可见光-红外跨模态行人重识别研究综述

励志勇[1,2],姜 伟[2*],刘浩杰[1]
(1. 浙江大学控制科学与工程学院,浙江杭州 310027;2. 浙江水利水电学院计算机科学与技术学院,浙江杭州 310018)

摘 要: 行人重识别(Person Re-identification,ReID)作为智能视频监控系统的核心技术,其核心任务是在非重叠视域的摄像头网络中实现对特定目标行人的高效检索与匹配. 然而,传统仅依赖可见光图像的方法在夜间或低照度等复杂光照条件下性能显著下降. 为应对这一挑战,可见光-红外行人重识别(Visible-Infrared Person Re-identification,VI-ReID)应运而生,旨在实现可见光图像与红外图像之间的交叉检索. 该任务不仅继承了单模态行人重识别中姿态变化、视角差异和遮挡等固有难题,更需克服由成像机理不同所导致的巨大跨模态鸿沟. 本文对近年来基于深度学习的可见光-红外跨模态行人重识别方法进行了系统性梳理、归纳与评述,将现有主流方法划分为三大核心类别:(1)基于跨模态网络结构设计的方法,通过精心构造网络架构以提取模态不变的身份特征,具体包括双流特征提取网络、身份信息解耦模块、细粒度特征对齐,以及利用网络结构搜索等设计方法;(2)生成式学习方法,旨在通过模态转换或数据增强弥合模态间差距,涵盖单向或双向图像生成、构建统一中间模态,以及在特征层面进行生成与补偿等策略;(3)基于跨模态相似度学习的方法,聚焦于损失函数与度量学习的设计,通过拉近跨模态正样本对的距离并推开负样本对,主要包括基于样本或中心(代理)的对比学习,以及针对测试阶段优化的跨模态重排序算法. 此外,考虑到实际应用中标注成本高昂且标签可能存在噪声或缺失,本文进一步深入探讨了非完全有监督学习范式下的研究进展,系统总结了噪声标签学习、半监督学习及无监督学习等方向所面临的独特挑战与代表性解决方案. 为全面评估各类算法的性能,本文在SYSU-MM01、RegDB和LLCM公开数据集上,对不同监督范式下的代表性算法进行了统一的性能对比与分析. 最后,本文立足于当前研究的技术瓶颈,对未来发展趋势进行了展望,指出构建更贴近真实场景的多样化数据集、缓解模态数据不平衡问题、推动模型轻量化部署、探索可持续或终身学习机制,以及拓展至视频级或多源异构信息融合的行人重识别等方向将是该领域极具潜力的研究热点,旨在为后续学者提供有价值的理论参考与技术指引.

关键词: 跨模态行人重识别;模式识别;深度学习;表征学习;网络结构设计;生成式学习;智能监控
基金项目: 浙江省自然科学基金(No.LZ24F030004)
中图分类号: TP391.4;TP18 **文献标识码:** A **文章编号:** 0372-2112(2025)12-4811-22
电子学报 URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20250800

A Survey on Visible-Infrared Cross-Modality Person Re-Identification

LI Zhi-yong[1,2], JIANG Wei[2*], LIU Hao-jie[1]
(1. *College of Control Science and Engineering*, *Zhejiang University*, *Hangzhou*, *Zhejiang* 310027, *China*;
2. *School of Computer Science and Technology*, *Zhejiang University of Water Resources and Electric Power*, *Hangzhou*, *Zhejiang* 310018, *China*)

Abstract: Person re-identification (ReID) is a core technology in intelligent video surveillance systems, with the fundamental objective of efficiently retrieving and matching a specific pedestrian across camera networks with non-overlapping fields of view. However, traditional approaches that rely solely on visible images suffer severe performance degradation under challenging illumination conditions such as nighttime or low-light environments. To address this limitation, visible-infrared person re-identification (VI-ReID) has emerged, aiming to enable cross-modal retrieval between visible and infrared images. This task not only inherits classic challenges from unimodal ReID—such as pose variations, viewpoint changes, and occlusions—but also faces a significant cross-modal discrepancy arising from the intrinsic differences in imaging mechanisms. This paper provides a systematic survey, comprehensive synthesis, and critical review of recent deep learning-based methods for VI-ReID. We categorize existing mainstream approaches into three major groups: (1) cross-modal net-

收稿日期:2025-09-13;录用日期:2025-12-05;责任编辑:覃怀银
*通讯作者:姜伟

work architecture design, which constructs specialized network structures to extract modality-invariant identity features, including dual-stream feature extraction networks, identity disentanglement modules, fine-grained feature alignment strategies, and architecture search-based designs; (2) generative learning methods, which seek to bridge the modality gap through modality translation or data augmentation, covering unidirectional or bidirectional image generation, construction of unified intermediate modalities, and feature-level generation and compensation techniques; (3) cross-modal similarity learning, which focuses on designing loss functions and metric learning schemes to pull together positive cross-modal pairs while pushing apart negative ones, primarily encompassing sample- or proxy-based contrastive learning and test-time optimized cross-modal re-ranking algorithms. Moreover, recognizing the high cost of annotation and the prevalence of noisy or incomplete labels in real-world applications, this survey further investigates advances under non-fully-supervised learning paradigms, systematically summarizing the unique challenges and representative solutions in noisy-label learning, semi-supervised learning, and unsupervised learning. To offer a holistic performance evaluation, we conduct unified comparisons and analyses of representative algorithms under different supervision settings on widely adopted public benchmarks—SYSU-MM01, RegDB, and LLCM. Finally, grounded in current technical bottlenecks, we outline promising future directions, including the development of more realistic and diverse datasets, mitigation of modality imbalance, lightweight model deployment, exploration of sustainable or lifelong learning mechanisms, and extension toward video-based or multi-source heterogeneous information-fused ReID. This survey aims to serve as a valuable theoretical reference and technical guide for future researchers in the field.

Key words: cross-modality person re-identification; pattern recognition; deep learning; representation learning; network architecture design; generative learning; intelligent surveillance

Foundation Item(s): Zhejiang Province Natural Science Foundation of China (No.LZ24F030004)

1 引言

行人重识别(Person Re-identification,ReID)作为智能视频监控领域的核心技术之一,其核心任务是在视野非重叠的摄像头网络中实现目标行人的精确检索与匹配[1]. 具体而言,给定待查询目标行人的单帧图像,行人重识别算法通过对不同摄像头采集的图像数据进行智能分析与比对,从非重叠视野的多摄像头网络中检索出与该查询图像属于同一身份的行人图像,从而实现目标行人在时空维度上的有效关联与轨迹重建. 这项技术在智能安防领域具有重要的应用价值,不仅为犯罪嫌疑人的跨摄像头追踪提供了技术支持,显著提高了案件侦破效率,同时在寻找失踪老人与儿童等民生领域也发挥着重要作用.

近年来,随着深度学习技术的快速发展,基于深度神经网络的解决方法在行人重识别领域取得了突破性进展. 深度学习方法通过构建端到端的特征学习框架,有效解决了传统方法在处理目标行人姿态变化、光照差异和视角多样性等问题时的局限性,显著提升了系统识别的准确率与鲁棒性. 早期的行人重识别研究主要集中在可见光波段的单模态识别问题上,相关算法的训练与测试主要依赖RGB图像数据集. 然而,在现实应用场景中,由于可见光摄像头在夜间或低照度环境下的成像质量显著下降,而红外摄像头能够在不依赖外部光源的条件下有效采集夜间行人信息,因此实际监控系统往往同时配备可见光和红外双模态成像设备.

为了应对双模态数据环境下的行人重识别需求,研究者提出了可见光-红外行人重识别(Visible-Infrared Person Re-identification,VI-ReID)这一新兴研究方向,帮助进行跨模态行人重识别任务. Wu等人[2]于2017年首次系统性地定义并阐述了跨模态行人重识别的任务范式与解决框架,同时构建并发布了大规模可见光-红外行人图像数据集,为后续研究奠定了基础. 如图1所示,VI-ReID的主要挑战在于实现可见光与红外模态图像间的交叉检索,即给定查询的可见光(或红外)行人图像,算法需要从另一模态的红外(或可见光)图像库中准确检索出具有相同身份的目标图像. 这一任务不仅继承了传统行人重识别中应对姿态变化、视角差异等挑战,还需要解决跨模态数据间的深度特征对齐与相似性度量等关键问题.

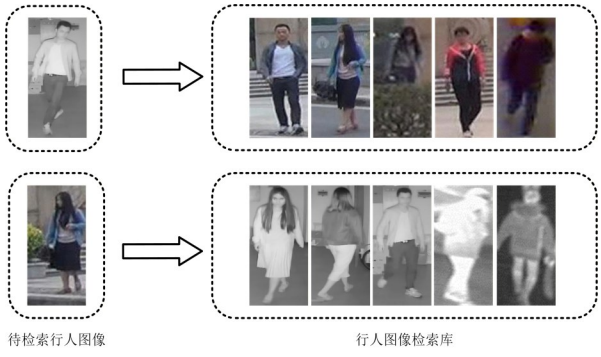


图1 跨模态行人重识别检索示例

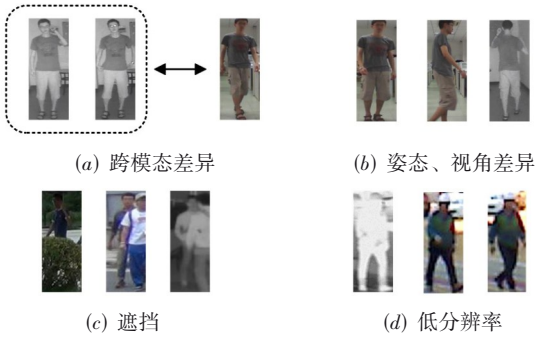
与传统的可见光行人重识别任务相比,跨模态行人重识别任务存在更大的困难与挑战,主要包括以下四个方面.

(1)跨模态差异. 由于可见光和红外图像天然存在着巨大差异,包括光照条件、相机色彩编码等信息的差异,导致同一行人在不同模态下的差异远大于不同行人在同一模态下的差异,使得行人聚类出错,如图2(*a*)所示.

(2)姿态、视角差异. 由于监控摄像头是在连续进行拍摄,并且拍摄的行人对象又是不断运动的,不同时刻拍下的行人图像会存在行人姿态的差异;监控摄像头往往存在多个角度拍摄行人,这会导致拍摄视角存在差异. 姿态、视角的差异会对特征对齐带来难度,如图2(*b*)所示.

(3)遮挡问题. 前景、物体或其他行人造成的遮挡,导致只有部分身体能够用于识别,模型首先需要判断未被遮挡的部分,并且通过未被遮挡的部分识别出正确的行人,如图2(*c*)所示.

(4)低分辨率问题. 在跨模态行人重识别问题中,许多图像是通过红外热成像手段拍摄的,这就造成成像质量较差、分辨率较低的问题,如图2(*d*)所示.



(*a*) 跨模态差异　(*b*) 姿态、视角差异　(*c*) 遮挡　(*d*) 低分辨率

图2　跨模态行人重识别面临的挑战

2　跨模态行人重识别研究方法

针对当前跨模态行人重识别的研究思路与方法设计,本文主要从以下四个分支对跨模态行人重识别研究方法进行详细综述.

(1)基于跨模态网络结构设计的方法. 针对不同模态的特点,需要对齐进行分支网络或专有学习模块的设计,以此来消除模态差异并提取身份不变特征. 该类网络结构设计方法可以细分为双流特征提取网络设计、身份信息解耦模块设计、细粒度特征对齐设计以及网络结构搜索设计.

(2)基于生成式学习的方法. 通过跨模态图像生成进行数据增广以及消除模态差异,包括单向生成图片、双向生成图片、中间模态生成以及特征级生成.

(3)基于跨模态相似度学习的方法. 该类方法关注跨模态学习中的损失函数设计以及相似度度量问题,主要可以细分为跨模态相似度对比学习、跨模态中心(代理)相似度对比学习以及跨模态样本相似度排序方法.

(4)区别于有监督学习范式,本文额外关注非完全有监督的跨模态行人重识别方法. 该部分主要介绍当标签错误或者缺失情况下的跨模态行人重识别方法,主要分为有监督学习范式、噪声标签学习范式、半监督学习范式以及无监督学习范式.

2.1　基于跨模态网络结构设计的方法

可见光与红外图像之间存在显著的模态差异,因此传统的单模态行人重识别方法难以有效处理跨模态数据,从而造成明显的精度下降. 研究者需要针对跨模态任务改进特征提取网络,以提取鲁棒的、模态不变且具有身份判别力的特征. 本节将总结并介绍基于跨模态网络结构设计的跨模态行人重识别方法,主要包括双流特征提取网络结构设计,其设计的双分支分别提取两模态的独立特征;身份信息解耦模块设计,通过设计特殊模块将身份信息从模态信息中分离出来;细粒度特征对齐,主要利用局部特征、姿态特征或辅助特征对跨模态图像进行细粒度对齐;网络结构搜索设计,通过神经网络自动搜索的方法,选择最适宜跨模态任务的网络.

