

中图分类号: TP391.4 文献标识码: A 文章编号: 1006-8961 (2023) 06-XXXX-XX

论文引用格式: Zhang Yongfei, Yang Hangyuan, Zhang Yujia, et al. 2023. Recent progress in person re-ID. Journal of Image and Graphics, XX(XX):XXXX-XXXX (张永飞, 杨航远, 张雨佳等, 等. 2023. 行人再识别技术研究进展. 中国图象图形学报, XX(XX):XXXX-XXXX [DOI:]

行人再识别技术研究进展

张永飞^{1,2*}, 杨航远¹, 张雨佳¹, 豆朝鹏^{3,4}, 廖胜才⁵, 郑伟诗⁶, 张史梁⁷, 叶茫⁸,
晏轶超⁹, 李俊杰⁹, 王生进^{3,4*}

1. 北京航空航天大学计算机学院, 北京 100191; 2. 北京航空航天大学虚拟现实技术与系统全国重点实验室, 北京 100191;
3. 清华大学电子系, 北京 100084; 4. 北京信息科学与技术国家研究中心, 北京 100084;
5. 起源人工智能研究院, 阿联酋阿布扎比 5151; 6. 中山大学计算机学院, 广东省广州市, 邮编 510006;
7. 北京大学计算机学院, 北京 100871; 8. 武汉大学计算机学院 湖北省武汉市 430072;
9. 教育部人工智能重点实验室, 上海交通大学人工智能研究院, 上海 200240

摘要: 行人再识别通过大时空范围内跨摄像机下目标行人图像的检索与匹配, 可实现人脸等生物特征失效情况下的行人关联, 已成为智能视频监控系统的关键环节和支撑技术, 并在智慧公安、智慧城市等国民经济建设中发挥了重要作用。近几年, 行人再识别技术吸引了越来越多的关注, 并取得了快速发展与进步。本文在对行人再识别技术进行简介的基础上, 面向行人再识别的技术发展和落地应用需求与挑战, 将总结分析遮挡行人再识别、无监督行人再识别、虚拟数据生成、域泛化行人再识别、换装行人再识别、跨模态行人再识别和行人搜索等热点方向的前沿进展, 归纳其发展现状和存在问题, 最终将对行人再识别技术的发展趋势进行展望。希望通过总结和分析, 能够为相关研究人员开展行人再识别相关研究、推动行人再识别技术进步提供参考。

关键词: 智能视频监控; 遮挡行人再识别; 无监督行人再识别; 虚拟数据生成; 域泛化行人再识别; 换装行人再识别; 跨模态行人再识别; 行人搜索。

Recent progress in person re-ID

Zhang Yongfei^{1,2*}, Yang Hangyuan¹, Zhang Yujia¹, Dou Zhaopeng^{3,4}, Liao Shengcai⁵,
Zheng Wei-Shi⁶, Zhang Shiliang⁷, Ye Mang⁸, Yan Yichao⁹, Li Junjie⁹, Wang Shengjin^{3,4*}

1. School of Computer Science and Engineering, Beihang University, Beijing 100191, China
2. State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China
3. Tsinghua University (EE), Beijing 100084, China
4. Beijing National Research Center for Information Science and Technology, Beijing 100084, China
5. Inception Institute of Artificial Intelligence, Abu Dhabi 5151, United Arab Emirates
6. School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510006, China
7. School of Computer Science, Peking University, Beijing 100871, China
8. School of Computer Science, Wuhan University, Wuhan, Hubei 430072, China
9. MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University, Shanghai 200240, China

收稿日期: ; 修回日期: ;

所有作者为共同一作; *通信作者: yfzhang@buaa.edu.cn, wgsgj@tsinghua.edu.cn

基金项目: 国家十三五重点研发计划项目(2021YFF0602103, 2021YFF0602102)和国家自然科学基金项目(61936011, 62072022, 62176188, 62201342, U20B2052)

Supported by: National key development program in 14th Five-Year(2021YFF0602103, 2021YFF0602102) and National Natural Science Foundation of China (61936011, 62072022, 62176188, 62201342, U20B2052).

Abstract: Person Re-Identification (person re-ID) aims to solve the problem of association and matching of target person images across multiple cameras within a camera network of a surveillance system, especially in the case of face, iris and other biometrics recognition failure under non-cooperative application scenarios, and has become the key component and supporting technique of intelligent video surveillance systems in numerous applications in intelligent public security and smart cities, etc. Recently, person re-ID has attracted more and more attention from both academia and industry, and has made rapid development and progress. Facing the technical challenges and application needs of person re-ID in practical scenarios, this paper will first give a brief introduction to the development history, major datasets and evaluation metrics of person re-ID, and then summarize and analyze the cutting-edge progress in hot research topics of person re-ID, including occluded person re-ID, unsupervised person re-ID, virtual data generation for person re-ID, domain generalization person re-ID, cloth-changing person re-ID, cross-modal person re-ID and person search. More specifically, to address the problem of impact of possible occlusions on the performance of person re-ID, recent progress in occluded person re-ID is first reviewed in Section 2.1, with a brief introduction of the major datasets for occluded person re-ID in 2.1.1, and review of the two major categories of occluded person re-ID models are presented in 2.1.2. Facing the challenges of low-efficiency and high-cost data annotation and great impact of training data on the performance of person re-ID, unsupervised person re-ID and virtual data generation for person re-ID emerges as two hot topics in person re-ID. Section 2.2 elaborates the recent advances on unsupervised person re-ID, which are classified into three major categories, namely, pseudo label generation-based models, domain transfer-based models and other models, which take into consideration of the extra information besides person image, like time-stamps, camera labels. Section 2.3 surveys the-state-of-the-art works on virtual data generation for person re-ID, with detailed introduction, as well as performance comparisons of major virtual datasets, including SOMAset, UnrealPerson, ClonePerson and WePerson. Despite the impressive progress in the supervised and unsupervised person re-ID settings where new models need to be trained for a new domain by accessing data, most of the exiting models suffer from drastic performance decline on unseen domains, i.e., new scenes in real-world applications, in which data might be unavailable and data re-training might be impossible. Thus, domain generalization person re-ID emerges, aiming at learning a model that has good generalization ability to unseen domains. The recent advances on domain generalization person re-ID will be reviewed in Section 2.4, which are classified into five categories, namely batch/instance normalization models in 2.4.1, domain invariant feature learning models in 2.4.2, deep-learning-based explicit image matching models in 2.4.3, models based on mixture of experts in 2.4.4 and meta-learning-based models in 2.4.5. Since most current person re-ID models largely depend on the color appearance of persons clothes, which might be not reliable when a person changes the clothes, cloth-changing person re-ID becomes a challenging setting in which person images with clothes change exhibit large intra-class variation and small inter-class variation. Typical cloth-changing person re-ID datasets will be introduced in 2.5.1 and the recent progress will be reviewed in 2.5.2, in which models in the first category explicitly introduces extra cloth-appearance-independent features like contour and face while the second try to decouple the cloth features and person ID features. To compensate the drawbacks of conventional person re-ID of visible light/RGB images in actual complex scenes like poor lighting conditions in the night, cross-modal person re-ID aims to explore the problem through other heterogeneous data besides visible RGB images. Section 2.6 surveys the-state-of-the-art cross-modal person re-ID, with commonly used cross-modal person re-ID datasets in 2.6.1 and four sub-categories models in 2.6.2 according to the different modalities employed, namely RGB-infrared image person re-ID, RGB image-Text person re-ID, RGB image-sketch person re-ID and RGB-depth image person re-ID, respectively. Since existing person re-ID benchmarks and methods mainly focus on matching cropped person images between queries and candidates and is different from practical scenarios where the bounding box annotations of persons are usually unavailable and the target person needs to be searched from a gallery of whole scene images, person search, which jointly considers person detection and person re-ID in a single framework, becomes a new hot research topic. The typical datasets and recent progress on person search will be presented in 2.7.1 and 2.7.2, respectively. Finally, the existing challenges and development trend of person re-ID techniques will be discussed. It is hoped that the summary and analysis can provide reference for relevant researchers to carry out research on person re-ID and promote the progress of person re-ID techniques and applications.

Key words: Intelligent video Surveillance, occluded person re-ID, unsupervised person re-ID, virtual data generation, domain

generalization, cloth-changing person re-ID, cross-modal person re-ID, person search.

OCIS:

0 引言

监控视频由于其实时性强、信息量大、直观等特点，在智慧城市、智慧社区、智慧公安等诸多领域发挥着无可替代的重要作用。近年来，我国监控摄像机安装数量持续稳定增长，仅2020年国内安防摄像机出货量就达到了4.1亿台，预计到2025年将突破8亿台。对于海量的监控视频数据，得益于近几年人工智能研究热潮的再度兴起和计算机视觉技术的迅猛发展，图像与视频中的目标检测、跟踪、分类、识别等技术为监控视频的智能化处理与应用提供了新的契机。

传统的人员识别与关联主要以人脸识别、指纹识别等技术为主，并已广泛应用。然而，随着应用场景的不断扩大和应用需求的不断提升，特别是在新冠疫情发生后，人脸识别、指纹识别等技术由于需要目标的配合，其应用范围受到了极大的限制。比如人脸识别技术，需要提取人员的正面脸部特征信息，在戴口罩、或者灯光黑暗、拍摄角度不佳、帽子遮挡等情况下，人脸识别技术难以发挥效用。行人再识别应运而生。

行人再识别(Person Re-Identification, Re-ID)，也被称为行人再识别、行人再识别或“**跨镜追踪**”，旨在基于行人的穿着、体态、发型、携带物等信息，实现大时空范围内人脸等生物特征失效情况下的跨摄像机下目标行人图像的检索与匹配，并借助摄像机的地理位置和拍摄时间，确定该行人出现的时间和地点，并最终形成行人的运动轨迹。行人再识别作为人脸识别、指纹识别等生物特征识别的补充手段，特别是开放环境、非配合场景下上述识别技术失效时的替代手段，已成为智能视频监控系统的关键环节和支撑技术，在智慧公安、智慧城市等国民经济建设中发挥了重要作用。例如，在智慧公安领域，借助行人再识别可实现嫌疑人员长时间、大范围的追踪定位和轨迹还原功能，快速筛查嫌疑人员并进一步锁定出行区域，辅助精准预防和打击；在智慧城市与智慧治理领域，通过行人识别技术，不仅可以实现人流信息的统计，甚至可以实现大场景的人流轨迹还原以及人员的比对与查询，为各种资源的实时管理和优化配置提供支撑；在疫情防控方面，基于行人再识别技术，可以还原行人活动轨

迹并关联追踪到密接、次密接人群，在减轻基层工作人员负担的同时，以最快速度锁定传播范围，极大提升疫情追踪、防控效率；在智能园区和智能交通领域，行人再识别可用于陌生人员的轨迹分析，有效保障园区安全；在人流密集的机场、车站、公园、商场等大型公共场所，行人再识别能够帮助实现不慎走失的儿童、老人的迅速查找；在智能商业领域，行人再识别可以实时动态跟踪顾客的行为轨迹，以此了解用户在商场的兴趣所在，帮助商家挖掘更多的商业价值，同时优化用户体验。此外，行人再识别在智能交通、自动驾驶等领域也已应用或具有重要应用前景。

近几年，随着智慧社会治理、智慧公安等国家战略、需求的推进，特别是深度学习和计算机视觉理论与方法的不断深入，行人再识别技术吸引了越来越多的关注，经过十多年从多摄像机追踪到再识别的起步阶段、基于传统特征的再识别、基于深度学习的再识别三个主要阶段（详见1.2节），取得了快速发展与进步（Ye等，2022）。但同时，实际应用场景下仍面临遮挡、标注成本高、训练样本不足、跨场景跨模态差异大、跨域泛化性能差等诸多挑战，这都导致当前行人再识别算法实际应用场景下的性能还比较低，距离实际可用还有很长的距离。基于此，本文面向行人再识别的技术发展和落地应用需求，总结整理遮挡行人再识别、无监督行人再识别、虚拟数据生成、域泛化行人再识别、换装行人再识别、跨模态行人再识别和行人搜索等方面的前沿进展，归纳其发展现状、存在问题和发展趋势。这对推动行人再识别技术进步，进而提升视频监控分析的智能化水平和实用性具有积极意义和重要作用。其中，遮挡行人再识别是现有再识别问题的一种特殊和挑战场景；而无监督行人再识别、虚拟数据生成、域泛化行人再识别则是主要缓解行人再识别数据收集与标注成本高的挑战；在此基础上，换装和跨模态行人再识别是面向长时间尺度、可能存在白天（可见光）、黑夜（红外）以及行人换装等实际场景下的行人再识别挑战；最后，行人搜索则旨在将行人检测和再识别统一到同一个端到端的框架中，以提升系统的性能和效率。

下面将首先对行人再识别技术进行简介，然后从遮挡行人再识别、无监督行人再识别、虚拟数据生成、域泛化行人再识别、换装行人再识别、跨模

态行人再识别和行人搜索等方面系统地总结和分析行人再识别前沿进展, 最终对行人再识别技术的发展趋势进行展望。

1 行人再识别技术简介

1.1 行人再识别的引入

图1给出了一个典型的智能视频监控系统应用示意图。无论对于嫌疑人员追踪或是走失老人/孩子找回等实际应用, 都需要基于采集所有摄像机下的原始监控视频, 然后基于目标检测算法检测原始监控视频中的行人并框选截图形成行人图像库(也称为注册图, Gallery)。然后, 对于要追踪的嫌疑目标或者要找回的走失老人/孩子, 即可利用其在某个监控摄像机下的图像作为查询图像(Probe或者Query), 基于行人再识别技术, 在保存的海量行人图像库中检索到同一个人的其它图像, 并借助摄像机的地理位置和拍摄时间, 确定该行人出现的时间和地点, 并最终形成行人的运动轨迹以及最后出现的时间地点, 从而实现嫌疑目标追踪或走失老人/孩子找回。

行人再识别算法的一般技术流程通常包括以

下三个步骤, 如图2所示。1) 数据集构建: 在校园等一定可控的场景下数据收集目标行人的监控视频, 并基于YOLO(you only look once)或者Fast RCNN(fast region convolutional neural network)等目标检测算法进行行人检测和是否同一行人目标的标注, 形成带标注的行人图像数据集; 2) 模型训练: 构建行人相似性判别模型, 并基于1)中得到的行人图像数据集进行训练, 得到合适的衡量行人图像之间相似性/距离的判别方法; 3) 行人检索: 最后, 对一个摄像机拍摄的某个目标行人的图像, 将其与其他摄像机拍摄的大量行人图像进行匹配, 找到与其他相似性最高或距离最小的行人图像, 从而实现目标行人的再识别。

1.2 行人再识别技术发展历程

纵观行人再识别的发展历程, 主要经历了从多摄像机追踪到再识别的起步阶段、基于传统特征的再识别、基于深度学习的再识别三个主要阶段。

早在1996、1997年, 就有学者关注行人或车辆的再识别问题(Cai等, 1996; T. Huang和S. Russell, 1997)。在2005年, 阿姆斯特丹大学(University of Amsterdam)的Zajdel等人提出“行人再识别”的概念(Zajdel, 2005)。

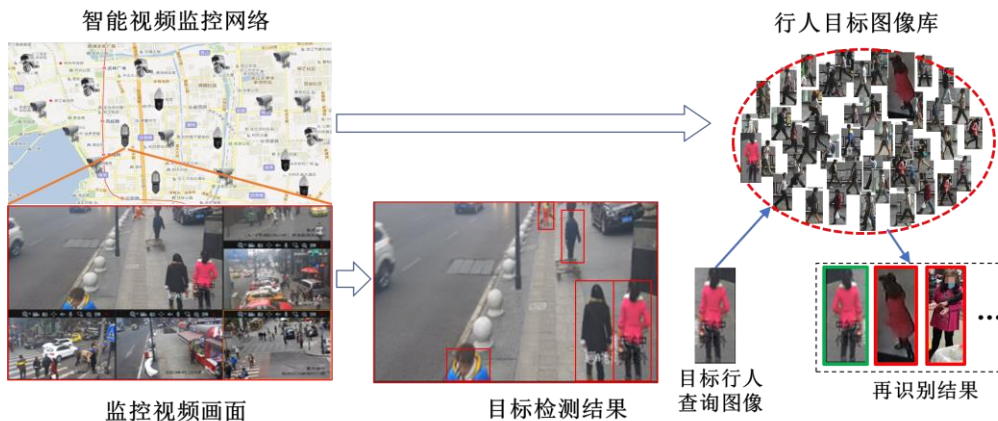


图1 典型智能视频监控系统应用示意图

Fig. 1. Illustration of a typical intelligent video surveillance system



图2 行人再识别系统流程(Ye等, 2022)

Fig. 2. The flow of a practical person Re-ID system

2006年, Gheissari等人在计算机视觉和模式识别会议CVPR(IEEE Conference on Computer Vision and Pattern Recognition)上首次将依靠行人的外观特征进行行人目标匹配的问题定义为行人再识别(Gheissari等), 标志着行人再识别成为一个独立的视觉任务。2007年Gray提出一个对于行人再识别的研究具有重大意义的数据库VIPeR(Gray等2008)。此后, 越来越多的学者开始关注行人再识别的研究, 相关的研究成果也不断涌现。

在2014年以前, 行人再识别主要利用颜色、纹理(Farenzena等, 2010; Xiang等, 2012; Zhao等, 2013; Yang等, 2014)、形状特征如方向梯度直方图(histogram of oriented gradients, HOG)(Oreofejd等, 2010)、SIFT(Scale-invariant feature transform)特征等局部不变量描述符(Kai, 2010; Zhao等, 2015)、Haar(haar-like feature)和DCT(discrete cosine transform feature)特征(Bak S等, 2010)等手工设计的传统底层特征来描述行人的表现信息, 结合度量学习方法(Koestinger, 2012; Xiong等, 2014), 对行人图像的特征距离加以学习。Layne在2012年首次提出将了15种手工标注的基于底层特征的行人属性特征用于行人再识别的方法(Layne等, 2012)。这一阶段还有行人再识别技术发展进程中的两个标志性事件, 其一是在2012年的领域顶会欧洲计算机视觉国际会议ECCV(European Conference on Computer Vision)上召开第一个行人再识别研讨会(First International Workshop on Re-Identification: <http://profs.sci.univr.it/~cristanm/reid2012/>), 其二就是2013年, 伦敦大学玛丽皇后学院的龚少刚教授(Shaogang Gong)等出版了第一本行人再识别的专著(Gong和Cristani等, 2013), 分别从学术交流和研究方面为行人再识别提供了坚实的基础和更大的舞台, 有效促进了行人再识别技术的快速发展。然而, 但是, 由于行人姿态变化大, 人体表现特征光照、复杂背景等因素影响, 如图3所示, 仅靠传统特征无法覆盖所有情况, 手工特征无法满足大规模、复杂场景的行人再识别应用需求。

2014年之后, 随着AlexNet(Alex network)(Krizhevsky等, 2014)的出现, 以卷积神经网络为代表的深度学习以其高效精准的特征提取能力和强大的端到端学习能力, 在图像分类等计算机视觉任务上取得了远超传统方法的精度。香港中文大学团队的相关工作DeepReID(Li等, 2014)是最早使用深度学习的行人再识别的工作之一, 标志着行人

再识别进入了基于深度学习的新的发展阶段, 基于深度学习的行人再识别也成为了学界研究的主流方向, 并迎来了井喷式的增长。同时, 相关行人再识别数据集也在不断地扩充(详见1.3节), 有效促进了行人再识别的广泛研究和快速发展。在各大计算机视觉会议中出现了动辄几十篇的相关论文, 行人再识别技术迅猛发展, 日新月异, 在各个数据集上的结果也获得很大提升。图4是2008年到2022年间计算机视觉三大顶级会议CVPR, ICCV(IEEE International Conference on Computer Vision)和ECCV上接受的行人再识别论文情况。此外, 在TPAMI(IEEE Trans. on Pattern Analysis and Machine Intelligence), TIP(IEEE Trans. On Image Processing), IJCV(International Journal of Computer Vision), TMM(IEEE Trans. on Multimedia), TCSVT(IEEE Trans. on Circuits and Systems for Video Technology)等相关国际顶级刊物上, 也发表了大量行人再识别的研究成果。行人再识别技术快速发展, 已成为计算机视觉的一个热点方向。

根据作者前期综述论文(Ye和Shen等, 2022)和最新调研, 结合已有总结(赵才荣等, 2021)等, 2017年前的行人再识别研究主要集中在行人的特征表达和相似性度量上, 而2018年到2020年结合注意力的研究促进了行人再识别性能的提升。而近年来, 面向大规模复杂实际应用场景, 遮挡、跨域、换装、跨模态、无监督、虚拟数据生成和行人检索等研究逐渐成为开放域行人再识别的趋势。

