. 因此, 可见光 – 红外行人再识别技术拥有极高的应用价值. 现有的可见光 – 红外行人再识别方法通过提取两种模态的共

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

享特征以及利用生成模型实现两种模态间的转换, 都有助于解决突出的模态异质问题.

4.5 场景行人搜索

场景行人搜索又名行人搜索, 其将行人检测和行人再识别任务联合起来, 只需输入原始场景图像, 即可完成对目标行人的查找和匹配 [161~163]. 与基于裁剪图片的行人再识别相比, 行人搜索更符合现实场景下的视频监控需求, 即从原始的监控场景中查询目标行人. 2014 年, Xu 等 [161] 首次提出了行人搜索概念, 并提出基于手工特征的模型, 采用滑动窗口和 Fisher 向量编码的策略. 随着深度学习的兴起, 大量基于深度学习的行人搜索方法被提出, 这些方法大致可分为两类: 将检测和识别联合起来的端到端方法和先检测再识别的两阶段方法.

• **端到端行人搜索.** Xiao 等 [162] 提出首个端到端行人搜索模型, 该模型基于 Faster R-CNN [164], 并提出 OIM (online instance matching) 在线实例匹配机制代替 Softmax, 以计算识别过程中的损失. 此外, 该工作还建立了一个用于训练和评估的行人搜索数据集 CUHK-SYSU. 2017 年, Zheng 等 [163] 提出了 PRW (person re-identification in the wild) 数据集, 丰富了搜索场景. Munjal 等 [165] 提出查询行人引导的端到端行人搜索模型, 该方法采用孪生 Faster R-CNN 结构, 通过查询行人辅助搜索, 取得了更好的性能. Yan 等 [166] 建立图卷积学习框架, 提出了上下文扩展模块, 使用相对注意力机制来搜索和过滤场景中的上下文信息. Chen 等 [167] 考虑在极坐标系中分解行人向量, 通过向量范数区分背景和行人, 通过向量角度实现行人身份识别. Zhong 等 [168] 针对现实遮挡场景, 提出了 LSPS 数据集, 该数据集包含室内和室外场景, 场景复杂多变, 对应地提出了 APNet 网络用于解决样本对齐和部分遮挡问题. Dong 等 [169] 提出了一种双向交互网络, 并利用裁剪的行人补丁来减少冗余上下文的影响. Li 等 [170] 提出一种序列端到端网络, 将检测和再识别任务视为一个渐进的过程, 并依次处理两个子网络. 考虑到大多数端到端行人搜索方法都是在 Faster R-CNN 基础上开展的, 虽然拥有很高的性能, 但是计算开销太大, Yan 等 [171] 提出无锚框行人搜索框架 AlignPS (feature-aligned person search network), 并利用可变形卷积和特征融合来改进特征金字塔网络, 以克服行人区域和尺度不对齐问题.

• **两阶段行人搜索.** 不同于联合行人检测和行人再识别的端到端方法, 一些研究 [172, 173] 认为, 检测和识别两个任务之间存在冲突: 检测关注行人的共性, 而识别关注行人之间的差异, 产生的错误也会在两个任务间累积, 因此端到端框架并不是一个好的解决办法. Chen 等 [172] 认为将行人搜索中的检测任务和识别任务分开进行特征提取可以获得更好的性能, 采取对前景人物和原始图像分别进行建模的方式, 从两个独立的 CNN 流中获取丰富的特征表示, 并通过加入掩膜融合特征, 增强行人特征的判别性. Lan 等 [174] 提出利用来自行人识别网络的多级特征来解决多尺度匹配问题. Han 等 [173] 在检测网络和识别网络间加入一个 ROI Transformer 结构, 并通过行人识别网络的损失优化检测器生成的框排除背景和其他行人的干扰. Wang 等 [175] 致力于解决行人检测和行人再识别之间的一致性问题, 通过计算检测器检测出的行人框和查询图片的相似度, 去除相似度低的框, 可以为后续行人识别任务提供数量更少、精确度更高的行人框. 此外, 将人工裁剪的行人图像和检测器检测出的行人图像混合, 共同用于行人识别网络的训练, 可以使行人识别网络更加适应检测器. Li 等 [176] 分析了行人搜索的时间瓶颈, 为了实现实时的行人搜索, 提出了更高效的行人搜索通道 (fast person search pipeline, FPSP) 方法, 相比于之前的工作, 速度提升了 10 倍左右.

综上所述, 现有的行人搜索方法中, 端到端行人搜索注重于检测识别一体化实现, 将行人再识别网络嵌入检测网络中, 再使用各自的损失函数进行联合优化. 相比之下, 两阶段行人搜索先做检测再做识别, 这种两阶段方案可以克服两种不同优化目标带来的冲突, 即行人检测关注行人之间的共性而行人再识别关注行人之间的差异.

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

4.6 对抗鲁棒行人再识别

一些研究表明, 当前大多数行人再识别模型易受到对抗性样本攻击[177~179], 通过在行人图像上添加人类无法感知的扰动, 会导致模型性能的大幅下降, 这种 “脆弱” 的模型会给实际监控系统带来极大的安全隐患.

对抗样本研究被广泛应用于分类领域, 但是极少在度量学习中被研究. Bai 等[177] 首次提出对抗性度量攻击, 通过攻击行人再识别中的度量系统产生对抗样本, 这是一种与现有对抗分类攻击平行的方法, 可以有效应用于其他依赖距离度量的任务中. 考虑到现有的对抗攻击行人再识别方法面临两个基本问题: 对抗攻击方法必须为每张输入图像生成定制的对抗样本; 跨模型的对抗扰动会导致攻击效果急剧下降. 为了克服这些局限, Ding 等[179] 提出一种通用对抗性扰动方法, 通过列表式攻击目标函数直接破坏相似度排名, 此外还提出了一种模型不敏感的跨模型攻击机制. 当前的对抗性攻击研究大多是针对封闭集设定下的分类和识别任务, 而在开放集设定中, 所有训练类别和测试类别是不相交的. Yang 等[180] 提出一种基于虚拟引导元学习的行人再识别通用攻击算法, 通过对抗性扰动在未知域上误导模型. Gong 等[178] 针对开放集行人再识别任务中的对抗鲁棒性进行了系统研究, 对比了两种黑盒攻击方式: 基于标准封闭集迁移的开放黑盒攻击和基于随机搜索的黑盒攻击, 大量实验表明开放集识别模型同样易受到对抗性攻击的干扰.

4.7 快速检索 —— 哈希行人再识别

近期, 哈希在大规模图像检索领域被广泛应用. 通过将高维实值特征映射为简短的二值编码, 可以在低维的汉明 (Hamming) 空间中快速检索. Lin 等[181] 提出一种无监督哈希网络, 可以同时保证编码的均匀分布以及量化损失的最小化, 而 Lai 等[182] 则设计了一个可以同时进行特征学习和哈希编码的框架用于最大程度保持映射过程中的相似度. 受这些方法启发, 近期多种基于监督哈希的行人再识别方法被提出, 并且这些方法的实验结果证明通过结合哈希方法可以较大程度改善现有行人再识别系统的效率. 现有的基于哈希的行人再识别方法可以大致分为两类: 基于传统哈希和基于深度哈希的行人再识别方法.

• **基于传统哈希的行人再识别.** 基于传统哈希的行人再识别方法[183,184] 旨在学习一个子空间映射方法和二进制编码策略从而通过多个映射矩阵将高维手工实值特征映射到汉明空间中. 这些方法大多把每个相机下捕捉到的图像视为单个模态从而利用不同表征之间的关联解决行人再识别中的跨视角问题. 由 Zheng 等[184] 提出的跨视角二值身份方法 (cross-view binary identities, CBI) 通过最小化同一行人类内的汉明距离, 同时最大化协方差值, 构建两个用于不同视角的由哈希函数组成的集合, 而随后由 Chen 等[183] 提出的跨相机语义二值化策略 (cross-camera semantic binary transformation, CSBT) 意图从缓解内在的跨相机变化入手提升特征的判别力.

• **基于深度哈希的行人再识别.** 基于深度哈希的行人再识别方法大多通过在网络的末端插入哈希层以生成近似二值化编码, 所使用的哈希层一般是使用 tanh 作为激活函数的全连接层. 其中, Zhang 等[185] 提出的比特扩展深度哈希 (deep regularized similarity comparison hashing, DRSCH) 设计了一种基于相对相似度比较的三元组模型用于生成比特可扩展的哈希编码. 而随后 Zhu 等[186] 则将分块模型融入到深度哈希网络框架中以增强特征在视觉匹配阶段的判别力. 然而这些方法都是使用显式的二值化策略, Liu 等[187] 认为这样的策略难以消除输出维度之间的关联, 同时它们需要采用诸如欧氏度量损失这类较强约束的损失函数以约束输出特征的形式, 因此难以生成高质量的哈希码. 针对这些问题, Liu 等[187] 提出一种基于对抗学习的二进制转换策略 (adversarial binary coding, ABC), 通过

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

将生成对抗网络的目标设置为标准二值分布形式, 隐式地将特征转化为哈希编码, 可以较大程度上保证特征二值化过程中的潜在相似度信息. 而为了进一步提升匹配阶段的效率, Wang 等 [188] 提出一种由粗到细的搜索策略, 同时通过自蒸馏训练的方式改善编码的质量, 可以取得较大的性能提升. Zhao 等 [189] 提出一种显著性引导的迭代非对称互学习哈希方法 (salience-guided iterative asymmetric mutual hashing, SIAMH) 用于行人再识别系统的加速. 通过设计的显著性引导的自蒸馏分支 (salience-guided self-distillation branch, SSB), SIAMH 可以从图像中最显著的区域生成哈希编码从而显式降低编码间的信息冗余, 提升哈希编码的性能.