2.1.1　双流特征提取网络结构设计

在单模态行人重识别中,主流方法通过一个单分支深度学习神经网络提取图像的深层特征表示;但在跨模态任务中,由于可见光与红外模态存在的巨大模态差异,使用传统行人重识别中的单分支网络难以提取有效的统一身份特征表示,因此,研究者首先针对跨模态行人重识别问题提出双流特征提取网络设计的结构. 卷积神经网络(Convolutional Neural Network,CNN)由于其强大的特征提取能力与高效的特征编码能力等优势,成为行人重识别技术的首选模型. 因此,早期的一些工作围绕双流卷积神经网络构建完成. Ye等人[3]提出使用独立双流学习网络来解决跨模态行人重识别问题,分类器使用参数共享的策略来学习跨模态共享身份特征. Feng等人[4]认为仅仅使用分类器共享分支会造成大量信息损失,因此在分类器阶段依然使用了模态独立分支学习独立特征. Ye等人[5]提出的MSTN方法认为过去在分类器阶段讨论模态共享与独立特征,实际上在基准卷积神经网络中间层就可以提取共享特征. 当浅层卷积网络学习模态独立的语义表达时,就可以使用深层网络共享的手段学习模态共享特征. 当前典型双流特征提取网络结构如图3所示,其在浅层网络通过独立参数提取到模态独立的表征信息,而在深层网络与分类器阶段采用共享参数结构以提取共享的身份信息. 为改进卷积神经网络受到的视野限制,Ye

等人[6]将非局部注意力(non-local)模块添加到双流学习网络中,使得网络摆脱局部视野的限制,同时基于此构建了跨模态行人重识别基线学习网络 AGW. Liu 等人[7]进一步探究了 Resnet50 网络中,模态共享卷积层的起始层数对识别精度的影响. Lu 等人[8]在设计模态独立分支和模态共享分支的基础上,提出了一种特征迁移算法,将模态独立特征和模态共享特征融合到一个统一的表示空间中,从而实现了模态独立特征和模态共享特征的有效结合.

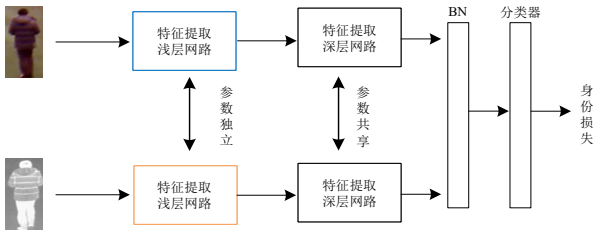


图3 典型双流特征提取网络结构示意图

基于卷积神经网络的残差网络模型(Resnet)在过去很长时间一直作为跨模态行人重识别的核心模型. 近年来,随着深度自注意力网络 Transformer 在视觉图像领域的广泛应用,一些学者开始基于 Vision Transformer 设计双流学习网络. Zhao 等人[9]在双流 ResNet 网络的每一个阶段后添加 Transformer 模块,用以学习长距离的位置联系. Jiang 等人[10]在使用双流卷积神经网络提取两模态图像特征后,利用 Transformer 的编码器与解码器以及互注意力机制对齐模态特征. Chai 等人[11]首次使用纯双流 Transformer 代替双流卷积神经网络,在图像输入时进行切片,双流 Transformer 浅层学习参数独立,深层共享学习参数. Transformer 在跨模态特征的学习仍有很大的潜力. Hua 等人[12]认为双流网络主要用于提取共性特征,却忽略了不同模态特有的语义信息,因此扩展双流网络为四分支网络,在每个模态中分出全局特征提取器与通道特征提取器,可以同时保留模态特有信息与共有信息进行学习. Qiu 等人[13]利用 CNN 捕获短程特征,使用 Transformer 捕获长程特征,通过高阶结构学习模块融合两种网络优势.

2.1.2 身份信息解耦模块设计

双流网络结构设计的主要原理是通过网络参数共享与独立来分离模态信息与身份信息的学习阶段,以提取更有效的身份信息表示. 另一种思路是设计身份信息解耦模块,以提升跨模态模型学习性能. Pu 等人[14]使用一种基于高斯变化的自动编码器,原本两个模态的特征重新映射到身份相关特征空间与身份无关特征空间. Kansal 等人[15]认为以前的方法忽略了频谱信息的干扰,因此使用了两个频谱提取分支单独提取频谱特征,目的是将身份信息和频谱信息解耦. Hao 等人[16]的方法刻意训练网络混淆网络对模态信息的敏感程度,即强迫网络不去区分不同模态,从反方向促进有效身份信息的提取. Ren 等人[17]使用两个深层模块分别提取模态共享特征和模态独立特征,通过训练模态混淆器与模态分类器使得两种特征得以分离. Wu 等人[18]使用单分支学习网络,但是基于通道注意力机制保护与身份相关的特征通道,使得身份信息被解耦的同时消除频谱信息带来的模态差异. Zhang 等人[19]基于实例归一化层(Instance Normalization, IN)进行特征解耦,设计了模态恢复模块与模态补偿模块从 IN 层移除的信息中提取身份信息与模态特征. Ding 等人[20]设计了一个层次化解耦模块,分别从解耦相机无关特征和模态不变身份特征,渐进式消除模态内差异与跨模态差异. 如图4所示,通过身份解耦模块可以使模型通过较小的参数量变化学习到模态无关的身份特征用于训练.

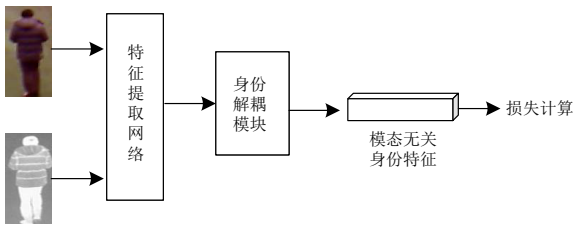


图4 身份信息解耦模块设计示意图

2.1.3 细粒度特征对齐设计

双流特征提取网络结构设计以及身份信息解耦模块设计主要针对全局特征进行设计,然而跨模态学习中的图像亦存在视角、姿态等差异,直接对齐全局特征容易导致共享信息提取产生偏差,因此一些研究者开始关注细粒度特征对齐方法. 该类方法主要包括局部特征对齐、姿态特征对齐、辅助特征对齐.

局部特征对齐. 相较于全局特征,局部特征可以更加关注到图像的不同细节区域. 传统的局部特征学习方法是将全局特征图进行水平切块,学习每一个水平块单独的特征表示,再联合地送入分类器中. Sun 等[21]在行人重识别问题中提出了基于水平切分的局部特征学习基准网络,在之后的工作中被广泛应用为局部特征学习基准网络. 除了水平切分的方法外,一些研究者着力寻找更具有身份判别力的局部特征图生成方法,同时针对跨模态行人重识别设计局部特征联合的学习方法. Zhang 等人[22]的方法使用特征图响应值更高的区域划分身体部分,对不同分割粒度的局部特征块与全局特征进行跨模态融合与联合学习. 针对局部特征不对齐的问题,Park 等人[23]基于可见光-红外的密集相似度建立了跨模态细粒度对齐手段. Kim 等人[24]通过混合两个模态的局部特征,促进跨模态特征局部对齐,同时避免模型过度依赖单一判别性部位. Ye 等人[25]改进

non-local模块,进一步设计了基于分块局部特征注意力的权重聚合模块以联合细粒度特征. Yang等人[26]联合了全局平均和的分块以及分块后的最大池化分支,增强了模态内的局部特征表达. Wang等人[27]将特征图沿通道维度分组,分别进行高度池化、宽度池化和3×3卷积,捕获不同维度的局部信息. 跨模态局部特征对齐学习的结构示意图如图5所示.

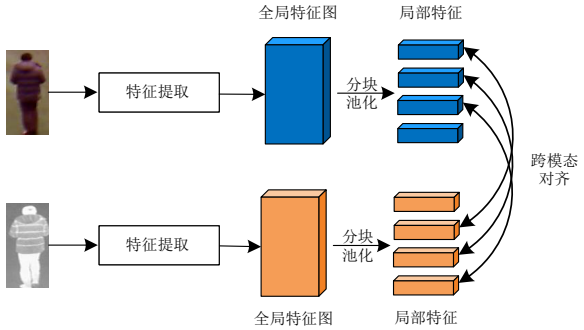


图5　典型局部特征对齐示意图

姿态特征对齐. 姿态估计器在行人重识别中有着关键作用,其针对复杂场景中的姿态变化能够有效提取模型的鲁棒性. Miao等人[28]的方法使用一个姿态估计器辅助人体特征的提取,因为姿态估计是不受模态影响的,所以这种方法更容易提取模态共享的局部特征. Chen等人[29]的方法使用姿态估计器提取结构性姿态特征,用Transformer学习跨模态结构特征之间的联系. Liu等人[30]利用人体关键点增强特征表示,从特征分布对齐和层次聚合两个角度减少模态差异. 孙锐等人[31]基于局部的异构聚合图卷积网络,构建人体拓扑图结构,使用节点向量同时学习模态内的结构信息与模态间关系信息,其将姿态特征转换为图结构问题并通过排列损失对齐跨模态的图节点. 姿态特征对齐学习示意图如图6所示,其与上段所述局部特征对齐的方法主要区别在于细粒度特征提取方式的差异.

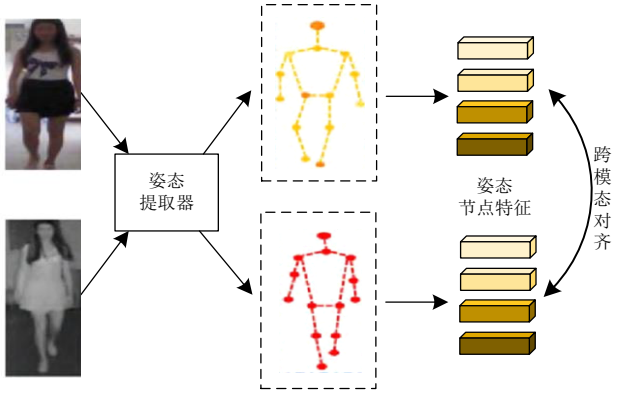


图6　典型姿态特征对齐示意图

辅助特征对齐. 一些研究者使用一些辅助的方法辅助细粒度特征的学习. 一类方法是利用手工标注的属性,Lin等人[32]提出使用属性标注的方法辅助学习行人的局部特征,但仅适用于可见光数据. Zheng等人[33]利用两个模态的属性标签并根据身份相关性为属性特征分配权重. 另一类方法是文本信息进行辅助学习,Du等人[34]利用行人特征的文本描述辅助特征提取与跨模态检索,通过红外图像与颜色相关文本信息结合消除模态差异,同时使用两个编码器分别提取纹理与颜色信息达到模态解耦目的,然而手工标注文本描述信息需要消耗大量人力. Yu等人[35]利用可学习的提示词模板构建模态特定语言描述,基于图文预训练模型CLIP(Contrastive Language - Image Pre-training)对齐语言与图像语义一致性. Hu等人[36]用微调后的大语言模型(Bootstrapping Language-Image Pre-training,BLIP)为可见光-红外图像生成描述,利用文本特征引导双模态图像特征的对齐.

2.1.4　网络结构搜索设计

目前大多数方法都是基于手工设计的深度学习网络提取模态不变的特征,这些方法大都依赖研究者的经验以及想法. Liu等人[7]通过手工设计比较了Resnet50双流网络各层在跨模态行人重识别问题的参数共享与独立策略,然而受限于人为设置的对照组,其比较策略具有局限性. Fu等人[37]在研究该问题时遍历了Resnet50结构中所有的53个BN层的独立与共享策略,显然,253个组合通过手工设计对照是不现实的,而神经架构搜索[38](Neural Architecture Search,NAS)的方法希望网络能够通过迭代的方法不仅学习最好的参数,同时寻找最优的网络结构. Fu等人[37]提出的跨模态神经网络架构搜索,让网络通过自我迭代的方法判断BN层的共享与独立策略,其基本结构如图7所示. Chen等人[39]受NAS方法的启发,提出了将特征图作为网络搜索的对象,以让网络通过自我迭代,寻找更有效的特征表示. 目前的网络结构搜索仅针对Resnet50做设计,但随着Transformer以及其他有效的骨干网络的提出,研究者可以尝试设计针对多骨干网络的网络结构搜索策略.

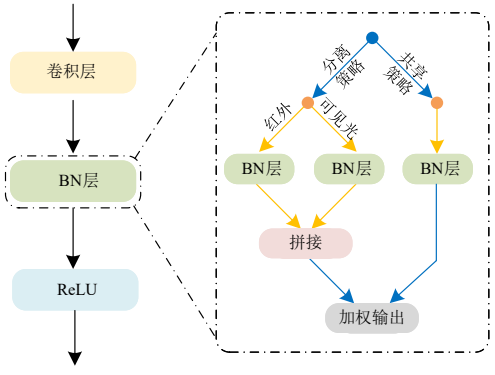


图7　跨模态网络结构搜索基本结构示例

综上，基于跨模态网络结构设计的方法通过不同的网络结构设计帮助网络学习与模态无关的鲁棒行人身份信息. 双流特征提取网络作为跨模态学习任务的主干网络现已被广泛使用并作为基线模型，主要的改进点在于对于输入的预处理、网络参数共享策略以及对不同网络架构(如Resnet、Transformer)的适应. 相较于双流网络会有较大的参数量提升，身份信息解耦模块的设计往往更加轻量化，通过更小的模块实现模态信息分离，一些工作仍使用单分支网络取得了不错的效果. 相较于一般视觉识别任务，行人重识别任务更加关注于细粒度的特征，包括局部特征、姿态特征、辅助特征等，通过细粒度特征的辅助能够提升身份判别特征的鲁棒性. 网络结构搜索设计对未来更复杂的网络设计有一定指导作用，避免了冗余的人工手动实验部分而自动寻找最优结构. 基于此，未来研究者可以参考这些角度进一步提升跨模态行人重识别学习精度：基于更好的视觉模型改进网络结构，学习更丰富的细粒度特征表示(如环境语义信息等)以及进一步挖掘模态信息分离手段(如从可见光红外频谱信息差异入手).