1.3 行人再识别数据集

行人再识别面临着光照、视角、姿态、摄像参数、遮挡、背景变化等诸多因素影响。为了更好的模拟这些因素、训练更好的行人再识别模型以及评估不同行人再识别算法的性能, 从2007年Gray提出第一个仅包含2个摄像头下、分别率为 128×48 的632个行人的1264张图像的行人再识别数据集VIPeR(Gray等, 2008), 再到第一个足以支撑深度学习的行人重新识别数据集CUHK03(Li和Zhao等, 2014), 不断有研究者构建出考虑更复杂场景因素、更大规模的数据集, 极大地促进了行人再识别技术的发展。鉴于多篇参考文献(赵才荣等, 2021; Ye等, 2022; Gou等, 2022)中已对行人再识别数据集进行了较为的介绍, 本文重点介绍目前最常用的三个行人再识别数据集, 即Market1501(Zheng和Shen等, 2015)、DukeMTMC-reID(Ristani等, 2016)和MSMT17(Wei和Zhang等, 2018)。此外, 针对

遮挡、换装、跨模态、行人搜索等特殊场景, 本文 将在第二部分相应章节进行介绍。



图 3 行人再识别面临的挑战

Fig.3 Challenges faced by Person Re-Identification

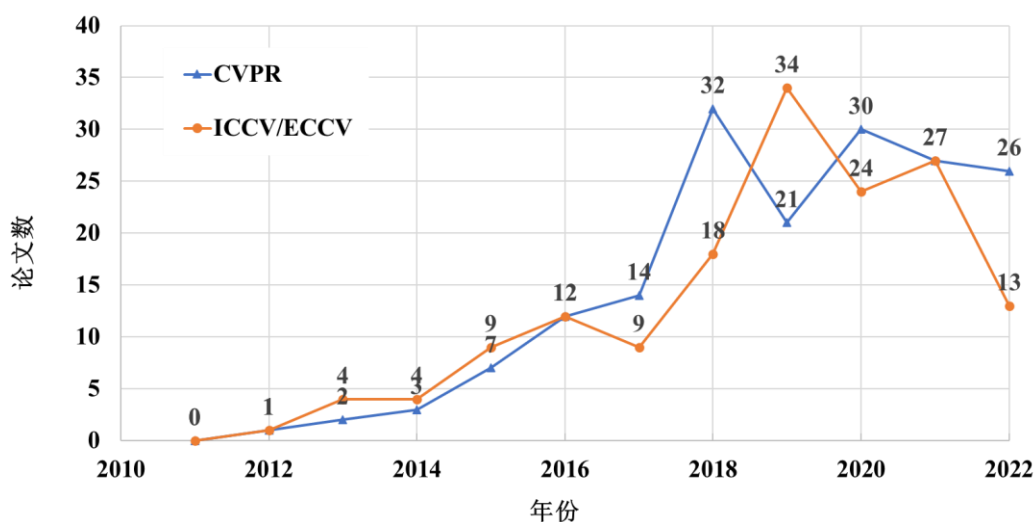


图 4 计算机视觉三大顶级会议上的行人再识别论文情况

Fig. 4 Number of Re-ID papers on three top conferences of computer vision

1) Market1501数据集 (Zheng和Shen等, 2015)。

Market1501数据集是由清华大学研究团队在2015年发布的行人再识别数据集。该数据集采集自6个摄像机, 包括5个高分辨率摄像机和1个低分辨率摄像机。该数据集总共包含有1501个人身份类别的32668张行人图片, 其中训练集包含751个行人类别和12936张行人图像, 测试集中包含750个行人类别和19732张行人图像。在测试集中, 查询图像采用人工标注的检测边界框, 而候选图像采用DPM(Deformable Parts Model)生成的检测边界框, 行人检测器的框不对齐、遮挡情况等更加严重。之后, 为了模拟大规模检索, 作者还加入了约50万张干扰图像, 包含了失败的检测框、不在测试集当中的行人等, 使得这个数据集更贴近真实场景。

2) DukeMTMC-reID数据集 (Ristani等, 2016)。

DukeMTMC-reID数据集是用于多目标跨镜头跟踪的数据集, 由美国杜克大学研究团队在2016年构建并发布。研究者使用8个摄像机来拍摄8段85分钟的高分辨率视频。基于这个数据集, 研究者构造了规模和Market-1501相似的行人再识别数据集DukeMTMC-reID, 其中的行人图像是手动框选出来的, DukeMTMC-reID数据集包含来自8个摄像机的1812个行人的36411张图像, 其中有702个行人被用于训练, 702个行人被用于测试集。

3) MSMT17数据集 (Wei和Zhang等, 2018)。

MSMT17数据集由北京大学的研究团队于2017年构建并发布的行人再识别数据集。MSMT17数据集是目前规模最大的、最接近真实场景的基于图像的行人

人再识别数据集。其原始数据采集自15个摄像机，包含12个户外摄像机和3个室内摄像机。为了提高数据集的多样性，该数据集拍摄的时间跨度大，在一个月内采集了4天，分别拍摄上午、中午、下午三个时间段的行人视频，每天拍摄3个小时的视频数据，因此所有视频数据的总时长为180小时。该数据集有复杂的场景和丰富的光照变化，即在不同的天气、不同的光照条件下拍摄采集。行人框使用Faster RCNN (Ren等, 2015) 检测器产生。最终收集的数据集包含了4101个行人类别和126441幅行人图像。其中，训练集包含1401个行人类别和32621幅行人图像，测试集包含3060个行人类别和93820幅行人图像。

表1比较了这三个行人再识别数据集的发布时间、图片数量、行人身份数量、摄像机数量信息。

1.4 行人再识别评测指标

行人再识别任务的常用评测指标主要有Rank-N指标、累计匹配性能曲线(Cumulated Matching Characteristics, CMC)(Moon和Phillips, 2001)、平均精度均值(mean Average Precision, mAP)(Zheng和Shen等, 2015)和平均逆负样本惩罚率(mean Inverse Negative Penalty, mINP)(Ye和Shen等, 2022)等。

1) **Rank-N**。Rank-N指按照查询图与所有注册图间的特征距离对检索得到的图像从小到大进行排序后，前N张排序结果中包含正确目标行人图像的概率，概率越大表示再识别算法的效果越好。Rank-N是一个最直观的指标，常用的有Rank-1、Rank-5、Rank-10和Rank-20，分别代表在第1张、前5张、前10张和前20张注册图图像中包含正确目标行人图像的概率。

2) **累计匹配性能曲线 (Cumulated Matching Characteristics, CMC)**。虽然Rank-N简单直观，但是每个Rank-N指标只能体现在给定N值情况下的性能。为此，Moon和Phillips (2001) 提出了累计匹配性能曲线(Cumulated Matching Characteristics, CMC)同时表示不同N值情况下的性能。CMC曲线首先使用再识别模型对查询图和注册图提取特征并计算特征之间距离，并按照特征距离从小到大对注册图进行排序；然后，取排在前K张的注册图作为查询结果，如果该结果中包含正确的目标行人图像，则设

置结果概率为1，否则设置结果概率为0；最后，对所有查询图的结果概率计算平均值。因此，随着K值的增大，CMC曲线呈上升趋势，它反映了不同长度候选列表中被检索的目标出现的概率。CMC是行人再识别领域最常用的性能评估方法之一。

3) **平均精度均值mAP**。CMC曲线可以有效表示不同查询下的精确率(Precision)性能，但是难以反映查询结果的召回率(Recall)性能。因此，便引入了平均精度均值mAP。mAP中的AP即平均精度(Average Precision) (Zheng和Shen等, 2015)，通过计算每个查询下精确率-召回率曲线(Precision-Recall Curve, PRC) 曲线下的面积，同时兼顾了精确率和召回率性能。而mAP通过计算多个查询图像的平均准确率的均值来综合评价算法的性能，因此，mAP是一个比CMC曲线更全面的评价指标。其计算公式为：

$$mAP = \frac{1}{m} \sum_{i=1}^m AP_i \quad (1)$$

式中， m 为查询集中查询图像的数量， AP_i 为第 i 个查询图像的平均精度。

4) **平均逆负样本惩罚率 (mean Inverse Negative Penalty, mINP)**。理想的ReID算法应该使得所有query到的图片排得尽可能靠前而CMC和mAP都无法体现这一点。为此，Ye和Shen等(2022)提出了平均逆负样本惩罚率(mean Inverse Negative Penalty, mINP)指标，其定义和计算公式如下：

$$mINP = \frac{1}{m} \sum_{i=1}^m \frac{|G_i|}{R_i^{hard}} \quad (2)$$

式中， m 为查询集中查询图数量， R_i^{hard} 指最困难正样本在匹配列表中的索引rank值，即最后一个正确匹配结果的索引位置， $|G_i|$ 指正确匹配样本的总数。

通过考虑正确匹配样本，特别是最难匹配正样本在匹配列表中的索引rank值，mINP能够弥补CMC和mAP难以衡量模型在最难匹配正样本上性能的问题。

需要说明的是，自从Market-1501等含有多个摄像机数据的大型数据集被提出后，研究者基本都只关注跨摄像机再识别的准确性，因此在计算上述指标时会忽略来自同一个摄像机的正确匹配图像。

表 1 行人再识别三大常用数据集信息统计表

Table 1 Statistics of the Three Most Commonly used Re-ID Datasets

数据集	发布时间	时间跨度	空间跨度	人数	图像数	摄像机数	说明
Market1501	2015	-	清华大学校园	1,501	32,668	6	含干扰图像，接近真实场景

数据集	发布时间	时间跨度	空间跨度	人数	图像数	摄像机数	说明
DukeMTMC-reID	2016	-	杜克大学校园	1,812	36,411	8	行人图像由手动框选
MSMT17	2018	一个月	北京大学校园	4,101	126,441	15	规模大、接近真实场景

2 行人再识别前沿进展

由1.2节可知,近年来,面向大规模复杂实际应用场景,跨域、遮挡、换装、跨模态、无监督、虚拟数据生成和行人检索等研究逐渐成为开放域行人再识别的趋势。因此,本节面向行人再识别的技术发展和落地应用需求,将从域泛化行人再识别、虚拟数据生成、遮挡行人再识别、换装行人再识别、无监督行人再识别、跨模态行人再识别和行人搜索等方面,梳理总结行人再识别技术的前沿进展,归纳其发展现状、存在问题和发展趋势。

2.1 遮挡行人再识别

在现实视频感知场景中,镜头下的行人图像存在着普遍的遮挡问题。该问题严重影响了现有行人再识别算法在实际场景中的性能。近三年来很多学者开始关注遮挡行人再识别算法研究。近年来已有研究通过提取可见的行人特征、特征配准、补全缺失行人信息等手段有效地降低了遮挡问题的影响,提升了遮挡行人再识别的性能。本部分将首先介绍遮挡行人再识别常用的数据集,然后重点阐述近几年遮挡行人再识别研究取得的进展和结果。

2.1.1 遮挡行人再识别数据集简介

为了更全面地检验遮挡行人再识别方法的有效性,除了1.3小节中介绍的Market-1501、DukeMTMC-reID等全身数据集外,常用的遮挡行人再识别数据集有两大类共四个:遮挡数据集Occluded-Duke(Miao等,2019)和Occluded-REID(Zhuo等,2018);部分成像数据集Partial-REID(Zheng, W.S.等,2015)和Partial-iLIDs(Zheng等,2011)。

1) 遮挡数据集。 Occluded-Duke 是从 DukeMTMC-reID 数据集中重构出来的。它删除了 DukeMTMC-reID 中一些重复的图片,并且挑选受到遮挡的行人图像作为待检图像。Occluded-Duke 包含 15618 张训练图像,2210 张待检图像(query)以及 17661 张数据库的图像(gallery)。目前,它是遮挡行人再识别问题中最大的且最具挑战的遮挡数据集。Occluded-REID 包含来自于 200 个行人的 2000 张图像。每个行人有 5 张全身的图像和 5 张被遮挡图像。

其中,被遮挡图像作为待检图像,全身图像作为数据库的图像。由于 Occluded-REID 数据集规模较小,且没有训练集,所以一般在 Market-1501 上进行训练,在 Occluded-REID 上进行测试。

2) 部分成像数据集。除了以上两个遮挡数据集之外,由于部分成像行人再识别和遮挡行人再识别面临的核心问题相近,两个部分成像数据集 Partial-REID 和 Partial-iLIDs 也常用来测试遮挡行人再识别方法的性能。Partial-REID 包含来自 60 个人的 600 张图像。每个行人有 5 张全身图像和 5 张部分成像图像。其中,部分成像图像作为待检图像,全身图像作为数据库图像。Partial-iLIDs 包含从机场拍摄的 119 个行人的 238 张图像。每人有 2 张图像,一张作为待检图像,另一张作为数据库图像。同样地,Partial-REID 和 Partial-iLIDs 没有训练集,所以一般在 Market-1501 上进行训练。

2.1.2 遮挡行人再识别研究进展

在遮挡行人再识别的问题中,待检索图像中的目标行人通常受到遮挡,造成目标行人表观信息缺失的问题。该问题给遮挡下的行人再识别带来了两个主要的挑战:一是遮挡物给提取的特征中引入了噪声,影响检索过程中的特征配准和相似性计算。二是被遮挡区域的行人信息缺失,使得提取的行人特征鉴别性降低。为了解决以上两种挑战,目前解决遮挡行人再识别的方法可以被大致地划分为两种:基于可见区域的遮挡行人再识别方法和基于补全的遮挡行人再识别方法。

1) 基于可见区域的遮挡行人再识别方法。这类方法的主要思路是,精确定位图像中的行人可见区域,仅仅依靠可见的行人部分进行再识别,忽略遮挡物和背景的影响。Miao 等人(2019)提出一种基于行人关键点信息的特征对齐方法。他们首先利用一个额外的人体关键点检测网络,得到图像中人体关键点的位置以及关键点是否被遮挡的信息。然后,提取关键点的特征作为行人的特征表示。在检索过程中,仅仅考虑没有被遮挡的关键点特征之间的距离,从而实现了特征对齐。He 等人(2018)提出了一种基于特征重建的方式,在求解重建矩阵的过程中隐式地实现了特征对齐。然而,求解重建矩阵需要的计算复杂度高,不适用大规模的再识别场景。

史维东等人(2020)通过擦除空间注意力高响应对应的区域而仅保留低响应的特征区域,模拟行人遮挡样本,改善了网络对遮挡样本的能力。Wang和Yang等人(2020)将行人的关键点特征看作图中节点,利用图卷积网络学习关键点之间的高阶关系和拓扑信息,进而提取鉴别力强的特征表示。田维怡等人(2021)借助额外的行人解析模型在遮挡环境下精确定位行人图像的前景,并且建模像素特征不确定性来提高网络的抗遮挡能力。李吉林等人(2021)借助注意力机制,根据人体姿态信息计算行人图像的注意力图,用于指导网络过滤背景遮挡信息的干扰,并借助于特征金字塔缓解了行人图像尺度差异较大的问题。任雪娜等人(2021)以行人的语义掩膜作为监督信息,通过全局和局部的语义引导来提取行人的全身和局部特征。在推理时,仅仅考虑共享的可见人体区域特征,提高了模型对遮挡的容忍能力。Zheng和Lan等人(2021)同样借助于关键点的特征,他们提出一种基于知识蒸馏的特征引导学习方法来学习不同身体部位的对齐表征。受益于知识蒸馏和交互学习,在该方法中,关键点检测器可以在测试时丢弃,从而提高检索的效率。这几个方法需要借助于额外的关键点检测模型或者姿态估计模型,由于这些模型的训练数据和行人再识别的数据之间存在着一定的域差异,导致在行人再识别数据上提取的关键点的信息容易出错,影响最终的行人再识别性能。考虑到这个问题,Yang和Zhang等人(2021)提出将姿态信息离散化为身体部位的可见性标签,以减轻噪声姿态信息在关键点特征提取过程中以及特征配准过程中的影响。Cheng和Jia等人(2022)通过集成多个在不同的数据上训练的行人解析模型来生成更可靠的行人语义信息,帮助提取更准确的行人特征。Chen和Liu等人(2021)提出了OAMN网络,通过给图像中人为地添加遮挡物,对训练数据进行增强,同时得到遮挡物精确位置标签。基于该标签,OAMN学习一个注意力模块,关注于行人的可见区域,有效减小了遮挡物带来的影响。Jia和Cheng等人(2022)提出了一个基于Transformer的解耦表示网络DRL-Net(Deep Reinforcement Learning-Network)来解决遮挡行人再识别问题,该网络不需要严格的特征对齐和额外的信息。基于Transformer架构,DRL-Net通过对被遮挡行人的局部特征进行全局推理来实现无需对齐的检索方法。在学习过程中,Transformer的queries启发式地解耦了抽象的语义部件。在检索过程中,直

接利用queries来计算图像的相似性,无需进行精确的特征对齐。Dou等人(2022)受到人类视觉系统中“自顶向下”的处理机制的启发,提出了为行人提取多粒度特征的方法,将行人的特征表示拓展到像素级别,并给不同粒度的特征设计了基于对齐的相似性度量方式,有效地解决了检索过程中的特征配准的问题。Tan等人(2022)提出了一个多头自注意力网络来剪枝图像中不重要的信息以及捕获行人关键的局部信息,有效减小了遮挡物的影响。

2) 基于补全的行人再识别方法。这类方法的主要思路是,利用时空上下文来补全图片中遮挡区域缺失的行人信息,提高特征的鉴别性。Iodice等人(2018)提出了部分匹配网络PMN(Partial Matching Network),它可以检测人体关键点,对齐部分视图并且根据视频帧中的可见行人信息和特有的行人模型来幻觉出行人缺失的部分,然后将对齐的视图和重建的视图组合成一个联合表示用于图像匹配。Hou等人(2019)提出STCnet(Spatio-Temporal Completion network)来利用视频中的空间连续性和时间连续性来补全某一帧被遮挡的行人区域,从而提高特征的鉴别力,然而这种方法仅适用于视频行人再识别,在图像行人再识别中,时序信息是无法获得的。为了解决这个问题,Xu和He等人(2022)提出了补全图像中缺失部件特征的方法。给定一张待检索的遮挡图像,他们首先为其提取部件特征,然后进行检索,得到图像库中的与之 k 张最相似的图像。然后用这 k 张图像的部件特征来恢复遮挡图像中缺失的部件特征,实现进一步检索。Hou等人(2021)设计了一种即插即用的模块,利用非遮挡区域的空间文来补全遮挡区域的行人特征,同时利用时间序列的上下文来进一步提纯补全的特征。该方法可以同时适用于视频行人再识别和图像行人再识别问题。

遮挡行人再识别问题受到越来越多的关注,在具有挑战性的遮挡数据集Occluded-Duke上,Rank-1指标已经由最开始提出时的51.4%不断提升,目前的已经达到了70+%。这其中,基于可见区域的方法和基于补全的方法均取得了相当的结果。具体来说,目前性能最好的基于可见区域的方法是Cheng和Jia等人(2022)提出的MSDPA(multi-source dynamic parsing attention),在Occluded-Duke上的Rank-1指标为70.4%。性能最好的基于补全的方法是Xu和He等人(2022)提出的FRT(Feature Recovery Transformer),在Occluded-Duke上的Rank-1指标为70.7%。

2.2 无监督行人再识别

利用人工在多路视频中查找和标注相同行人难度高、效率低, 严重制约了有监督行人再识别方法的应用。针对该问题, 近三年很多学者开始关注无监督行人再识别算法研究。相关方法通过引入预训练模型、数据增强策略、设计伪标签生成方法等方式显著降低了人工标注需求, 提升了无监督行人再识别性能, 推进了行人再识别系统的落地应用。本部分将对近三年无监督行人再识别领域的进展进行分类和总结。

