

可见光红外跨模态行人重识别方法综述

范慧杰¹, 郁航^{1,2}, 赵颖畅^{1,3}, 唐延东¹

1. 中国科学院沈阳自动化研究所机器人学国家重点实验室, 辽宁 沈阳 110016;

2. 沈阳化工大学信息工程学院, 辽宁 沈阳 110142;

3. 沈阳理工大学自动化与电气工程学院, 辽宁 沈阳 110158

基金项目: 国家自然科学基金项目(62273339, U20A20200)

通信作者: 范慧杰, fanhuijie@sia.cn 收稿/录用/修回: 2024-02-02/2024-07-23/2024-08-26

摘要

可见光红外跨模态行人重识别技术因不受夜间限制, 可实现全天候监控而受到广泛关注。本文希望通过分析现有可见光红外跨模态行人重识别研究方法及其适用场景、算法优缺点来帮助研究人员根据研究需求找到合适的解决方案。同时, 寻找该领域难点和困境, 以此探讨可见光红外跨模态行人重识别未来方向。首先介绍行人重识别的概念, 回顾发展历程, 介绍可见光红外跨模态行人重识别的意义; 其后将可见光红外跨模态行人重识别研究方法分为基本方法、辅助模型方法、无监督方法、基于视频的方法, 并对每种方法的适用场景、优缺点和未来研究方法展开分析; 再对当前可见光红外跨模态行人重识别的评价指标以及现有数据集进行介绍, 并对每种数据集的优劣进行分析; 最后讨论了可见光红外跨模态行人重识别未来发展。

关键词

跨模态行人重识别

计算机视觉

无监督

辅助模型

中图法分类号: TP18

文献标志码: A

Review of Visual-infrared Cross-modal Person Re-identification Methods

FAN Huijie¹, YU Hang^{1,2}, ZHAO Yingchang^{1,3}, TANG Yandong¹

1. State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China;

2. School of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, China;

3. School of Automation and Electrical Engineering, Shenyang Ligong University, Shenyang 110158, China

Abstract

Visible-infrared cross-modality person re-identification technology has garnered wide attention for its ability to provide round-the-clock monitoring regardless of nighttime conditions. We aim to help researchers find suitable solutions by analyzing existing methods, their applicable scenarios, and the advantages and disadvantages of different algorithms. Additionally, We seek to identify current challenges and explores future directions for visible-infrared cross-modality person re-identification. We first introduce the concept of person re-identification, review its development, and highlight the significance of visible-infrared cross-modality person re-identification. Subsequently, it categorizes research methods into basic methods, auxiliary model, unsupervised methods, and video-based approaches, analyzing their applicable scenarios, advantages, disadvantages, and potential future research directions. We also discuss current evaluation metrics and existing datasets

Keywords

cross-modal person

re-identification;

computer vision;

unsupervised;

auxiliary model;

for visible-infrared cross-modality person re-identification, analyzing the strengths and weaknesses of each dataset. Finally, We discuss the future prospects in this area.

0 引言

行人重识别(Person Re-Identification, ReID)作为计算机视觉领域的研究热点之一,旨在从非重叠的摄像机视域中检索目标行人,被认为是图像检索任务中的一个子任务^[1]。如图1所示,给出一张待检测行人图像 Probe,行人重识别模型从不同相机采集得到的行人图像库 Gallery 中找出所有的目标行人图像。该任务起源可以追溯到1997年^[2],当时是与多相机跟踪任务相联系,其后在2005年由ZAJDEL等^[3]首次提出行人重识别概念。随着深度学习的发展,在2014年YI等^[4]开始将深度学习应用到行人重识别领域,直至今日深度学习仍是解决行人重识别问题的主流方法^[5-10]。

近些年来,由于行人重识别技术在智能监控、智能安防等领域的广泛应用^[11-13],逐渐成为一个独立的计算机视觉任务。但现有的方法大都是基于可见光图像的单模态行人重识别^[14-17],用于训练和测试的数据集往往都是单模态的可见光图像。然而由于可见光(RGB)摄像机不能在光照差的环境

下(如夜间)提供有用的外观信息,使 ReID 在真实场景中的适用性受到限制。例如,犯罪嫌疑人经常在夜间进行犯罪活动,仅仅依靠可见光成像设备很难捕获行人的外观信息。相比之下,红外(IR)成像设备不受光线条件的影响,可以在夜间收集更清晰的行人图像,同时大部分监控摄像头都能在黑暗环境下自动从可见光模式切换到红外模式,以此实现全天候的视频监控,这就提出了可见光红外跨模态行人重识别。该问题于2017年在行人重识别领域首次被WU等^[18]提出定义,他们提出了一种深度零填充的深度学习网络结构以处理可见光图像和红外图像之间的差异,并提供了一个公开的大规模可见光红外跨模态行人数据集,名为SYSU Multiple Modality ReID(SYSU-MM01)。跨模态行人重识别的目的是在给定目标人物的可见光(或红外)图像后,找到由其它相机视角下捕捉到的同一人的相应红外(或可见光)图像。近些年来,基于可见光红外的跨模态行人重识别受到越来越多研究者关注,已成为行人重识别任务的一个热点问题。