2.2 基于生成式学习的方法

多模态学习问题的另一个重要解决手段就是跨模态生成方法，随着生成式对抗网络(Generative Adversarial Networks，GAN)的发展，图像生成的方法也更加受到重视. 本节将综述基于生成式的跨模态学习方法，主要包括：单向生成，如将红外图像转换为可见光图像，或者将可见光图像转换为红外图像；双向生成，指将两模态的图像进行相互转换；中间模态，将两个模态的图像都转化为一个统一的中间模态图像，以此得到统一的特征表示；特征级生成，生成辅助特征，以此来避免图像生成带来的噪声干扰.

2.2.1 单向生成

单向生成的方法旨在将一个模态的图像转换为另一个模态的图像，以此消除模态差异性. 红外图像转换为可见光图像可以看作是红外图像的上色行为. Zhong等人[40]使用GAN将红外图像转化为可见光图像，同时通过约束使得身份信息保持不变. 基于这篇工作，Zhong等人[41]认为红外到可见光没有像素级对齐的训练集样本，因此将可见光生成灰度图像辅助红外图像的着色. 相比于上色行为带来的颜色信息对于身份信息的干扰，部分研究者认为可见光去颜色化的行为更为可靠. Wang等人[42]通过GAN将可见光转换为像素级对齐的伪红外图像后与真实红外图像进行特征级对齐.

Liu等人[43]认为直接从可见光到红外的GAN网络生成会对图像的结构造成损失，因此直接采用可见光生成与红外图像更相似的灰度图. 然而灰度图本质上仍是可见光波段的，其成像原理也与红外图像存在巨大差异. 因此，Liu等人[44]在一年后使用的模型中将灰度化的图像再使用GAN与红外图像进行相互生成以进一步消除生成图像之间的差异. 单向生成辅助训练的结构如图8所示，可见光图像通过生成器生成与真实红外图像相似的图像一起输入特征提取网络，由于输入图像模态一致，因此可以省去跨模态网络设计而使用通用网络学习，通过判别器的监督使得生成图像更为真实.

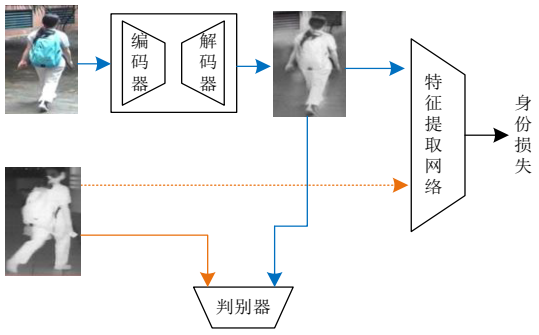


图8 可见光生成红外图像辅助训练结构示例

2.2.2 双向生成

跨模态单向生成一般的作用是由一个模态向另一个模态过渡，使用对抗损失监督这一过程. 而双向生成的方法通常将同一身份的两模态图像映射到一个统一空间，再由统一特征空间生成伪造的对应模态图像，而这个统一特征空间则用于提取身份不变特征. 张玉康等人[45]通过CycleGAN对可见光图像与红外图像进行相互迁移，同时构建了一个基于GAN的中间模态生成器辅助图像迁移. Wang等人[46]提出基于统一特征空间的跨模态双向图像生成的方法. 关于信息编码方式，不同研究者采用了不同的方式：Wang等人[47]用编码器编码模态独立特征与模态不变特征，再进行跨模态生成；Choi等人[48]改进了编码器编码身份信息以及姿态、光照信息，保留身份信息的同时交换姿态、光照等信息进行跨模态双向图像生成，然而这种解耦的生成方式会导致生成图像质量不佳. 典型的双向生成模型结构如图9所示，其可以在编码阶段对模态信息与身份信息进行解耦，在特征学习阶段，可以利用生成图像辅助消除模态差异，也可以利用解耦的身份特征直接训练特征学习网络.

2.2.3 中间模态

一些方法通过将两个模态图像转换为一个统一的中间模态(或者第三模态)的方式来消除模态差异. 中间模态图像表示一般介于两模态图像之间. Li等人[49]首先由可见光图像生成一个X模态的图像来辅助模态差异的消除，X模态介于红外与可见光之间，起到桥梁的作用. Ye等人[50]使用灰度图像作为中间模态的图像，虽然灰度图像确实和红外图像有差异，但其可以作为一种有效的中间模态. Zhang等人[51]的方法使用通道

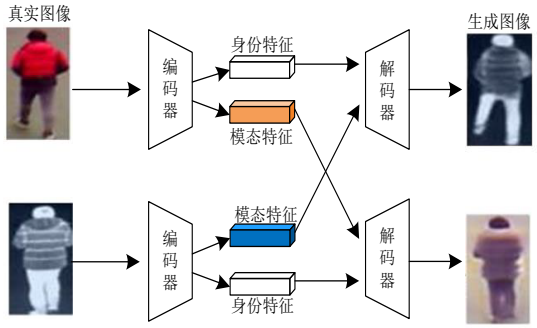


图9　双向生成模型结构示例

全连接器生成中间模态的图像. 相较于只通过可见光图像生成的灰度图像,这种方法更能生成一种统一的模态表示. Liu 等人[52]通过随机重组 RGB 的三通道灰度图生成5种新样式模态图像输入特征网络,以增强网络对不同模态图像的适应性. Ye 等人[53]将之改进为通道增强手段,同时提出通道随机剔除方法生成增强图像,在完全不引入新结构的情况下取得了较大的精度提升. Cui 等人[54]将图像转换到 HSV 空间,通过改变随机部分的亮度(V)、色相(H)以及饱和度(S)获得增强图像.

这些生成方法只做通道级别的改变,相比于 GAN 的图像质量更高,但依赖 CNN 提取有效特征表示. 中间模态图像生成示例如图 10 所示,由于其只在通道级进行变化,因此对于图像的细节破坏更少.

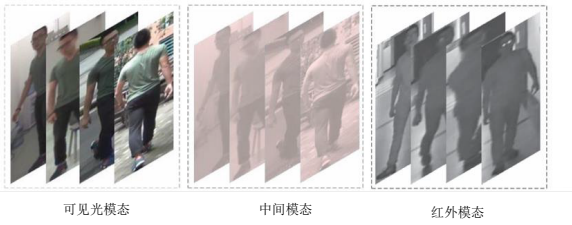


图10　中间模态生成示例

2.2.4　特征级生成

图像级的生成方法旨在生成跨模态或者中间模态的图像,从而达到数据增广及减小模态差异的目的,然而生成的图像常常面临分辨率不高以及噪声影响. 特征级生成的方法主要通过对深度学习网络提取的两个模态的特征向量进行融合、补偿、分离等操作,获得新的特征向量表示. Zhang 等人[55]使用编码器将两模态的特征分别分离为模态共享特征与模态独立特征,通过一个模态的共享特征生成另一个模态的独立特征,以此拉近两模态之间的距离. Zhang 等人[56]通过一个多样特征嵌入扩展模块分别由可见光和红外特征生成新的中间特征,生成的新的中间特征相比原特征在特征空间中距离更加靠近. 相似地,Yu 等人[57]通过两个模态内学习机以及一个跨模态学习机生成模态统一的中间特征用以辅助学习. Kim 等人[24]通过混合不同模态局部特征来生成增强样本,提高了对于跨模态共享局部特征的关注度. Liu 等人[58]通过在卷积层不同阶段插入随机通道变化以生成辅助特征,改善了红外模态相较于可见光数据更少的问题. Li 等人[59]设计了多特征生成模块,从原始特征生成一组分布紧密且多样化的辅助特征,以丰富行人表征,减少模态差异. 特征生成辅助训练结构示例如图 11 所示,不同的研究在中间特征生成方法上有所差异,生成的中间特征一定程度上起到桥梁作用,辅助跨模态对比学习. 特征级生成的方法减少了图像生成引入的噪声,达到了更高的识别精度.

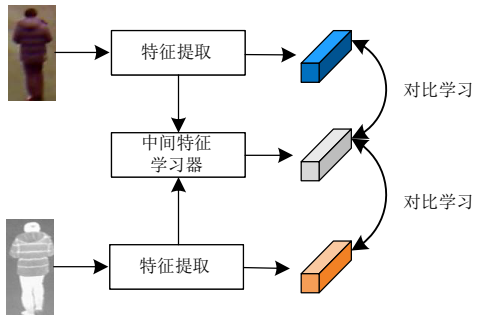


图11　特征生成辅助训练结构示例

综上,相较于跨模态网络结构设计的方法通过网络结构的设计克服模态差异,基于生成式的方法更加关注从样本处理层面来获得模态统一的表征,其优势在于表观上获得了克服模态差异的样本用于学习,并且增广了训练数据. 不足之处在于对于生成图像的质量要求较高,由于 ReID 任务对于细粒度特征的要求较高,而早期基于 GAN 网络的生成方式其图像质量普遍偏低,阻碍了精确度的提升. 通道层面图像增强以及特征级增强一定程度缓解了生成图像质量不佳带来的问题. 未来研究与提升可以考虑利用更好的生成网络,如 Diffusion 结构等作为跨模态图像生成网络,生成质量更好的辅助图像. 另一方面,生成网络与识别网络的端到端统一模型也是值得研究的方向.

2.3　基于跨模态相似度学习的方法

传统网络设计的方法主要通过分类器输出计算交叉熵损失,本节介绍基于跨模态相似度学习的方法,聚焦于模型提取到的特征之间的相似度度量,以此对跨模态特征进行损失计算或度量排序. 该场景中研究者通常需要联合模态内与跨模态的相似度来缓解不同程度的模态内与跨模态差异. 本节主要分为:跨模态样本相似度对比学习,主要基于样本间相似度计算对比损失;跨模态中心(代理)相似度对比学习,主要基于中心或构建代理计算对比损失;跨模态样本相似度排序方法,主要在测试阶段基于样本相似度改进排序算法.

2.3.1 跨模态样本相似度对比学习

样本相似度对比学习旨在通过拉近正样本之间的距离及推开负样本间距离构建损失,在跨模态行人重识别中,还需要综合考虑模态内与模态间的差异. Ye等人[3]在双流学习网络的基础上提出了跨模态匹配的模型,这个模型通过对比损失同时优化模态共享和模态独立模块,其设计了双向的对比损失,分别计算从可见光与红外模态样本出发的跨模态对比损失[60]. Wu等人[61]通过对比学习减小模态之间的差异性,以学习跨模态图像之间的相似性.

三元组损失是对比损失的一种推广形式,希望在拉近正样本对的同时推开负样本对. 在跨模态行人重识别中,来自不同模态的统一身份的行人特征向量应聚合在一起. Ye等人[60]基于跨模态问题设计了一种双向双约束的跨模态三元损失,通过将样本之间的距离替换为样本与样本中心的距离对该损失进行了优化[62]. Ye等人[63]提出了在超球面计算三元损失,即通过约束特征向量之间的角度来避免特征向量模的影响. Cai等人[64]提出的DCTL方法通过跨模态难样本挖掘的三元损失来挖掘难样本以及减少计算复杂度. Hao等人[65]的HSME方法将特征向量映射到一个统一的超球面上,使得身份分类界限更加明晰,并基于角度差异进行对比学习减少向量长度差异带来的影响,其使用基于三元损失改进的双向排序损失来同时解决跨模态与模态内差异. 四元组损失在三元组损失的基础上加入了负样本对之间的相对约束,给正负样本对之间的距离一个第三参考系. Jia等人[66]提出的方法通过图推理的方式优化了四元组的学习.

2.3.2 跨模态中心(代理)相似度对比学习

样本对比学习在实例级别对模型进行约束,但易受到难样本或噪声的影响,不一定能直接对应到相应的身份特征学习. 因此另一种思路是构建身份对应的中心特征或代理,通过拉近正样本与中心的距离学习到对应身份特征. 在此情景下,如何联合两模态构建的中心成为研究者方法设计的重点. Liu等人[67]通过分类器输出构建模态内代理,通过计算跨模态样本到代理之间的单向对比损失更加显式地拉近减小模态差异. Zhu等人[68]通过拉近跨模态正样本中心之间的距离来减少类内跨模态变化. Hao等人[16]改进中心损失使得跨模态样本不严格聚集到中心,而是保持合理距离避免强制聚合中心可能导致的过拟合,同时基于相机中心对比损失增强相机域判别性. Ye等人[62]提出了一种双向中心约束统一跨模态与模态内的中心对比损失,能够同时应对模态内与模态间的差异. Liu等人[7]提出了一种基于中心到中心的三元损失,来代替传统的样本到样本的三元损失,以解决噪声样本的影响以及降低计算量. 周非等人[69]通过角度度量计算跨模态的异构中心三元损失. Wang等人[70]通过拉近样本与其对应中心的距离构建中心聚合损失,同时通过跨模态的异构中心三元组损失缓解模态间差异. Kong等人[71]提出了动态混合中心的混合两模态聚类中心作为优化的动态参考点,指导其提出的混合模态特征的空间分布.