1) **基于伪标签生成的方法。**很多无监督行人再识别方法首先在无标注数据上利用标签预测算法来获取每个数据的伪标签, 然后利用这些伪标签来训练模型。当前获取伪标签的主要方法包括基于K近邻的方法(Zhong等, 2019; Yu等, 2019; Zhong等, 2020)、基于K均值聚类的方法(Ge等, 2020; Huang等, 2019; Zhai等, 2020)以及基于密度空间聚类(Density-Based Spatial Clustering of Applications with Noise, DBSCAN(Ester等, 1996)(Fu等, 2019; Dai等, 2021)等。例如Zhong等人(2019)使用KNN(K-NearestNeighbor)方法获取正样本, 即每个样本的前K个近邻样本作为其正样本, 其余样本为负样本。Zhong等人(2020)进一步利用图神经网络优化特征距离远近关系, 获得更准确的K近邻正样本。Yu等人(2020)提出在使用特征间欧氏距离之外, 额外使用对目标图像的分类预测得分来共同计算样本间相似度, 提升了K近邻正样本的准确率。也有研究者设计了新的聚类算法来预测正负样本(Ding等, 2019; Lin等, 2019)。例如, Lin等人(2019)提出了一种自底向上的聚类方法(Bottom-Up Clustering, BUC)。BUC首先将每个单独的样本看作一个簇, 然后逐渐融合相近的簇为同一个簇。融合过程中, BUC使用两个簇内最小的样本距离作为两个簇的距离, 并融合距离最近的簇。他们还设计了簇内样本个数的正则化项以优先融合样本个数较少的簇, 从而平衡不同簇内的样本数量, 并最终较好地平衡了簇内样本的相似性和差异性。Ding等人(2019)则进一步优化了BUC中簇间距离计算方式, 提升了簇内样本的准确性。这些聚类方法都有助于在无标注数据上获取更为准确的伪标签, 用于监督模型的学习。

在对无标注数据预测伪标签的基础上, 可以使用分类损失函数训练网络。为提升特征的判别力, 也有工作在特征空间引入三元组损失函数共同训

练模型(Fu等, 2019; Lin等, 2020)。但是, 基于伪标签训练模型会受到伪标签中噪声的干扰, 因此直接使用伪标签作为分类或三元组损失难以达到较高的性能。为应对伪标签中噪声的影响, 研究者们提出了多种方法来提升训练过程对标签噪声的鲁棒性, 主要手段为模型协同训练。这类方法(Yu等, 2019; Han等, 2018; Ke等, 2019)使用多个模型协同训练以提升对噪声的鲁棒性。例如, Han等人(2018)提出使用两个网络互相为对方选择训练样本并预测伪标签以降低噪声标签的影响。随后, Tarvainen等人(2017)提出了均值教师模型用于模型协同训练, 并应用于半监督训练任务中。均值教师模型是指在训练中, 利用动量更新的方式累加各次梯度更新后的模型参数, 并将其作为教师模型, 指导当前模型的训练。

在Tarvainen等人(2017)工作基础上, Jiang等人(2018)发现均值教师模型能够减小噪声标签对训练的影响。He等人(2020)将均值教师模型用于自监督特征学习, 并有研究者做了许多改进工作(Grill等, 2020)。在此启发下, Ge等人(2020)提出了相互均值学习方法(Mutual Mean-Teaching, MMT), 将均值教师模型和双模型协同训练方法相结合, 使用两个均值教师模型同时监督模型训练, 并应用于无监督目标再识别任务中。Zhai等人(2020)则在MMT的基础上进一步拓展, 提出了多模型训练方法(Multiple Expert Brainstorming Network, MEB-Net), 使用多个不同网络结构的三个均值教师模型联合训练。

2) **基于域迁移的方法。**为解决目标场景无标注数据的问题, 另一些方法试图利用其他场景的标注数据, 并将其中的信息迁移到目标场景之中。有研究者尝试利用GAN模型将有标注的其他场景图像迁移到目标场景所在的数据分布之中, 从而减小光照、背景等图像分布差异, 以辅助目标场景的模型训练(Wei和Zhang等, 2018; Chen等, 2019; Liu, Zha和Chen等, 2019)。例如, Wei和Zhang等人提出的行人图像风格迁移模型(Person Transfer GAN, PTGAN)将有标注的行人图像迁移到无标注行人图像的风格之中, 并提出行人目标的前景一致性约束来提高生成图像的质量。图5中举例展示了四组PTGAN方法所生成的迁移风格后的图像, 每组中第一列为来自于CUHK03数据集的原始标注图像, 第二列为将其风格迁移到PRID数据集所在场景之后的图像。另一类域迁移的方法(Mekhzazni等, 2019;

Liu等, 2020) 通过拉近不同场景数据在特征空间的

分布距离, 使得模型在标注图像上学习到的表达能

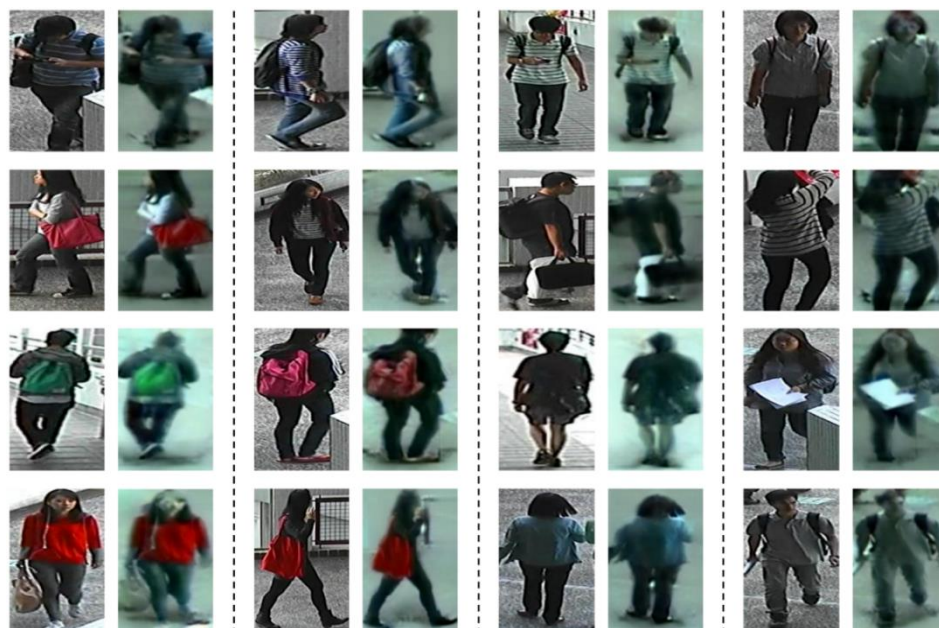


图5 行人目标图像风格迁移效果举例 (Wei 等, 2018)

Fig. 5 Illustration of Results by Person Transfer GAN (Wei 等, 2018)

力可以更有效地推广到无标注场景。例如, Liu和Zhang等人(2020)提出基于对抗训练的场景不变映射, 要求不同场景的数据在判别器层面是不可分的, 从而拉近特征空间中不同场景下特征分布。

3) 其他方法。行人再识别任务一般针对于监控摄像机数据, 在数据收集的过程中, 除了图像本身, 还可以得到对应图像的时间戳信息, 摄像机的编号等信息。一些方法通过使用这些额外信息, 进一步提高了无监督行人再识别方法的性能。

Li等人(2020)在常规行人外观特征的基础上, 根据对应图像的时间戳信息, 提出了基于视觉时空一致性相似度的聚类方法。该方法通过对不同图像间的时间采样间隔进行分布建模, 并根据分布模型对基于行人外观特征的图像相似度进行修正, 从而生成更加准确的伪标签。为了优化网络, Li等人进一步提出了局部和全局分类损失函数。局部分类损失函数将每个训练批次中的图片看作不同类别, 并拉开它们特征间的距离。全局分类损失函数根据视觉相似度和时空一致性相似度对整个训练集进行聚类的结果, 通过多类别分类损失函数将同一类别图像的特征拉近不同类别图像的特征推远。除了训练阶段, 所提出的时空一致性相似度可以用于一般行人再识别方法的后处理来进一步提高模型性能。

与大多数直接使用样本相似度进行聚类来生成伪标签的方法不同, Xuan等人(2021)根据图像

被拍摄摄像机的标签, 将相似度计算划分为摄像机内部和跨摄像机两个阶段。摄像机内部阶段通过计算摄像机内部样本间的相似度进行聚类生成伪标签来训练网络, 从而提高网络对于摄像机内部样本外观差异的鲁棒性。跨摄像机阶段为了减少摄像机差异所引起的相似度降低, 提出了基于域无关相似度的跨摄像机相似度计算方法, 并使用跨摄像机相似度进行聚类生成伪标签训练网络, 提高了网络对于样本间跨摄像机外观差异的鲁棒性。

2.3 虚拟数据行人再识别

数据匮乏一直制约着行人再识别的发展, 除了标注困难之外, 近年来对监控视频数据隐私敏感性和伦理问题的关注更是让这个问题雪上加霜, 已经有一些公开数据集因此而下架了, 如DukeMTMC (Ristani等, 2016), MS-Celeb-1M (Guo等, 2016), TinyFace (Cheng等, 2018) 等数据集。与之相反, 虚拟数据几乎无需标注、没有隐私敏感问题, 且能够弥补现实世界的不足, 创造出全新、独立的图像。为此, 研究人员开始探索通过游戏引擎产生虚拟数据以辅助行人再识别。并且, 一些方法通过生成大规模虚拟数据库, 证明了其能显著提升可泛化行人再识别的能力。本部分将全面阐述近些年行人再识别虚拟数据的构建方法和取得的进展和结果 (近年来虚拟行人再识别数据集的总结见表2, 数据集示例图像见图6)。SOMAs (somatotype networks)

(Barbosa等, 2018)是最早创建虚拟数据辅助行人再识别的工作, 其使用Makehuman三维人物模型和

表 2 近年来虚拟行人再识别数据集的总结

Table 2 Summary of the virtue re-ID datasets

数据集	人数	摄像机数	图像数	场景数	模拟监控	真实衣服纹理	全身搭配	MSMT17***	
								Rank1	mAP
SOMAsset	50	250*	100,000	1	否	否	否	1.4	0.3
SyRI	100	280*	56,000	140**	否	否	否	16.4	4.4
PersonX	1,266	6*	273,456	2**	否	否	否	11.7	3.6
RankPerson	8,000	19	1801,816	11	是	否	否	20.0	6.3
UnrealPerson	6,799	34	1256,381	9	是	是	否	38.5	15.3
ClonePerson	5,621	24	887,766	6	是	是	是	51.6	20.8
WePerson	1,500	40	4,000,000	14	是	否	否	46.4	18.9

注: *该数值为单人多角度和/或多场景摆拍。**该数值为独立无关场景。***跨库泛化到 MSMT17 上的测试结果, 其中 PersonX 和 SOMAsset 的结果来自 RandPerson 论文, SyRI 的结果来自 UnrealPerson 论文。

Blender渲染引擎创建了一个有50名行人、11种布料、8套服装和250个姿势的虚拟Re-ID数据集,如图6(1)所示。但为了专注前景,其将合成行人放置在单个场景中,缺少了场景的多样性。此外,Bak等(2018)提出了一个包含100个虚拟行人的数据集SyRI。通过采集140万张各种光照的HDR(High Dynamic Range)环境图、使用Adobe Fuse CC制作三维人体、Unreal Engine 4渲染三个步骤,构建了包含100个虚拟行人的数据集,如图6(2)所示。但SyRI只有100个行人,背景和人物外貌多样性有限。进一步地,Sun等人(2019)首先基于Unity创建了一个包含1266个三维人体模型的灵活的合成数据引擎PersonX,可以根据输入的三维人物模型随意调节参数合成各种行人图片,如图6(3)所示。在此基础上,该文定量分析了摄像头角度对于re-ID准确率的影响。

然而,以上工作主要采用的是公开获取的或手工制作的三维人物模型,因此其人物数量很难增长到一定的量级,很难再进行扩展。例如,上述三个虚拟行人数据集中最大的是PersonX,但也仅有1266人。此外,这些数据集都是每个人物独立导入场景中并在同一时刻只有一个摄像机进行拍照或录像的,这与真实监控场景下多摄像机同时拍摄一个大场景下所有人的活动有很大区别,最明显的区别是缺乏人与人之间的遮挡和摄像机之间的转移规律。因此,为了解决上述这些问题,Wang和Shao等人(2020)提出一种通过随机组合颜色和纹理产生大量UV纹理图的方法,并由此创建了大量的三维衣服模型。随后,针对MakeHuman的人物格式,作者提出一种通过Python代码随机自动产生大量三维人物模型的方法,使得创建的三维人物在数量上具有很

好的可扩展性,且在身高、肤色、衣服颜色和纹理、配饰等方面拥有丰富的多样性。其次,该工作在Unity3D的多个室内外场景下定制了多摄像机环境、光照、视角、视距和行走路径等,并导入大量三维人物同时进行行走穿插和多摄像机录像。这在一定程度上模拟了真实监控,使得产生的数据在背景、光照、分辨率、视角、姿态、遮挡等方面也拥有丰富的多样性,从而有力地支持了训练出可泛化的行人再识别模型。由此,作者最终得到一个包含全自动标注的虚拟行人数据集,称为RandPerson,它包含8,000个虚拟人物,11个场景,19个摄像机,38段密集行人的视频,1,801,816张切好的行人图片,示例如图6(4)所示。通过训练行人再识别模型并直接跨库测试泛化性能,该文作者首次证明了虚拟数据上训练的行人再识别模型可以更好地泛化到未知场景的真实图像上:对比之下RandPerson一举超越了CUHK03, Market-1501, DukeMTMC-reID和几乎MSMT17在内的四大主流行人再识别数据集。

然而,RandPerson的随机纹理图跟真实衣服纹理还是有比较大的差别,且场景规模不够大,背景不够丰富。在此启发下,Zhang等人(2021)进一步构建了一个包含3,000个行人(实际6,799人)和120,000个实例的包含真实衣服纹理的虚拟数据集UnrealPerson。根据RandPerson修改UV纹理图的方法,该文作者首先采用clothing-co-parsing和DeepFashion数据集生成来自真实图像的人体服装纹理,从而也通过Makehuman直接替代UV纹理图生成大量三维人体。其次,作者在虚幻引擎4上模拟场景,采用UnrealCV为虚拟相机捕获的每张图像收集像素级实例分割标注,最后生成UnrealPerson数据集,

如图6(5)所示。实验表明,当直接跨库到MSMT17测试时,该方法达到了38.5%的rank-1精度。

虽然UnrealPerson和RandPerson这两个数据集比以前的数据集更加灵活多样,取得了更好的模型训练效果,但是,RandPerson的随机颜色和纹理组合生成虚拟人物的策略使其人物偏卡通化;而尽管UnrealPerson使用了真实衣服纹理来生成人物,由于衣服纹理的尺寸问题,生成的虚拟人物和真实人物依然有较大的差异,且未考虑上下半身的协调穿搭。

鉴于此,Wang和Liang等人(2022)进一步提出一种自动从单角度人物照片中克隆整套衣服穿搭并产生三维人物模型的方案,并在游戏环境里模拟真实监控进行渲染,由此得到一个更真实的大规模虚拟行人数据集,并最终提升行人再识别模型的泛化表现。不同于已有的通过视频或多视角照片重建的方法,作者立足于克隆单张照片中的整套衣服生成虚拟人物。这样可以使用互联网上大量人物照片进行服装克隆,由此具有很低的限制条件达到规模化。另外,由于该文提出的方法将照片中人物的整套衣服克隆到虚拟人物身上,有效解决了现有虚拟数据库服装搭配与现实生活不匹配的问题。

具体地,该文设计了衣服配准映射和均匀布料扩展两种方法来克隆衣服纹理。衣服配准映射主要针对目标UV纹理图衣服结构清晰的人物模型,根据正面人物照片和模型纹理图中衣服对应关键点的位置进行投影变换,从而保留衣服纹理的清晰度和纹理结构。均匀布料扩展则主要针对人物背面(在正面照片中背面纹理不可见)和不规则的UV纹理图,通过找到并扩展衣服区域面积最大的同质纹理实现自动填充人物背面纹理和不规则的UV纹理图。除此之外,该文提出一个相似性-多样性人物扩展策略,通过聚类的方法使生成的虚拟人物既具有相似性,又具有多样性,从而形成最大效用的大规模虚拟行人数据集。该方法通过克隆同一聚类的人物照片生成相似人物,同时通过克隆不同聚类的人物照片生成多样性的人物。最后,作者通过在Unity3D的场景下进行渲染得到一个包含自动标注的虚拟行人数据集,称为ClonedPerson,如图6(6)所示。它包含5621个虚拟人物,6个场景,24个摄像机,168段密集行人的视频,887,766张切好的行人图片。该论文通过实验证明了ClonedPerson虚拟数据集可以应用在多个任务中,如行人再识别及其跨域泛化、无监督域适应、无监督学习和人体关键点检测。其中,跨库泛化性测试显著超越了三大主流行人再识别数据

集(CUHK03, Market-1501和MSMT17)和当前流行的虚拟行人再识别数据集(RandPerson和UnrealPerson),例如,在MSMT17上Rank-1达到了51.6%。

除上述工作外,Li等人(2021)也探索了利用已有游戏引擎产生ReID虚拟数据的方法,主要解决的是天气环境的多样性。该方法建立在热门游戏GTA5(Grand Theft Auto V)游戏引擎的基础上,其中包含了数字城市,天气和照明模拟系统,以及各种角色模型与多方面的装扮。该文提出了一个Weather Person的生成流程,它可以自动生成具有不同天气、场景和自然光照条件的合成Re-ID数据集。WePerson数据集中有400万张1500个行人的图片,同时包含比已有工作更多的环境因素,如图6(7)所示。实验结果表明,在具有挑战性的MSMT17数据集上,WePerson将Rank-1精确度从38.5%提高到了46.4%。然而,虽然因为GTA5是个商业引擎从而其人物质量非常高,但因为它是个闭源系统,无法导入自定义的人物,因此其人物数量是固定的,无法实现规模化构建虚拟数据。

另一方面,虚拟数据毕竟跟真实数据有较大的差别。因此,在利用虚拟数据方面,Wang和Liao等(2021)提出了一个更具有实际应用价值的行人再识别任务“A+B→C”:即如何利用大规模有标签的合成数据集A和无标签的真实数据集B训练出能泛化到未知场景C的模型。该任务不再依赖于对真实数据的手工标注,因此可以扩展到更大规模、更多样化的真实数据上,从而提高模型的泛化能力。在实现“开箱即用”的行人再识别方法中,该任务是更具潜力且成本低廉的方案。针对该任务,作者提出了一个新的方法,即DomainMix。所提出的方法可以很好地从有标签的虚拟数据和无标签的真实数据中学习。实验表明,这种完全不需要人工标注的方法可以在泛化能力上跟需要手工标注的方法相媲美。具体而言,该方法首先对未标记的真实世界图像进行聚类,并选择可靠的聚类。在训练过程中,针对两个域之间存在较大域差的问题,提出了一种域不变特征学习方法,该方法引入了一种新的损失,即域平衡损失,在域不变特征学习和域判别之间进行对抗性学习,同时学习一个判别特征用于行人再识别模型。此外,针对现有虚拟行人图像数据生成方法未考虑目标域数据特点、跨域性能有待提升的问题,蔡益武等(Cai和Zhang等,2022)提出了一种面向跨域行人再识别的虚拟数据生成与选择算

法。首先利用目标域前景信息如行人着装颜色分布

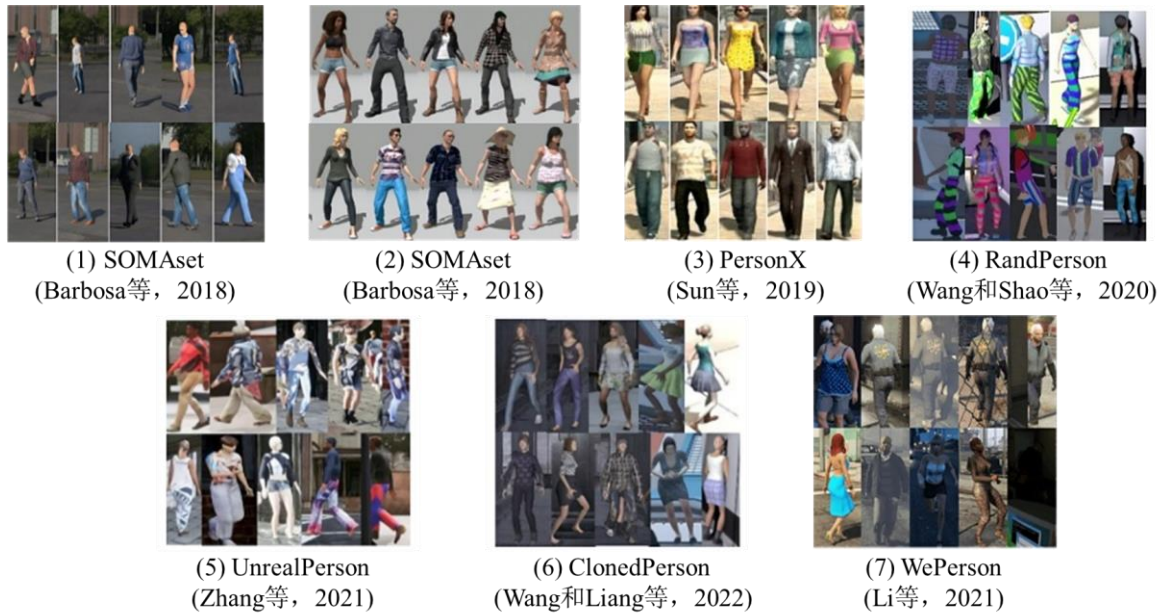


图 6. 已有虚拟数据库示例

Fig. 6 Illustration of Existing Virtual Datasets

指导虚拟人体模型生成以及部分替换目标域背景信息达到在像素级上提高数据质量, 然后根据 Wasserstein Distance 做分布度量, 在特征级上挑选与目标域最接近的源域训练子集用以模型训练。实验结果表明, 该方法可以提升行人再识别模型的跨域泛化性能。