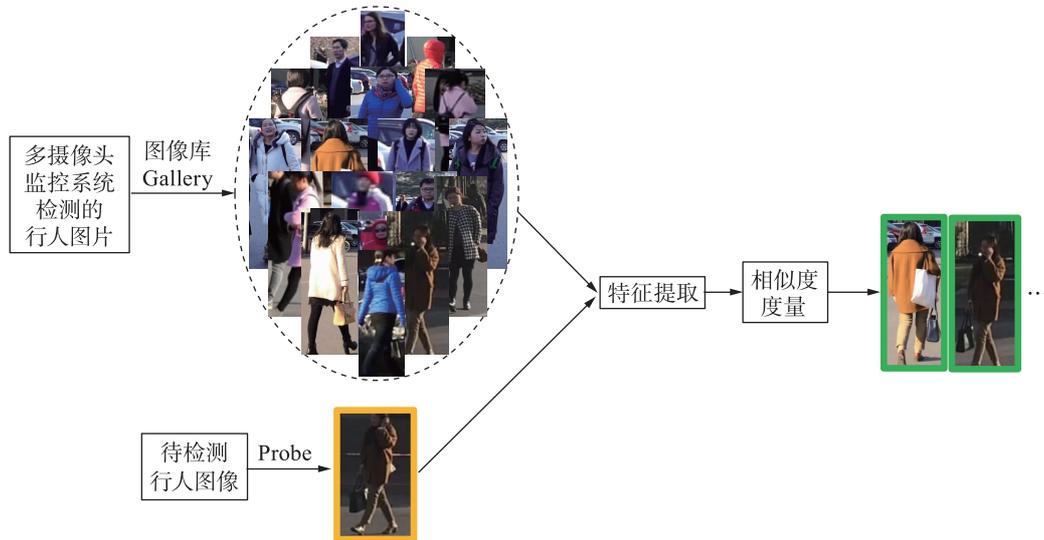


图1 行人重识别

Fig.1 Person re-identification

可见光红外跨模态行人重识别除了要面临单模态的角度不同^[19]、分辨率不同^[20]、姿态变化^[21]和遮挡^[22]等问题外,还面临一个更大挑战是模态间的巨大差异,甚至同一个人的模态间差异(类内差

异)大于不同行人之间的身份特征差异(类间差异)^[23],如图2所示。因此,可见光红外跨模态行人重识别要解决的关键问题是模态差异的消除和模态不变特征的提取。自从2017年首次提出可见光

红外跨模态行人重识别之后,一系列的优秀算法模型被提出,根据模型方法和解决的问题不同可分为基本方法、辅助模型方法、无监督方法和基于视频的方法。基本方法主要包括图片级方法、特征级方法和基于 Transformer 的方法。在图片级经过模态互转和模态融合的方法来映射不同模态的图像到同一像素空间,实现模态的统一,减少模态间差异,例如文[24]中采用生成式对抗网络来分别为可见光图像生成红外图像、为红外图像生成可见光图像减少模态差异;在特征级通过特征层面的细微纹理提取、对齐和互补融合等方式实现模态间的匹配和转换,目的也是减少模态间差异,实现模态间的图像匹配,例如文[25]中提出一种基于双分类器的注意力融合机制,将各模态的自有特征赋予较大的融合权重,共有特征赋予较小的融合权。文[26]采用姿态估计聚焦于局部特征的学习,这有助于行人之间的对齐和干扰噪声的去除。图片级方法和特征级方法都是采用 CNN 网络作为图像编码器,最近由于 Transformer 强大的特征提取能力^[27-30],基于 Transformer 的方法也被用于跨模态行人重识别任务。辅助模型方法是利用姿态估计、分割模型、身体解析等其它模型调整模型关注到更有辨别性的区域,使模型对环境干扰更加鲁棒。上述都是监督方法的可见光红外跨模态行人重识别,由于其需要大量的人工标注,实际应用价值被限制,之后无监督的方法被提出解决标注问题,在不需要任何标签的情况下实现模态间的行人检索,此任务更具挑战。在实际视频监控中,摄像头直接采集的是视频序列,所以基于视频方法的可见光红外跨模态行人重识别可直接利用视频序列进行行人匹配,同时视频