跨模态样本相似度对比学习与跨模态中心(代理)相似度对比学习方法如图12所示,其中相似度或距离度量的选择可以有多种(如欧氏距离、余弦距离等),跨模态样本对比学习总体目标是拉近正样本对,同时推开负样本对. 对于中心(代理)相似度学习,其损失构建逻辑不变,而中心或代理的构建手段在不同研究方法中有所不同.

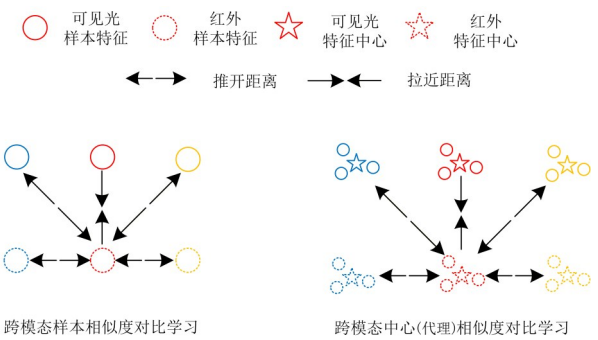


图12 跨模态样本相似度对比学习与跨模态中心(代理)相似度对比学习示例(不同颜色表征不同行人身份)

2.3.3 跨模态样本相似度排序方法

行人重识别任务在测试时通过一般比较查询样本与库样本特征之间的欧氏距离来对查询结果进行排序,一些研究者针对查询过程提出了基于Jaccard距离与局部查询扩展的重排序算法[72],优化了查询结果. 在跨模态行人重识别中,由于查询样本与库样本的模态差异,导致其距离分布与单模态场景存在较大差异,研究者也针对跨模态任务改进了测试排序方法. Jia等人[66]提出跨模态相似性推理度量,为改进测试阶段仅通过特征相似度计算的距离,其基于图推理与跨模态互最近邻样本计算查询样本与库样本之间的距离,优化了测试排序方法. Liang等人[73]基于跨模态K互邻样本集计算跨模态样本间的Jaccard距离,与距离加权组合构成测试时的相似度度量距离. 针对传统K互逆重排序算法中跨模态正样本占比低以及模态间距离分布不一致的问题,Du等人[34]设计了跨模态K互逆重排序算法,通过扩展策略整合原始邻样本与跨模态邻样本,并通过模态感知局部查询扩展在局部扩展时考虑了跨模态邻居,得到了更优的跨模态Jaccard距离计算方法. 区别于前两节的跨模态相似度学习方法,本节的跨模态样本相似度排序方法并不直接用于训练阶段的损失

计算,而是优化跨模态检索时的距离度量方法,使得测试阶段的排序结果更优.

综上,基于跨模态样本相似度对比学习的方法从特征相似度角度设计跨模态行人重识别学习方法,其基本思想是提升相同身份特征之间相似度,并降低不同身份特征的相似度.其中的特征表征包括样本特征或者中心(代理)特征.跨模态样本相似度对比学习以及中心(代理)相似度对比学习通常与特征提取方法结合使用,以作为模型学习目标约束手段.跨模态任务的损失函数设计应平衡模态内差异与跨模态差异学习的权重,为了应对较大的跨模态差异,研究者一般更着重于跨模态相似度的对比学习.跨模态样本相似度排序方法主要用于测试阶段的排序优化,能够有效提升测试精度.未来研究者可以从统一跨模态与模态内的相似度度量算法,简化损失函数计算策略的角度进行考量.

2.4　非完全有监督的跨模态行人重识别方法

跨模态行人重识别往往要求比可见光单模态行人重识别更多的训练样本,然而现实中对于跨模态行人重识别数据的标注更加困难.首先,在海量数据集中人工寻找到出现在不同时间地点特定身份的行人本来就是耗费时间的事情;其次,对于某些没有颜色信息的红外图像,可能标注者也难以区分其身份而导致无法标注的情况出现.针对这些问题,研究者提出应对于不同场景的不同学习范式的跨模态行人重识别方法,大致分为4类:有监督学习范式、噪声标签学习范式、半监督学习范式以及无监督学习范式.当获得两个模态图像的正确标签时,其对应于有监督学习范式,如图13(*a*)所示,这也是之前所介绍的大部分方法使用的学习范式.下面将着重介绍其他三种非完全有监督的学习范式.

(*a*) 有监督学习范式　(*b*) 噪声标签学习范式

(*c*) 半监督学习范式　(*d*) 无监督学习范式

图13　不同监督范式下的跨模态行人重识别

2.4.1　噪声标签学习范式

如图13(*b*)所示,当在两个模态中出现部分错误标签时,模型的性能会受到不正确标签的影响而下降.Yang等人[74]首次将此问题归纳为跨模态双噪声问题,不同于单模态的噪声问题,双噪声问题不仅带来模态内标签的不一致性,还会导致跨模态匹配的不一致性,从而影响跨模态不变特征的学习.由于噪声样本的分类损失一般会明显大于正确样本,因此他们提出的DART模型使用双模型联合建模,并通过高斯混合模型预测噪声样本的概率,对噪声样本与正确样本进行分类分别计算不同的度量损失,如图14所示.之后,Yang等人[75]在此基础上进行改进,根据同质性对的距离和标签信息调整优化目标,使得损失函数能够灵活地适应不同的三元组组合.Zhang等人[76]具体化了真实世界噪声的定义,基于样本特性生成了更接近真实场景的噪声,提出了一种鲁棒混合损失函数,将数据分为干净、模糊以及明显噪声子集并分别进行处理.

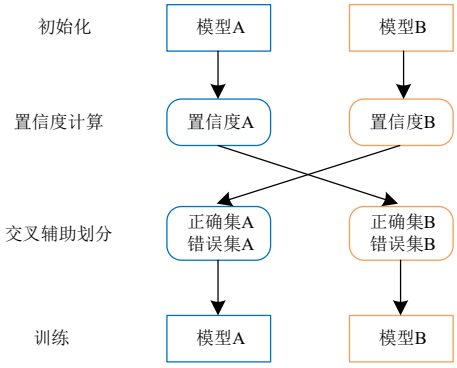


图14　双模型噪声预测训练模式示例

2.4.2　半监督学习范式

由于在实际应用中,可见光相机拍摄的照片一般会多于红外相机,并且红外相机的标注难度也更高,因此仅存在可见光相机标签的半监督学习范式被提出.该学习范式的关键在于将已有的可见光行人标签迁移到红外图像上.Huang等人[77]通过无监督训练的扩散模型,利用高通图和低通图分别保留轮廓与模态信息,生成跨模态的红外图像,以此来获得有标签的红外图像进行进一步训练.然而由于扩散模型生成的红外图像的细节质量欠佳,该方法识别准确度较低.Wang等人[78]基于最优运输策略(Optimal Transport,OT)将可见光标签迁移至红外标签,该方法原理是基于可见光样本与红外样本的特征相似度,其中相似度更高的跨模态样本的标签一致性可能性更大.Shi等人[79]在此基础上利用高斯混合模型估计迁移标签噪声的概率,并加入相应的惩罚项进行约束.Zheng等人[80]通过动态生成中间模态特征弥合跨模态特征差异,并通过基于置信度的加权身份损失,减少红外伪标签噪声带来的负面影响.

总体来说,半监督学习范式与噪声学习范式的方法具有一定的相似性与共通性,但由于模型相较于完

全有监督学习效果更差,并且假设条件相较于无监督学习方法更多,因此研究者更加偏向于无监督学习范式的研究.

2.4.3 无监督学习范式

相对于有标签的学习任务,在实际应用中存在大量的无标签行人数据,因此不依赖人工标签的无监督跨模态行人重识别也是值得研究的重要问题. 模态无监督 ReID 面临巨大的模态差异带来的困难,同时无法通过身份相同的跨模态图像对消除模态差异. 相比于半监督学习,无监督跨模态行人重识别无法确定两个模态的行人身份数量,导致其具有更大的难度. Liang 等人[73]和 Yang 等人[81]提出了先在模态内部使用单模态无监督学习,为图像添加伪标签,再进行跨模态学习的方法,不过这些方法需要利用有标签的单模态可见光数据集做预训练. 孙锐等人[82]提出了一个将大型无标签单模态数据集预训练模型迁移到跨模态数据集的自监督学习框架. 对于不依赖外部数据集的无监督跨模态行人重识别方法,Yang 等人[83]提出了两阶段的聚类加聚合方法 ADCA 框架,在单模态聚类完成后,通过跨模态特征相似度聚合跨模态的中心,在训练过程中动态更新聚类中心,该方法不再依赖可见光数据集的预训练模型. Wu 等人[84]在 ADCA 基础上,利用渐进式的图匹配方法寻找跨模态的相同身份特征中心,提高了跨模态聚类中心匹配的准确度. Li 等人[85]使用跨模态相机中心以及跨模态双向匹配的方法,进一步消除模态内以及模态间的差异. Shi 等人[86]利用距离聚类中心较远的困难样本构成困难聚类中心,对应于有监督的难样本挖掘,基于困难聚类中心的聚合学习取得了较好的效果. Pang 等人[87]将聚类硬标签转换为软标签,进一步减少了噪声对标签结果的影响. 典型两阶段跨模态无监督行人重识别训练框架如图 15 所示,第一阶段会进行单模态训练样本的聚类与对比学习,然后第二阶段基于单模态聚类结果进行模态聚合匹配,并进行跨模态对比学习. 当前研究者研究重点主要在于如何构建更优的模态聚合匹配模块以提高跨模态匹配精度. 无监督跨模态行人重识别难点在于模态内的精确聚类以及跨模态正标签对的匹配问题,同时,如何有效区分困难样本以及噪声聚类样本,也是研究者需要面对的问题.

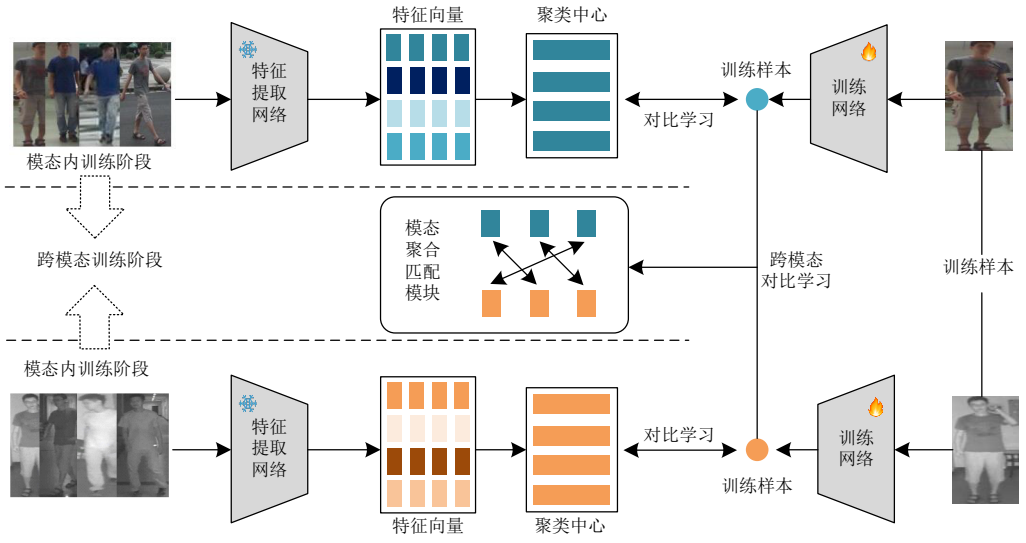


图 15 典型两阶段跨模态无监督行人重识别训练框架

综上,本节总结的非完全有监督的跨模态行人重识别方法是面对实际应用中标签错误或缺失问题的解决方法,对应研究相对较少,具备未来长期研究价值. 目前研究者设计非完全有监督学习范式数据集的方法主要通过改变有监督数据集标签达成,如随机改变样本标签、剔除有监督标签的方法,且这些方法尚存在改进空间:(1)随机改变标签与实际中的标签错误可能存在差距,研究者可以针对性采集新的适合非完全有监督学习范式相对应的数据集;(2)提出统一的学习基线与进一步规范评价准则.

3 典型算法比较

3.1 数据集

目前公认的跨模态行人重识别数据集主要包含可见光-近红外双模态数据集 SYSU-MM01[1]、可见光-热红外双模态数据集 RegDB[88]以及低光照环境下可见光-近红外数据集 LLCM[56],其中部分图片如图 16 所示.

SYSU-MM01 数据集[1]是 2017 年为研究跨模态行

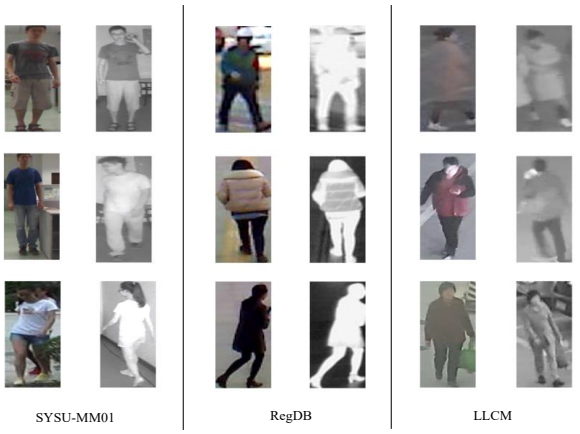


图16　跨模态行人重识别数据集

人重识别问题而提出的公开数据集,包含了在白天的4个用于捕捉RGB图像的摄像视域以及在黑暗环境中用于捕捉红外图像的摄像视域,共有491个不同身份行人的287 628张RGB图像和15 792张红外图像. 在测试时,遵循原始的数据集评测协议,随机划分查询集合和候选集合并重复10次,计算最终的平均得分. 其难点在于姿态、视角、遮挡等问题. 由于其提出较早,包含场景多样,现已成为跨模态行人重识别中最为常用的数据集.