近年来, 多个大规模行人再识别数据集被提出用于模拟真实场景, 而大多基于高维实值特征的行人再识别方法会极大增加行人再识别系统的存储和时间成本, 导致这些方法难以部署到实际监控场景中. 相比之下, 基于哈希的行人再识别方法展现出巨大的速度优势, 通过把高维实值映射到二值编码, 即可在低维汉明空间中快速完成检索. 值得一提的是, 虽然基于哈希的行人再识别方法可以显著提升匹配阶段的效率, 但它们目前的性能仍然和现有的最优实值特征方法存在一定的差距.

5 行人再识别常用数据集

实际监控场景涉及很多复杂的环境因素, 例如光照、视角、姿态、摄像参数、遮挡、背景变化等, 这使得行人再识别问题极具挑战性. 为了有效模拟这些环境因素和评估行人再识别算法性能, 近年来许多行人再识别数据集被提出. 如表 1 [3, 48, 60, 61, 85, 86, 92, 96, 103, 141, 149, 162, 163, 190~194] 所示, 包括针对静态图像和视频序列的标准行人再识别数据集以及针对各种开放问题的开放行人再识别数据集.

目前常用的图像行人再识别数据集有 VIPeR [3], CUHK03 [61], Market-1501 [96], DukeMTMC-reID [60], MSMT17 [103]. VIPeR [3] 是较早公开的行人再识别数据集, 创建于 2007 年. 该数据集包含两个摄像头, 632 个行人, 共计 1264 幅图像, 其中涉及大幅度姿态变化以及轻微光照变化. CUHK03 [61] 是香港中文大学于 2014 年发布的数据集, 由 6 个不同监控摄像机拍摄收集, 每个行人都至少出现在两个摄像视角下, 共计 28192 张图片, 其中含有 1467 个行人. 该数据集中存在光照、遮挡、身体部分不匹配等环境因素造成的影响. Market-1501 [96] 数据集采集自清华大学的校园中, 包括 6 个摄像头 (5 个高清和 1 个低清). 该数据集包含 32688 张图片, 1501 个行人, 其中的行人边界框由 DPM [195] 自动检测得到. DukeMTMC-reID [60] 是大规模多目标跟踪数据集 DukeMTMC [97] 中的一个子集, 采集自杜克大学 (Duke University) 校园内, 于 2016 年公布. 该数据集包含 1404 个行人的 34183 张图像, 其中 702 个行人 16522 张图像用作训练, 剩下 702 个行人 17661 张图像用作测试. MSMT17 [103] 是目前最大规模图像行人再识别数据集之一, 由 12 个室外摄像机和 3 个室内摄像机在不同时段、不同气候条件下拍摄得到, 共计 126441 张行人图像, 4101 个行人. 该数据集中的行人边界框由 Faster R-CNN [164] 检测器检测得到.

不同于仅需要单帧图像的图像行人再识别方法, 基于视频的行人再识别需要的是连续视频帧的数据集. 常用的视频行人再识别数据集主要有: PRID-2011 [190], iLIDS-VID [191], MARS [192], DukeMTMC-Video [193], LS-VID [48]. PRID-2011 是 Hizer 等 [190] 于 2011 年提出的行人再识别数据集. 该数据集由两台摄像机分别拍摄 385 条和 749 条行人轨迹组成, 其中有 200 个行人在这两台摄像头中都出现过. 该数据集的背景比较干净并且遮挡较少, 因此相对简单. 此外, 该数据集还有一个由随机选择的快照组成的单镜头版本. iLIDS-VID 是 Wang 等 [191] 于 2014 年发布的行人再识别数据集. 该数据集拍摄于监控航空接站大厅, 包含由两个视野不相交的摄像头拍摄到的 300 个行人的 600 条行人轨迹. 该数据集由于严重遮挡和复杂的背景等原因, 很具有挑战性. MARS 是 Zheng 等 [192] 于 2016 年发布的

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

表 1 行人再识别常用数据集统计

Table 1 Statistics of common datasets for person ReID

	Direction	Dataset	Year		Direction	Dataset	Year
Standard ReID	Image	VIPeR [3]	2007	Open ReID	Occluded	P-ETHZ [86]	2008
		CUHK03 [61]	2014			Partial-REID [92]	2015
		Market-1501 [96]	2015			P-DukeMTMC [86]	2016
		DukeMTMC-reID [60]	2017			Occluded-REID [85]	2018
		MSMT17 [103]	2018			Occluded-Duke [86]	2019
	Video	PRID-2011 [190]	2011		Person search	CUHK-SYSU [162]	2017
		iLIDS-VID [191]	2014			PRW [163]	2017
		MARS [192]	2016		Text-based	CUHK-PEDES [141]	2017
		DukeMTMC-Video [193]	2018		RGB-Infrared	SYSU-MM01 [149]	2017
		LS-VID [48]	2019			RegDB [194]	2017

行人再识别数据集, 是视频行人再识别数据集中首个大规模数据集. MARS 是 Marker-1501 [96] 数据集的视频拓展版, 它包含 1261 个行人的大约 20000 个轨迹. 相比之前的其他数据集, 更加贴近现实场景. DukeMTMC-Video [193] 是 DukeMTMC [97] 的子集, Wu 等 [134] 于 2018 年将其从 DukeMTMC 中提取出, 专门用于基于视频的行人再识别. Wu 等 [134] 每秒从视频中裁剪行人图像 12 帧, 以生成行人轨迹. 该数据集包含用于训练的 2196 条轨迹的 369656 帧和用于测试和干扰的 2636 条轨迹的 445764 帧. LS-VID 是 Li 等 [48] 于 2019 年提出的大规模视频行人再识别数据集. LS-VID 具有更长的轨迹序列, 是目前最大的视频行人再识别数据集, 同时更加面向现实场景和具有挑战性.

开放行人再识别数据集的构建主要为了解决遮挡、行人搜索、跨模态等问题. 针对遮挡行人再识别, Partial-REID [92], Occluded-REID [85], Occluded-Duke [86], P-ETHZ [86], P-DukeMTMC [86] 数据集被提出. Partial-REID [92] 采集自大学校园中, 包含 60 个行人的 600 张图片, 每个行人有 5 张全身图片和 5 张遮挡图片, 该数据集存在复杂的视角、背景变化和严重遮挡情况. Occluded-REID [85] 由校园内的移动摄像设备拍摄, 包含 200 个行人的 2000 张标注图像, 类似于 Partial-REID 的设置, 每个行人由 5 张全身图片和 5 张不同遮挡图片组成. Occluded-Duke [86] 由 DukeMTMC-reID [60] 中的遮挡图像组成, 是迄今为止最大的遮挡行人数据集, 其训练集包含 702 个行人的 15618 张图像, 测试集包含 1110 个行人的 17661 张图库图像和 2210 张待查询图像. 该数据集存在多种视角变化和丰富的遮挡类型. 一些研究者分别从 ETHZ 和 DukeMTMC-reID 数据集中挑选出遮挡行人图像组成 P-ETHZ [86] 和 P-DukeMTMC [86] 数据集. 在 P-ETHZ 中, 包含 85 个行人的 3897 张图像. 在 P-DukeMTMC 中, 训练集包含 665 个行人的 12927 张图像, 测试集包含 634 个行人的 2163 张待查询图像和 9053 张图库图像.

针对行人搜索, 常用的两个数据集是 CUHK-SYSU [162] 和 PRW [163] 数据集. 此外还有针对场景行人遮挡的 LSPS [168] 数据集被提出. CUHK-SYSU 收集自手持摄像机拍摄的街景和电影中的场景, 共有 18184 张包含行人的场景图像, 96143 个行人边界框, 8432 个行人身份. PRW 采集自大学校园内的 6 台摄像设备, 共有 11816 张场景图像, 43110 个行人边界框, 932 个行人身份. LSPS 由 17 个摄像设备拍摄的室内室外场景组成, 拥有复杂的背景、视角、光照、行人密度变化. 此外, 还具有大量不完整的行人框, 该数据集包含 51836 张场景图像, 60433 个行人边界框, 4067 个行人身份.

在通过自然语言描述查询图库行人的文本 – 图像行人再识别方面, CUHK-PEDES [141] 是该领域

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

目前唯一的数据集, 相关研究人员从 CUHK03, Market-1501, SSM, VIPER 和 CUHK01 数据集中挑选出 13003 个行人的 40206 张图像, 每张图像都有两段自然语言描述.

可见光 – 红外行人再识别的两个常用数据集分别是由 Wu 等 [149] 提出的 SYSU-MM01 和由 Nguyen 等 [194] 提出的 RegDB 数据集. SYSU-MM01 包含 6 个摄像头, 其中 4 个是在白天工作的可见光摄像头, 2 个是在黑暗环境中工作的红外摄像头. 该数据集共有 491 个行人身份的 287628 张可见光图像和 15792 张红外图像. RegDB 数据集共有 432 个行人身份, 每个行人包含 10 张可见光图像和 10 张红外图像, 且只有正面和背面两种拍摄视角.

近年来, 随着行人再识别研究的不断深入, 行人数据集规模快速增加, 摄像机位增多, 拍摄视角更加多变, 数据集中涉及的开放问题也更加丰富, 行人再识别数据集越来越接近真实监控场景. 然而, 行人再识别数据集标注成本十分昂贵, 为了缓解这一问题, 一些学者致力于数据集标注方法的改进. Xu 等 [196] 提出主动冗余削减框架 (active redundancy reduction, ARR), 以最少的标注工作量, 实现全监督环境下训练有效的行人再识别模型, 其目标是在构建大规模行人再识别数据集时减少冗余样本. 通过估计样本的不确定性和内部多样性, 主动选择信息丰富和多样化的样本进行标注. 实验表明 ARR 可以在 Market1501, MSMT17 和 CUHK03 上分别减少 57%, 63% 和 49% 的标注工作量. 当前的行人再识别数据集大多涉及隐私等敏感问题, 为此, 一些学者致力于构建虚拟行人数据集 [53, 197, 198], 在虚拟数据集上预训练一个行人再识别模型, 并迁移到真实数据上进行测试.