然而, 以上工作都是将虚拟数据作为大规模和多样化的训练数据源来提升行人再识别的泛化能力。为了完全摆脱隐私敏感问题的影响, 采用大规模虚拟数据集作为测试集来对行人再识别算法进行基准测试也同样重要和有价值。这便引出另一个关键问题: 虚拟数据集是否能可靠地用于行人再识别算法基准测试? 换句话说, 在虚拟数据上测试得到的结果, 其算法排序是否跟在真实数据库上测试得到的结果一致? 为了解决这个问题, Kang (2022) 设计了一种称为 PRA (Pairwise Ranking Analysis) 的方法来定量测量排序相似性并做了是否独立同分布的统计检验。具体地, 作者采用 Kendall 秩相关系数来评估不同数据集上的算法排名之间的相似程度。然后, 作者采用了一个非参数二样本检验方法 Kolmogorov-Smirnov (KS) 用于判断虚拟数据集和真实数据集之间的算法排名相关性跟那些仅在真实数据集之间的算法排名相关性是否是独立同分布的。作者实验了 10 个代表性算法, 3 个主流的真实世界行人再识别数据集, 以及 3 个最近发布的大规模虚拟数据集, 通过 PRA 方法得出的结论是, 最

近的一个大规模虚拟数据集 ClonedPerson 可以可靠地用于可泛化行人再识别的基准测试, 其上得到的结果跟用真实数据集测试在统计意义上是完全一致的。因此, 该研究保证了虚拟数据集 (至少是 ClonedPerson) 既可以用于作为训练集, 也可以作为测试集, 因此行人再识别的研究完全可以摆脱真实世界的监控数据, 从而完全没有隐私问题。此外, 该研究也可能会启发未来虚拟数据集的设计。

2.4 域泛化行人再识别

行人再识别随着深度学习的发展取得了很大的进步, 但已有模型在不同场景下的泛化能力依然较差。这一方面是因为该任务本身的复杂性, 涉及很多的变化因素和困难。另一方面, 当前也缺乏大量的有标签训练数据。很多研究人员很早就意识到了跨场景这个问题, 为此提出了迁移学习 (或无监督域适应) 作为解决方案, 用于增强模型在新场景下的适应性, 目前已经发展成为行人再识别的一个主流研究方向。但是这种应用模式需要在每个应用场景都采集数据进行深度学习训练, 虽然无需人工标注数据, 但仍然是费时费力的。

针对上述问题, 受人脸识别应用实践的启发, Yi 等人 (2014) 最先考察了深度学习行人再识别模型的泛化性问题, 并发现在一定的训练数据条件下, 通过深度度量学习能训练出泛化性更好的模型, 超越了迁移学习的结果。随后, Hu 等人 (2014) 专门研究了行人再识别的跨库测试和泛化问题, 并发现

即使训练数据集人数不多，但只要摄像机数量足够多，即场景和光照环境足够丰富，仍然能训练出泛化性更好的行人再识别模型。

然而，上述研究取得的跨库泛化结果仍然是比较低的。为此，面向实际应用迫切需要进一步研究开箱即用（ready-to-use）的行人再识别——即域泛化行人再识别。本部分将根据域泛化行人再识别算法思路的不同，分别从批归一化和实例归一化方法、域不变特征学习方法、显式深度图像匹配方法、多专家混合方法和元学习方法等五方面，全面阐述域泛化行人再识别的前沿进展与发展趋势。

2.4.1 批归一化和实例归一化方法

批归一化（Batch Normalization, BN）和实例归一化（Instance Normalization, IN）及其组合是研究人员最先关注的影响和改进深度模型泛化能力的模块。Jia等人（2019）提出了一个基于归一化的域泛化（DG）行人再识别方法，通过批归一化和实例归一化的结合，减轻了域间风格和内容偏差，从而促进了行人再识别深度模型的泛化和可移植性。

Jin等人（2020）提出了样式归一化和补偿（style-normalization and restitution, SNR）模块，利用实例归一化的优点，过滤掉来自样式变化的干扰，并恢复被实例归一化丢弃的身份相关特征。此外，论文进一步设计了双重因果损失约束，以实现身份相关和不相关特征的解耦。

Zhou等人（2021）提出了一种轻量级的CNN(convolutional neural networks)体系结构OSNet(omni-scale network)，用以学习行人再识别的全尺度特征表示。与现有的行人再识别CNN网络相比，OSNet可以明确学习每个构件内部的多尺度特征，并通过统一聚合门动态融合多尺度特征，生成全尺度特征。为了改进跨域泛化，论文进一步通过可微架构搜索为OSNet配备了实例归一化，设计了一个称为OSNet-AIN(omni-scale network instance normalisation)的域自适应变体。

Choi等人（2021）提出了一个称为元批量实例正则化（Meta batch-instance, MetaBIN）的Re-ID框架，通过在元学习流程中预先模拟不成功的泛化场景来泛化归一化层，将可学习的批处理实例归一化层与元学习相结合，并通过元序列丢失和循环内部更新方式使虚拟仿真多样化以提高泛化能力。

Jiao等人（2022）提出了动态转换实例归一化（Dynamically Transformed Instance Normalization, DTIN）方案，以缓解实例归一化的缺陷。该方法使

用动态卷积来允许非归一化特征控制归一化特征到新表示的转换，从而确保网络具有足够的灵活性，并在消除不相关特定域特征和适应单个域或实例之间取得适当的平衡。此外，作者进一步利用多任务学习策略来训练模型，自适应地生成任意域的鉴别特征表示。

Liu和Huang等人（2022）提出了一种通过高斯过程（gaussian debiased batch normalization, GDNorm）的去偏批量归一化方法，将来自特定领域的BN层的特征统计估计（参数）建模为动态自精炼的高斯过程，以适应域的真实超分布，从而提高对未知目标域的泛化能力。它采用自优化机制进一步优化源域的域特定BN层的学习参数，并通过它们估计更准确的高斯过程作为回报，从而趋向于真实的数据分布。此外，GDNorm直接利用高斯过程的平均路径作为未见目标域的最佳BN估计，与集成学习相比，不需要额外的计算成本。

针对不同数据集图像明暗变化风格对模型泛化能力的影响，于明等人（2022）设计了一种由一个实例归一化层（即IN层）和一个修正线性单元（即Rectified Linear Units, ReLU）层构成的风格归一化模块（style normalization module）。具体的，首先由即IN层计算归一化统计量。然后，通过ReLU函数来增加特征的非线性。区别于跨样本在单通道上计算归一化统计量的BN层和在单样本上跨通道计算归一化统计量的LN层，风格归一化模块中的IN层通过在单通道单样本数据上计算归一化统计量，因此可以避免通道数和批大小的影响。此外，论文还通过叠加卷积层的瓶颈层设计，实现了对包含不同视野域的多尺度特征的提取，并利用特征融合注意力模块，实现了对多尺度特征的加权动态融合，从而提升了域泛化行人再识别性能。

2.4.2 域不变特征学习方法

域泛化最直观的一个目标是学习所有源域之间的域不变特征。除了归一化特征外，研究人员还采用各种方式进行域不变特征学习。例如，Lin和Li等人（2020）将人员再识别定义为一个多数据集领域泛化问题，提出了一种多数据集特征泛化网络（multi-dataset feature generalization network, MMFA-AAE）。该网络基于对抗性自动编码器，以学习具有最大平均差异（Maximum Mean Discrepancy, MMD）度量的广义域不变潜在特征表示。它能够从多个标记数据集学习通用的域不变特征表示，并将其推广到“看不见”的摄像机系统。

Chen 和 Dai 等人 (2021) 提出了一个端到端双分布对齐网络 (Dual distribution alignment network, DDAN) 来学习域不变特征。它具有两个约束条件: 域智能对抗特征学习和身份智能相似增强。在域层面, 该方法将外围域对准中心域, 以最小的分布偏移减少域差异。在身份级别, 该方法通过跨域的 ID 池捕获身份方面的相似性, 从而减少了域偏移。它实现了一个理想的场景, 即任何一组视觉上相似的 ID (尽管来自不同的域) 都比来自同一域的不相似的 ID 更接近。

Ang 等人 (2021) 提出了域嵌入扩展 (Domain Embedding Expansion, DEX) 模块。DEX 在训练过程中基于人员和领域标签动态操作和增强深度特征, 显著提高了行人再识别模型对不可见领域的泛化能力和鲁棒性。作者还开发了 DEX 的轻量级版本 DEXLite, 应用负样本采样技术扩展到更大的数据集, 并减少多分支网络的内存使用。作者提出的 DEX 和 DEXLite 可以以即插即用的方式与许多现有的方法相结合 (如 BagofTricks、多粒度网络和基于部分的卷积基线)。

杨晓峰等人 (2021) 提出一种基于深度胶囊网络的跨域行人再识别方法, 通过视角分类训练任务, 模型可以学习图像中行人的有效特征, 这些特征可以直接迁移到行人再识别任务中, 缓解了行人再识别泛化能力不足的问题。

沈贤文 (2021) 提出了一种基于前景特征学习的行人再识别方法 FFL, 利用人体解析算法生成的掩模图像对行人区域和背景区域加以区分, 使用掩模图像可以使网络在学习过程中更加侧重于学习行人本身的特征, 从而减少域之间行人图像背景差异带来的影响, 进一步提高了模型的跨域泛化能力。

2.4.3 显式深度图像匹配方法

已有基于深度学习的行人再识别方法主要关注单张图的特征学习, 但其训练好的模型是固定的, 因此在面对未见过场景时缺乏自适应性。对于图像分类来说, 它只涉及单张输入图片, 因此分类可以看作是输入图像跟学到的模型之间的一个匹配, 而模型仅能判断已知的类别。但是图像匹配任务不一样, 它涉及一对输入图像, 而且通常是开放类别的问题, 即是说, 测试的图像通常是新的未知类别, 如人脸识别或行人再识别。然而, 大部分已有方法并不直接考虑待匹配的两张图像之间的关系, 它们依然把图像匹配当作图像分类来做, 把每张图像独立对待, 只不过抛弃了分类器, 应用学到的模型从

中提取一个固定的特征表达。然后, 图像匹配就变成对两个特征向量简单地应用传统的距离度量, 如欧式距离或余弦距离, 而忽略两张图像的实际内容之间的直接关系。然而, 一旦在一个数据集上训练完成, CNN 的卷积核就是固定的, 它只能代表对训练数据固化的记忆。如果应用场景跟训练场景不一样, CNN 学到的固定卷积核就有可能对没见过的图像模式缺乏响应, 从而导致泛化性不够好。

有鉴于此, 在特征学习之外, Liao 等人 (2020) 重点研究了在两张待匹配图像的深度特征图上直接做点对点图像匹配, 以提升行人再识别泛化性能。具体地, 作者提出一种查询图自适应的卷积 (Query-Adaptive Convolution, QAConv), 通过查询图的深度特征图实时生成新的卷积核, 并在注册图的深度特征图上做卷积和全局最大池化实现图像对应点匹配。这样, 这种匹配过程和结果是可解释的, 而且这种显式匹配也比特征学习更容易泛化到未知场景, 如未知的错位、姿态或视角变化等。为了有效地端到端训练 QAConv, 作者进一步提出一个类别记忆模块用来缓存每个类别新近样本的深度特征图, 从而计算图像匹配的损失函数进行度量学习。

然而, 在大规模数据上学习深度图像匹配的效率尚未得到充分研究。虽然使用分类参数或类别记忆模块进行学习是一种流行的方式, 但它会产生大量的内存和计算成本。相比之下, 小批量 (mini batch) 内的成对深度度量学习将是更好的选择。然而, 主流的随机采样方法, 如 PK 采样器, 对于深度度量学习来说样本过于平凡因此学习效率不高。虽然在线难例挖掘在一定程度上提高了学习效率, 但在随机采样后的小批量内做难例挖掘仍然是有限的。为此, Liao 等人 (2022) 进一步提出了一种高效的小批量采样方法, 称为图采样 (Graph Sampling, GS), 用于大规模深度度量学习。其基本思想是在每个 epoch 开始时为所有类别构建最近邻关系图。然后, 每个小批量由随机选择的锚类及其最近的相邻类组成, 以便为学习提供信息丰富且具有挑战性的示例。由此, GS 可以仅在小批量内部做成对匹配的度量学习, 而不再需要低效的类别记忆模块。实验表明该方法大幅提升了可泛化行人再识别的性能, 同时, 在具有 8,000 个类别的 RandPerson 数据集上训练时, 训练时间显著减少, 从 25.4 小时减少到 2 小时。

GS 方法的提出赋能了更多更复杂的深度图像匹配方法。例如, Liao 等人 (2021) 提出了一种新的基于 Transformer 的深度图像匹配方法, 称为

TransMatcher, 用于可泛化的行人再识别。具体地, 该方法简化了Transformer的解码器, 移除了基于softmax加权的原始的Transformer注意力模块, 而只保留了其中的查询-键值相似度计算。此外, 作者进一步采用了全局最大池化和多层感知器(Multilayer Perceptron, MLP)用于解码匹配结果。这样, 简化的解码器在计算上更高效, 同时对图像匹配更有效。

2.4.4 多专家混合方法

训练多专家系统(模型)并以一定的方式进行融合也是提高模型泛化能力的一种方法。例如, Dai等人(2021)提出了关联感知专家混合(relevance-aware mixture of experts, RaMoE)方法, 使用一种基于投票的混合机制来动态地利用源域的多样性特征来提高模型的泛化性能。具体而言, 作者提出了一种去相关损失, 以使源域网络(专家)保持各个域特征的多样性和可辨别性。其次, 作者设计了一个投票网络, 将所有专家的特征自适应地集成到具有领域相关性的更通用的聚集特征中。考虑到目标域在训练过程中的不可见性, 作者进一步提出了一种新的学习-学习算法, 结合关系对齐损失来更新投票网络。

Xu和Liang等人(2022)提出了一种自适应聚合的模拟嵌入方法(Mimic Embedding via Adaptive Aggregation, META)用于可泛化的行人再识别。为了避免模型规模过大, META专家不为每个源域采用分支网络, 而是共享除批量归一化层之外的所有参数。除了多个专家之外, META还利用实例归一化, 并将其引入到一个全局分支中, 以跨域追求不变的特性。同时, META通过归一化统计来考虑未知目标样本和源域的相关性, 并开发了一个聚合模块, 以自适应地集成多个专家来模拟未知目标域。

余世杰(2022)提出了一种多领域专家协同学习算法。该算法中每个源域拥有一位独有的领域专家以及共享同一位全面专家, 并通过领域专家间协同学习以及全面-领域专家协同学习两种方式使得领域专家和全面专家相互学习促进。

2.4.5 元学习方法

另一类方法是采用元学习进行模型学习, 关键思想是在训练期间模拟源-目标域转移, 并引入元知识提高模型泛化能力。例如, Song等人(2019)提出了一种新的域不变映射网络(domain-invariant mapping network, DIMN)。DIMN设计用于学习人物图像与其身份分类器之间的映射, 即使用单个样本生成分类器。为了使模型域保持不变, 作者遵循

元学习流程, 并在每个训练集期间对源域训练任务的子集进行采样。

Bai等人(2021)提出了一个二元泛化网络(dual-meta generalization network, DMG-Net), 以利用元学习在训练过程和度量空间学习中的优点。具体而言, 作者设计了一个“先学习后概括评估”的元训练过程和一个元判别损失, 以增强模型的概括和判别能力。

Zhao和Zhong等人(2021)提出了一个基于记忆的多源元学习(Memory-based Multi-Source Meta-Learning, M3L)框架, 用于多源训练的可泛化行人再识别。所提出的元学习策略使模型能够在训练过程中模拟DG的训练测试过程, 从而有效地提高了模型在未知域上的泛化能力。此外, 还引入了一个基于内存的模块和MetaBN, 以充分利用元学习的优势并获得进一步的改进。

然而, Zhao和Zhao等人(2022)认为元学习的重复小段训练会导致过学习问题。为此, 作者提出了两种随机行为, 即从数据分布的角度来看随机分裂滑动采样器和从优化过程的角度来看方差变化的梯度丢失。所提出的采样器在两个相邻的训练集中保持各种源优势, 并且所提出的梯度下降帮助模型走出局部最小值并通过不断改变其梯度方差来搜索最优解。所提出的方法可以提高在相对平衡或极度不平衡的源域上的泛化能力。

2.5 换装行人再识别

在现实监控场景中, 在长期监控中或逃犯通过伪装躲避追捕等情况下, 行人目标可能换装, 这就打破了现有面向短期监控场景的行人再识别方法对于行人衣服不变的假设。与行人不更换衣服的再识别相比, 换衣场景下的行人衣服颜色和纹理失去身份判别性, 换衣再识别对细粒度特征提取的难度更大, 因而现有方法的精度仍不理想。本部分将在换装行人再识别数据集简介的基础上, 全面阐述近几年换装行人再识别的前沿进展与发展趋势。

2.5.1 换装行人再识别数据集简介

自2019年起, 为促进换装行人再识别的发展, 研究者们针对换装场景采集了多个公开数据集。现有的主流标准数据集是从短期监控中采集的, 并不适用于换装行人再识别的研究。代表性的换装行人数据集如表3所示。PRCC(Yang和Wu等, 2021)数据集在真实的室内监控场景中采集, 包含3个摄像头下221个行人的33698张图像, 其中换装后的衣服厚度没有太大变化。LTCC(Qian等, 2020)数据集

也是在室内场景中采集, 包含12个摄像头下152个行人的17138张图像, 其中光照、姿态和遮挡的变化比PRCC的数据更多。Celeb-reID (Huang等, 2019) 是一个由谷歌、必应和百度网站上获得的名人街拍, 包含1052个行人的34036张图像。VC-Clothes (Wan等, 2020) 是用游戏引擎渲染的虚拟数据集, 从4角度渲染得到512个行人的19060张图像。2021年至今, 公开换装数据集的数据规模进一步增大。DeepChange (Xu等, 2021a) 是目前规模最大的真实监控场景换装行人数据集, 包含17个摄像头下1121个行人的178407张图像, 比较接近实际应用场景。LaST (Shu等, 2021) 数据集从电影和电视剧中不同角色的截图中捕获, 包含10862个人物超过228000张图像。CCVID (Gu等, 2022) 是视频换装行人数据集, 包含226个行人2856段序列的数据。

2.5.2 换装行人再识别研究进展

在换装行人数据集的支持下, 研究的换装行人再识别方法主要有两类: 第一类方法通过显式引入特定的衣物无关线索来提取细粒度特征, 如利用行人的三维信息或者行人的轮廓提取身体形状特征; 第二类方法通过解耦行人图像的衣服信息和身份信息提取换衣不变特征, 如设计对抗学习损失分离衣服相关特征与衣服无关特征。

1) 显式引入衣服无关线索的特征提取方法

与一般场景中的跨摄像头视角变化和光照变化相比, 服装变化是难以直接建模的。因此, 这类方法通过引入不受服装变化影响的行人外观线索, 来引导服装无关特征的提取。Yang等人 (2021) 对行人的轮廓图使用空间极坐标变换, 然后对极坐标下的轮廓图使用可学习的采样方式提取轮廓图中最具判别性的部位, 以此学习与人体形状有关的特征。3DSL(3D Shape Learning) (陈坤峰等, 2022)