RegDB数据集[88]共有412个不同身份的行人,分为254个女性和158个男性,每个人分别对应10张RGB图像和10张红外图像,其中拍摄到156个行人的正面,256个行人的背面. 该数据集总共有4 120张RGB图像和4 120张的红外图像,难点在于低分辨率等问题.

LLCM数据集[56]为最新提出的低光照的跨模态数据集. 使用可见光和红外摄像头各9个,拍摄了1 064个不同身份行人的46 767张图像. 该数据集包含了光照变化、遮挡、视角变化、低分辨率等多种挑战,不过目前还未得到广泛应用.

3.2　评价指标

3.2.1　CMC

累计匹配特性(Cumulated Matching Characteristics, CMC):对于每一张查询图像,分别判断与查询图像属于同一人的候选图像是否包含在排序列表的前k项中,并计算查询集合中所有图像的排序列表前k项中包含属于同一个人的正确匹配的比率. 对于查询集合中的第i个查询图像q_i,其排序列表的前k个样本中是否包含正确匹配项可以用式(1)表示:

$$r(q_i,k)=\begin{cases}1, & \text{如果前}k\text{个候选样本包含查询图像}\\ 0, & \text{其他}\end{cases} \tag{1}$$

常见的k的取值有1、5、10和20等,得到rank-k准确率,即CMC(k)的值,表示排序列表的前k项中包含正确匹配的概率.

3.2.2　mAP

每一个查询样本,计算其准确率-召回率曲线下面的面积,即平均精度(Average Precison, AP),然后求查询集合中所有图像的平均精度的均值,得到平均精度均值(mAP). 平均精度可以用式(2)表示:

$$\mathrm{AP}=\frac{1}{M}\sum_{i\in\{i_1,\,i_2,\cdots,i_M\}}\frac{M_i}{i} \tag{2}$$

其中,M是候选集合中与查询图像属于同一个人的所有正确匹配的总数,M_i是候选排序列表前i个样本中包含正确匹配的数量,$i\in\{i_1,\,i_2,\cdots,i_M\}$是$M$个正确匹配在排序列表中的位置索引. 此时mAP可以表示为

$$\mathrm{mAP}=\frac{1}{N}\sum_{j=1}^{N}\mathrm{AP}(j) \tag{3}$$

其中,N是查询集合中的样本总数,AP(j)是第j个查询样本对应的平均精度,而mAP则是最终的平均精度均值.

另外,Ye等人[6]还提出了反向负样本惩罚均值(mean Inverse Negative Penalty, mINP)指标,用于衡量行人重识别算法检索出最难的正确匹配样本的能力,作为另一个行人重识别算法的补充评价指标,不过尚未得到广泛的应用.

3.3　算法比较

本文统计了近年来发表于顶级期刊、会议的比较有代表性的跨模态行人重识别方法,并统一在公认度较高的数据集上比较其精确度. 根据其训练时的监督范式,将结果分为四类:有监督学习范式、标签噪声学习范式、半监督学习范式及无监督学习范式.

3.3.1　完全有监督学习范式

在完全有监督学习范式中,训练阶段两个模态数据集的行人标签均完整提供,我们使用rank-1、mAP指标在SYSU数据集、RegDB数据集以及LLCM数据集上进行精度比较.

在SYSU-MM01数据集测试中,按照第二节的分类标准将测试结果所用方法也分为基于跨模态网络结构设计的方法、基于生成式学习的方法以及基于跨模态相似度学习的方法. 结果如表1所示,我们对每类方法的每个指标最优结果进行了加粗. 对于所有方法的精确度,在全搜索模式下,rank-1、mAP的最高值为84.90%、81.74%;而在室内搜索模式下,rank-1最高值为89.06%,mAP最高值为89.37%. 注意这里的指标均是在single shot模式下统计得到的.

表1 SYSU-MM01数据集上完全有监督学习范式实验结果比较 单位:%

分类	方法	全搜索模式		室内搜索模式		方法概述	发表
		Rank-1	mAP	Rank-1	mAP		
基于跨模态网络结构设计的方法	MSTN[5]	51.64	50.11	57.35	64.79	双流网络	TIP′20
	AGW[6]	47.50	47.65	54.17	62.97	双流网络+non-local	TMAMI′21
	cm-SSFT[8]	61.60	63.20	70.50	72.60	双流网络	CVPR′20
	TransVI[11]	71.36	68.63	77.40	81.31	双流网络	TCSVT′23
	HOS-Net[13]	75.60	74.20	84.20	86.70	双流网络	AAAI′24
	MSCMNet[12]	78.53	74.20	83.00	85.54	四分支网络	PR′25
	IDKL[17]	81.42	79.85	87.14	**89.37**	身份解耦+模态混淆	CVPR′24
	MPANet[18]	70.58	68.24	76.74	80.95	身份解耦+局部特征	CVPR′21
	DMiR[89]	50.54	49.29	53.92	62.49	身份解耦+域适应	TCSVT′22
	FSDCC[90]	68.79	65.72	75.32	78.58	身份解耦	KNOSYS′22
	HDNet[20]	82.12	77.62	85.63	92.20	层次身份解耦	ACM MM′25
	DDAG[25]	54.75	53.02	61.02	67.98	局部特征对齐	ECCV′20
	PartMix[24]	77.78	74.62	81.52	84.38	局部特征对齐	CVPR′23
	MSCLNet[91]	76.99	71.64	78.49	81.17	局部特征对齐	ECCV′22
	DDFC[27]	78.08	73.78	82.12	84.79	局部特征对齐	Neurocomputing′25
	SGIEL[92]	77.12	72.33	82.07	82.95	姿态特征对齐	CVPR′23
	SPOT[29]	65.34	62.25	69.42	74.63	姿态+Transformer	TIP′22
	TVI-LFM[36]	**84.90**	**81.74**	**89.06**	88.39	文本特征辅助	NIPS′24
	CSDN[35]	76.70	73.00	84.50	86.80	文本特征辅助	TMM′25
	CM-NAS[37]	61.99	60.02	—	—	网络结构搜索	CVPR′21
基于生成式学习的方法	AlignGAN[42]	42.40	40.70	45.90	54.30	单向生成	ICCV′19
	GECNet[41]	53.37	51.83	60.60	62.89	单向生成	TCSVT′22
	AGMNet[44]	69.63	66.11	74.68	78.30	单向生成	JSTSP′23
	D2RL[46]	28.90	29.20	—	—	双向生成	CVPR′19
	JSIA[47]	38.10	36.90	43.80	52.90	双向生成	AAAI′20
	Hi-CMD[48]	34.94	35.94	—	—	双向生成	CVPR′20
	TSME[93]	64.23	61.21	64.80	71.53	双向生成+灰度图	TCSVT′22
	MMN[51]	70.60	66.90	76.20	79.60	中间模态	ACM MM′21
	CAJ[53]	69.88	66.89	76.26	80.37	中间模态+通道增强	ICCV′21
	SMCL[94]	67.39	61.78	68.84	75.56	中间模态	ICCV′21
	STAR[95]	76.07	72.73	83.47	85.76	中间模态	TMM′23
	DMA[54]	74.57	70.41	82.85	85.10	中间模态+局部增强	TIFS′24
	FMCNet[55]	66.34	62.51	68.15	74.09	特征级生成	CVPR′22
	MUN[57]	76.24	73.81	79.42	82.06	特征级生成	ICCV′23
	DEEN[56]	74.70	71.80	80.30	83.30	特征级生成	CVPR′23
	CoMix[58]	74.62	70.22	80.55	82.11	特征级生成	TSMCS′25
	PDM[59]	**79.30**	**76.30**	**88.70**	**89.80**	特征级生成	ICASSP′25
基于跨模态相似度学习的方法	BDTR[60]	17.01	19.66	—	—	双向样本三元损失	IJCAI′18
	HSME[65]	20.68	23.12	—	—	超球面三元损失	AAAI′19
	expAT[63]	38.57	38.61	—	—	余弦三元损失	TIP′21
	DTCL[64]	54.14	54.14	—	—	难样本三元损失	KBS′21
	eBDTR[62]	27.82	28.42	—	—	双向中心损失	IFS′20
	HCTL[7]	**61.68**	57.51	63.41	64.26	中心三元损失	TMNF′21
	MAUL[67]	61.59	**59.96**	**67.07**	**73.58**	代理对比损失	CVPR′22
	HMML[96]	63.63	60.44	—	—	混合损失	ACM TMCCA′22

续表

分类	方法	全搜索模式		室内搜索模式		方法概述	发表
		Rank-1	mAP	Rank-1	mAP		
	SIM[66]	57.47	53.75	—	—	测试重排序	IJCAI′20

注:加粗数字表示每个指标最优结果.

可以看到,最优方法的思路依然是使用跨模态网络结构设计,尤其是近年来通过更好的身份解耦模块设计使得跨模态学习精度有了很大的提升,同时局部特征对齐等模块也对跨模态学习有较大帮助. 生成式学习的方法中,其中基于GAN网络的单向生成和双向生成的方法普遍表现不佳,可能是由于GAN生成的质量会对细粒度特征产生较大损害,基于灰度图与通道增强构成的中间模态以及特征级生成的方法则取得了更好的效果. 基于跨模态相似度学习的方法中,早期的单纯基于损失函数改进的方法取得的效果较差,近年来通过结合更优模型设计的方法取得了一定的精度提升,其中基于三元组的跨模态中心损失方法效果较好,该类方法通常与网络结构设计或者生成式方法结合以取得更好的效果.

RegDB数据集测试精度如表2所示,可见光到红外模式下,rank-1最高值为96.66%,mAP最高值为91.22%. 红外到可见光模式下,rank-1最高值为96.30%,mAP最高值为91.21%. 不同分类方法在RegDB数据集上的表现优劣与SYSU-MM01数据集上基本一致,不再复述. 大多数跨模态行人重识别方法在SYSU-MM01数据集上的精度低于RegDB数据集,这是因为RegDB数据集虽然清晰度更低,但缺少姿态、视角的变化,因此特征更为对齐. 而SYSU-MM01数据集中包含了相机视角、姿态、遮挡等多重影响特征对齐的因素.

表2　RegDB数据集上完全有监督学习范式实验结果比较　　单位:%

分类	方法	可见光到红外		红外到可见光		方法概述	发表
		Rank-1	mAP	Rank-1	mAP		
基于跨模态网络结构设计的方法	MSR[4]	48.43	70.32	79.95	48.67	双流网络	TIP′19
	AGW[6]	70.05	66.37	—	—	双流网络+non-local	TMAMI′21
	cm-SSFT[8]	72.30	72.90	71.00	71.70	双流网络	CVPR′20
	TransVI[11]	**96.66**	**91.22**	**96.30**	**91.21**	双流网络	TCSVT′23
	MSCMNet[12]	90.40	81.20	87.70	78.20	四分支网络	PR′25
	HOS-Net[13]	94.70	90.40	93.30	89.20	双流网络	AAAI′24
	IDKL[17]	94.72	90.19	94.22	90.43	身份解耦+模态混淆	CVPR′24
	MPANet[18]	83.70	80.90	82.80	80.70	身份解耦+局部特征	CVPR′21
	HDNet[20]	92.19	81.08	87.48	77.92	层次身份解耦	ACM MM′25
	DDAG[25]	69.34	63.46	68.06	61.80	局部特征对齐	ECCV′20
	AMC-Net[96]	91.21	81.61	89.03	79.85	局部特征学习	Neurocomputing′21
	PartMix[24]	84.93	82.52	85.66	82.27	局部特征对齐	CVPR′23
	DDFC[27]	92.34	86.87	91.16	87.43	局部特征对齐	Neurocomputing′25
	SGIEL[92]	91.07	85.23	92.18	86.59	姿态特征对齐	CVPR′23
	SPOT[29]	80.35	72.46	79.37	72.26	姿态+Transformer	TIP′22
	CSDN[35]	95.40	87.70	98.00	85.50	文本特征辅助	TMM′25
	CM-NAS[37]	84.54	80.32	82.57	78.31	网络结构搜索	CVPR′21
基于生成式学习的方法	AlignGAN[42]	57.90	53.60	56.30	53.40	单向生成	ICCV′19
	GECNet[41]	82.33	78.45	78.93	75.58	单向生成	TCSVT′22
	AGMNet[44]	88.40	81.45	85.34	81.19	单向生成	JSTSP′23
	JSIA[47]	48.10	48.90	48.50	49.30	双向生成	AAAI′20
	Hi-CMD[48]	70.93	66.04	—	—	双向生成	CVPR′20
	MMN[51]	91.60	84.10	87.50	80.50	第三模态	ACM *MM*′21
	CAJ[53]	85.03	79.14	84.75	77.82	中间模态+通道增强	ICCV′21
	CPN[52]	51.29	49.37	—	—	中间模态+通道增强	APIN′22
	STAR[95]	94.09	**88.75**	**93.30**	**88.20**	中间模态	TMNF′23
	DMA[54]	93.30	88.34	91.50	86.80	中间模态+局部增强	TIFS′24

续表

分类	方法	可见光到红外		红外到可见光		方法概述	发表
		Rank-1	mAP	Rank-1	mAP		
	FMCNet[55]	89.12	84.43	88.38	83.86	特征级生成	CVPR′22
	DEEN[56]	91.10	85.10	89.50	83.40	特征级生成	CVPR′23
	MUN[57]	**95.19**	87.15	91.86	85.01	特征级生成	ICCV′23
	CoMix[58]	91.23	85.99	88.78	85.45	特征级生成	TSMCS′25
基于跨模态相似度学习的方法	BDTR[60]	33.47	31.83	—	—	双向样本对比损失	IJCAI′18
	HSME[65]	50.85	47.00	—	—	超球面三元损失	AAAI′19
	expAT[63]	44.71	32.20	—	—	余弦三元损失	TIP′21
	DTCL[64]	30.56	32.45	—	—	难样本三元损失	KBS′21
	eBDTR[62]	31.83	33.18	—	—	双向中心损失	IFS′20
	MAUL[67]	83.39	78.75	81.07	78.89	代理对比损失	CVPR′22
	DGTL[92]	83.90	73.78	63.11	69.20	样本+中心三元损失	SPL′21
	HCTL[7]	**91.05**	**83.28**	**89.30**	**81.46**	中心三元损失	TMNF′21
	LCCRF[97]	80.97	79.92	79.27	77.69	样本余弦三元损失	WWW′22
	HMML[96]	82.97	77.56	—	—	混合对比损失	TMCCA′22
	SIM[66]	60.88	56.93	—	—	测试重排序	IJCAI′20

注:加粗数字表示每个指标最优结果.