6 行人再识别代表性方法比较

近年来, 行人再识别问题受到越来越多的关注, 当前的行人再识别算法在多个大型数据集上取得了巨大的性能突破, 接下来将通过比较图像行人再识别和视频行人再识别的代表性算法及其性能回顾行人再识别发展历程.

6.1 图像行人再识别代表性方法比较

行人再识别通常可以视为图像检索的子任务, 因此一般延用图像检索任务的评价指标, 包括平均精度均值 (mean average precision, mAP) 和累计匹配特性曲线 (cumulated matching characteristics, CMC). 对于查询集中的一个目标行人, 通过行人再识别算法得到排序结果, 并根据其真实身份标签计算它的平均精度 AP 值 (精确率 – 召回率曲线下方面积). mAP 即为查询集中所有目标行人 AP 的均值. CMC 曲线以 Rank-k 识别率作为纵坐标, 而 Rank-k 是指正确目标排在候选列表前 k 个位置的概率统计结果.

在图像行人再识别领域, 我们选择最常用的 Market-1501 和 DukeMTMC-reID 作为方法比较的基准数据集, 并选择了 2015~2021 年发表在顶级会议和顶级刊物上的代表性文章进行分析, 如表 2 [9, 18, 20, 22, 74, 87, 91, 96, 163, 199~216] 所示. 其中 BoW+KISSME [96] 是 2015 年提出的基于手工特征的方法, 其在两个大规模数据集上的表现上远远低于基于深度学习的行人再识别算法, IDE [163] 作为基于深度学习的行人再识别方法基线, 而后 2017 和 2018 两年是行人再识别算法性能的爆发阶段. 在 Market-1501 和 DukeMTMC-reID 上, mAP 每年都有近 10 个点的提升, Rank-1 有近 20 个点的提升, 这一时期的研究主要集中在行人表征学习和度量学习, 以 PCB [9] 为代表的局部特征学习方法和结合局部特征和全局特征的 MGN [199] 在当时获得了很高的关注. 此外, 三元组损失及其改进版本也被广泛应用于行人再识别算法. 而后 2019 年的方法大多结合注意力机制的研究: 混合高阶注意力网络 MHN [203]、多样注意力 ABD-Net [22]、二阶非局部注意力网络 SONA [216], 结合强化学习的自我批评注

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

表 2 图像行人再识别方法在最常用的两种数据集上的性能对比

Table 2 Performance comparison of image-based person ReID methods on two most commonly used datasets

Method	Market-1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
BoW+KISSME [96](15 ICCV)	44.4	20.8	25.1	12.1
IDE [163](16 arXiv)	72.5	46	–	–
Quadruplet [74](17 CVPR)	81.5	64.9	73.5	54.3
PDC [18](17 ICCV)	84.1	63.4	–	–
HA-CNN [20](18 CVPR)	91.2	75.7	80.5	63.8
PCB+RPP [9](18 ECCV)	93.8	81.6	83.3	69.2
MGN [199](18 ACMM)	95.7	86.9	88.7	78.4
Auto-ReID [200](19 ICCV)	94.5	85.1	88.6	73.5
BagTricks [201](19 CVPRW)	94.5	85.9	86.4	76.4
OSNet [202](19 ICCV)	94.8	84.9	86.6	73.5
MHN-6(PCB) [203](19 ICCV)	95.1	85	89.1	77.2
ABD-Net [22](19 ICCV)	95.6	88.3	89	78.6
SFT [204](19 ICCV)	93.5	90.6	88.3	83.3
P2Net [205](19 ICCV)	95.2	85.6	86.5	73.1
SONA [216](19 ICCV)	95.6	88.8	89.6	78.2
SCAL [66] (19 ICCV)	95.8	89.3	89	79.6
IANet [206](19 CVPR)	94.4	83.1	87.1	73.4
ConsAtt [67] (19 ICCV)	96.1	84.7	86.3	73.1
DenseS [207](19 CVPR)	95.7	87.6	86.2	74.3
DG-Net [208](19 CVPR)	94.8	86	86.6	74.8
Pyramid-Net [11](19 CVPR)	95.7	88.2	89	79
SCSN [209](CVPR2020)	95.7	88.5	91	79
RGA-SC [210](CVPR2020)	96.1	88.4	–	–
ISP [211](2020 ECCV)	95.3	88.6	89.6	80
HOReID [87](2020 CVPR)	94.2	84.9	86.9	75.6
MoS [212](2021 AAAI)	95.4	89.0	90.6	80.2
PAT [91](2021 CVPR)	95.4	88	88.8	78.2
TransReID [213](2021 ICCV)	95.2	89.5	90.7	82.6
CDNet [214](2021 CVPR)	95.1	86	88.6	76.8
InSTD [215](2021 ICCV)	97.6	90.8	95.7	89.1

意力网络 SCAL [66] 和一致性注意力正则网络 ConsAtt [67] 等. 这些结合注意力机制的算法在 Market-1501 和 DukeMTMC-reID 上性能达到新高, 并且在 Market-1501 上趋于瓶颈. 除结合注意力之外, 也有很多工作继续关注于表征学习以及网络架构的优化并且也达到了很高的性能: 采用神经网络架构搜索的 Auto-ReID [200]、采用全局特征结合多种网络训练技巧的 BagTricks [201] 基线方法、全尺度网络 OSNet [202]、金字塔网络 Pyramid-Net [11] 等. 2020 年的方法在两个数据集上的性能提升不大, 有一些结合先验知识的行人再识别方法被提出: 人体解析网络 ISP [211]、结合人体关键点检测和图匹配的

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

HOReID [87]. 也有工作继续聚焦于注意力: 关系感知全局注意力网络 RGA-SC [210]、显著性引导级联抑制网络 SCSN [209]. 2021 年, 计算机视觉领域的一大热潮是以视觉 Transformer 为代表的方法成功挑战卷积神经网络的地位. 卷积神经网络的局限性在于其局部相关性, 而采用自注意力机制堆叠的 Transformer 结构在全局信息上有着更好的表示. 在行人再识别领域, 同样有一些基于 Transformer 的行人再识别方法被提出, 并且突破了卷积网络架构下的性能瓶颈: TransReID [213] 在 ViT [217] 网络基础上, 将摄像视角信息嵌入 Transforemer 框架, 并且结合简单的移位和置换操作, 使得该框架同时拥有出色全局和局部信息表征能力; PAT [91] 并不是一个纯 Transformer 架构, 其首先使用卷积神经网络提取特征, 再通过一个基于 Transformer 的 Encoder 和 Decoder 架构进一步挖掘局部信息. 值得一提的是, TransReID 和 PAT 对于遮挡场景同样十分有效. 一些研究发现单独训练 Transformer 容易陷入局部最优, 为此, Peng 等 [218] 设计了结合卷积的混合结构 Conformer, 能够同时耦合 CNN 局部表征和 Transformer 全局表征.

综上所述, 图像行人再识别的研究历经多个阶段: 早期的手工特征方法在大规模数据集上表现十分不理想, 深度学习时期表征学习和度量学习的飞速发展促进行人再识别算法性能的突飞猛进, 基于注意力机制的研究进一步提升行人再识别性能, 基于视觉 Transformer 的行人再识别算法性能突破了卷积架构的瓶颈.

6.2 视频行人再识别代表性方法比较

在视频行人再识别领域中, 最常用于性能对比的数据集为 PRID-2011, iLIDS-VID 和 MARS. 之后的数据集虽然规模和质量上更加优秀, 但由于公布的时间在 2018 年之后, 因此之前的方法没有相关的实验数据. DukeMTMC-VideoReID 虽然公布时间也较晚, 但在近 3 年来常与 MARS 一起被使用来验证方法在大规模视频行人再识别任务上的有效性. 然而, LPW 和 LS-VID 极少被其他文献使用. 因此, 本文只列出在最常用的 PRID-2011, iLIDS-VID, MARS 和近期在大规模视频行人再识别方法中常使用的 DukeMTMC-VideoReID 4 组数据集的性能对比结果, 如表 3 和 4 [35, 40, 44, 48, 50, 54, 56, 193, 219, 220] 所示. 首先, 从表中可以很明显观察到的一个结论是基于人工设计特征的方法的性能远低于基于深度学习的方法, 证明了深度学习在视频行人再识别这一领域的优秀性能. 因此, 下面将主要分析基于深度学习的各种方法在不同数据集下的表现.