在缺乏3D真实数据的情况下, 通过辅助的三维重建任务, 从单张2D图像重建了人体的3D网格形态, 将3D形状信息与视觉纹理解耦, 获得容易受服装等纹理影响辨别的人的可靠且易区分的3D形状特征, 从而根据人体的3D形状判定行人身份。FSAM(Fine-Grained Shape-Appearance Mutual) (Hong等, 2021) 提出了一个细粒度的形状-外观互学习框架, 分为形状流与外观流。形状流中, 在身份的指导下学习细粒度的鉴别掩模, 并通过一个明确姿态的多分支网络提取细粒度的体型特征。为了补充外观流中与衣服无关的体型知识, 通过在低级和高级特征之间执行密集的交互式互学习, 将形状流中的体型知识转移到外观流中, 以补充外观特征中衣物无关的知识。推理时, 外观流能够独立使用。Qian等人 (2020) 则使用人体的关键点蒸馏出人体形状特征, 以此减少颜色信息的影响。Wan等人 (2020) 则在检测出人脸后提取人脸特征, 因为人脸同样是在衣服改变情形下仍然不变的信息。Yu等人 (2020) 使用自注意力机制, 鼓励网络去聚焦于行人的身体形状和脸部等这些与衣服无关的信息。CC-ReID(Clothes-Changing Person Re-identification) (Jin等, 2022) 利用步态信息来协助基于图像的换衣行人再识别表示学习。Li等人 (2022) 提出利用衣服模板提取额外特征与行人特征进行融合。Jia等人 (2022) 提出一种互补数据增强策略模拟行人换衣的。另外, 细粒度特征的提取方法也对衣服高度相似场景下的行人再识别有帮助。Yin等人 (2020) 提取两种彼此互补的局部动态姿态特征进行融合, 以此提取具有身份判别性的细粒度特征。陈和潘等人 (Chen和Pan等, 2021) 采取双目叠加效应的仿生思想, 提出一个自注意力孪生网络提取同一行人不同衣着的图像利用孪生架构实现叠加效应。

表 3 换装行人再识别数据集信息统计表

Table 3 Statistics of the Cloth-changing Re-ID Datasets

数据集	发布时间	人数	图像数	摄像机数	说明
Celeb-reID	2019	1052	34036	-	百度、谷歌等网站的名人街拍
LTCC	2020	152	17138	12	真实室内监控场景
VC-Clothes	2020	512	19060	4	游戏引擎渲染的虚拟数据集
DeepChange	2021	1121	178407	17	目前规模最大真实监控场景数据集
LaST	2021	10862	228000	-	电影/电视剧中不同角色截图
PRCC	2021	221	33698	3	真实室内监控场景
CCVID	2022	226	347833	-	视频换装行人数据集

显式引入衣服无关线索可有效利用人对于判

别换装行人的先验知识, 但在一些方法中也容易使

特征学习受限于人为指定的因素如身体形状等，而忽略其他细粒度特征的提取。

2) 针对衣服信息和身份信息特征解耦方法

特征解耦方法希望通过隐式学习的方式将衣物特征与身份信息独立表示，从而使得模型能够获得换衣不变的判别特征。下面把特征解耦方法分两类介绍。第一类特征解耦方式希望模型学习到行人身份相关的模式从而间接排除换衣的影响。ReIDCaps(Re-identification capsules) (Huang和Xu等, 2019) 利用了胶囊网络发掘特征之间关系且不同维向量蕴含不同模式的特性进行行人特征提取。进一步，刘乾等人 (Liu和Wang等, 2021) 在ReIDCaps的基础上引入软嵌入注意力机制和特征系数表示加强换衣不变特征的学习。RCSANet(Regularization via Clothing Status Awareness Network) (Huang和Wu等, 2021) 利用外观特征指导相同身份的特征聚类，使得不同身份的特征能在高维中分离。UCAD(Universal Clothing Attribute Disentanglement) (Yan和Yu等, 2022) 构建了一个服装相关-服装过滤的两分支网络，通过把服装相关分支学习到的知识从服装过滤分支过滤，来学习服装无关的身份特征。CAL(Clothes-based Adversarial Loss)方法 (Gu等, 2022) 提出了一种基于服装的对抗损失函数，通过惩罚服装分类模型的判别能力，从原始RGB图像中挖掘与服装无关的特征。CAL在PRCC、LTCC、LaST和DeepChange上达到当前的领先性能，rank-1准确率分别为55.2、40.1、73.7和54.0。

第二类特征解耦方式希望模型能将特征中的衣服相关信息去除。CASE-Net(Clothing Agnostic Shape Extraction Network) (Li 和Weng等, 2021) 提取灰度图和RGB图相似的特征以排除颜色的干扰，并在模型中嵌入了一个循环生成对抗网络结构，通过不同姿势的相同身份行人RGB图像和原灰度图的特征重建原RGB图片，以达到特征解耦的目的。CASE-Net在Celeb-reID数据集达到当前领先的66.4的rank-1准确率。AFD-Net(Adversarial Feature Disentanglement Network) (Xu和Liu等, 2021) 进一步发展了CASE-Net的工作，不仅通过类内重建加强了特征关于行人身份的鲁棒性，还加入了类间对抗，重建不同身份行人互换衣物的图片，使得服装特征与身份特征的分离更加彻底。Yang和Kang等人 (2022) 提出一种采样独立的特征解耦方法SirNet(sampling independent robust feature representation network)，从随机选择的样本中学习

换衣无关特征。SirNet在VC-Clothes上达到当前的领先性能，rank-1精度为92.3。

2.6 跨模态行人再识别

跨模态行人再识别旨在探索可见光图像与其它异构数据间的行人检索问题。该任务可以有效弥补可见光行人再识别技术在实际复杂场景下的不足。本节针对可见光图像与红外图像、素描图像、文本描述、深度图等各类数据异构场景，总结近年跨模态再识别的数据集、研究进展和难点，同时讨论未来复杂多变场景下的多模态行人再识别任务的发展趋势和方向。

2.6.1 跨模态行人再识别数据集

跨模态行人再识别常用数据集按模态不同可分为可见光-红外行人再识别数据集、文本-图像行人再识别数据集、素描-照片行人再识别数据集和深度图-可见光行人再识别数据集等，常用跨模态行人再识别数据集的信息统计表如表4所示。

1) 可见光-红外行人再识别数据集

可见光-红外行人再识别主要有SYSU-MM01数据集和RegDB数据集。

SYSU-MM01数据集 (Wu和Zheng等, 2017)：SYSU-MM01是目前最大和最具挑战性的可见光-红外跨模式人像识别数据集。它由491个身份的29,033张可见光图像和15,712张红外图像组成，这些图像由4台可见光相机和2台红外相机从室内和室外采集。训练集包含395个身份的22258张可见光图像和11909张红外图像。测试集包含96个身份的图像，这些测试身份的3,803张红外图像被用来形成查询集。对应于单次拍摄或多次拍摄的设置，每个可见光相机下的每个身份的1或10张图像被随机选取，形成图库集。此外，还有两种测试模式：全搜索模式是用室内和室外图像进行评估，而室内搜索模式只用室内图像进行评估。

RegDB数据集 (Nguyen等, 2017)：RegDB是由双对齐的可见光和远红外相机收集的，包括412个身份的图像。每个身份有10张可见光图像和10张远红外图像。数据集被随机分为训练和测试两部分，206个身份的图像用于训练，其余206个身份用于测试。

2) 文本-图像行人再识别数据集

文本-图像行人再识别主要有CUHK-PEDES数据集、ICFG-PEDES数据集和RSTPReid数据集。

CUHK-PEDES数据集 (Li等和Xiao等, 2017)：CUHK-PEDES包含40,206张图片和80,412个文字

描述, 涉及13, 003个身份, 每张图片有两个说明。每张图片有两个标题, 对13, 003个身份的描述。每个文本描述中平均有23.5个字。该数据集被分成三个子集, 分别用于训练、验证和测试, 不存在同一人的ID重叠。训练集包括11, 003人, 34, 054张图片和68, 108句描述。验证集和测试集分别包含3, 078张和3, 074张图片, 它们都有1, 000人。所有的实验都是基于这个训练-测试分割进行的。

ICFG-PEDES数据集(Ding和Ding等, 2021): 该数据集包含了更多以身份为中心的、细粒度的文字描述。它总共包含了4102个身份的54, 522个行人图像。所有的图像都是从MSMT17数据库中收集的。每张图片有一个标题, 每个描述平均有37.2个字。该数据库共包含5, 554个独特的词。与原始MSMT17数据库的协议类似, ICFG-PEDES分为训练集和测试集: 前者包括3102人的34674个图像-文本对, 而后者包含其余1000人的19848个图像-文本对。

RSTPReid数据集(Zhu和Wang等, 2021): RSTPReid包含20505张来自15个摄像头的4101人的图像。每个人都有5张由不同相机拍摄的相应图像,

每张图像都有2个文本描述的注释。对于数据划分, 分别利用3701、200和200个身份进行训练、验证和测试。每个句子不短于23个字。剔除出现少于两次的单词后, 单词数为2204。

3) 素描-照片行人再识别数据集

素描-照片行人再识别主要有PKU-Sketch数据集(Pang和Wang等, 2018)。该数据集包含200个人, 每个人都有一张素描和两张照片。为了确保创建的数据集能够用于现实的监控系统, 每个人的照片都是在白天由两个交叉视角的摄像机拍摄的。原始图像(或视频帧)经过了手工裁剪, 以确保每张照片都包含一个特定的人。

4) 深度图-可见光行人再识别数据集

深度图-可见光行人再识别主要有BIWI数据集和RobotPKU数据集。

BIWI数据集(Munaro和Basso等, 2014): 包含78个实例, 有22, 038张RGB和深度模式的图像。随机选择32个实例进行训练。8人进行验证, 38人进行测试。RobotPKU数据集(Liu和Liang等, 2017): 该数据集由90个行人组成, 有16, 512张深度和RGB

表 4 跨模态行人再识别数据集信息统计表

Table 2 Statistics of the Cross-modal Re-ID Datasets

模态	数据集	时间	身份	图像	视角	评价指标
可见光+	SYSU-MM01	2017	491	303420	6	CMC&mAP&mINP
红外	RegDB	2017	412	8240	1	CMC&mAP&mINP
可见光+ 文本	CUHK-PEDES	2017	13003	40206	-	Top-k accuracy
	ICFG-PEDES	2021	4102	54522	-	Top-k accuracy
	RSTPReid	2021	4101	20505	15	Top-k accuracy
可见光+素描	PKU-Sketch	2018	200	600	5	CMC
可见光+ 深度图	BIWI	2014	78	22038	-	CMC&mAP
	RobotPKU	2017	90	16512	-	CMC&mAP

模式的图像。随机分为40人进行训练, 10人进行验证, 其余40人进行测试。

2.6.2 跨模态行人再识别研究进展

1) 可见光-红外行人再识别

在低光照条件(夜晚)下, 监控系统中的摄像设备通常自动启用红外模式, 拍摄行人的红外图像。因此, 有研究者提出可见光-红外行人再识别, 旨在实现白天可见光图像和夜间红外图像之间的交叉模态匹配。由于成像原理的不同, 两种模态间存在巨大模态差异。现有的可见光-红外行人再识别算法可以大致分为以下两类: 基于域不变表征学习的方法(Wu和Zheng等, 2017; Ye等, 2018; Ye等, 2020;

Zhu等, 2020; Liu等, 2020; 叶和刘等, 2021; 王凤随等, 2022)和基于图像生成的方法(Wang和Wang等, 2019; Wang和Zhang等2019; Choi等, 2020)。其中, 域不变特征学习方法又从特征表示学习(Wu和Zheng等, 2017; Ye等, 2018; Ye等, 2020)和度量学习(Zhu等, 2020; Liu等, 2020; Ye等, 2021)两个方面来挖掘跨模态共享表征。Wu和Zheng等人(2017)首次尝试解决这个问题, 提出了一个深度零填充框架, 以自适应地学习模态可共享表征。Ye等人(2020)设计了一种双流网络, 以对模态共享和模态特定信息进行建模, 同时解决模态内和模态间的变化。Zhu等人(2020)设计一种异构中心损失

用来约束可见光模态中心和红外模态中心间的距离,从而缩小类内差异。文献(Ye和Liu等,2021)中提出一种双向指数角度三元组损失,该损失函数通过学习角度可分离的共享表征空间来减轻模态差异的影响。文献(Ye和Shen等,2020)中提出了一种双注意聚合学习方法,以捕捉多级关系。Chen和Ye等人(2022b)提出利用不同模态下的行人结构信息来指导跨模态共享表征的学习。最近,文献(Zhang和Lu等,2018)首次提出利用属性信息来弥补不同模态间的信息差异,设计一种粒度语义信息与全局信息融合网络,通过建模模态属性间关系,增强模态特征判别力。此外,基于图像生成的方法又可以具体分为模态转换(Wang和Wang等,2019; Wang和Zhang等2019; Choi等,2020)和中间模态生成(Ye和Ruan等,2021)。文献(Wang和Wang等,2019; Wang和Zhang等2019)采用生成对抗技术生成跨模态人物图像,以减少图像和特征级别的跨模态差异。文献(Choi,2020)中对分层交叉模态解纠缠因子进行了建模。文献(吴岸聪等,2022)提出利用易获取的可见光图像进行单模态自监督学习,从而学习先验知识指导跨模态表征提取。

2) 文本-图像行人再识别

在真实场景中,有时会出现无法获得待查询目标行人的视觉图像情况,工作人员通常利用目击者的文本描述来实现目标搜索,即文本-图像行人再识别。现有方法主要可以划分为非注意力的跨模态方法(Zheng和Zheng,2020; Zhang和Lu等,2018; 莫承见,2022)和基于注意力的跨模态方法(Li和Xiao等,2017; Chen和Li等,2018; Liu,Zha和Hong等,2019)。第一类方法主要通过设计网络结构和损失函数来实现两种模态在共享空间中的对齐,计算过程高效。相比之下,第二类方法则侧重于利用注意力方法挖掘区域块-单词、区域块-短语间的对应关系,实现了较高的检索性能。具体地,Li和Xiao等人(2017)提出利用带有递归神经网络的门控神经注意模型来学习文本描述和人物图像之间的共享特征,实现了文本到图像行人检索的端到端训练。Chen和Li等人(2018)提出了一种全局识别图像语言关联学习方法,在重建过程中捕获身份识别信息和局部重建图像语言关联。文献(Liu和Zha,2019)中利用图关系挖掘设计了一个深度对抗图注意力卷积网络。最近,Shao和Zhang等人(2022)首次分析了特征粒度上的模态差异,即相似但不同的图像区域可能共享相同的文本描述,并提出一种粒度统

一的表示学习算法。然而,文本描述和视觉图像之间的巨大语义差距仍使得该项任务颇具挑战性。文献(陈琳等,2021)提出利用不同模态的行人属性信息来减轻文本和图像模态间差异影响,提取语义共享的跨模态表征。姜定等(2022)基于语言-图像预训练双流Transformer模型实现仅全局特征的图像文本语义特征对齐,并设计一种温度缩放跨模态投影匹配损失函数进行图像文本跨模态特征匹配,其实验结果大幅度超过现有基于局部特征学习的方法。

3) 素描-照片行人再识别

考虑到文本描述的粗粒度性,有研究者提出利用行人的素描图像实现目标人物检索,即素描-照片行人再识别。有谚语说“一图胜千言”,相比于文本描述,素描图像更加直观,且包含丰富的行人细节信息。Pang和Wang等人(2018)提出了一种域相关对抗框架,该框架使用全身专业素描图像作为查询集,收集域敏感信息并学习域不变特征。Yang和Wu等人(2020)试图通过结合领域自适应来提高草图-照片模型的泛化能力。Gui和Zhu等人(2020)探索了草图和照片的多层次特征,并使用梯度反向层来弥补域差异。与上述方法不同,文献(Chen和Du等,2022a)设计了一种新的解纠缠原型和动态合成学习方法来处理跨模态差异。该方法在辅助素描模态的指导下,将照片表征分为素描相关线索和素描无关线索,并利用知识迁移技术补充素描模态中缺失信息,形成信息对称的交叉模态空间。黄和李等人(2022)考虑了真实场景中素描-照片匹配的跨模态身份不一致问题,并设计一种基于交叉分类的素描行人重识算法。为了充分利用文本描述和素描表述的优势,如何将文本与手绘素描图像相结合也是值得未来研究的问题。

4) 深度图-可见光行人再识别

深度图表示摄像设备到场景中各点距离,可以描述行人的身体形态和骨骼信息。与可见光图像相比,深度图对光照变化和行人衣着变化具有较强的鲁棒性。随着激光雷达技术的发展,微软Kinect摄像机拍摄的深度图像可以更方便地获取。近年来,有研究者提出深度图-可见光行人再识别,该任务旨在实现深度图和可见光图片间的跨模态匹配。由于深度图仅仅包含行人的结构信息,缺失行人颜色和纹理信息,该跨模态任务也极具挑战。文献(Zhang和Xu等,2019)利用传统手工模态特征来实现两种模态间的匹配。文献(Frank等,2019)借助蒸馏的

思想设计一种两阶段训练模式来减轻模态差异的影响, 提高跨模态识别的精度。不同于这些方法, Wu等人(Wu和Jiang等, 2022)提出一种全局和局部特征联合学习网络, 通过挖掘行人外观细节和空间关系, 增强深度网络中的特征表达能力。

2.7 行人搜索

行人搜索技术旨在将行人检测和再识别统一到同一框架中, 在确定行人位置的同时提供其身份特征, 以提升系统的性能和效率。近年来, 行人搜索方法不仅探索了新的网络结构, 而且更加注重开放世界的实际需求, 涌现出了基于文本的多模态行人搜索、仅使用位置标签的弱监督行人搜索、域自适应的行人搜索等一系列探索工作。本节将首先介绍行人搜索常用的数据集, 然后总结分析行人搜索方向的最新进展和发展趋势。

2.7.1 行人搜索常用数据集简介

Zheng和Zhang等人(2017)建立了首个专门面向行人搜索的数据集PRW, 并在此基础上评估了不同行人检测模型组合行人再识别模型的性能表现。PRW数据集从六个固定位置的摄像机中采集, 其训练集共有5,704张图片, 包括482个行人的18,048个实例。Xiao等人(2017)从街景照片和电影视频中抽取图片, 建立的CUHK-SYSU数据集拥有更大的数据规模、更加复杂的场景, 并涵盖了视角变化、光照变换、多分辨率、遮挡等复杂场景。CUHK-SYSU总计有18,184张图片, 包括8,432个不同身份行人的96,143个不同实例。Zheng和Qin等人(2021)从MovieNet数据集中采样制作了用于行人搜索的MovieNet-CS数据集, 包含了3,087个不同身份人物的160,816张图片和274,274个实例, 是现有的最大规模行人搜索数据集。换装行人再识别数据集信息统计表如表5所示。

2.7.2 行人搜索研究进展

从模型结构上来讲, 目前主流的行人搜索框架可以分为单阶段模型和两阶段模型两类。

两阶段模型将行人检测和再识别两个子任务

分开执行, 先进行目标检测, 再将检测结果裁剪, 并对剪裁的图像进行再识别。这一类框架的主流研究工作致力于解决检测结果和再识别模型所需要的区域不匹配、不对齐的问题。Lan等人(2018)分析了行人搜索任务中普遍存在的多尺度匹配问题, 从这一角度出发提出了跨层级的语义对齐方法。Han等人(2019)将再识别子任务作为主任务, 从再识别任务的需求出发, 设计了感兴趣区域(ROI)变换层对检测器提取的边界框进行二次优化, 以获得更可靠的边界框用于后续的再识别模型。Wang和Ma等人(2020)设计了一种任务连贯的两阶段行人搜索框架, 通过引入目标身份信息来过滤检测结果, 同时设计自适应的再识别模型使其匹配检测器输出的结果, 该方法在两个子任务的评价指标上都取得了较好的性能。

单阶段的行人搜索模型在一个端到端的框架中联合完成行人检测和再识别子任务, 因为其相比于两阶段模型更加简单, 且在效率上存在明显优势, 目前受到了越来越多研究者的关注。Xiao等人(2017)提出了首个基于Faster R-CNN(Ren等, 2015)的单阶段行人搜索框架, 通过在检测特征之后加入了一个并行的再识别特征层, 并使用在线实例匹配损失来约束识别任务, 使得检测和再识别任务能够在同一个框架中进行端到端地训练。目前的大部分单阶段行人搜索工作都在此框架的基础上进行改进。如Kim等人(2021)指出了行人搜索中普遍存在姿势等不对齐的情况, 并使用动态更新原型的方式学习更有辨识度的行人特征。Chen和Zhang等人(2020)考虑了联合框架中检测和再识别任务的关系和优化过程中的内在矛盾, 通过改进模型结构解耦二者使用的特征, 使得两个子任务能更好的配合。Li和Miao等人(2021)则考虑了两个任务间存在的顺序性依赖关系, 指出单阶段模型中低质量的检测中间结果可能会对再识别产生负面影响, 在此基础上设计了一种优化行人检测中间结果的单阶段模型。

表 5 换装行人再识别数据集信息统计表

Table 5 Statistics of the Cloth-changing Re-ID Datasets

数据集	发布时间	人数	图像数	实例数	说明
PRW	2017	482	5,704	18,048	从六个固定摄像机中采集
CUHK-SYSU	2017	8,432	18,184	96,143	从街景照片和电影视频抽取, 含视角/光照/分辨率变化、遮挡等复杂场景
MovieNet-CS	2021	3,087	160,816	274,274	从 MovieNet 数据集中采样, 是现有的最大规模行人搜索数据集