在LLCM数据集上测试精度如表3所示,可见光到红外模式下,rank-1最高为64.3%,mAP最高值为66.6%. 红外到可见光模式下,rank-1最高值为70.2%,mAP最高值为65.8%. 由于LLCM是2023年新提出的数据集,其应用范围并不广泛,且其分辨率普遍较低,并且相比其他数据集包含了更多的姿态、视角、相机以及服装的变化,目前的方法在该数据集上的准确率相比其他数据集有较为明显的下降,说明目前跨模态行人重识别方法在多数据集上的稳定性亟待提高.

表3 LLCM数据集上完全有监督学习范式实验结果比较 单位:%

方法	可见光到红外		红外到可见光		方法概述	发表
	Rank-1	mAP	Rank-1	mAP		
AGW[6]	43.6	51.8	51.5	55.3	双流网络+non-local	CVPR′21
MSCMNet[12]	**64.3**	**66.6**	56.5	63.5	四分支网络	PR′25
MRCN[18]	51.3	58.3	65.3	49.5	身份解耦	AAAI′23
DDAG[25]	40.3	48.4	48.0	52.3	局部特征对齐+图注意力	ECCV′20
LBA[23]	43.8	53.1	50.8	55.6	局部特征对齐	SPL′21
CM2GT[98]	52.1	58.3	65.9	50.3	局部特征对齐	PR′25
CAJ[53]	48.8	56.6	56.5	59.8	中间模态+通道增强	ICCV′21
MMN[51]	52.5	58.9	59.9	62.7	中间模态	ACM *MM*′21
DEEN[56]	54.9	62.9	62.5	**65.8**	特征级生成	TNNLS′23
FDNM[99]	56.6	62.7	**70.2**	55.8	频域特征生成	Arxiv′24

注:加粗数字表示每个指标最优结果.

3.3.2 标签噪声学习范式

在标签噪声学习范式中,训练集的标签被添加随机百分比的噪声. 以SYSU-MM01数据集为例,在实验中使用0%、20%、50%的噪声,结果如表4所示,其中DART和LCNL是针对标签噪声学习范式设计的方法. 可以发现,针对完全有监督学习范式设计的方法在标签存在噪声时有很大幅度下降,并且在细粒度特征上有更多设计的方法精度下降更多,如LbA. 而DART以及LCNL通过双模型交叉验证以及高斯混合模型区分噪声样本的设计,在不同噪声比例下均取得了较为鲁棒的结果.

3.3.3 半监督学习范式

在半监督学习范式中,训练集仅有一个模态具有标签(研究者通常假设为可见光模态具有标签),而另一模态标签缺失. 以SYSU-MM01数据集为例,对比结果如表5所示. 基于置信度引导机制的MUCG[81]方法取得了最优的结果,半监督学习方法的关键在于将有

标注的模态图像标签迁移到无标注模态同时避免引入噪声,由于具有单模态正确标签并可以忽略聚类步骤,其精度相比于无监督学习范式略高. 然而由于实际应用中该学习范式较少出现(即仅有单模态标签),因此对应研究方法较少,当前更多研究者转为研究无监督学习范式.

表 4　SYSU-MM01 数据集上标签噪声学习范式实验结果比较　　单位:%

方法	0% 噪声		20% 噪声		50% 噪声		方法简述	发表
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP		
DDAG[25]	54.8	53.0	14.6	14.0	6.7	7.5	局部特征对齐	ECCV′20
AGW[6]	47.5	47.7	17.7	18.2	7.9	9.8	双流网络+non-local	TPAMI′21
LbA[23]	55.4	54.1	9.9	10.2	2.7	4.2	损失函数设计	ICCV′21
MPANet[18]	**70.6**	**68.2**	21.6	21.2	7.0	8.2	注意力+局部特征学习	CVPR′21
CAJ[53]	69.9	66.9	25.4	23.7	8.0	10.8	通道数据增强	ICCV′21
DART[74]	68.7	66.3	66.3	64.1	60.3	58.7	第三模态	CVPR′23
LCNL[75]	70.2	68.0	**67.2**	**64.9**	**62.4**	**59.8**	局部特征学习	Comput Vis′24

注:加粗数字表示每个指标最优结果.

表 5　SYSU-MM01 数据集上半监督学习范式实验结果比较　　单位:%

方法	全搜索模式		室内搜索模式		简述	发表
	Rank-1	mAP	Rank-1	mAP		
VI-Diff[77]	24.69	24.97	30.88	40.54	扩散模型生成	Arxiv′23
OTLA[78]	48.20	43.90	47.40	56.80	最优传输策略	ECCV′22
DPIS[79]	58.40	55.60	63.00	70.00	最优传输+噪声标签优化	ICCV′23
MUCG[80]	**68.80**	**65.90**	**77.40**	**81.00**	置信度引导机制	MM′24

注:加粗数字表示每个指标最优结果.

3.3.4　无监督学习范式

在无监督学习范式中,训练集中的行人图像均不具有身份标签,同时也没有跨模态的行人配对信息. 我们统计了近年来的跨模态无监督行人重识别算法结果,如表6所示,可以发现近年来无监督跨模态行人重识别围绕跨模态聚类中心匹配、聚类标签优化提出了多种算法,也取得了明显的精度提升. 尽管如此,无监督跨模态行人重识别仍然有许多需要考虑与改进的问题,例如如何平衡难样本与错误标签的关系,进一步优化跨模态匹配的策略. 另外,现阶段的无监督跨模态行人重识别主要通过去除有监督学习数据集(如SYSU-MM01)中的标签得到数据集,其中大部分无监督学习方法均假设两模态的数据是均衡且身份相同的,然而实际应用中两模态的身份数可能无法对应,因此应在

表 6　SYSU-MM01 数据集上无监督学习范式实验结果比较　　单位:%

方法	全搜索模式		室内搜索模式		简述	发表
	Rank-1	mAP	Rank-1	mAP		
H2H[73]	30.15	29.40	-	-	两阶段无监督训练	TIP′21
OTLA[78]	29.98	27.13	29.80	38.80	最优传输策略	ECCV′22
ADCA[83]	45.51	42.73	50.60	59.11	跨模态中心聚合	MM′22
CHCR[100]	47.72	45.34	50.12	42.17	多频谱聚类	TCSVT′23
DOTLA[101]	50.36	47.36	53.47	61.73	邻域标签优化	MM′23
MBCCM[102]	53.14	48.16	55.21	61.98	跨模态中心匹配策略	MM′23
PGM[84]	57.27	51.78	56.23	62.74	跨模态中心匹配策略	CVPR′23
GUR[103]	63.51	61.63	71.11	76.23	相机信息+伪标签迁移	ICCV′23
SCA-RCP[85]	51.41	48.52	56.77	64.19	跨模态中心匹配	TKDE′24
PCLMP[86]	64.40	58.70	69.50	74.40	动态难样本中心学习	Arxiv′24
MULT[104]	65.03	58.62	65.35	66.60	伪标签迁移	Arxiv′24
SDCL[105]	64.49	63.24	**71.37**	76.90	多层次特征联合	CVPR′24
ASM[87]	**65.07**	**63.37**	71.08	**76.91**	软标签动态更新策略	ICCV′25

注:加粗数字表示每个指标最优结果.

设计无监督学习数据集中考虑这一因素.

总体来说,非完全有监督类方法当前的研究相较于有监督方法偏少,且精度差距较大(无监督方法与有监督方法的最优mAP差距在20%左右),因此未来研究者可以将研究重心更多转移至非完全有监督跨模态行人重识别.

4 未来研究方向

4.1 丰富跨模态行人重识别数据集

尽管跨模态行人重识别的识别精度相对于其刚提出时有较为明显的提升,但目前可用数据集依然较少,且缺乏多样性问题,如遮挡、姿态、服饰变化、年龄、场景变化等问题. 当前大部分跨模态行人重识别方法专注于缓解跨模态差异,而忽略了模态内可能产生的变化. 一方面,这些方法在提取全局特征或局部特征时往往忽略了遮挡、姿态变化等因素;另一方面,在数据集设计上也缺少这样的样例图片. 目前广泛应用的SYSU-MM01数据集主要为室内红外图像与校内的学生图像,并非对应实际街道监控视频;另一广泛使用的RegDB数据集缺少遮挡、姿态等变化;LLCM数据集采集了低光照下的图像并提供了街道行人图像,然而其单一行人姿态和场景变化较少,缺少跨摄像头数据的采集. 未来设计者可以考虑针对性地采集更多样化的数据集,应对这些实际应用中的变化.

4.2 模态不平衡问题研究

在研究跨模态问题时,两模态的数据量以及数据分布均是不平衡的,网络往往会侧重于某一个模态的学习而忽略了另一模态. 一方面,由于预训练的ImageNet网络使用的大部分是可见光图像,因此其对红外图像的解析能力会更差;另一方面,由于夜间行人的活动较少,因此跨模态行人重识别数据集的可见光图像明显多于红外图像. Xia等人[106]在其论文中首次提出了两模态不平衡的问题,并通过生成跨模态图像来试图缓解不平衡问题. Liu等人[68]针对网络偏向于可见光图像的学习提出对于红外模态的图像做更多的数据增强手段. 研究者可进一步研究模态部分缺失或完全缺失的跨模态行人重识别场景.

4.3 网络轻量化与数据集压缩

网络轻量化一直是深度学习模型在实际应用中需要考虑的问题,特别是需要部署在嵌入系统中进行快速识别的模型. 当前研究者的思路主要围绕增加各种辅助网络模块以提升识别精度,从而忽略了网络的轻量化问题. 随着监控行人待检索数据规模的增加,如何在不损失过多精度的情况下提升识别速度,或者设计使用更加轻量化网络骨干架构,值得研究者关注. 同时,数据集压缩有助于网络在节省存储空间与计算开支并同时保护行人隐私的情况下学习到身份辨识能力,在CIFAR、ImageNet等分类数据集上应用较为广泛. 数据集压缩对于行人细粒度特征破坏较大,如何针对行人特征设计有效的压缩方法是未来研究方向之一.

4.4 可持续的跨模态行人重识别

持续学习是ReID模型部署时需要考虑的关键问题,当前跨模态行人重识别测试方法大多采用源域学习、源域测试的方法,而模型部署后,需要面对持续采集的新目标域数据,如何通过类增量方法学习目标域知识同时避免对源域数据产生灾难性遗忘,是可持续的跨模态行人重识别需要考虑的问题. Chen等人[107]提出无监督可见光-红外跨域学习方法,基于聚类中心扩展实现模型在新域上的持续学习,但未考虑到对于旧数据集的遗忘问题. Xing等人[108]通过存储回放数据的方法在每个训练阶段训练存储的部分旧域样本,达到抗遗忘的终身学习目的,然而其存放的旧域数据偏多并且随着训练阶段增加其回放所需内存逐渐增大,同时存储大量回放样本也会给模型带来隐私方面的问题. 未来研究方向包括减少对于回放数据的依赖,通过数据集风格学习或者数据集压缩等手段避免对于旧域数据的遗忘.

4.5 视频跨模态行人重识别

目前主流的跨模态行人重识别均关注图像到图像的匹配,然而实际应用过程中视频监控数据反而占据主导位置,因此基于视频的跨模态行人重识别需要得到进一步重视. 视频相对于图像数据包含了更多的时间以及空间信息,可以从一个人的动作以及步伐等信息综合判断跨模态行人的身份. Lin等人[109]构建了HITSZ-VCM数据集,如图17所示,包括12个可见光以及红外摄像头采集到的927个行人的视频数据集,基于视频中的图像序列构建双流特征学习网络. Li等人[110]生成每帧图像的浮雕图像作为中间模态,通过双向的时间-空间信息挖掘学习到模态不变的信息. 视频跨模态行人重识别更加有利于海量监控数据的学习.