PRID-2011 是一个提出很早的视频行人再识别数据集, 因此数据难度在现在看来相对简单, 规模小. 在各个方向最先进的工作均在 PRID-2011 上取得了 90% 以上的优秀性能, 而基于注意力机制的 JAFN 取得了目前最领先的性能, 在 Rank-1 上达到 97.4%, 在其他指标上达到 100%. 同时, 从分布上看, 基于注意力机制的方法明显优于其他深度学习方法, 仅有基于图的方法性能表现与之接近. 目前看来, 基于注意力机制的方法在小规模和遮挡较少的视频行人再识别上具有最好的性能表现. iLDS-VID 虽然规模同样较小, 但具有严重的遮挡问题. 因此, 可以据此观察不同方法对于遮挡问题的表现. RNN-CNN 结构和 3D CNN 结构由于仅在网络构架上注重学习到时序级或视频级的特征, 在遮挡问题上表现一般. 注意力机制由于能学习到哪一帧或者帧上哪些空间区域具有更显著的特征, 从而可以忽视部分遮挡的帧或者帧上遮挡的区域来缓解部分遮挡问题. 因此基于注意力机制的方法在 iLID-VID 上较为优秀, 其中 ASTA-Net 取得了最优秀的性能, 在 Rank-1 上达到 88.1%. VRSTC 虽然采用补全遮挡区域来解决遮挡问题, 但在性能上并没有优于其他方法. MARS 和 DukeMTMC-VideoReID 规模相对较大, 因此可结合两大数据集分析不同方法对于大规模视频行人再识别的表现. 基于图的方法的 MGH 和基于注意力机制的 TCLNet 在两个大规模数据集上分别达到了目前最领先的性能. 基于图的方法通常将图模型与人体结构信息相结合, 将图像级的特征水平划块学习全局与局部特征. 性能对比结果证

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

表 3 视频行人再识别方法在最常用的 3 种数据集上的性能对比

Table 3 Performance comparison of video person ReID methods on three most commonly used datasets

Method	PRID-2011			iLIDS-VID			MARS			
	Rank-1	Rank-5	Rank-20	Rank-1	Rank-5	Rank-20	Rank-1	Rank-5	Rank-20	mAP
DVR [191] (14 ECCV)	28.9	55.3	82.8	23.3	42.4	68.4	–	–	–	–
RCN [29] (16 CVPR)	70	90	97	58	84	96	–	–	–	–
SFT [41] (17 CVPR)	79.4	94.4	99.3	55.2	86.5	97	70.6	90	97.6	50.7
TSSCN [24] (17 ICCV)	78	94	99	60	86	97	–	–	–	–
ASTPN [36] (17 ICCV)	77	95	99	62	86	98	44	70	81	–
QAN [26] (17 CVPR)	90.3	98.2	100	68	86.8	97.4	–	–	–	–
AMOC [32] (IEEE TCSVT)	68.7	94.3	99.3	83.7	98.3	100	68.3	81.4	90.6	52.9
RQEN [27] (18 AAAI)	91.8	98.4	99.8	77.1	93.2	99.4	77.8	88.8	94.3	71.1
SDM [25] (18 CVPR)	85.2	97.1	99.6	60.2	84.7	95.2	71.2	85.7	94.3	–
EUG [193] (18 CVPR)	–	–	–	–	–	–	62.7	74.9	96.1	67.4
TRL [31] (IEEE TIP)	87.8	97.4	99.3	57.7	81.7	94.1	80.5	91.8	96	69.1
Co-Att [42] (18 CVPR)	93	99.3	100	85.4	96.7	99.5	86.3	94.7	98.2	76.1
SCAN [49] (IEEE TIP)	95.3	99	100	88	96.7	100	87.2	95.2	98.1	77.2
JAFN+DB [37]	97.4	100	100	76	94.7	100	89.3	98.7	100	74.9
ADFD [28] (19 CVPR)	93.9	99.5	100	86.3	97.4	99.7	87	95.4	98.7	78.2
GTLR [48] (19 ICCV)	95.5	100	–	86	98	–	87	95.8	98.2	78.47
COSAM [44] (19 ICCV)	–	–	–	79.6	95.3	–	84.9	98.5	97.9	79.9
STA [35] (19 AAAI)	–	–	–	–	–	–	86.3	85.7	98.1	80.8
AGRL [54] (IEEE TIP)	94.6	99.1	100	84.5	96.7	99.5	89.8	96.6	97.8	81.9
MGH [41] (20 CVPR)	94.8	99.3	100	85.6	97.1	99.5	90	96.7	98.5	85.8
STGCN [56] (20 CVPR)	–	–	–	–	–	–	89.9	96.4	98.3	83.7
TCLNet [50] (20 ECCV)	–	–	–	86.6	–	–	89.8	–	–	85.1
ASTA-Net [45] (20 ACM MM)	96.4	100.0	100.0	88.1	98.6	–	90.4	97.0	98.8	84.1
RGST [40] (20 AAAI)	93.7	99.0	100.0	86.0	98.0	99.4	89.4	96.9	98.3	84.0
GTF [43] (20 AAAI)	95.8	–	–	87.7	–	–	87.1	–	–	85.2

明, 学习行人结构信息和注意力机制有利于提升模型在大规模视频行人再识别数据集上的性能.

综上所述, 基于深度学习的方法在近 5 年来在视频行人再识别任务上获得了巨大的提升, 基本解决了小规模和少量遮挡场景下的行人再识别问题. 不过, 严重遮挡问题和大规模场景仍然具有挑战性. 对于遮挡问题, 忽视和补全遮挡部分是两大主流方法. 对于大规模应用场景, 基于图的方法提供了全新的思想, 仍有提升的潜力. 基于三维卷积的方法虽然目前性能并不是最优秀的, 但本文建议可将注意力机制或者人体结构信息与三维卷积相结合来提升性能. 此外, PRID-2011 和 DukeMTMC-VideoReID 数据集对于目前方法来说已经不具有挑战性, 近期各种方法之间的性能差距常少于 1%. MARS 和 iLIDS-VID 仍有提升的空间, LS-VID 的规模最大, 场景更接近现实. 本文建议之后的方法可在这 3 种数据集上进行实验对比, 从而真正解决视频行人再识别任务在实际中遇到的问题.

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

表 4　视频行人再识别方法在 DukeMTMC-VideoReID 数据集上的性能对比

Table 4　Performance comparison of video person ReID methods on DukeMTMC-VideoReID dataset

Method	Rank-1	Rank-5	Rank-20	mAP
EUG [193] (18 CVPR)	83.6	94.6	97.6	78.3
VRSTC [219] (19 CVPR)	95	99.1	99.4	93.5
GLTR [48] (19 ICCV)	96.3	99.3	99.7	93.7
COSAM [44] (19 ICCV)	95.4	99.3	99.8	94.1
STA [35] (19 AAAI)	96.2	99.3	99.6	94.9
AGRL [54] (IEEE TIP)	97	99.3	99.9	95.4
STGCN [56] (20 CVPR)	97.3	99.3	99.7	95.7
TCLNet [50] (20 ECCV)	96.9	–	–	96.2
AP3D [220] (20 ECCV)	96.3	–	–	95.6
RGST [40] (20 AAAI)	97.2	99.4	99.9	95.8

7　总结与展望

行人再识别旨在解决跨摄像头跨场景下目标行人的关联与匹配, 是智能视频监控系统的关键技术, 对维护公共安全具有重要作用. 目前, 标准行人再识别研究日趋成熟, 其性能在多个行人再识别数据集上已经很高. 但是开放行人再识别研究仍然面临诸多挑战, 未来的研究可能有以下趋势.

• **完整行人再识别系统.** 完整行人再识别系统应当包括行人检测、行人追踪和行人再识别 3 个阶段. 现有的行人再识别算法大多基于裁剪好的行人边界框, 即假设行人检测完全正确, 且图像和行人身份一一对应. 但在实际监控场景中, 各种错检、漏检情况都会存在. 因此如何从原始监控视频中开始处理, 将目标行人从背景中定位出来, 并进行后续跟踪和识别, 是一个重大挑战.

• **隐私与安全.** 隐私安全是行人再识别面临的重要伦理问题, 同时也是一个技术难点. 隐私问题随着视频技术的广泛使用而增加, 为了保护隐私安全, 需要在不影响视频动作和背景内容的前提下, 采用自动化的方法来遮盖个人身份. 去身份化涉及对个人图像或视频的检测和改造, 使其无法识别个人身份, 同时不影响动作和其他背景内容. 通过将个人周围的保守区域替换成黑色像素, 很容易隐藏个人的身份. 然而, 这隐藏了大部分关于该空间内发生何种人类活动的信息, 使得视频监控等应用失去意义. 因此, 如何在不影响监控视频内容的前提下, 去除行人身份、维护隐私安全是一个亟须解决的问题.

• **开放问题集成.** 开放场景下的行人再识别面临诸多困难, 如光照、姿态、遮挡、衣着相似、背景干扰等. 现有的研究都是只针对单一的问题, 而现实监控系统中同时包含上述多个问题. 因此, 如何将解决不同开放问题的行人再识别方法集成到同一个模型中是一项待解决的任务.

• **多源信息融合.** 当前基于图像的行人再识别大多是建立在 RGB 上的, 这种单一模态数据易受环境因素干扰, 而多源数据 [221] 相比于单一 RGB 图像更有优势. 例如, 红外图像可以不受光照因素制约、深度图像蕴含行人轮廓特征、文本数据中有着丰富的行人属性信息. 因此, 在同一行人再识别过程中, 如何利用多源数据获取更多的行人身份识别信息有待深入研究.

• **半监督与无监督学习.** 在实际视频监控系统中, 目标场景下不同相机拍摄的图片易于获取, 但这些图片往往是没有明确行人标签信息的 (人工标注费时费力, 且有一定识别难度), 因此很难直接用来监督训练行人再识别网络. 近年来, 研究利用无标签数据的无监督和半监督方法逐渐兴起, 这些方法致

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

力于解决无标签数据或少量标签数据下的行人再识别模型训练问题, 促进了行人再识别实际应用, 然而现有的无监督和半监督行人再识别性能距离有监督方法还存在一定差距, 仍然是一个待解决问题.

• **视觉 Transformer.** 由于卷积神经网络的局部相关性, 即使结合许多注意力方法, CNN 对全局的感知仍然十分有限. Transformer 得益于自注意力机制和并行化训练结构, 可以对全局信息有更好的表示. Transformer 原本是自然语言处理领域最受认可的网络架构, 随着 ViT [217] 在视觉领域的成功尝试, Transformer 被广泛应用于计算机视觉的多个基础领域, 并且性能上也取得突破. 虽然 Transformer 在视觉领域取得成功, 但目前仍然有着许多不足之处: 首先是其计算开销远高于当前卷积神经网络, 当前的 Transformer 对图像信息处理优化有限, 这导致模型训练上的困难. 此外, 一些视觉 Transformer 工作也只是浅尝辄止, 并没有足够深入地对视觉难点问题作进一步攻坚. 因此, 作为计算机视觉领域的热点问题, Transformer 未来在行人再识别上的研究工作是值得期待的.