2.7.3 行人搜索的前沿动态

近年来, 主流的单阶段行人搜索方法在模型设计上还涌现出了一些基于 DETR(Detection Transformer) (Carion 等, 2020), FCOS(Fully convolutional one-stage) (Tian 等, 2019) 等新型检测器的探索性工作。Yan 等人 (2021) 提出了一种不需要锚点的单阶段行人搜索模型, 这种设计减小了 Faster-RCNN 中大量提取候选框带来的计算开销, 该方法在两个子任务上都取得了比此前方法更高性能, 且进一步提升了行人搜索模型的效率。Cao 等人 (2022) 近期提出了一种基于 DETR 的端到端行人搜索模型, 通过在 DETR 中引入一个再识别解码器, 并结合使用不同层级的主干网络和检测解码器的输出来提供多层次监督。该方法在不同的主干网络上都取得了目前最优的性能表现, 并在推理效率上领先于基于 Faster-RCNN 的单阶段行人搜索方法。Yu 和 Du 等人 (2022) 基于 Transformer 在行人再识别领域中的成功经验, 提出了使用三层级联的 Transformer 作为识别分支, 并通过特征混淆的方式来模拟行人搜索中的遮挡场景, 该方法有效地提高了识别分支特征的鲁棒性。

除了模型结构设计, 近年来还有一些研究者探索如何让行人搜索模型在开放环境下也能满足实际场景的需求。在实际使用中, 行人的位置和身份标签的获取难度是存在差异的, 获取行人位置框标注的难度要明显低于获取身份标签的难度。因此, 一些研究者探索了只依赖行人位置框标签的弱监督行人搜索任务。Yan 和 Li 等 (2022) 错误!未定义

书签。) 通过聚类方法为再识别子任务提供动态更新的身份伪标签, 并且充分利用了图像中的环境上下文、记忆上下文等信息来辅助行人身份特征学习, 该方法在性能上甚至超过了一些全监督的行人搜索方法。Han 等 (2021) 也按聚类产生身份伪标签这一思路, 通过设计一个孪生网络来对齐实例的特征表示, 以提高弱监督设定下的行人特征的鲁棒性。Jia 和 Luo 等 (2022) 同样利用了环境上下文来辅助身份特征学习和聚类, 并将聚类产生的大量长尾孤例作为训练负例以提升特征的鲁棒程度。

此外, 考虑到现有的行人搜索方法大多基于单一场景设定, 直接将现有模型迁移到新场景往往会导致性能不佳, 而在实际使用中针对每个场景都采集大量带有标注的训练数据成本高昂。因此, Li 和

Yan 等人 (2022) 提出了一种具有域自适应能力的行人搜索方法, 在目标场景上仅使用图像数据, 无需任何行人位置和身份的标注。该工作对齐了不同域的特征表示, 并随着训练过程动态地生成行人检测实例和对应的身份伪标签, 通过这些设计, 该工作在没有标注的目标域上取得了接近弱监督甚至部分全监督方法的性能。

还有一些研究者考虑到开放世界中使用非图像输入进行检索的可能, 提出了一种基于文本的多模态行人搜索设定。这一设定中, 推理时的输入不再是给定行人的图像, 而是关于该人的文字描述。Zhang 和 Long 等 (2021) 利用了文本信息来引导区域提议网络关注文本所描述的区域。Yan 和 Tang 等 (2022) 的工作中考虑到文字和图像的信息不对等问题, 提出模型抑制图像信息中的背景和环境信息, 并设计了一个局部对齐模块来对齐描述文字和行人图像。

3 总结与展望

行人再识别旨在基于行人的穿着、体态、发型、携带物等信息, 实现大时空范围内人脸等生物特征失效情况下的跨摄像机下目标行人图像的关联与匹配, 已成为智能视频监控系统的关键环节和支撑技术, 在智慧公安、智慧城市等国民经济建设中发挥了重要作用。经过十多年从多摄像机追踪到再识别的起步阶段、基于传统特征的再识别、特别是近年来基于深度学习的再识别的快速发展与进步, 再多个数据集上了非常好、甚至超过人类的性能。但同时, 面向开放域实际场景的应用, 仍然面临诸多挑战, 这也引领着未来行人再识别技术的发展趋势。下面将结合本文涉及的七个方面, 对未来一段时间的行人再识别技术发展趋势进行分析和展望。

1) 遮挡行人再识别。遮挡行人再识别问题中, 对于基于可见区域的这类方法, 需要依赖于额外的模型来提供行人的语义信息进而实现特征配准。如何保证额外模型在行人再识别数据上的可靠性, 以及减小其带来的计算开销是需要考虑的一个主要问题。其次, 目前的特征配准仅考虑到部件特征, 如何高效地实现更细粒度的特征配准 (如像素级) 是一个值得探索的方向。对于基于补全的遮挡行人再识别方法, 如何保证补充的区域信息具有原信息表征且其他区域不具备的鉴别性是主要的问题, 特别是在图像行人再识别中无法利用时序信息的场

合。除此之外, 现有的数据集中, 训练集中的遮挡行人图像严重不足, 这极大的限制了现有方法的性能, 探究合理的遮挡图像生成方式来丰富训练样本, 以及如何有效地利用生成样本进行训练, 提升遮挡行人再识别方法的泛化能力, 是未来的一个研究发展趋势。

2) 无监督行人再识别。近几年, 无监督行人再识别性能显著提升, 很多近期工作在学术界常用数据集上所取得的性能已经超越了3年前部分有监督行人再识别算法。与有监督行人再识别相比, 无监督行人再识别可以显著降低数据标注成本, 从而使行人再识别系统易于部署, 具备更好的环境适应能力。未来, 无监督行人再识别仍将是一个具有研究意义和应用价值的研究方向。现有的无监督行人再识别算法难以对易混淆样本进行准确的自动标注, 使其性能受到制约。未来, 通过引入主动学习将少量的人机交互引入无监督学习过程, 有望在不显著增加标注成本的同时, 显著提升无监督行人再识别性能。此外, 随着大模型研究的进展以及大数据的有效利用, 如何利用大模型知识指导无监督行人再识别模型的学习与训练将会成为已有价值的研究方向。

3) 虚拟数据行人再识别。近年来虚拟数据用于行人再识别也取得了显著的进步, 特别是用于训练深度模型并提升泛化性能。此外, 部分研究也已证实了虚拟数据可以像真实数据一样用于可靠地测试算法。因此, 行人再识别的研究将不会再受制于隐私保护的影响, 而有了可靠的数据保障。未来在行人再识别虚拟数据的研究主要在于如何进一步扩大规模和多样性, 并进一步提升行人再识别的泛化性能。虽然直觉上数据越多越好, 但在虚拟数据领域虽然扩大规模很容易, 但如何有效地扩展数据仍然是个瓶颈问题。例如, 虽然UnrealPerson总共创建了6,799个人, 但其实验发现随机选择3,000个人实验效果是最好的, 因此仍未能发挥更多人的优势。相应地, ClonedPerson提出一个相似性-多样性人物扩展策略, 通过聚类的方法使生成的虚拟人物既具有相似性, 又具有多样性, 从而形成最大效用的大规模虚拟行人数据集, 这对未来虚拟数据的设计有一定的借鉴意义。

4) 域泛化行人再识别。近年来研究人员从批归一化和实例归一化、域不变特征学习、显式深度图像匹配、多专家混合、元学习等多方面开展了域泛化行人再识别的研究, 并显著提升了行人再识别模

型的泛化性能。这些研究进一步明确了在实际应用条件下已有方法较差的跨库泛化性能以及提升泛化性能的难度。因此行人再识别的研究必须重视跨库测试的评估, 而不能乐观于单个数据库上训练测试得到的几近饱和的性能。由于域泛化行人再识别的研究仍然处于起步阶段, 因此以上各子方向仍然有比较大的潜力和发展空间。另一方面, 探索更多的泛化思路以及更深入理解深度网络的过拟合现象和泛化机理, 仍然有很长的道路要走。此外, 除了算法外, 数据本身的作用也是提升泛化能力不可或缺的, 而受监控数据隐私保护的影响, 大规模虚拟数据用于行人再识别的训练和测试将是大势所趋。

5) 换装行人再识别。换装行人再识别是现实监控场景中的难点问题, 近年来受到越来越多的关注。随着研究领域的发展, 公开换装行人数据集的数据规模越来越大, 场景更接近实际监控情况。对比短期不换装的再识别, 现有换装再识别方法在现实场景数据集中的性能仍不理想。研究领域目前仍存在问题, 有待未来研究的解决。一方面, 方法的发展仍受制于换装行人数据难以标注的限制, 未来可以考虑针对换装场景的虚拟数据合成利用与弱监督学习; 另一方面, 大多数研究忽略了现实应用中换装行人与非换装行人同时出现的情况, 未来可以考虑学习统一的特征提取方法。

6) 多模态行人再识别。由于现实场景的复杂性和多样性, 结合现有前沿进展, 跨模态行人再识别研究仍有如下几个研究方向值得研究者进一步深入研究: 1) 跨模态联邦学习。现有的跨模态数据集比较少且规模小, 由于隐私受限问题, 大量真实场景数据未被利用。如何借助联邦学习联合多方数据, 提升模型性能值得进一步研究。2) 多模态数据泛化问题。现有的跨模态模型通常只是针对两两模态间的匹配问题, 无法应用模态缺失问题。自适应多模态数据间的跨模态匹配, 是一个智能监控系统不可或缺的能力。

7) 行人搜索。目前行人搜索研究的根本目的在于提升行人检测和再识别两个子任务的性能和效率, 而在实际使用中, 各种复杂因素导致我们无法轻易完成这一目标。首先, 从两个子任务关系的角度, 检测任务需要学习行人共性特征, 再识别任务则关注不同行人的身份信息, 二者之间存在天然冲突。而识别任务又依赖于检测结果的质量, 因此如何处理好两个子任务间的平衡是一个具有挑战性

的问题。其次,从数据角度,由于获取有标签数据的成本高昂,行人搜索研究逐渐从全监督设定转向了弱监督、域自适应设定。如何在仅有部分甚至完全没有标签的情况下为模型提供更有效的监督是许多研究者关注的热点。除了获取训练数据的困难,模型本身的鲁棒性也是一个大问题,在跨场景或复杂背景(背景遮挡、行人重叠、行人衣着类似)的情况下现有模型往往会出现明显的性能下降,如何训练鲁棒、泛化能力强的行人搜索网络也是值得探讨的研究课题。总体而言,行人搜索正向利用更少的标签解决更加贴近实际问题这一道路上发展。

参考文献 (References)

- Ang E.P., Shan L., & Kot A. C. 2021. DEX: Domain Embedding Expansion for Generalized Person Re-identification. [EB/OL]. [202-10-22]. <https://arxiv.org/pdf/2110.11391v1.pdf>. [DOI: 10.48550/arXiv.2110.11391]
- Bai Y., Jiao J., Ce W., Liu J., Lou Y., Feng X., & Duan L.Y. 2021. Person30k: A dual-meta generalization network for person re-identification//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2123-2132. [DOI: 10.1109/CVPR46437.2021.00216]
- Bak S., Corvee E., Bremond F., et al. 2010. Person re-identification using haar-based and DCD-based signature//Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance. Boston, US.1-8. [DOI:10.1109/AVSS.2010.68]
- Bak S., Carr P., & Lalonde J.F. 2018. Domain adaptation through synthesis for unsupervised person re-identification//Proceedings of the European conference on computer vision, 189-205. [DOI:10.1007/978-3-030-01261-8_12]
- Barbosa I. B., Cristani M., Caputo B., Rognhaugen A., & Theoharis T. 2018. Looking beyond appearances: Synthetic training data for deep cnns in re-identification. Computer Vision and Image Understanding, 167, 50-62. [DOI: 10.1016/j.cviu.2017.12.002]
- Cai Q, Aggarwal J K.1996.Tracking human motion using multiple cameras//Proceedings of the International Conference on Pattern Recognition.Vienna, Austria, 68-72. [DOI: 10.1109/ICPR.1996.546796]
- Cai Y., Zhang Y. & Zhang Y. 2022. Generation and Selection of Virtual Data for Cross-Domain Person Re-Identification. Journal of Graphics. Accepted. (蔡益武, 张雨佳, 张永飞. 面向跨域行人再识别的虚拟数据生成与选择, 图学学报. 已录用.)
- Cao J, Pang Y, Anwer R.M, Cholakkal H, Xie J, Shah M and Khan F.S. 2022. PSTR: End-to-End One-Step Person Search with Transformers//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.New Orleans, LA, USA: IEEE: 9448-9457. [DOI: 10.1109/CVPR52688.2022.00924]
- Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A and Zagoruyko, S. 2020. End-to-end object detection with transformers//Proceedings of the European Conference on Computer Vision.Glasgow, UK: Springer: 213-229. [DOI: 10.1007/978-3-030-58452-8_13]
- Chen C, Ye M, Qi M, Du B, et al. 2022a. Sketch Transformer: Asymmetrical Disentanglement Learning from Dynamic Synthesis//Proceedings of the 30th ACM International Conference on Multimedia.4012-4020. [DOI: 10.1145/3503161.3547993]
- Chen C, Ye M, Qi M, et al. 2022b. Structure-Aware Positional Transformer for Visible-Infrared Person Re-Identification[J]. IEEE Transactions on Image Processing, 31: 2352-2364. [DOI: 10.1109/TIP.2022.3141868]
- Chen D, Li H, Liu X, et. al. 2018. Improving deep visual representation for person re-identification by global and local image-language association//Proceedings of the in European Conference on Computer Vision.pp.54-70. [DOI: 10.1007/978-3-030-01270-0_4]
- Chen D, Zhang S, Yang J and Schiele B. 2020. Norm-aware embedding for efficient person search// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Seattle, WA, USA: IEEE: 12615-12624. [DOI: 10.1109/CVPR42600.2020.01263]
- Chen H., Lagadec B., & Bremond F. 2020. Learning discriminative and generalizable representations by spatial-channel partition for person re-identification//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2483-2492. [DOI:10.1109/WACV45572.2020.9093541]
- Chen J, Jiang X, Wang F, Zhang J, Zheng F, Sun X and Zheng W S. 2022. Learning 3D Shape Feature for Texture-Insensitive Person Re-Identification//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Virtual: IEEE: 8146-8155 [DOI: 10.1109/CVPR46437.2021.00805]
- Chen K., Pan Z., Wang J., Shi L. and Zhang J. 2022. Moderate Clothes-Changing Person Re-identification Based on Bionics of Binocular Summation. Computer Science., 49(8): 165-171. (陈坤峰, 潘志松, 王家宝, 施蕾, 张锦. 2022. 基于双目叠加仿生的微换衣行人再识别. 计算机科学, 49(8): 165-171.) [DOI: 10.11896/jsjx.210600140]
- Chen L., Gao Z., Song X., Wang Y. and Nie L. 2022. A Cross-modal Pedestrian Re-Id Algorithm Based on Dual Attribute Information, Journal of Beijing University of Aeronautics and Astronautics.48(04): 647-656.(陈琳, 高赞, 宋雪萌, 王英龙, 聂礼强. 2022. 基于双重属性信息的跨模态行人再识别算法.北京航空航天大学学报, 48(04):647-656.) [DOI:10.13700/j.bh.1001-5965.2020.0614.]
- Chen P., Dai P., Liu J., Zheng F., Xu M., Tian Q., & Ji R. 2022. Dual distribution alignment network for generalizable person re-identification//Proceedings of the AAAI Conference on Artificial Intelligence, Vol.35, No.2, 1054-1062. [DOI: 10.1609/aaai.v35i2.16190]
- Chen P., Liu W., Dai P., Liu J., Ye Q., Xu M., ...& Ji R. 2022. Occlude them all: Occlusion-aware attention network for occluded person re-id//Proceedings of the IEEE/CVF International Conference on Computer Vision, 11833-11842. [DOI: 10.1109/iccv48922.2021.01162]
- Chen Y, Zhu X, Gong S. 2019. Instance-guided context rendering for

- cross-domain person re-identification//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).232-242. [DOI: 10.1109/ICCV.2019.00032]
- Cheng X., Jia M., Wang Q., & Zhang J. 2022. More is better: Multi-source Dynamic Parsing Attention for Occluded Person Re-identification//Proceedings of the 30th ACM International Conference on Multimedia, 6840-6849. [DOI: 10.1145/3503161.3547819]
- Cheng Z., Zhu X., & Gong S. 2018. Low-resolution face recognition//Proceedings of the Asian Conference on Computer Vision.605-621. [DOI: 10.1007/978-3-030-20893-6_38]
- Choi S, Lee S, Kim Y, Kim T, and Kim C. 2020. Hi-cmd: Hierarchical cross-modality disentanglement for visible-infrared person re-identification//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.10257–10266. [DOI: 10.1109/CVPR42600.2020.01027]
- Choi S., Kim T., Jeong M., Park H., & Kim C. 2022. Meta batch-instance normalization for generalizable person re-identification//Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition, 3425-3435. [DOI: 10.1109/CVPR46437.2021.00343]
- Dai Y., Li X., Liu J., Tong Z., & Duan L.Y. 2022. Generalizable person re-identification with relevance-aware mixture of experts//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 16145-16154. [DOI: 10.1109/CVPR46437.2021.01588]
- Dai Z, Wang G, Yuan W, et al. 2022. Cluster Contrast for Unsupervised Person Re-Identification [EB/OL]. [2022-10-22]. <https://arxiv.org/pdf/2103.11568.pdf>. [DOI: 10.48550/arXiv.2103.11568]
- Ding G, Khan S H, Tang Z, et al. 2019. Dispersion based Clustering for Unsupervised Person Re-identification//Proceedings of the British Machine Vision Conference (BMVC), 2, 7.1-12. [DOI: 10.1016/j.patrec.2020.04.039]
- Ding Z, Ding C, Shao Z, et al. 2022. Semantically self-aligned network for text-to-image part-aware person re-identification [EB/OL]. [2022-10-2]. <https://arxiv.org/pdf/2107.12666v2.pdf>. [DOI: 10.48550/arXiv.2107.12666]
- Dou Z., Wang Z., Li Y., & Wang S. 2022. Progressive-Granularity Retrieval Via Hierarchical Feature Alignment for Person Re-Identification//Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2714-2718. [DOI: 10.1109/ICASSP43922.2022.9747234]
- Ester M, Kriegl H P, Sander J, et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise//Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD), 226-231. [DOI: 10.5120/739-1038]
- Farenzena M, Bazzani L, Perina A, et al. 2010. Person re-identification by symmetry-driven accumulation of local features// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.2360-2367. [DOI:10.1109/CVPR.2010.5539926]
- Felzenszwalb P. F., Girshick R. B., McAllester D., & Ramanan D. 2010. Object Detection with Discriminatively Trained Part-based Models. IEEE transactions on pattern analysis and machine intelligence.32(9): 1627-1645. [DOI: 10.1109/TPAMI.2009.167]
- Fu Y, Wei Y, Wang G, et al. 2019. Self-similarity Grouping: A Simple Unsupervised Cross Domain Adaptation Approach for Person Re-identification//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 6112-6121. [DOI: 10.1109/ICCV.2019.00621]
- Gheissari N, Sebastian T, and Hartle R. 2006. Person reidentification using spatio temporal appearance//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2006:1528-1535. [DOI: 10.1117/1.JEI.27.5.051215]
- Gong S, Cristani M, Yan S and Loy C. C. 2013. Person Re-Identification, New York, NY, USA: Springer. (<https://link.springer.com/book/10.1007/978-1-4471-6296-4>)
- Gou M. 2022. Person Re-identification Datasets [EB/OL], <https://github.com/NEU-Gou/awesome-reid-dataset>.
- Gray D, Tao H. 2008. Viewpoint invariant pedestrian recognition with an ensemble of localized features//Proceedings of European Conference on Computer Vision, Marseille.262–275. [DOI:10.1007/978-3-540-88682-2_21]
- Gu X, Chang H, Ma B, Bai S, Shan S and Chen X. 2022. Clothes-Changing Person Re-identification with RGB Modality Only//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, USA: IEEE: 1060-1069 [DOI: 10.1109/CVPR52688.2022.00113]
- Gui S, Zhu Y, Qin X, and Ling X. 2020. Learning multi-level domain invariant features for sketch re-identification. Neurocomputing, vol.403, pp.294-303. [DOI: 10.1016/j.neucom.2020.04.060]
- Guo Y., Zhang L., Hu Y., He X., & Gao J. 2016. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition//Proceedings of the European conference on computer vision, 87-102. [DOI:10.1007/978-3-319-46487-9_6]
- Hafner F. M., Bhuiyan A., Kooij J. F. P., and Granger E. 2019. RGB-depth cross-modal person re-identification//Proceedings of the 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS'19). IEEE, 1–8. [DOI: 10.1109/AVSS.2019.8909838]
- Han B, Yao Q, Yu X, et al. 2018. Co-teaching: Robust training of deep neural networks with extremely noisy labels//Proceedings of the Conference on Neural Information Processing Systems (NeurIPS).8536–8546. [DOI: 10.5555/3327757.3327944]
- Han C, Su K, Yu D, Yuan Z, Gao C, Sang N, Yang Y and Wang C. 2022. Weakly supervised person search with region siamese networks// Proceedings of the IEEE/CVF International Conference on Computer Vision.Montreal, QC, Canada: IEEE: 12006-12015. [DOI: 10.1109/ICCV48922.2021.01179]
- Han C, Ye J, Zhong Y, Tan X, Zhang C, Gao C and Sang N. 2019. Re-id driven localization refinement for person search//Proceedings of the IEEE/CVF International Conference on Computer Vision.Seoul, Korea (South): IEEE: 9813-9822. [DOI: 10.1109/ICCV.2019.00991]
- He L., Liang J., Li H., & Sun Z. 2018. Deep spatial feature reconstruction