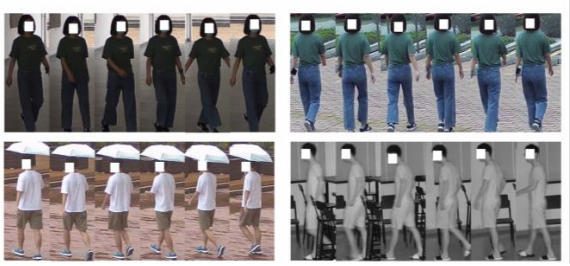


图17 视频跨模态行人重识别数据集

4.6 多模态行人重识别

跨模态行人重识别旨在通过可见光和红外图像跨模态搜索寻找到特定行人,然而在现实场景中,例如对

于犯罪嫌疑人的搜索中,可能还需要更多模态的信息,例如文字描述信息、深度图像信息,甚至素描图像信息. 这些信息组合构成了多模态行人重识别问题.

Hafner等人[111]提出使用深度相机拍摄的照片与可见光图像之间的跨模态行人重识别问题,深度图像包含的人物信息相比红外模态要更少,因此该识别问题挑战更大. Liu等人[112]提出了基于图像与文字的跨模态行人重识别问题,使用BERT(Bidirectional Encoder Representations from Transformers)对文字进行特征提取,并与图像一同输入到卷积神经网络映射到高维空间. Zhai等人[113]提出了基于文字描述信息、素描图像以及可见光图像的三模态行人重识别,如图18所示. 对于每一种模态的信息,使用了单独的编码器将其映射到高维的可以对齐的特征空间中.

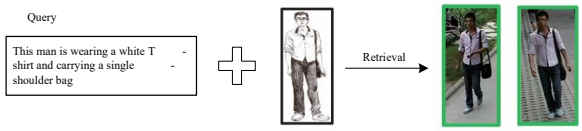


图18　多模态行人重识别示例

未来研究者可以尝试构建多模态行人重识别系统,包含文字信息、素描图像、可见光图像、红外图像以及人脸图像. 如何寻找多模态统一的特征表示是该系统建立的难点,研究者可以考虑结合发展迅速的多模态大模型进行下游微调训练作为其中一个研究思路.

5　结论

行人重识别是计算机视觉领域的一个热门研究课题,而深度学习的发展极大地促进了该领域的研究. 由行人重识别引申的跨模态行人重识别问题近年来受到广泛关注. 本文总结了跨模态行人重识别问题所遇到的难点,将研究方法分为基于跨模态网络结构设计的方法、基于生成式学习的方法以及基于跨模态相似度学习的方法,还进一步总结了非完全有监督学习范式下的跨模态行人重识别算法研究进展.

基于跨模态网络结构设计的方法通过设计适用于跨模态行人重识别的网络结构来缓解模态差异,主要包括双流特征提取网络、身份信息解耦模块、细粒度特征对齐设计以及网络结构搜索. 基于生成式学习的方法旨在通过图像生成达到模态不变特征提取以及数据增广的目的,包括单向生成、双向生成、中间模态等方法. 基于跨模态相似度学习的方法针对跨模态特征相似度设计合适的度量函数,包括跨模态样本相似度对比学习以及跨模态中心(代理)相似度对比学习,另外也介绍了用于优化测试度量的跨模态样本相似度排序方法. 总结了三种非完全有监督学习范式,包括噪声标签学习范式、半监督学习范式以及无监督学习范式,这些学习范式的研究开展较少,其精度有待提升.

在公开数据集上,集中比较了当前具有代表性的跨模态行人重识别方法在不同学习范式下的识别精度,对所有方法采用相同的评价准,集中比较有助于后来研究者判断更有价值的研究方向. 最后,对未来的研究方向做出了可能的展望,包括模态不平衡问题、可持续跨模态行人重识别问题等,希望能对其他研究者有所帮助.

参考文献

[1] 罗浩, 姜伟, 范星, 等. 基于深度学习的行人重识别研究进展[J]. 自动化学报, 2019, 45(11): 2032-2049.
LUO H, JIANG W, FAN X, et al. A survey on deep learning based person re-identification[J]. Acta Automatica Sinica, 2019, 45(11): 2032-2049. (in Chinese)

[2] WU A C, ZHENG W S, YU H X, et al. RGB-infrared cross-modality person re-identification[C]//2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 5390-5399.

[3] YE M, LAN X Y, LI J W, et al. Hierarchical discriminative learning for visible thermal person re-identification[C]//AAAI'18/IAAI'18/EAAI'18: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence. New York: ACM, 2018: 7501-7508.

[4] FENG Z X, LAI J H, XIE X H. Learning modality-specific representations for visible-infrared person re-identification[J]. IEEE Transactions on Image Processing, 2020, 29: 579-590.

[5] YE M, LAN X Y, LENG Q M, et al. Cross-modality person re-identification via modality-aware collaborative ensemble learning[J]. IEEE Transactions on Image Processing, 2020, 29: 9387-9399.

[6] YE M, SHEN J B, LIN G J, et al. Deep learning for person re-identification: A survey and outlook[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(6): 2872-2893.

[7] LIU H J, TAN X H, ZHOU X C. Parameter sharing exploration and hetero-center triplet loss for visible-thermal person re-identification[J]. IEEE Transactions on Multimedia, 2021, 23: 4414-4425.

[8] LU Y, WU Y, LIU B, et al. Cross-modality person re-identification with shared-specific feature transfer[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 13376-13386.

[9] ZHAO J Q, WANG H Z, ZHOU Y, et al. Spatial-channel enhanced transformer for visible-infrared person re-identi-

fication[J]. IEEE Transactions on Multimedia, 2023, 25: 3668-3680.

[10] JIANG K Z, ZHANG T Z, LIU X, et al. Cross-modality transformer for visible-infrared person re-identification[C]// Computer Vision - ECCV 2022. Cham: Springer, 2022: 480-496.

[11] CHAI Z H, LING Y G, LUO Z M, et al. Dual-stream transformer with distribution alignment for visible-infrared person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(11): 6764-6776.

[12] HUA X C, CHENG K, LU H, et al. MSCMNet: Multi-scale semantic correlation mining for visible-infrared person re-identification[J]. Pattern Recognition, 2025, 159: 111090.

[13] QIU L X, CHEN S, YAN Y, et al. High-order structure based middle-feature learning for visible-infrared person re-identification[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(5): 4596-4604.

[14] PU N, CHEN W, LIU Y, et al. Dual Gaussian-based variational subspace disentanglement for visible-infrared person re-identification[C]//Proceedings of the 28th ACM International Conference on Multimedia. New York: ACM, 2020: 2149-2158.

[15] KANSAL K, SUBRAMANYAM A V, WANG Z, et al. SDL: Spectrum-disentangled representation learning for visible-infrared person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(10): 3422-3432.

[16] HAO X, ZHAO S Y, YE M, et al. Cross-modality person re-identification via modality confusion and center aggregation[C]//2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2022: 16383-16392.

[17] REN K J, ZHANG L. Implicit discriminative knowledge learning for visible-infrared person re-identification[C]// 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 393-402.

[18] WU Q, DAI P Y, CHEN J, et al. Discover cross-modality nuances for visible-infrared person re-identification[C]// 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 4328-4337.

[19] ZHANG Y K, YAN Y, LI J, et al. MRCN: A novel modality restitution and compensation network for visible-infrared person re-identification[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(3): 3498-3506.

[20] DING H, SUN J, LONG R, et al. Visible-infrared person re-identification based on feature decoupling and refinement[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2025, 21(9): 1-16.

[21] SUN Y F, ZHENG L, YANG Y, et al. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)[C]//Computer Vision - ECCV 2018. Cham: Springer, 2018: 501-518.

[22] ZHANG L Y, DU G D, LIU F, et al. Global-local multiple granularity learning for cross-modality visible-infrared person reidentification[J]. IEEE Transactions on Neural Networks and Learning Systems, 2025, 36(3): 4209-4219.

[23] PARK H, LEE S, LEE J, et al. Learning by aligning: Visible-infrared person re-identification using cross-modal correspondences[C]//2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2022: 12026-12035.

[24] KIM M, KIM S, PARK J, et al. PartMix: Regularization strategy to learn part discovery for visible-infrared person re-identification[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 18621-18632.

[25] YE M, SHEN J B, CRANDALL D J, et al. Dynamic dual-attentive aggregation learning for visible-infrared person re-identification[C]//Computer Vision - ECCV 2020. Cham: Springer, 2020: 229-247.

[26] YANG X, LIU H L, WANG N N, et al. Bidirectional modality information interaction for Visible-Infrared Person re-identification[J]. Pattern Recognition, 2025, 161: 111301.

[27] WANG R, PI D C, YU R, et al. Dimension-driven feature complementation for visible-infrared person re-identification[J]. Neurocomputing, 2025, 653: 131162.

[28] MIAO Y Q, HUANG N C, MA X, et al. On exploring pose estimation as an auxiliary learning task for visible-Infrared Person re-identification[J]. Neurocomputing, 2023, 556: 126652.

[29] CHEN C Q, YE M, QI M B, et al. Structure-aware positional transformer for visible-infrared person re-identification[J]. IEEE Transactions on Image Processing, 2022, 31: 2352-2364.

[30] LIU M, SUN Y Q, WANG X P, et al. Pose-guided modality-invariant feature alignment for visible-infrared object re-identification[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 5017610.

[31] 孙锐, 张磊, 余益衡, 等. 基于局部异构聚合图卷积网络的跨模态行人重识别[J]. 电子学报, 2023, 51(4): 810-825.
SUN R, ZHANG L, YU Y H, et al. Cross-modality person re-identification based on locally heterogeneous polymerization graph convolutional network[J]. Acta Electronica Sinica, 2023, 51(4): 810-825. (in Chinese)

[32] LIN Y T, ZHENG L, ZHENG Z D, et al. Improving person re-identification by attribute and identity learning[J]. Pattern Recognition, 2019, 95: 151-161.

[33] ZHENG A H, FENG M Y, PAN P, et al. Attributes based visible-infrared person re-identification[C]//Pattern Recognition

and Computer Vision. Cham: Springer, 2022: 254-266.

[34] DU Y H, ZHAO Z C, SU F. YYDS: Visible-infrared person re-identification with coarse descriptions[EB/OL]. (2024-03-07)[2025-09-30]. https://arXiv.org/abs/2403.04183.

[35] YU X Y, DONG N, ZHU L H, et al. CLIP-driven semantic discovery network for visible-infrared person re-identification[J]. IEEE Transactions on Multimedia, 2025, 27: 4137-4150.

[36] HU Z Y, YANG B, YE M. Empowering visible-infrared person re-identification with large foundation models[C]// Advances in Neural Information Processing Systems 37. San Diego: NeurIPS, 2024: 117363-117387.

[37] FU C Y, HU Y B, WU X, et al. CM-NAS: Cross-modality neural architecture search for visible-infrared person re-identification[C]//2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2022: 11803-11812.

[38] ZOPH B, LE Q V. Neural architecture search with reinforcement learning[EB/OL]. (2017-02-15) [2025-09-30]. https://arXiv.org/abs/1611.01578.

[39] CHEN Y, WAN L, LI Z H, et al. Neural feature search for RGB-infrared person re-identification[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 587-597.

[40] ZHONG X, LU T Y, HUANG W X, et al. Visible-infrared person re-identification via colorization-based Siamese generative adversarial network[C]//Proceedings of the 2020 International Conference on Multimedia Retrieval. New York: ACM, 2020: 421-427.

[41] ZHONG X, LU T Y, HUANG W X, et al. Grayscale enhancement colorization network for visible-infrared person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(3): 1418-1430.

[42] WANG G A, ZHANG T Z, CHENG J, et al. RGB-infrared cross-modality person re-identification via joint pixel and feature alignment[C]//2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2020: 3622-3631.

[43] LIU H J, MA S, XIA D X, et al. SFANet: A spectrum-aware feature augmentation network for visible-infrared person reidentification[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(4): 1958-1971.

[44] LIU H J, XIA D X, JIANG W. Towards homogeneous modality learning and multi-granularity information exploration for visible-infrared person re-identification[J]. IEEE Journal of Selected Topics in Signal Processing, 2023, 17(3): 545-559.

[45] 张玉康, 谭磊, 陈靓影. 基于图像和特征联合约束的跨模态行人重识别[J]. 自动化学报, 2021, 47(8): 1943-1950. ZHANG Y K, TAN L, CHEN J Y. Cross-modality person re-identification based on joint constraints of image and feature[J]. Acta Automatica Sinica, 2021, 47(8): 1943-1950. (in Chinese)

[46] WANG Z X, WANG Z, ZHENG Y Q, et al. Learning to reduce dual-level discrepancy for infrared-visible person re-identification[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 618-626.

[47] WANG G A, ZHANG T Z, YANG Y, et al. Cross-modality paired-images generation for RGB-infrared person re-identification[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12144-12151.

[48] CHOI S, LEE S, KIM Y, et al. Hi-CMD: Hierarchical cross-modality disentanglement for visible-infrared person re-identification[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 10254-10263.

[49] LI D G, WEI X, HONG X P, et al. Infrared-visible cross-modal person re-identification with an X modality[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(4): 4610-4617.