致谢 作者感谢吕心钦和博佳悦两位同学的协助.

参考文献

1 Wang X. Intelligent multi-camera video surveillance: a review. Pattern Recogn Lett, 2013, 34: 3–19
2 Zajdel W, Zivkovic Z, Krose B J A. Keeping track of humans: have I seen this person before? In: Proceedings of the 2005 IEEE International Conference on Robotics and Automation, 2005. 2081–2086
3 Gray D, Tao H. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: Proceedings of European Conference on Computer Vision, Marseille, 2008. 262–275
4 Ye M, Shen J, Lin G, et al. Deep learning for person re-identification: a survey and outlook. IEEE Trans Pattern Anal Mach Intell, 2021. doi: 10.1109/TPAMI.2021.3054775
5 Zhao C R, Chen K, Zang D, et al. Uncertainty-optimized deep learning model for small-scale person re-identification. Sci China Inf Sci, 2019, 62: 220102
6 Zhao C, Wang X, Wong W K, et al. Multiple metric learning based on bar-shape descriptor for person re-identification. Pattern Recogn, 2017, 71: 218–234
7 Zhao C, Wang X, Miao D, et al. Maximal granularity structure and generalized multi-view discriminant analysis for person re-identification. Pattern Recogn, 2018, 79: 79–96
8 Zhao C, Chen K, Wei Z, et al. Multilevel triplet deep learning model for person re-identification. Pattern Recogn Lett, 2019, 117: 161–168
9 Sun Y, Zheng L, Yang Y, et al. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline). In: Proceedings of the European Conference on Computer Vision (ECCV), Munich, 2018. 480–496
10 Fu Y, Wei Y, Zhou Y, et al. Horizontal pyramid matching for person re-identification. In: Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, 2019. 8295–8302
11 Zheng F, Deng C, Sun X, et al. Pyramidal person re-identification via multi-loss dynamic training. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, 2019. 8514–8522
12 Bai X, Yang M, Huang T, et al. Deep-person: learning discriminative deep features for person re-identification. Pattern Recogn, 2020, 98: 107036
13 Zhao C, Wang X, Zuo W, et al. Similarity learning with joint transfer constraints for person re-identification. Pattern Recogn, 2020, 97: 107014
14 Zheng L, Huang Y, Lu H, et al. Pose-invariant embedding for deep person re-identification. IEEE Trans Image Process, 2019, 28: 4500–4509
15 Zhao H, Tian M, Sun S, et al. Spindle Net: person re-identification with human body region guided feature decomposition and fusion. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017. 1077–1085
16 Wei L, Zhang S, Yao H, et al. GLAD: global-local-alignment descriptor for scalable person re-identification. IEEE

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

Trans Multimedia, 2019, 21: 986–999
17 Xu J, Zhao R, Zhu F, et al. Attention-aware compositional network for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018. 2119–2128
18 Su C, Li J, Zhang S, et al. Pose-driven deep convolutional model for person re-identification. In: Proceedings of the IEEE International Conference on Computer Vision, Venice, 2017. 3960–3969
19 Liu H, Feng J, Qi M, et al. End-to-end comparative attention networks for person re-identification. IEEE Trans Image Process, 2017, 26: 3492–3506
20 Li W, Zhu X, Gong S. Harmonious attention network for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018. 2285–2294
21 Song C, Huang Y, Ouyang W, et al. Mask-guided contrastive attention model for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018. 1179–1188
22 Chen T, Ding S, Xie J, et al. ABD-Net: attentive but diverse person re-identification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, 2019. 8351–8361
23 Zhang W, Hu S, Liu K, et al. Learning compact appearance representation for video-based person re-identification. IEEE Trans Circ Syst Video Technol, 2019, 29: 2442–2452
24 Chung D, Tahboub K, Delp E J, et al. A two stream siamese convolutional neural network for person re-identification. In: Proceedings the IEEE International Conference on Computer Vision, Los Alamitos, 2017. 1992–2000
25 Zhang J, Wang N, Zhang L, et al. Multi-shot pedestrian re-identification via sequential decision making. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018. 6781–6789
26 Liu Y, Yan J, Ouyang W. Quality aware network for set to set recognition. In: Proceeedings of IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017. 4694–4703
27 Song G, Leng B, Liu Y, et al. Region-based quality estimation network for large-scale person re-identification. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence, New Orleans, 2018. 7347–7354
28 Zhao Y R, Xu S, Zhongming J, et al. Attribute-driven feature disentangling and temporal aggregation for video person re-identification. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Los Alamitos, 2019. 4908–4917
29 McLaughlin N, Rincon J M D, Miller P. Recurrent convolutional network for video-based person re-identification. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, 2016. 1325–1334
30 McLaughlin N, Rincon J M D, Miller P. Video person re-identification for wide area tracking based on recurrent neural networks. IEEE Trans Circ Syst Video Technol, 2019, 29: 2613–2626
31 Dai J, Zhang P P, Wang D, et al. Video person re-identification by temporal residual learning. IEEE Trans Image Process, 2019, 28: 1366–1377
32 Liu H, Jie Z, Jayashree K, et al. Video-based person re-identification with accumulative motion context. IEEE Trans Circ Syst Video Technol, 2018, 28: 2788–2802
33 Dosovitskiy A, Fischer P, Ilg E, et al. FlowNet: learning optical flow with convolutional networks. In: Proceedings of IEEE International Conference on Computer Vision, Santiago, 2015. 2758–2766
34 Li S, Yu H, Hu H. Appearance and motion enhancement for video-based person re-identification. In: Proceedings of the AAAI Conference on Artificial Intelligence, New York City, 2020. 11394–11401
35 Fu Y, Wang X, Wei Y, et al. STA: spatial-temporal attention for large-scale video-based person re-identification. In: Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI 2019), Honolulu, 2019. 8287–8294
36 Xu S, Cheng Y, Gu K, et al. Jointly attentive spatial-temporal pooling networks for video-based person re-identification. In: Proceedings of the IEEE International Conference on Computer Vision, Venice, 2017. 4743–4752
37 Chen L, Yang H, Gao Z. Joint attentive spatial-temporal feature aggregation for video-based person re-identification. IEEE Access, 2019, 7: 41230–41240
38 Chen G, Lu J, Yang M, et al. Spatial-temporal attention-aware learning for video-based person re-identification. IEEE Trans Image Process, 2019, 28: 4192–4205
39 Liu Y, Yuan Z, Zhou W, et al. Spatial and temporal mutual promotion for video-based person re-identification. In: Proceedings of AAAI Conference on Artificial Intelligence, Honolulu, 2019. 8786–8793

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

40 Li X, Zhou W, Zhou Y, et al. Relation-guided spatial attention and temporal refinement for video-based person re-identification. In: Proceedings of AAAI Conference on Artificial Intelligence, 2020. 11434–11441

41 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. In: Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017. 6776–6785

42 Chen D, Li H, Xiao T, et al. Video person re-identification with competitive snippet-similarity aggregation and co-attentive snippet embedding. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018

43 Jiang X, Gong Y, Guo X, et al. Rethinking temporal fusion for video-based person re-identification on semantic and time aspect. In: Proceedings of AAAI Conference on Artificial Intelligence, New York, 2020. 11133–11140

44 Subramaniam A, Nambiar A, Mittal A. Co-segmentation inspired attention networks for video-based person re-identification. In: Proceedings of IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, 2019. 562–572

45 Zhu X, Liu J, Wu H, et al. ASTA-Net: adaptive spatio-temporal attention network for person re-identification in videos. In: Proceedings of ACM International Conference on Multimedia, Seattle, 2020. 1706–1715

46 Liu J, Zha Z-J, Zhu X, et al. Co-saliency spatio-temporal interaction network for person re-identification in videos. In: Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI), Yokohama, 2020. 1012–1018

47 Zhang W, He X, Yu X, et al. A multi-scale spatial-temporal attention model for person re-identification in videos. IEEE Trans Image Process, 2020, 29: 3365–3373

48 Li J, Wang J, Gao W, et al. Global-local temporal representations for video person re-identification. In: Proceedings of the IEEE International Conference on Computer Vision, Seoul, 2019. 3957–3966

49 Zhang R, Li J, Sun H, et al. SCAN: self-and-collaborative attention network for video person re-identification. IEEE Trans Image Process, 2019, 28: 4870–4882

50 Gu X Q, Ma B P, Chang H, et al. Temporal knowledge propagation for image-to-video person re-identification. In: Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Los Alamitos, 2019. 9646–9655

51 Hou R, Chang H, Ma B, et al. Temporal complementary learning for video person re-identification. In: Proceedings of European Conference on Computer Vision (ECCV), Glasgow, 2020. 388–405

52 Hou R, Chang H, Ma B, et al. BiCnet-TKS: learning efficient spatial-temporal representation for video person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021. 2014–2023

53 Jiang X, Qiao Y, Yan J, et al. SSN3D: self-separated network to align parts for 3D convolution in video person re-identification. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2021. 1691–1699

54 Wu Y, Bourahla O E F, Li X, et al. Adaptive graph representation learning for video person re-identification. IEEE Trans Image Process, 2020, 29: 8821–8830

55 Yan Y, Qin J, Chen J, et al. Learning multi-granular hypergraphs for video-based person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020. 2899–2908