- for partial person re-identification: Alignment-free approach//Proceedings of the IEEE conference on computer vision and pattern recognition, 7073-7082. [DOI: 10.1109/cvpr.2018.00739]
- Hong P, Wu T, Wu A Han X and Zheng W S. 2022. Fine-Grained Shape-Appearance Mutual Learning for Cloth-Changing Person Re-Identification//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 10513-10522. [DOI: 10.1109/CVPR46437.2021.01037]
- Hou R., Ma B., Chang H., Gu X., Shan S., & Chen X. 2019. Vrsc: Occlusion-free video person re-identification//Proceedings of the IEEE conference on computer vision and pattern recognition, 7183-7192. [DOI: 10.1109/cvpr.2019.00735]
- Hou R., Ma B., Chang H., Gu X., Shan S., & Chen X. 2021. Feature completion for occluded person re-identification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(9): 4894 – 4912. [DOI: 10.1109/tpami.2021.3079910]
- Hu Y., Yi D., Liao S., Lei Z., & Li S.Z. 2014. Cross dataset person re-identification//Proceedings of the Asian Conference on Computer Vision.650-664. [DOI:10.1007/978-3-319-16634-6_47]
- Huang Bo., Li F. and Wang S. 2022. Sketch Person Re-identification With Cross Classification Towards Cross Modal Identity Inconsistent Scenes[J/OL]. Laser & Optoelectronics Progress. (黄勃淳, 李凡, 汪淑娟. 2022. 跨模态身份不一致的交叉分类素描行人重识别[J/OL]. 激光与光电子学进展.)
- Huang T and Russell S. 1997. Object identification in a Bayesian context//Proceedings of International Joint Conference on Artificial Intelligence, vol.97, 12761282. [DOI: 10.5555/1622270.1622340]
- Huang Y, Peng P, Jin Y, et al. 2019. Domain Adaptive Attention Model for Unsupervised CrossDomain Person Re-Identification [EB/OL]. [2022-10-22]. <https://arxiv.org/pdf/1905.10529.pdf>. [DOI: 10.48550/arXiv.1905.10529]
- Huang Y, Wu Q, Xu J, Zhang Z. 2022. Clothing Status Awareness for Long-Term Person Re-Identification//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 11895-11904. [DOI: 10.1109/ICCV48922.2021.01168]
- Huang Y, Xu J, Wu Q, Zhong Y, Zhang P and Zhang Z X. 2019. Beyond scalar neuron: Adopting vector-neuron capsules for long-term person re-identification. IEEE Transactions on Circuits and Systems for Video Technology, 30(10): 3459-3471. [DOI: 10.1109/TCSVT.2019.2948093]
- Iodice S., & Mikolajczyk K. 2018. Partial person re-identification with alignment and hallucination//Proceedings of the Asian Conference on Computer Vision, 101-116. [DOI: 10.1007/978-3-030-20876-9_7]
- Jia C, Luo M, Yan C, Chang X, and Zheng Q. 2022. CGUA: Context-Guided and Unpaired-Assisted Weakly Supervised Person Search [EB/OL]. [2022-03-27]. <https://arxiv.org/pdf/2203.14307.pdf> [DOI: 10.48550/arXiv.2203.14307]
- Jia J., Ruan Q., & Hospedales T.M. 2019. Frustratingly easy person re-identification: Generalizing person re-id in practice[EB/OL].[2022-10-22].<https://arxiv.org/pdf/1905.03422.pdf>. [DOI:10.48550/arXiv.1905.03422]
- Jia M., Cheng X., Lu S., & Zhang J. 2022. Learning disentangled representation implicitly via transformer for occluded person re-identification. IEEE Transactions on Multimedia. [early access] [DOI: 10.1109/tmm.2022.3141267]
- Jia X, Zhong X, Ye M, Liu W and Huang W. 2022. Complementary Data Augmentation for Cloth-Changing Person Re-Identification. IEEE Transactions on Image Processing, 31: 4227-4239. [DOI: 10.1109/TIP.2022.3183469]
- Jiang D. and Ye M. 2022. Temperature-scaled Projection Matching with CLIP for Text-to-Image Person Re-identification. Journal of Image and Graphics.2022(12).(姜定, 叶茫. 2022. 温度投影匹配结合 CLIP 的文本到图像行人重识别, 中国图象图形学报, (12)) [DOI: 10.11834/jig.220620]
- Jiang L, Zhou Z, Leung T, et al. 2018. Mentornet: Learning data-driven curriculum for very deep neural networks on corrupted labels//Proceedings of the International Conference on Machine Learning (ICML), 2304-2313. [DOI: 10.48550/arXiv.1712.05055]
- Jiao B., Liu L., Gao L., Lin G., Yang L., Zhang, S., and Zhang Y. 2022. Dynamically Transformed Instance Normalization Network for Generalizable Person Re-Identification//Proceedings of the European Conference on Computer Vision, 285-301. [DOI:10.1007/978-3-031-19781-9_17]
- Jin X, He T, Zheng K, et. al. 2022. Cloth-changing person re-identification from a single image with gait prediction and regularization//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE: 14278-14287 [DOI: 10.1109/CVPR52688.2022.01388]
- Jin X., Lan C., Zeng W., Chen Z., & Zhang L. 2020. Style normalization and restitution for generalizable person re-identification//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 3143-3152. [DOI: 10.1109/CVPR42600.2020.00321]
- Kai J, Bodensteiner C, Arens M. 2010. Person re-identification in multi-camera networks//Proceedings of the Computer Vision and Pattern Recognition Workshops. Colorado, USA.709-716. [DOI: 10.1109/CVPRW.2011.5981771]
- Kang C. 2022. Is Synthetic Dataset Reliable for Benchmarking Generalizable Person Re-Identification? //Proceedings of the 2022 International Joint Conference on Biometrics, Oct.10-13. [DOI: 10.48550/arXiv.2209.05047]
- Ke Z, Wang D, Yan Q, et al. 2019. Dual student: Breaking the limits of the teacher in semi-supervised learning//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 6728-6736. [DOI: 10.1109/ICCV.2019.00683]
- Kim H, Joung S, Kim I.J and Sohn, K. 2022. Prototype-Guided Saliency Feature Learning for Person Search//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.virtual: IEEE: 4865-4874. [DOI: 10.1109/CVPR46437.2021.00483]
- Koestinger M, Hirzer M, Wohlhart P, et al. 2012. Large scale metric learning from equivalence constraint//IEEE Conference on Computer Vision and Pattern Recognition. 2288–2295. [DOI:10.1109/CVPR.2012.6247939]

- Krizhevsky A, Sutskever I, Hinton G E. 2012. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60: 84-90. [DOI:10.1145/3065386]
- Lan X, Zhu X and Gong S. 2018. Person search by multi-scale matching//*Proceedings of the European Conference on Computer Vision*. Munich, Germany: Springer: 553-569. [DOI: 10.1007/978-3-030-01246-5_33]
- Layne R, Hospedales T M, Gong S, et al. 2012. Person re-identification by attributes// *Proceedings of the British Machine Vision Conference*. Vol.2.2012: 8. [DOI:10.5244/C.26.24]
- Li H., Ye M., & Du B. 2021. Weperson: Learning a generalized re-identification model from all-weather virtual data//*Proceedings of the 29th ACM International Conference on Multimedia*, 3115-3123. [DOI:10.1145/3474085.3475455]
- Li J, Yan Y, Wang G, et al. 2022. Domain Adaptive Person Search[C]//*Proceedings of the European Conference on Computer Vision*. Springer, Cham.302-318. [DOI: 10.1007/978-3-031-19781-9_18]
- Li J, Zhang S. 2020. Joint visual and temporal consistency for unsupervised domain adaptive person re-identification//*Proceedings of the European Conference on Computer Vision (ECCV)*, 483-499. [DOI: 10.1007/978-3-030-58586-0_29]
- Li J. 2021. Research On Adaptive Occluded Person Re-Identification Method based on Human Pose. Chengdu: Xihua University. (李吉林. 2022. 基于人体姿态的自适应遮挡行人再识别方法研究.成都:西华大学.) [DOI: 10.27411/d.cnki.gscgc.2021.000452.]
- Li S, Chen H, Yu S, He Z, Zhu F, Zhao R, Chen J and Qiao Y. 2022. COCAS+: Large-Scale Clothes-Changing Person Re-identification with Clothes Templates. *IEEE Transactions on Circuits and Systems for Video Technology*. IEEE: early access [DOI: 10.1109/TCSVT.2022.3216769]
- Li S, Xiao T, Li H, et al. 2017. Person search with natural language description//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.1970-1979. [DOI: 10.1109/CVPR.2017.551]
- Li S, Xiao T, Li H, et al. 2017. Person search with natural language description//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1970-1979.
- Li S, Xiao T, Li H, Zhou B, Yue D, and Wang X. 2017. Person search with natural language description//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.1970-1979. [DOI: 10.1109/CVPR.2017.551]
- Li W, Zhao R, Xiao T, et al. 2014. Deepreid: Deep filter pairing neural network for person re-identification//*Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.152-159. [DOI:10.1109/CVPR.2014.27]
- Li W., Zhao R., Xiao T., & Wang X. 2014. Deepreid: Deep filter pairing neural network for person re-identification//*Proceedings of the IEEE conference on computer vision and pattern recognition*. [DOI:152-159.10.1109/CVPR.2014.27]
- Li Y J, Weng X, Kitani K M. 2022. Learning Shape Representations for Person Re-Identification under Clothing Change//*Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 2432-2441 [DOI: 10.1109/WACV48630.2021.00248]
- Li Z and Miao D. 2022. Sequential end-to-end network for efficient person search// *Proceedings of the AAAI Conference on Artificial Intelligence*.2011-2019. [DOI: 10.1609/aaai.v35i3.16297]
- Liao S and Shao L. 2020. Interpretable and generalizable person re-identification with query-adaptive convolution and temporal lifting//*Proceedings of the In European Conference on Computer Vision*, 456-474. [DOI:10.1007/978-3-030-58621-8_27]
- Liao S, and Shao L. 2022. TransMatcher: Deep Image Matching Through Transformers for Generalizable Person Re-identification//*Proceedings of the Advances in Neural Information Processing Systems*, 34, 1992-2003. [DOI:10.48550/arXiv.2105.14432]
- Liao S, and Shao L. 2022. Graph sampling based deep metric learning for generalizable person re-identification//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7359-7368. [DOI: 10.1109/CVPR52688.2022.00721]
- Lin S., Li C.T., & Kot A. C. 2020. Multi-domain adversarial feature generalization for person re-identification. *IEEE Transactions on Image Processing*, 30, 1596-1607. [DOI: 10.1109/TIP.2020.3046864]
- Lin Y, Dong X, Zheng L, et al. 2019. A bottom-up clustering approach to unsupervised person re-identification//*Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 33, 01, 8738-8745. [DOI: 10.1609/aaai.v33i01.33018738]
- Lin Y, Xie L, Wu Y, et al. 2020. Unsupervised person re-identification via softened similarity learning//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3390-3399. [DOI: 10.1109/CVPR42600.2020.00345]
- Liu H, Tan X, Zhou X. 2020. Parameter sharing exploration and hetero-center triplet loss for visible-thermal person re-identification[J]. *IEEE Transactions on Multimedia*, 23(12): 4414-4425. [DOI: 10.1109/TMM.2020.3042080]
- Liu H., Liang H., & Ma L. 2017. Online RGB-D person re-identification based on metric model update. *CAAI Trans.Intelligence Technology* 2.1(2017): 48-55. [DOI: 10.1016/j.trit.2017.04.001]
- Liu J, Zha Z J, Chen D, et al. 2019. Adaptive transfer network for cross-domain person re-identification//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.2019: 7202-7211. [DOI: 10.1109/CVPR.2019.00737]
- Liu J, Zha Z J, Hong R, et. al. 2019. Deep adversarial graph attention convolution network for text-based person search//*Proceedings of the 27th ACM international Conference on Multimedia*, 2019, pp.665-673. [DOI: 10.1145/3383184]
- Liu J., Huang Z., Li L., Zheng K., & Zha Z.J. 2022. Debaised batch normalization via gaussian process for generalizable person re-identification [EB/OL]. [2022-10-22]. <https://arxiv.org/pdf/2203.01723.pdf>. [DOI:10.48550/arXiv.2203.01723]
- Liu Q., Wang H., Cao L., Sun B., Xiao Y. and Zhang J. 2022. Cloth-changing person re-identification based on joint loss capsule

- network, *Journal of Computer Applications*, 41(12): 3596-3601. (刘乾, 王洪元, 曹亮, 孙博言, 肖宇, 张继. 2021 基于联合损失胶囊网络的换衣行人重识别. *计算机应用*, 41(12): 3596-3601.) [DOI: 10.11772/j.issn.1001-9081.2021061090]
- Liu X, Zhang S. 2020. Domain adaptive person re-identification via coupling optimization // *Proceedings of the 28th ACM international conference on multimedia*. 547-555. [DOI: 10.1145/3394171.3413904]
- Mekhzazi D, Bhuiyan A, Ekladios G, et al. 2020. Unsupervised domain adaptation in the dissimilarity space for person re-identification // *Proceedings of the European Conference on Computer Vision (ECCV)*. 159-174. [DOI: 10.1007/978-3-030-58583-9_10]
- Miao J., Wu Y., Liu P., Ding Y., & Yang Y. 2019. Pose-guided feature alignment for occluded person re-identification // *Proceedings of the IEEE international conference on computer vision*, 542-551. [DOI: 10.1109/iccv.2019.00063]
- Mo C. 2022. Text Pedestrian Cross-Modal Retrieval Based on Two-Way Convolutional Local Alignment. *Video Engineering*. 46(04): 25-30. (莫承见. 基于双路卷积局部对齐的文本行人跨模态检索. *电视技术*, 46(04): 25-30.) [DOI: 10.16280/j.videoe.2022.04.006.]
- Moon H and Phillips P.J. 2001. Computational and performance aspects of PCA-based face-recognition algorithms. *Perception*, 30(3): 3003-321. [DOI: 10.1068/p2896]
- Munaro M, Basso A, Fossati A, et al. 2014. 3D reconstruction of freely moving persons for re-identification with a depth sensor // *2014 IEEE international conference on robotics and automation (ICRA)*. 4512-4519. [DOI: 10.1109/ICRA.2014.6907518]
- Nguyen D T, Hong H G, Kim K W, et al. 2017. Person Recognition System Based on a Combination of Body Images from Visible Light and Thermal Cameras. *Sensor*. 17(3), 605. [DOI: 10.3390/s17030605]
- Oreofej O, Mehran R, Shah M. 2010. Human identity recognition in aerial images // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. San Francisco, USA. 709-716. [DOI: 10.1109/CVPR.2010.5540147]
- Pang L, Wang Y, Song Y Z, et al. 2018. Cross-domain adversarial feature learning for sketch re-identification // *Proceedings of the 26th ACM international conference on Multimedia*. 609-617. [DOI: 10.1145/3240508.3240606]
- Pang L., Wang Y., Song Y.-Z., Huang T., and Tian Y. 2018. Cross-domain adversarial feature learning for sketch re-identification // *Proceedings of the 26th ACM international conference on Multimedia*. pp. 609-617. [DOI: /10.1145/3240508.3240606]
- Qi L, Wang L, Huo J, et al. 2019. A Novel Unsupervised Camera-Aware Domain Adaptation Framework for Person Re-Identification // *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 8080-8089. [DOI: 10.1109/ICCV.2019.00817]
- Qian X L, Wang W X, Zhang L, Zhu F R, Fu Y W, Xiang T, Jiang Y G and Xue X Y. 2020. Long-Term Cloth-Changing Person Re-identification // *Asian Conference on Computer Vision (ACCV)*. Kyoto, Japan: Springer: 71-88 [DOI: 10.1007/978-3-030-69535-4_5]
- Ren S., He K., Girshick R., & Sun J. 2015. Faster RCNN: Towards Real-time Object Detection with Region Proposal Networks. *Advances in neural information processing systems*, 28. [DOI: 10.1109/TPAMI.2016.2577031]
- Ren X., Zhang D., Bao X. and Li B. 2022. Semantic guidance attention network for occluded person re-identification. 42(10), 106-116. (任雪娜, 张冬明, 包秀国 & 李冰. 2022. 语义引导的遮挡行人再识别注意力网络. *通信学报*, 42(10), 106-116.) [DOI: 10.11959/j.issn.1000-436x.2021184]
- Ristani E., Solera F., Zou R., Cucchiara R., & Tomasi C. 2016. Performance measures and a data set for multi-target, multi-camera tracking // *Proceedings of the European conference on computer vision*, 17-35. [DOI: 10.1007/978-3-319-48881-3_2]
- Shao Z., Zhang X., Fang M., et al. 2022. Learning Granularity-Unified Representations for Text-to-Image Person Re-identification // *Proceedings of the 30th ACM International Conference on Multimedia*. 5566-5574. [DOI: 10.1145/3503161.3548028]
- Shen X. 2022. Research on domain generalized feature learning network for pedestrian recognition. Nanjing: Nanjing University of Science and Technology (沈贤文. 2022. 面向行人再识别的域泛化特征学习网络研究. 南京: 南京理工大学.) [DOI: 10.27241/d.cnki.gnjgu.2021.001178]
- Shi W., Zhang Y., Liu S., Zhu S., and Bao J. 2020. Person re-identification based on deformation and occlusion mechanisms. 25(12), 2530-2540. (史维东, 张云洲, 刘双伟, 朱尚栋 & 暴吉宁. 2020. 针对形变与遮挡问题的行人再识别. *中国图象图形学报* 25(12), 2530-2540.) [DOI: 10.11834/jig.200016]
- Shu X, Wang X, Zang X, Zhang S, Chen Y, Li G and Tian Q. 2022. Large-Scale Spatio-Temporal Person Re-identification: Algorithms and Benchmark. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(7): 4390 - 4403 [DOI: 10.1109/TCSVT.2021.3128214]
- Song J., Yang Y., Song Y.Z., Xiang T., & Hospedales T.M. 2019. Generalizable person re-identification by domain-invariant mapping network. *IEEE/CVF conference on Computer Vision and Pattern Recognition*, 719-728. [DOI: 10.1109/CVPR.2019.00081]
- Sun X., & Zheng L. 2019. Dissecting person re-identification from the viewpoint of viewpoint. // *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 608-617. [DOI: 10.1109/CVPR.2019.00070]
- Tan H., Liu X., Yin B., & Li X. 2022. MHSA-Net: Multi-head self-attention network for occluded person re-identification. *IEEE Transactions on Neural Networks and Learning Systems*. [DOI: 10.1109/tnnls.2022.3144163]
- Tarvainen A, Valpola H. 2017. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results // *Proceedings of the Conference on Neural Information Processing Systems (NeurIPS)*, 1195-1204. [DOI: 10.48550/arXiv.1703.01780]
- Tian W. 2022. Improving Occluded Person Re-identification by Pixel-