[50] YE M, SHEN J B, SHAO L. Visible-infrared person re-identification via homogeneous augmented tri-modal learning[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 728-739.

[51] ZHANG Y K, YAN Y, LU Y, et al. Towards a unified middle modality learning for visible-infrared person re-identification[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM, 2021: 788-796.

[52] LIU J C, SONG W R, CHEN C H, et al. Cross-modality person re-identification via channel-based partition network[J]. Applied Intelligence, 2022, 52(3): 2423-2435.

[53] YE M, RUAN W J, DU B, et al. Channel augmented joint learning for visible-infrared recognition[C]//2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2022: 13547-13556.

[54] CUI Z Y, ZHOU J H, PENG Y X. DMA: Dual modality-aware alignment for visible-infrared person re-identification[J]. IEEE Transactions on Information Forensics and Security, 2024, 19: 2696-2708.

[55] ZHANG Q, LAI C Z, LIU J N, et al. FMCNet: Feature-level modality compensation for visible-infrared person re-identification[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 7339-7348.

[56] ZHANG Y K, WANG H Z. Diverse embedding expan-

sion network and low-light cross-modality benchmark for visible-infrared person re-identification[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 2153-2162.

[57] YU H, CHENG X, PENG W, et al. Modality unifying network for visible-infrared person re-identification[C]//2023 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2024: 11151-11161.

[58] LIU H J, GU J Y, LI Z Y, et al. CoMix: Collaborative mixed learning via style fuzzy normalization for visible-infrared person re-identification[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2025, 55(11): 8572-8586.

[59] LI J R, ZHEN Q, YANG Y L, et al. Prototype-driven multi-feature generation for visible-infrared person re-identification[C]//ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2025: 1-5.

[60] YE M, WANG Z, LAN X Y, et al. Visible thermal person re-identification via dual-constrained top-ranking[C]//Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. California: AAAI, 2018: 1092-1099.

[61] WU A C, ZHENG W S, GONG S G, et al. RGB-IR person re-identification by cross-modality similarity preservation[J]. International Journal of Computer Vision, 2020, 128(6): 1765-1785.

[62] YE M, LAN X Y, WANG Z, et al. Bi-directional center-constrained top-ranking for visible thermal person re-identification[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 407-419.

[63] YE H R, LIU H, MENG F Y, et al. Bi-directional exponential angular triplet loss for RGB-infrared person re-identification[J]. IEEE Transactions on Image Processing, 2021, 30: 1583-1595.

[64] CAI X, LIU L, ZHU L, et al. Dual-modality hard mining triplet-center loss for visible infrared person re-identification[J]. Knowledge-Based Systems, 2021, 215: 106772.

[65] HAO Y, WANG N N, LI J, et al. HSME: Hypersphere manifold embedding for visible thermal person re-identification[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 8385-8392.

[66] JIA M X, ZHAI Y P, LU S J, et al. A similarity inference metric for RGB-infrared cross-modality person re-identification[EB/OL]. (2020-07-03)[2025-09-30]. https://arXiv.org/abs/2007.01504.

[67] LIU J L, SUN Y F, ZHU F, et al. Learning memory-augmented unidirectional metrics for cross-modality person re-identification[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 19344-19353.

[68] ZHU Y X, YANG Z, WANG L, et al. Hetero-Center loss for cross-modality person re-identification[J]. Neurocomputing, 2020, 386: 97-109.

[69] 周非, 舒浩峰, 白梦林, 等. 生成对抗网络协同角度异构中心三元组损失的跨模态行人重识别[J]. 电子学报, 2023, 51(7): 1803-1811.
ZHOU F, SHU H F, BAI M L, et al. Cross-modal person re-identification based on generative adversarial network coordinated with angle based heterogeneous center triplet loss[J]. Acta Electronica Sinica, 2023, 51(7): 1803-1811. (in Chinese)

[70] WANG X J, CHEN C Q, ZHU Y, et al. Feature fusion and center aggregation for visible-infrared person re-identification[J]. IEEE Access, 2022, 10: 30949-30958.

[71] KONG J, HE Q B, JIANG M, et al. Dynamic center aggregation loss with mixed modality for visible-infrared person re-identification[J]. IEEE Signal Processing Letters, 2021, 28: 2003-2007.

[72] ZHONG Z, ZHENG L, CAO D L, et al. Re-ranking person re-identification with k-reciprocal encoding[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 3652-3661.

[73] LIANG W Q, WANG G C, LAI J H, et al. Homogeneous-to-heterogeneous: Unsupervised learning for RGB-infrared person re-identification[J]. IEEE Transactions on Image Processing, 2021, 30: 6392-6407.

[74] YANG M X, HUANG Z Y, HU P, et al. Learning with twin noisy labels for visible-infrared person re-identification[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 14288-14297.

[75] YANG M X, HUANG Z Y, PENG X. Robust object re-identification with coupled noisy labels[J]. International Journal of Computer Vision, 2024, 132(7): 2511-2529.

[76] ZHANG R H, CAO Z, HUANG Y, et al. Visible-infrared person re-identification with real-world label noise[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2025, 35(5): 4857-4869.

[77] HUANG H, HUANG Y, WANG L. VI-diff: Unpaired visible-infrared translation diffusion model for single modality labeled visible-infrared person re-identification[EB/OL]. (2023-10-06)[2025-9-30]. https://arXiv.org/abs/2310.04122.

[78] WANG J M, ZHANG Z Z, CHEN M G, et al. Optimal transport for label-efficient visible-infrared person re-identification[M]//Computer Vision - ECCV 2022. Cham: Springer, 2022: 93-109.

[79] SHI J M, ZHANG Y C, YIN X B, et al. Dual pseudo-labels interactive self-training for semi-supervised visible-

infrared person re-identification[C]//2023 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2024: 11184-11194.

[80] ZHENG X Y, ZHANG Y K, LU Y, et al. Semi-supervised visible-infrared person re-identification via modality unification and confidence guidance[C]//Proceedings of the 32nd ACM International Conference on Multimedia. New York: ACM, 2024: 5761-5770.

[81] YANG B, CHEN J, MA X Z, et al. Translation, association and augmentation: Learning cross-modality re-identification from single-modality annotation[J]. IEEE Transactions on Image Processing, 2023, 32: 5099-5113.

[82] 孙锐, 谢瑞瑞, 张磊, 等. 基于灾难性遗忘及组合叠加擦除的跨模态行人重识别预训练方法[J]. 电子学报, 2023, 51(10): 2925-2935.
SUN R, XIE R R, ZHANG L, et al. Cross-modal pedestrian re-identification pre-training method based on catastrophic forgetting and combination superimposed erasure[J]. Acta Electronica Sinica, 2023, 51(10): 2925-2935. (in Chinese)

[83] YANG B, YE M, CHEN J, et al. Augmented dual-contrastive aggregation learning for unsupervised visible-infrared person re-identification[C]//Proceedings of the 30th ACM International Conference on Multimedia. New York: ACM, 2022: 2843-2851.

[84] WU Z S, YE M. Unsupervised visible-infrared person re-identification via progressive graph matching and alternate learning[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 9548-9558.

[85] LI Z Y, LIU H J, PENG X T, et al. Inter-intra modality knowledge learning and clustering noise alleviation for unsupervised visible-infrared person re-identification[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(8): 3934-3947.

[86] SHI J M, YIN X B, ZHANG Y C, et al. Learning commonality, divergence and variety for unsupervised visible-infrared person re-identification[EB/OL]. (2024-10-24) [2025-10-10]. https://arxiv.org/abs/2402.19026.

[87] PANG Z Q, WANG C Y, ZHAO L L, et al. Augmented and softened matching for unsupervised visible-infrared person re-identification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2025: 11100-11109.

[88] NGUYEN D T, HONG H G, KIM K W, et al. Person recognition system based on a combination of body images from visible light and thermal cameras[J]. Sensors, 2017, 17(3): 605.

[89] HU W P, LIU B H, ZENG H T, et al. Adversarial decoupling and modality-invariant representation learning for visible-infrared person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(8): 5095-5109.

[90] LIU Q, HE X H, ZHANG M Z, et al. Feature separation and double causal comparison loss for visible and infrared person re-identification[J]. Knowledge-Based Systems, 2022, 239: 108042.

[91] ZHANG Y Y, ZHAO S Y, KANG Y H, et al. Modality synergy complement learning with cascaded aggregation for visible-infrared person re-identification[C]//Computer Vision - ECCV 2022. Cham: Springer, 2022: 462-479.

[92] FENG J W, WU A C, ZHENG W S. Shape-erased feature learning for visible-infrared person re-identification[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 22752-22761.

[93] LIU J N, WANG J L, HUANG N C, et al. Revisiting modality-specific feature compensation for visible-infrared person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(10): 7226-7240.

[94] WEI Z Y, YANG X, WANG N N, et al. Syncretic modality collaborative learning for visible infrared person re-identification[C]//2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2022: 225-234.

[95] WU J B, LIU H, SHI W, et al. Style-agnostic representation learning for visible-infrared person re-identification[J]. IEEE Transactions on Multimedia, 2024, 26: 2263-2275.

[96] WANG H Z, ZHAO J Q, ZHOU Y, et al. AMC-Net: Attentive modality-consistent network for visible-infrared person re-identification[J]. Neurocomputing, 2021, 463: 226-236.

[97] ZHANG L, GUO H Y, ZHU K, et al. Hybrid modality metric learning for visible-infrared person re-identification[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2022, 18(1s): 1-15.

[98] FENG Y J, CHEN F, SUN G Z, et al. Learning multi-granularity representation with transformer for visible-infrared person re-identification[J]. Pattern Recognition, 2025, 164: 111510.

[99] ZHANG Y K, LU Y, YAN Y, et al. Frequency domain nuances mining for visible-infrared person re-identification[EB/OL]. (2024-01-10)[2025-09-30]. https://arXiv.org/abs/2401.02162.

[100] PANG Z Q, WANG C Y, ZHAO L L, et al. Cross-modality hierarchical clustering and refinement for unsupervised visible-infrared person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(4): 2706-2718.

[101] CHENG D, HUANG X J, WANG N N, et al. Unsupervised

visible-infrared person ReID by collaborative learning with neighbor-guided label refinement[C]//Proceedings of the 31st ACM International Conference on Multimedia. New York: ACM, 2023: 7085-7093.

[102] CHENG D, HE L F, WANG N N, et al. Efficient bilateral cross-modality cluster matching for unsupervised visible-infrared person ReID[C]//Proceedings of the 31st ACM International Conference on Multimedia. New York: ACM, 2023: 1325-1333.

[103] YANG B, CHEN J, YE M. Towards grand unified representation learning for unsupervised visible-infrared person re-identification[C]//2023 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2024: 11035-11045.

[104] HE L F, CHENG D, WANG N N, et al. Exploring homogeneous and heterogeneous consistent label associations for unsupervised visible-infrared person ReID[J]. International Journal of Computer Vision, 2025, 133(6): 3129-3148.

[105] YANG B, CHEN J, YE M. Shallow-deep collaborative learning for unsupervised visible-infrared person re-identification[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 16870-16879.

[106] XIA D X, LIU H J, XU L L, et al. Visible-infrared person re-identification with data augmentation via cycle-consistent adversarial network[J]. Neurocomputing, 2021, 443: 35-46.

[107] CHEN Q S, QUAN Z Z, LI Y J, et al. An unsupervised domain adaption approach for cross-modality RGB-infrared person re-identification[J]. IEEE Sensors Journal, 2023, 23(24): 31399-31413.

[108] XING Y T, XIAO G Q, LEW M S, et al. Lifelong visible-infrared person re-identification via a tri-token transformer with a query-key mechanism[C]//Proceedings of the 2024 International Conference on Multimedia Retrieval. New York: ACM, 2024: 988-997.

[109] LIN X Y, LI J X, MA Z Y, et al. Learning modal-invariant and temporal-memory for video-based visible-infrared person re-identification[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 20941-20950.

[110] LI H F, LIU M H, HU Z X, et al. Intermediary-guided bidirectional spatial-temporal aggregation network for video-based visible-infrared person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(9): 4962-4972.

[111] HAFNER F M, BHUYIAN A, KOOIJ J F P, et al. Cross-modal distillation for RGB-depth person re-identification[J]. Computer Vision and Image Understanding, 2022, 216: 103352.

[112] LIU Q, HE X H, TENG Q Z, et al. BDNet: A BERT-based dual-path network for text-to-image cross-modal person re-identification[J]. Pattern Recognition, 2023, 141: 109636.

[113] ZHAI Y J, ZENG Y W, CAO D, et al. TriReID: Towards multi-modal person re-identification via descriptive fusion model[C]//Proceedings of the 2022 International Conference on Multimedia Retrieval. New York: ACM, 2022: 63-71.

作者简介

励志勇 男,1998年12月出生于四川省成都市. 现为浙江大学控制科学与工程学院博士研究生. 主要研究方向为计算机视觉与行人重识别.

E-mail: lizhiyong_zju@zju.edu.cn

姜 伟 男,1969年11月出生于黑龙江省哈尔滨市. 现为浙江水利水电学院教授. 主要研究方向为机器视觉、计算机图形学与机器学习.

E-mail: jiangwei_zju@zju.edu.cn

刘浩杰 男,1997年9月出生于江苏省南通市. 现为浙江大学控制科学与工程学院博士研究生. 主要研究方向为行人重识别、多模态大模型.

E-mail: liuhaojie@zju.edu.cn