56 Yang J, Zheng W-S, Yang Q, et al. Spatial-temporal graph convolutional network for video-based person re-identification. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Los Alamitos, 2020. 3286–3296

57 Zheng M, Karanam S, Wu Z, et al. Re-identification with consistent attentive siamese networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, 2019. 5735–5744

58 Fan X, Jiang W, Luo H, et al. SphereReID: deep hypersphere manifold embedding for person re-identification. J Visual Commun Image Represent, 2019, 60: 51–58

59 Wojke N, Bewley A. Deep cosine metric learning for person re-identification. In: Proceedings of 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, 2018. 748–756

60 Zheng Z, Zheng L, Yang Y. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In: Proceedings of the IEEE International Conference on Computer Vision, Venice, 2017. 3754–3762

61 Li W, Zhao R, Xiao T, et al. DeepReID: deep filter pairing neural network for person re-identification. In: Proceedings

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, 2014. 152–159
62 Zheng Z, Zheng L, Yang Y. A discriminatively learned CNN embedding for person reidentification. ACM Trans Multimedia Comput Commun Appl, 2018, 14: 1–20
63 Wang F, Zuo W, Lin L, et al. Joint learning of single-image and cross-image representations for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 2016. 1288–1296
64 Yu R, Dou Z, Bai S, et al. Hard-aware point-to-set deep metric for person re-identification. In: Proceedings of the European Conference on Computer Vision (ECCV), 2018. 188–204
65 Chen D, Xu D, Li H, et al. Group consistent similarity learning via deep CRF for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018. 8649–8658
66 Chen G, Lin C, Ren L, et al. Self-critical attention learning for person re-identification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, 2019. 9637–9646
67 Zhou S, Wang F, Huang Z, et al. Discriminative feature learning with consistent attention regularization for person re-identification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, 2019. 8040–8049
68 Ding S, Lin L, Wang G, et al. Deep feature learning with relative distance comparison for person re-identification. Pattern Recogn, 2015, 48: 2993–3003
69 Cheng D, Gong Y, Zhou S, et al. Person re-identification by multi-channel parts-based CNN with improved triplet loss function. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 2016. 1335–1344
70 Zhou S, Wang J, Wang J, et al. Point to set similarity based deep feature learning for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017. 3741–3750
71 Wang C, Zhang Q, Huang C, et al. Mancs: a multi-task attentional network with curriculum sampling for person re-identification. In: Proceedings of the European Conference on Computer Vision (ECCV), Munich, 2018. 365–381
72 Suh Y, Wang J, Tang S, et al. Part-aligned bilinear representations for person re-identification. In: Proceedings of the European Conference on Computer Vision (ECCV), Munich, 2018. 402–419
73 Zhao C, Lv X, Zhang Z, et al. Deep fusion feature representation learning with hard mining center-triplet loss for person re-identification. IEEE Trans Multimedia, 2020, 22: 3180–3195
74 Chen W, Chen X, Zhang J, et al. Beyond triplet loss: a deep quadruplet network for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017. 403–412
75 Guo R P. Research on reranking approaches for image-based person reidentification. Dissertation for Ph.D. Degree. Beijing: Beijing University of Posts and Telecommunications, 2020 [郭若沛. 基于图像的行人再识别重排序算法研究. 博士学位论文. 北京: 北京邮电大学, 2020]
76 Liu C, Loy C C, Gong S, et al. Pop: person re-identification post-rank optimisation. In: Proceedings of the IEEE International Conference on Computer Vision, Sydney, 2013. 441–448
77 Wang H, Gong S, Zhu X, et al. Human-in-the-loop person re-identification. In: Proceedings of European Conference on Computer Vision, Amsterdam, 2016. 405–422
78 Loy C C, Liu C, Gong S. Person re-identification by manifold ranking. In: Proceedings of IEEE International Conference on Image Processing, Sydney, 2013. 3567–3571
79 Bai S, Bai X, Tian Q. Scalable person re-identification on supervised smoothed manifold. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, 2017. 2530–2539
80 Leng Q, Hu R, Liang C, et al. Bidirectional ranking for person re-identification. In: Proceedings of IEEE International Conference on Multimedia and Expo (ICME), San Jose, 2013. 1–6
81 Zhong Z, Zheng L, Cao D, et al. Re-ranking person re-identification with k-reciprocal encoding. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017. 1318–1327
82 Bai S, Bai X. Sparse contextual activation for efficient visual re-ranking. IEEE Trans Image Process, 2016, 25: 1056–1069
83 Ye M, Chen J, Leng Q, et al. Coupled-view based ranking optimization for person re-identification. In: Proceedings of International Conference on Multimedia Modeling, Sydney, 2015. 105–117

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

84 Zheng W S, Wu A C. Asymmetric person re-identification: cross-view person tracking in a large camera network. Sci Sin Inform, 2018, 48: 545–563 [郑伟诗, 吴岸聪. 非对称行人重识别: 跨摄像机持续行人追踪. 中国科学: 信息科学, 2018, 48: 545–563]

85 Zhuo J, Chen Z, Lai J, et al. Occluded person re-identification. In: Proceedings of 2018 IEEE International Conference on Multimedia and Expo (ICME), San Diego, 2018. 1–6

86 Miao J, Wu Y, Liu P, et al. Pose-guided feature alignment for occluded person re-identification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, 2019. 542–551

87 Wang G A, Yang S, Liu H, et al. High-order information matters: learning relation and topology for occluded person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, 2020. 6449–6458

88 He L, Liang J, Li H, et al. Deep spatial feature reconstruction for partial person re-identification: alignment-free approach. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018. 7073–7082

89 Yoo D, Park S, Lee J-Y, et al. Multi-scale pyramid pooling for deep convolutional representation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, 2015. 71–80

90 Zhao C, Lv X, Dou S, et al. Incremental generative occlusion adversarial suppression network for person ReID. IEEE Trans Image Process, 2021, 30: 4212–4224

91 Li Y, He J, Zhang T, et al. Diverse part discovery: occluded person re-identification with part-aware transformer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021. 2898–2907

92 Zheng W-S, Li X, Xiang T, et al. Partial person re-identification. In: Proceedings of the IEEE International Conference on Computer Vision, Santiago, 2015. 4678–4686

93 He L, Wang Y, Liu W, et al. Foreground-aware pyramid reconstruction for alignment-free occluded person re-identification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, 2019. 8450–8459

94 Sun Y, Xu Q, Li Y, et al. Perceive where to focus: learning visibility-aware part-level features for partial person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, 2019. 393–402

95 Chen P, Liu W, Dai P, et al. Occlude them all: occlusion-aware attention network for occluded person Re-ID. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021. 11833–11842

96 Zheng L, Shen L, Tian L, et al. Scalable person re-identification: a benchmark. In: Proceedings of the IEEE International Conference on Computer Vision, Santiago, 2015. 1116–1124

97 Ristani E, Solera F, Zou R, et al. Performance measures and a data set for multi-target, multi-camera tracking. In: Proceedings of European Conference on Computer Vision, Amsterdam, 2016. 17–35

98 Panareda B P, Gall J. Open set domain adaptation. In: Proceedings of the IEEE International Conference on Computer Vision, Venice, 2017. 754–763

99 Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. 2014. ArXiv:14062661

100 Zhu J-Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, Venice, 2017. 2223–2232

101 Yi Z, Zhang H, Tan P, et al. DualGAN: unsupervised dual learning for image-to-image translation. In: Proceedings of the IEEE International Conference on Computer Vision, Venice, 2017. 2849–2857

102 Choi Y, Choi M, Kim M, et al. StarGAN: unified generative adversarial networks for multi-domain image-to-image translation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018. 8789–8797

103 Wei L, Zhang S, Gao W, et al. Person transfer GAN to bridge domain gap for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018. 79–88

104 Deng W, Zheng L, Ye Q, et al. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018. 994–1003

105 Zhai Y, Ye Q, Lu S, et al. Multiple expert brainstorming for domain adaptive person re-identification. In: Proceedings of the 16th European Conference on Computer Vision, Glasgow, 2020. 594–611

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

106 Wu G, Zhu X, Gong S. Tracklet self-supervised learning for unsupervised person re-identification. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2020. 12362–12369
107 Wang J, Zhu X, Gong S, et al. Transferable joint attribute-identity deep learning for unsupervised person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018. 2275–2284
108 Peng P, Xiang T, Wang Y, et al. Unsupervised cross-dataset transfer learning for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 2016. 1306–1315
109 Hu J, Lu J, Tan Y-P. Deep transfer metric learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, 2015. 325–333
110 Yang F, Yan K, Lu S, et al. Part-aware progressive unsupervised domain adaptation for person re-identification. IEEE Trans Multimedia, 2021, 23: 1681–1695
111 Jiang K, Zhang T, Zhang Y, et al. Self-supervised agent learning for unsupervised cross-domain person re-identification. IEEE Trans Image Process, 2020, 29: 8549–8560
112 Zhong Z, Zheng L, Li S, et al. Generalizing a person retrieval model hetero-and homogeneously. In: Proceedings of the European Conference on Computer Vision (ECCV), Munich, 2018. 172–188
113 Yu H-X, Wu A, Zheng W-S. Cross-view asymmetric metric learning for unsupervised person re-identification. In: Proceedings of the IEEE International Conference on Computer Vision, Venice, 2017. 994–1002
114 Yu H X, Wu A, Zheng W-S. Unsupervised person re-identification by deep asymmetric metric embedding. IEEE Trans Pattern Anal Mach Intell, 2020, 42: 956–973
115 Fan H, Zheng L, Yan C, et al. Unsupervised person re-identification. ACM Trans Multimedia Comput Commun Appl, 2018, 14: 1–18
116 Song L, Wang C, Zhang L, et al. Unsupervised domain adaptive re-identification: theory and practice. Pattern Recogn, 2020, 102: 107173
117 Fu Y, Wei Y, Wang G, et al. Self-similarity grouping: a simple unsupervised cross domain adaptation approach for person re-identification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, 2019. 6112–6121
118 Ester M, Kriegel H-P, Sander J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, Portland, 1996. 226–231
119 Zhang X, Cao J, Shen C, et al. Self-training with progressive augmentation for unsupervised cross-domain person re-identification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, 2019. 8222–8231
120 Xuan S, Zhang S. Intra-inter camera similarity for unsupervised person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021. 11926–11935
121 Yang F, Zhong Z, Luo Z, et al. Joint noise-tolerant learning and meta camera shift adaptation for unsupervised person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021. 4855–4864
122 Yu H-X, Zheng W-S, Wu A, et al. Unsupervised person re-identification by soft multilabel learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, 2019. 2148–2157
123 Zhong Z, Zheng L, Luo Z, et al. Invariance matters: exemplar memory for domain adaptive person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, 2019. 598–607
124 Yang Q, Yu H-X, Wu A, et al. Patch-based discriminative feature learning for unsupervised person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, 2019. 3633–3642
125 Huang H, Yang W, Chen X, et al. EANet: enhancing alignment for cross-domain person re-identification. 2018. ArXiv:181211369
126 Li J, Ma A J, Yuen P C. Semi-supervised region metric learning for person re-identification. Int J Comput Vis, 2018, 126: 855–874