- wise Semantic Annotation and Feature Uncertainty. Wuhan: Huazhong University of Science and Technology. (田维怡. 2021 基于像素语义注释与特征不确定性的遮挡行人再识别框架. 武汉: 华中科技大学. [DOI: 10.27157/d.cnki.ghzku.2021.003156.]
- Tian Z., Shen C., Chen H. and He T. 2019. Fcos: Fully convolutional one-stage object detection// Proceedings of the IEEE/CVF international conference on computer vision. Seoul, Korea (South): IEEE: 9627-9636. [DOI: 10.1109/ICCV.2019.00972]
- Wan F., Wu Y., Qian X., et al. 2020. When person re-identification meets changing clothes//IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, USA: IEEE: 830-831 [DOI: 10.1109/CVPRW50498.2020.00423]
- Wang C., Ma B., Chang H., Shan S. and Chen X. 2020. Tcts: A task-consistent two-stage framework for person search//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE: 11952-11961. [DOI: 10.1109/CVPR42600.2020.01197]
- Wang F., Yan T., Liu F., Qian Y. and Xu Y. 2022. Multi-scale Cross-Modality Person Re-identification Method Based on Shared Subspace Features[J/OL]. Journal of Electronics & Information Technology. (王凤随, 闫涛, 刘芙蓉, 钱亚萍, 许月. 2022. 融合子空间共享特征的多尺度跨模态行人重识别方法[J/OL]. 电子与信息学报.
- Wang G., Zhang T., Cheng J., Liu S., Yang Y., and Hou Z. 2019. "gb-infrared cross-modality person re-identification via joint pixel and feature alignment//Proceedings of the IEEE/CVF International Conference on Computer Vision.pp.3623-3632. [DOI: 10.1109/ICCV.2019.00372]
- Wang G. A., Yang S., Liu H., et. al. 2020. High-order information matters: Learning relation and topology for occluded person re-identification//Proceedings of the IEEE conference on computer vision and pattern recognition, 6449-6458. [DOI: 10.1109/cvpr42600.2020.00648]
- Wang W., Liao S., Zhao F., Kang C., and Shao L. 2022. DomainMix: Learning Generalizable Person Re-Identification Without Human Annotations//In British Machine Vision Conference. [DOI: 10.48550/arXiv.2011.11953]
- Wang W., Liao S., Zhao F., Kang C., & Shao L. 2020. Domainmix: Learning generalizable person re-identification without human annotations. [EB/OL]. [2022-10-22]. <https://arxiv.org/pdf/2011.11953.pdf>. [DOI: 10.48550/arXiv.2011.11953]
- Wang Y., Liang X., & Liao S. 2022. Cloning Outfits from Real-World Images to 3D Characters for Generalizable Person Re-Identification// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 4900-4909. [DOI: 10.1109/CVPR52688.2022.00485]
- Wang Y., Liao S., & Shao L. 2020. Surpassing real-world source training data: Random 3d characters for generalizable person re-identification//Proceedings of the 28th ACM international conference on multimedia, 3422-3430. [DOI: 10.1145/3394171.3413815]
- Wang Z., Wang Z., Zheng Y., Chuang Y.-Y., and Satoh S. 2019. Learning to reduce dual-level discrepancy for infrared-visible person re-identification//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition.pp.618-626. [DOI: 10.1109/CVPR.2019.00071]
- Wei L., Zhang S., Gao W., et al. 2018. Person transfer GAN to bridge domain gap for person re-identification//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).79-88. [DOI: 10.1109/CVPR.2018.00016]
- Wu A., Zheng W S, Yu H X, et al. 2017. RGB-infrared cross-modality person re-identification// Proceedings of the IEEE international conference on computer vision.5380-5389. [DOI: 10.1109/ICCV.2017.575]
- Wu A., Lin C. and Zheng W. 2022. Single-modality self-supervised information mining for cross-modality person re-identification, Journal of Image and Graphics.27(10): 2843-2859.(吴岸聪, 林城桔, 郑伟诗. 面向跨模态行人重识别的单模态自监督信息挖掘[J]. 中国图象图形学报, 2022, 27(10):2843-2859.) [DOI: 10.11834/jig.211050]
- Wu J., Jiang J., Qi M., et al. 2022. An End-to-end Heterogeneous Restraint Network for RGB-D Cross-modal Person Re-identification. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM).18(4): 1-22. [DOI: 10.1145/3506708]
- Xiang Z., Chen Q., Liu Y. 2012. Person re-identification by fuzzy space color histogram. Multimedia Tools and Applications, 73: 91-107. [DOI: 10.1007/s11042-012-1286-7]
- Xiao T., Li S., Wang B., Lin L. and Wang X. 2017. Joint detection and identification feature learning for person search//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE: 3415-3424. [DOI: 10.1109/CVPR.2017.360]
- Xiong F., Gou M., Camps O I., et al. 2004. Person re-identification using kernel-based metric learning methods//Proceedings of the European Conference on Computer Vision.1-16. [DOI: 10.1007/978-3-319-10584-0_1]
- Xu B., He L., Liang J., & Sun Z. 2022. Learning Feature Recovery Transformer for Occluded Person Re-Identification. IEEE Transactions on Image Processing, 31, 4651-4662. [DOI: 10.1109/tip.2022.3186759]
- Xu P., Zhu X. 2022. Deepchange: A large long-term person re-identification benchmark with clothes change. [EB/OL]. [2022-10-22]. <https://arxiv.org/pdf/2105.14685.pdf> [DOI: 10.48550/arXiv.2105.14685]
- Xu W., Liu H., Shi W., Miao Z., Lu Z. and Chen F. 2022. Adversarial Feature Disentanglement for Long-Term Person Re-identification//Proceedings of the International Joint Conference on Artificial Intelligence. Montreal, USA: AAAI: 1201-1207 [DOI: 10.24963/ijcai.2021/166]
- Xu, B., Liang, J., He, L., & Sun, Z. 2022. Mimic Embedding via Adaptive Aggregation: Learning Generalizable Person Re-identification//Proceedings of the European Conference on Computer Vision, 372-388. [DOI: 10.1007/978-3-031-19781-9_22]
- Xuan S., Zhang S. 2022. Intra-inter camera similarity for unsupervised

- person re-identification//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 11926-11935. [DOI: 10.1109/CVPR46437.2021.01175]
- Yan S., Tang H., Zhang L. and Tang J. 2022. Image-Specific Information Suppression and Implicit Local Alignment for Text-based Person Search. [EB/OL]. [2022-08-30]. <https://arxiv.org/pdf/2208.14365.pdf> [DOI: 10.48550/arXiv.2208.14365]
- Yan Y, Li J, Liao S, Qin J, Ni B, Lu K and Yang X. 2022. Exploring Visual Context for Weakly Supervised Person Search// Proceedings of the AAAI Conference on Artificial Intelligence.Virtual Event: AAAI Press: 3027-3035. [DOI: 10.1609/aaai.v36i3.20209]
- Yan Y, Li J, Qin J, Bai S, Liao S, Liu L, Zhu F and Shao L. 2022. Anchor-Free Person Search// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.virtual: IEEE: 7690-7699. [DOI: 10.1109/CVPR46437. 2022. 00760]
- Yan Y., Yu H., Li S., Lu Z., He J., Zhang H. and Wang R. 2022. Weakening the Influence of Clothing: Universal Clothing Attribute Disentanglement for Person Re-Identification//Proceedings of the International Joint Conference on Artificial Intelligence, Vienna, Austria: AAAI: 1523-1529 [10.24963/ijcai.2022/212]
- Yang F., Wu Y., Wang Z., Li X., Sakti S., and Nakamura S.2020 "Instance-level heterogeneous domain adaptation for limited-labeled sketch-to-photo retrieval," IEEE Transactions on Multimedia.23: 2347 – 2360. [DOI: 10.1109/TMM.2020.3009476]
- Yang J., Zhang J., Yu F., et al. 2022. Learning to know where to see: a visibility-aware approach for occluded person re-identification//Proceedings of the IEEE International Conference on Computer Vision, 11885-11894. [DOI: 10.1109/iccv48922.2021.01167]
- Yang Q Z, Wu A C and Zheng W S. 2022. Person re-identification by contour sketch under moderate clothing change. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(6): 2029-2046 [DOI: 10.1109/TPAMI.2019.2960509]
- Yang S, Kang B and Lee Y. 2022. Sampling Agnostic Feature Representation for Long-Term Person Re-Identification. IEEE Transactions on Image Processing, 31: 6412-6423 [DOI: 10.1109/TIP.2022.3207024]
- Yang X., Zhang L., Wang Z., et al. 2022. Cross-domain pedestrian re-identification based on capsule network. 2022. (09), 1591-1599.(杨晓峰, 张来福, 王志鹏, 萨旦姆, 邓红霞 & 李海芳. 2022. 基于胶囊网络的跨域行人再识别.计算机工程与科学(09), 1591-1599.) [DOI: 10.3969/j.issn.1007-130X.2021.09.009]
- Yang Y, Yang J, Yan J, et al. 2014. Salient color names for person re-identification//Proceedings of the European Conference on Computer Vision.536-551. [DOI:10.1007/978-3-319-10590-1_35]
- Ye H., Liu H., Meng F., Li X. 2022. Bi-directional exponential angular triplet loss for rgb-infrared person re-identification. IEEE Transactions on Image Processing, 30: 1583-1595. [DOI: 10.1109/TIP.2020.3045261]
- Ye M, Shen J, Crandall D. J., Shao L., and Luo J. 2020. Dynamic dual-attentive aggregation learning for visible-infrared person re-identification//Proceedings of the European Conference on Computer Vision.229-247. [DOI: 10.1007/978-3-030-58520-4_14]
- Ye M, Lan X, Li J, and Yuen P. C. 2018. Hierarchical discriminative learning for visible thermal person re-identification //Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence.pp.7501–7508. [DOI: 10.1609/aaai.v32i1.12293]
- Ye M, Lan X, Wang Z, and Yuen P. C. 2020. "Bi-directional center-constrained top-ranking for visible thermal person re-identification," IEEE Transactions on Information Forensics and Security (TIFS), vol.15, pp.407–419. [DOI: 10.1109/TIFS.2019.2921454]
- Ye M, Ruan W, Du B, et al. 2022. Channel augmented joint learning for visible-infrared recognition[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision.13567-13576. [DOI: 10.1109/ICCV48922.2021.01331]
- Ye M., Shen J., Lin G., Xiang T., Shao L., & Hoi S. C. 2022. Deep Learning for Person Re-Identification: A Survey and Outlook. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44, 2872-2893. [DOI: 10.1109/TPAMI.2021.3054775]
- Yi D., Lei Z., Liao S., & Li S.Z.2014, Deep metric learning for person re-identification// IEEE International conference on pattern recognition. IEEE.34-39. [DOI: 10.1109/ICPR.2014.16]
- Yin J H, Wu A C and Zheng W S. 2020. Fine-grained Person Re-identification.International Journal of Computer Vision, 128(6): 1654-1672 [DOI: 10.1007/s11263-019-01259-0]
- Yu H X, Zheng W S, Wu A, et al. 2019. Unsupervised Person Re-identification by Soft Multilabel Learning//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2148-2157. [DOI: 10.1109/CVPR.2019.00225]
- Yu M., Li X., & Guo Y. 2022. Domain generalization person re-identification based on attention mechanism. (07), 1721-1728. 2022. (于明, 李学博, 郭迎春. 2022. 融合注意力机制的域泛化行人再识别.控制与决策(07), 1721-1728.) [DOI: 10.13195/j.kzyjc.2020.1844]
- Yu R, Du D, LaLonde R, Davila D, Funk C, Hoogs A and Clipp B. 2022. Cascade Transformers for End-to-End Person Search// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.New Orleans, LA, USA: IEEE: 7267-7276. [DOI: 10.1109/CVPR52688.2022.00712]
- Yu S, Li S, Chen D, Zhao R, Yan J and Qiao Y. 2020. Cocas: A large-scale clothes changing person dataset for re-identification//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 3400-3409 [DOI: 10.1109/CVPR42600.2020.00346]
- Yu S. 2022. General Person Re-Identification: Methods and Application.Shenzhen: University of Chinese Academy of Sciences. (余世杰. 2022. 通用行人重识别方法与应用.深圳:中国科学院大学.)[DOI: 10.27822/d.cnki.gsxxj.2022.000135]
- Yu X, Han B, Yao J, et al. 2019. How does disagreement help generalization against label corruption? //Proceedings of the International Conference on Machine Learning (ICML), 7164-7173. [DOI:10.48550/arXiv.1901.04215]
- Zajdel W, Zivkovic Z, Krose B J A. 2005. Keeping track of humans: have

- I seen this person before? //Proceedings of the 2005 IEEE International Conference on Robotics and Automation.2081-2086. [DOI:10.1109/ROBOT.2005.1570420]
- Zhai Y, Ye Q, Lu S, et al. 2020. Multiple expert brainstorming for domain adaptive person re-identification//Proceedings of the European Conference on Computer Vision (ECCV).594-611. [DOI: 10.1007/978-3-030-58571-6_35]
- Zhang P, Xu J, Wu Q, Huang Y, and Zhang J. 2019. Top-push constrained modality-adaptive dictionary learning for cross-modality person re-identification. IEEE Transactions on Circuits and Systems for Video Technology, 30(12): 4554-4566. [DOI: 10.1109/TCSVT.2019.2939564]
- Zhang S, Long D, Gao Y, Gao L, Zhang Q, Niu K and Zhang, Y. 2022. Text-based Person Search in Full Images via Semantic-Driven Proposal Generation [EB/OL]. [2021-09-27]. <https://arxiv.org/pdf/2109.12965.pdf> [DOI: 10.48550/arXiv.2109.12965]
- Zhang T., Xie L., Wei L., Zhuang Z., Zhang Y., Li B., and Tian Q. 2022. Unrealperson: An adaptive pipeline towards costless person re-identification//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 11506-11515. [DOI:10.1109/CVPR46437.2021.01134]
- Zhang Y. and Lu H. 2018. Deep cross-modal projection learning for image-text matching//Proceedings of the European Conference on Computer Vision.686-701. [DOI: 10.1007/978-3-030-01246-5_42]
- Zhao C., Qi D., Dou S., et. al. 2022. Key technology for intelligent video surveillance: a review of person re-identification.SCIENTIA SINICA Informationis.51(12): 1979-2015) (赵才荣, 齐鼎, 窦曙光等. 2022. 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, 51(12): 1979-2015) [DOI: 10.1360/ssi-2021-0211]
- Zhao J., Zhao Y., Chen X., & Li J. 2022. Revisiting Stochastic Learning for Generalizable Person Re-identification//Proceedings of the ACM International Conference on Multimedia, 1758-1768. [DOI:10.1145/3503161.3547812]
- Zhao R, Ouyang W, Wang X R. 2013. Unsupervised salience learning for person re-identification//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.Portland, USA.3586-3593. [DOI:10.1109/CVPR.2013.460]
- Zhao R, Ouyang W, Wang X. 2013. Unsupervised Salience Learning for Person Re-identification// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.3586-3593. [DOI:10.1109/CVPR.2013.460]
- Zhao Y, Zhao X, Luo R, et al. 2015. Person re-identification by encoding free energy feature maps.Multimedia Tools and Applications, 75: 4795-4813. [DOI:10.1007/s11042-015-2503-y]
- Zhao Y., Zhong Z., Yang F., Luo Z., Lin Y., Li S., & Sebe N. 2022. Learning to generalize unseen domains via memory-based multi-source meta-learning for person re-identification//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 6277-6286. [DOI: 10.1109/CVPR46437.2021.00621]
- Zheng A, Pan P, Li H, et al. 2022. Progressive Attribute Embedding for Accurate Cross-modality Person Re-ID[C]//Proceedings of the 30th ACM International Conference on Multimedia.4309-4317. [DOI: 10.1145/3503161.3548336]
- Zheng K., Lan C., Zeng W., Liu J., Zhang Z., & Zha Z.J. 2022. Pose-guided feature learning with knowledge distillation for occluded person re-identification//In Proceedings of the 29th ACM International Conference on Multimedia, 4537-4545. [DOI: 10.1145/3474085.3475610]
- Zheng L, Zhang H, Sun S, Chandraker M, Yang Y and Tian Q. 2017. Person re-identification in the wild//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.Honolulu, HI, USA: IEEE: 3346--3355. [DOI: 10.1109/CVPR.2017.357]
- Zheng L., Shen L., Tian L., Wang S., Wang J., & Tian Q. 2015. Scalable person re-identification: A benchmark// Proceedings of the IEEE international conference on computer vision, 1116-1124. [DOI: 10.1109/ICCV.2015.133]
- Zheng P, Qin J, Yan Y, Liao S, Ni B, Cheng X and Shao L. 2022. Global-Local Context Network for Person Search. [EB/OL]. [2022-04-12]. <https://arxiv.org/pdf/2112.02500.pdf> [DOI: 10.48550/arXiv.2112.02500]
- Zheng W.S., Gong S., & Xiang T. 2011. Person re-identification by probabilistic relative distance comparison//Proceedings of the IEEE conference on computer vision and pattern recognition, 649-656. [DOI: 10.1109/cvpr.2011.5995598]
- Zheng W.S., Li X., Xiang T., Liao S., Lai J., & Gong S. 2015. Partial person re-identification//In Proceedings of the IEEE International Conference on Computer Vision, 4678-4686. [DOI: 10.1109/ICCV.2015.531]
- Zheng Z, Zheng L, Garrett M, et. al. 2020. Dual-Path Convolutional Image-Text Embeddings with Instance Loss. ACM Transactions on Multimedia Computing Communications and Applications, 16(2) 1-23. [DOI: 10.1145/3383184]
- Zheng Z., Zheng L., & Yang Y. 2017. Unlabeled Samples Generated by GAN Improve the Person Re-identification Baseline in Vitro //Proceedings of the IEEE international conference on computer vision, 3754-3762. [DOI: 10.1109/ICCV.2017.405]
- Zhong Z, Zheng L, Luo Z, et al. 2019. Invariance matters: Exemplar memory for domain adaptive person re-identification//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 598-607. [DOI: 10.1109/CVPR.2019.00069]
- Zhong Z, Zheng L, Luo Z, et al. 2022. Learning to adapt invariance in memory for person reidentification. IEEE transactions on pattern analysis and machine intelligence, 43(8): 2723 - 2738. [DOI: 10.1109/TPAMI.2020.2976933]
- Zhou K., Yang Y., Cavallaro A., & Xiang T. 2022. Learning generalizable omni-scale representations for person re-identification. IEEE Transactions on Pattern Analysis and Machine Intelligence. [DOI:10.1109/TPAMI.2021.3069237]
- Zhu A, Wang Z, Li Y, et al. 2022. DSSL: Deep Surroundings-person Separation Learning for Text-based Person Retrieval[C]//Proceedings of the 29th ACM International Conference on Multimedia.209-217. [DOI: 10.1145/3474085.3475369]
- Zhu Y., Yang Z., Wang L., Zhao S., Hu X., Tao D. 2020. Hetero-center

loss for cross-modality person re-identification. Neurocomputing. 386: 97-109. [DOI: 10.1016/j.neucom.2019.12.100]

Zhuo J., Chen Z., Lai J., & Wang G. 2018. Occluded person re-

identification//IEEE International Conference on Multimedia and Expo, 1-6. [DOI: 10.1109/icme.2018.8486568]

作者简介:



张永飞, 通信作者, 1982 年生, 男, 教授, 研究方向为计算机视觉、智能视觉监控。E-mail: yfzhang@buaa.edu.cn



杨航远, 1994 年生, 男, 博士研究生, 研究方向为计算机视觉、机器学习。E-mail: hyyang@buaa.edu.cn



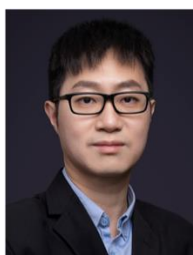
张雨佳, 2000 年生, 女, 博士研究生, 研究方向为计算机视觉、智能视觉监控。E-mail: yujia_zhang@buaa.edu.cn



豆朝鹏, 1996 年生, 男, 博士生, 研究方向为计算机视觉与机器学习。E-mail: dcp19@mails.tsinghua.edu.cn



廖胜才, 1982 年生, 男, 主任科学家, 研究方向为行人和人脸分析。E-mail: scliao@ieee.org



郑伟诗, 1981 年生, 男, 教授, 研究方向为协同与交互分析理论与方法, 解决(跨场景)追踪与行为感知处理问题。E-mail: zhwshi@mail.sysu.edu.cn



张史梁, 1984 年生, 男, 长聘副教授, 研究方向为人工智能与媒体计算。E-mail: slzhang.jdl@pku.edu.cn.



叶茫, 1991 年生, 男, 教授, 研究方向为计算机视觉。E-mail: yemang@whu.edu.cn



晏轶超, 1991 年生, 男, 助理教授, 研究方向为数字人生成与分析。E-mail: yanyichao@sjtu.edu.cn



李俊杰, 2000 年生, 男, 硕士研究生, 研究方向为行人搜索。E-mail: junjieli00@sjtu.edu.cn



王生进, 通信作者, 1962 年生, 男, 教授, 研究方向为计算机视觉与机器学习、多模态协作机器人。E-mail: wgsgj@tsinghua.edu.cn