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

127 Chang X, Ma Z, Wei X, et al. Transductive semi-supervised metric learning for person re-identification. Pattern Recogn, 2020, 108: 107569
128 Ding G, Zhang S, Khan S, et al. Feature affinity-based pseudo labeling for semi-supervised person re-identification. IEEE Trans Multimedia, 2019, 21: 2891–2902
129 Liu Y, Song G, Shao J, et al. Transductive centroid projection for semi-supervised large-scale recognition. In: Proceedings of the European Conference on Computer Vision (ECCV), Munich, 2018. 70–86
130 Xin X, Wang J, Xie R, et al. Semi-supervised person re-identification using multi-view clustering. Pattern Recogn, 2019, 88: 285–297
131 Xin X, Wu X, Wang Y, et al. Deep self-paced learning for semi-supervised person re-identification using multi-view self-paced clustering. In: Proceedings of IEEE International Conference on Image Processing (ICIP), Taipei, 2019. 2631–2635
132 Ye M, Ma A J, Zheng L, et al. Dynamic label graph matching for unsupervised video re-identification. In: Proceedings of the IEEE International Conference on Computer Vision, Venice, 2017. 5142–5150
133 Liu Z, Wang D, Lu H. Stepwise metric promotion for unsupervised video person re-identification. In: Proceedings of the IEEE International Conference on Computer Vision, Venice, 2017. 2429–2438
134 Wu Y, Lin Y, Dong X, et al. Exploit the unknown gradually: one-shot video-based person re-identification by stepwise learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018. 5177–5186
135 Wu Y, Lin Y, Dong X, et al. Progressive learning for person re-identification with one example. IEEE Trans Image Process, 2019, 28: 2872–2881
136 Zhu X, Jing X-Y, Wu F, et al. Learning heterogeneous dictionary pair with feature projection matrix for pedestrian video retrieval via single query image. In: Proceedings of AAAI Conference on Artificial Intelligence, San Francisco, 2017. 4341–4348
137 Wang G, Lai J, Xie X. P2SNet: can an image match a video for person re-identification in an end-to-end way? IEEE Trans Circuits Syst Video Technol, 2018, 28: 2777–2787
138 Zhang D, Wu W, Cheng H, et al. Image-to-video person re-identification with temporally memorized similarity learning. IEEE Trans Circ Syst Video Technol, 2018, 28: 2622–2632
139 Niu K, Huang Y, Wang L. Fusing two directions in cross-domain adaption for real life person search by language. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, 2019
140 Jing Y, Si C, Wang J, et al. Pose-guided multi-granularity attention network for text-based person search. In: Proceedings of the AAAI Conference on Artificial Intelligence, New York, 2020. 11189–11196
141 Li S, Xiao T, Li H, et al. Identity-aware textual-visual matching with latent co-attention. In: Proceedings of the IEEE International Conference on Computer Vision, Venice, 2017. 1890–1899
142 Chen D, Li H, Liu X, et al. Improving deep visual representation for person re-identification by global and local image-language association. In: Proceedings of the European Conference on Computer Vision (ECCV), Munich, 2018. 54–70
143 Aggarwal S, Radhakrishnan V B, Chakraborty A. Text-based person search via attribute-aided matching. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass Village, 2020. 2617–2625
144 Jing Y, Wang W, Wang L, et al. Cross-modal cross-domain moment alignment network for person search. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, 2020. 10678–10686
145 Wu Y, Yan Z, Han X, et al. LapsCore: language-guided person search via color reasoning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021. 1624–1633
146 Li H, Xiao J, Sun M, et al. Transformer based language-person search with multiple region slicing. IEEE Trans Circ Syst Video Technol, 2021. doi: 10.1109/TCSVT.2021.3073718
147 Zhu A, Wang Z, Li Y, et al. DSSL: deep surroundings-person separation learning for text-based person retrieval. In: Proceedings of the 29th ACM International Conference on Multimedia, 2021. 209–217
148 Wu A, Zheng W-S, Gong S, et al. RGB-IR person re-identification by cross-modality similarity preservation. Int J Comput Vis, 2020, 128: 1765–1785
149 Wu A, Zheng W-S, Yu H-X, et al. RGB-infrared cross-modality person re-identification. In: Proceedings of the IEEE

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

International Conference on Computer Vision, Venice, 2017. 5380–5389
150 Ye M, Wang Z, Lan X, et al. Visible thermal person re-identification via dual-constrained top-ranking. In: Proceedings of the 27th International Joint Conferences on Artificial Intelligence, Stockholm, 2018. 1092–1099
151 Dai P, Ji R, Wang H, et al. Cross-modality person re-identification with generative adversarial training. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence, Stockholm, 2018. 677–683
152 Wei X, Li D, Hong X, et al. Co-attentive lifting for infrared-visible person re-identification. In: Proceedings of the 28th ACM International Conference on Multimedia, 2020. 1028–1037
153 Lu Y, Wu Y, Liu B, et al. Cross-modality person re-identification with shared-specific feature transfer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, 2020. 13379–13389
154 Wang G A, Zhang T, Cheng J, et al. RGB-infrared cross-modality person re-identification via joint pixel and feature alignment. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, 2019. 3623–3632
155 Wang Z, Wang Z, Zheng Y, et al. Learning to reduce dual-level discrepancy for infrared-visible person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, 2019. 618–626
156 Wang G-A, Zhang T, Yang Y, et al. Cross-modality paired-images generation for RGB-infrared person re-identification. In: Proceedings of the AAAI Conference on Artificial Intelligence, New York City, 2020. 12144–12151
157 Zhao Z, Liu B, Chu Q, et al. Joint color-irrelevant consistency learning and identity-aware modality adaptation for visible-infrared cross modality person re-identification. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2021. 3520–3528
158 Li D, Wei X, Hong X, et al. Infrared-visible cross-modal person re-identification with an X modality. In: Proceedings of the AAAI Conference on Artificial Intelligence, New York City, 2020. 4610–4617
159 Chen Y, Wan L, Li Z, et al. Neural feature search for RGB-infrared person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021. 587–597
160 Liang W, Wang G, Lai J, et al. Homogeneous-to-heterogeneous: unsupervised learning for RGB-infrared person re-identification. IEEE Trans Image Process, 2021, 30: 6392–6407
161 Xu Y, Ma B, Huang R, et al. Person search in a scene by jointly modeling people commonness and person uniqueness. In: Proceedings of the 22nd ACM International Conference on Multimedia, Orlando, 2014. 937–940
162 Xiao T, Li S, Wang B, et al. Joint detection and identification feature learning for person search. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017. 3415–3424
163 Zheng L, Zhang H, Sun S, et al. Person re-identification in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017. 1367–1376
164 Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell, 2017, 39: 1137–1149
165 Munjal B, Amin S, Tombari F, et al. Query-guided end-to-end person search. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, 2019. 811–820
166 Yan Y, Zhang Q, Ni B, et al. Learning context graph for person search. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, 2019. 2158–2167
167 Chen D, Zhang S, Yang J, et al. Norm-aware embedding for efficient person search. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, 2020. 12615–12624
168 Zhong Y, Wang X, Zhang S. Robust partial matching for person search in the wild. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, 2020. 6827–6835
169 Dong W, Zhang Z, Song C, et al. Bi-directional interaction network for person search. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, 2020. 2839–2848
170 Li Z, Miao D. Sequential end-to-end network for efficient person search. In: Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2021
171 Yan Y, Li J, Qin J, et al. Anchor-free person search. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021. 7690–7699
172 Chen D, Zhang S, Ouyang W, et al. Person search via a mask-guided two-stream CNN model. In: Proceedings of the European Conference on Computer Vision (ECCV), Munich, 2018. 734–750
173 Han C, Ye J, Zhong Y, et al. Re-ID driven localization refinement for person search. In: Proceedings of the

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

IEEE/CVF International Conference on Computer Vision, Seoul, 2019. 9814–9823
174 Lan X, Zhu X, Gong S. Person search by multi-scale matching. In: Proceedings of the European Conference on Computer Vision (ECCV), Munich, 2018. 536–552
175 Wang C, Ma B, Chang H, et al. TCTS: a task-consistent two-stage framework for person search. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, 2020. 11952–11961
176 Li J, Liang F, Li Y, et al. Fast person search pipeline. In: Proceedings of IEEE International Conference on Multimedia and Expo (ICME), Shanghai, 2019. 1114–1119
177 Bai S, Li Y, Zhou Y, et al. Adversarial metric attack and defense for person re-identification. IEEE Trans Pattern Anal Mach Intell, 2021, 43: 2119–2126
178 Gong X, Hu G, Hospedales T, et al. Adversarial robustness of open-set recognition: face recognition and person re-identification. In: Proceedings of European Conference on Computer Vision, Glasgow, 2020. 135–151
179 Ding W, Wei X, Ji R, et al. Beyond universal person re-identification attack. IEEE Trans Inform Forensic Secur, 2021, 16: 3442–3455
180 Yang F, Zhong Z, Liu H, et al. Learning to attack real-world models for person re-identification via virtual-guided meta-learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2021. 3128–3135
181 Lin K, Lu J, Chen C-S, et al. Unsupervised deep learning of compact binary descriptors. IEEE Trans Pattern Anal Mach Intell, 2019, 41: 1501–1514
182 Lai H, Pan Y, Liu Y, et al. Simultaneous feature learning and hash coding with deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, 2015. 3270–3278
183 Chen J, Wang Y, Qin J, et al. Fast person re-identification via cross-camera semantic binary transformation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017. 3873–3882
184 Zheng F, Shao L. Learning cross-view binary identities for fast person re-identification. In: Proceedings of the 25th International Joint Conference on Artificial Intelligence, New York, 2016. 2399–2406
185 Zhang R, Lin L, Zhang R, et al. Bit-scalable deep hashing with regularized similarity learning for image retrieval and person re-identification. IEEE Trans Image Process, 2015, 24: 4766–4779
186 Zhu F, Kong X, Zheng L, et al. Part-based deep hashing for large-scale person re-identification. IEEE Trans Image Process, 2017, 26: 4806–4817
187 Liu Z, Qin J, Li A, et al. Adversarial binary coding for efficient person re-identification. In: Proceedings of IEEE International Conference on Multimedia and Expo (ICME), Shanghai, 2019. 700–705
188 Wang G, Gong S, Cheng J, et al. Faster person re-identification. In: Proceedings of European Conference on Computer Vision, Glasgow, 2020. 275–292
189 Zhao C R, Tu Y P, Lai Z H, et al. Salience-guided iterative asymmetric mutual hashing for fast person re-identification. IEEE Trans Image Process, 2021, 30: 7776–7789
190 Hirzer M, Beleznai C, Roth P M, et al. Person re-identification by descriptive and discriminative classification. In: Proceedings of Scandinavian Conference on Image Analysis (SCIA), Ystad, 2011. 91–102
191 Wang T, Wang S, Gong S, et al. Person re-identification by video ranking. In: Proceedings of European Conference on Computer Vision (ECCV), Zurich, 2014. 688–703
192 Zheng L, Bie Z, Sun Y, et al. MARS: a video benchmark for large-scale person re-identification. In: Proceedings of European Conference on Computer Vision (ECCV), Scottsdale, 2016. 868–884
193 Bazzani L, Cristani M, Perina A, et al. Multiple-shot person re-identification by HPE signature. In: Proceedings of International Conference on Pattern Recognition (ICPR), Istanbul, 2010. 1413–1416
194 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. Sensors, 2017, 17: 605
195 Felzenszwalb P F, Girshick R B, McAllester D, et al. Object detection with discriminatively trained part-based models. IEEE Trans Pattern Anal Mach Intell, 2010, 32: 1627–1645
196 Xu X, Liu L, Zhang X, et al. Rethinking data collection for person re-identification: active redundancy reduction. Pattern Recogn, 2021, 113: 107827
197 Wang Y, Liao S, Shao L. Surpassing real-world source training data: random 3D characters for generalizable person re-identification. In: Proceedings of the 28th ACM International Conference on Multimedia, 2020. 3422–3430
198 Zhang T, Xie L, Wei L, et al. UnrealPerson: an adaptive pipeline towards costless person re-identification.

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021. 11506–11515
199 Wang G, Yuan Y, Chen X, et al. Learning discriminative features with multiple granularities for person re-identification. In: Proceedings of the 26th ACM International Conference on Multimedia, Seoul, 2018. 274–282
200 Quan R, Dong X, Wu Y, et al. Auto-ReID: searching for a part-aware convnet for person re-identification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, 2019. 3750–3759
201 Luo H, Gu Y, Liao X, et al. Bag of tricks and a strong baseline for deep person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, 2019
202 Zhou K, Yang Y, Cavallaro A, et al. Omni-scale feature learning for person re-identification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, 2019. 3702–3712
203 Chen B, Deng W, Hu J. Mixed high-order attention network for person re-identification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, 2019. 371–381
204 Luo C, Chen Y, Wang N, et al. Spectral feature transformation for person re-identification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, 2019. 4976–4985
205 Guo J, Yuan Y, Huang L, et al. Beyond human parts: dual part-aligned representations for person re-identification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, 2019. 3642–3651
206 Hou R, Ma B, Chang H, et al. Interaction-and-aggregation network for person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, 2019. 9317–9326
207 Zhang Z, Lan C, Zeng W, et al. Densely semantically aligned person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, 2019. 667–676
208 Zheng Z, Yang X, Yu Z, et al. Joint discriminative and generative learning for person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, 2019. 2138–2147
209 Chen X, Fu C, Zhao Y, et al. Salience-guided cascaded suppression network for person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, 2020. 3300–3310
210 Zhang Z, Lan C, Zeng W, et al. Relation-aware global attention for person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, 2020. 3186–3195
211 Zhu K, Guo H, Liu Z, et al. Identity-guided human semantic parsing for person re-identification. In: Proceedings of the 16th European Conference on Computer Vision, Glasgow, 2020. 346–363
212 Jia M, Cheng X, Zhai Y, et al. Matching on sets: conquer occluded person re-identification without alignment. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2021. 1673–1681
213 He S, Luo H, Wang P, et al. TransReID: transformer-based object re-identification. 2021. ArXiv:210204378
214 Li H, Wu G, Zheng W-S. Combined depth space based architecture search for person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021. 6729–6738
215 Ren M, He L, Liao X, et al. Learning instance-level spatial-temporal patterns for person re-identification. 2021. ArXiv:210800171
216 Xia B N, Gong Y, Zhang Y, et al. Second-order non-local attention networks for person re-identification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, 2019. 3760–3769
217 Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale. 2020. ArXiv:201011929
218 Peng Z, Huang W, Gu S, et al. Conformer: local features coupling global representations for visual recognition. 2021. ArXiv:210503889
219 Hou R B, Ma B P, Chang H, et al. VRSTC: occlusion-free video person re-identification. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, 2019. 7176–7185
220 Gu X, Chang H, Ma B, et al. Appearance-preserving 3D convolution for video-based person re-identification. In: Proceedings of European Conference on Computer Vision (ECCV), Glasgow, 2020. 228–243
221 Ye Y, Wang Z, Liang C, et al. A survey on multi-source person re-identification. Act Autom Sin, 2020, 46: 1869–1884 [叶钰, 王正, 梁超, 等. 多源数据行人重识别研究综述. 自动化学报, 2020, 46: 1869–1884]

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

Key technology for intelligent video surveillance: a review of person re-identification

Cairong ZHAO[1*], Ding QI[1], Shuguang DOU[1], Yuanpeng TU[1], Tianli SUN[1], Song BAI[2], Xinyang JIANG[3], Xiang BAI[2*] & Duoqian MIAO[1*]

1. *School of Electronic and Information Engineering, Tongji University, Shanghai* 201804, *China*;
2. *School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan* 430074, *China*;
3. *Microsoft Research Asia (Shanghai), Shanghai* 200232, *China*
* Corresponding author. E-mail: zhaocairong@tongji.edu.cn, xbai@hust.edu.cn, dqmiao@tongji.edu.cn

Abstract Person re-identification (person ReID) aims to solve the association and matching of target person across cameras and scenes, which is a key link of intelligent video surveillance systems and plays a significant role in maintaining social public order. In order to understand the development status of person ReID technology and accelerate the implementation of person ReID research and applications, this paper provides statistics on the number of applications, funding intensity and geographic distribution of NSFC in this field, as well as a comprehensive review of person ReID research published in top international conferences and journals in the past decade. The paper starts with a standard person ReID algorithm process and details three key techniques: representation learning, metric learning, and re-ranking. Then, the main challenges faced in practical open scenarios are listed, and seven open person ReID tasks are outlined accordingly: occlusion, unsupervised, semi-supervised, cross-modal, end-to-end search, adversarial robustness and fast retrieval. In addition, representative datasets of standard person ReID and open person ReID are collated and some representative algorithms are compared. Finally, this paper provides an outlook on the future trends of person ReID.

Keywords person re-identification, intelligent video analysis, deep learning, feature representation learning, metric learning

Cairong ZHAO is currently a professor at Tongji University. He received his Ph.D. degree from Nanjing University of Science and Technology, M.S. degree from Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, and B.S. degree from Jilin University, in 2011, 2006, and 2003, respectively. His research interests include computer vision, pattern recognition, and visual surveillance.

Ding QI is currently working toward a Ph.D. degree with the College of Electronics and Information Engineering, Tongji University, Shanghai, China. His research interests include computer vision, deep learning, and person re-identification.

Xiang BAI received his B.S., M.S., and Ph.D. degrees from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2003, 2005, and 2009, respectively, all in electronics and information engineering. He is currently a professor at the School of Artificial Intelligence and Automation, HUST. His research interests include object recognition, shape analysis, and OCR.

Duoqian MIAO is currently a professor and a Ph.D. tutor at the College of Electronics and Information Engineering, Tongji University. His interests include machine learning, data mining, big data analysis, granular computing, artificial intelligence, and text image processing.

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net