中含有丰富的时空和运动信息可被挖掘以弥补单个图片信息的不足,提高行人匹配的准确性。

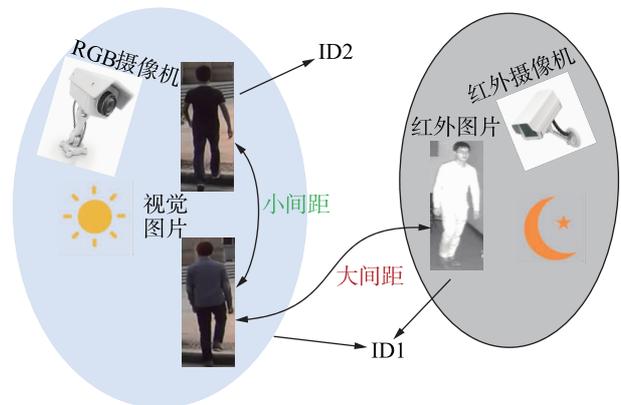


图2 模态差异示意图

Fig.2 Schematic diagram of modality differences

1 可见光红外跨模态行人重识别方法

1.1 基本方法

可见光红外跨模态行人重识别主要面临的问题是两种模态之间差异巨大。可见光图像拥有红、绿、蓝三个颜色通道,而红外图像只有一个通道,从成像原理来看,二者来自不同波长范围。因此大多数方法致力于缩小模态间的差距,并更好地提取模态间的不变特征。主流的解决模态巨大差异的骨干网络是基于 CNN 的双流结构^[31-33],如图 3 所示。可见光图像和红外图像分别经过网络的模态特定层提取不同模态图像的具体特征,属于不同模态的特定层的特征提取网络权重不共享,随后经过权重共享的模态共享层获取模态间的一致性特征,减少模态间的差异。大量研究^[34-36]基于此骨干网络提出了一系列解决方法,以进一步消除模态间差

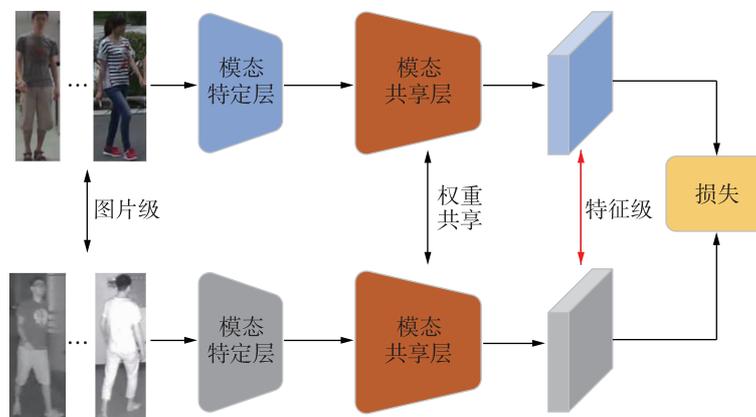


图3 可见光红外跨模态行人重识别模型骨干网络

Fig.3 Backbone network of visual-infrared cross-modal person re-identification model

异, 提高跨模态行人重识别检索的准确性, 主要分为基于图片级的方法和基于特征级的方法。近年来另一种基于 Transformer 方法的跨模态行人重识别方法^[37-39]也得到了广泛研究, 这种方法依赖于 Transformer 强大的特征提取能力, 展现出显著的效果。

基于图片级的方法是在图片水平上通过模态转换或模态融合实现模态统一, 减少模态间的差异。模态转换方法^[40-44]通过将一种模态转换为另一种模态来消除模态间的差异。模态融合方法^[40-41]则是在像素层面协调模态间的关系, 利用可见光图像和红外图像共同作用生成新的融合或混合模态。随着 GAN 的发展, 文[41]提出了一种双级差异减少学习(D2RL)方案, 通过图片级子网络将红外图像转换为可见光图像或反之, 以减少模态差距。然而, 训练一个可靠的图像生成器需要大量的计算成本, 生成的图像通常包含不可避免的噪声, 并且这种基于 GAN 的方法在通道和空间维度上重构信息, 破坏了原始的空间结构信息。相比之下, 文[42]提出了一种同质增强三模态(HAT)方法, 通过从同质的可见光图像生成辅助的灰度图像, 使模型能够更好地捕获结构信息, 并且不需要复杂的图像生成过程, 从而避免了潜在的噪声引入。文[43]则提出了 X-红外-可见三模态学习框架, 使用轻量级网络通过自监督学习生成 X 模态图像, 避免了 D2RL 中的病态红外到可见光生成, 并且可以直接使用大多数现有的行人重识别主干, 而不需要额外的工作量来改变结构和微调, 但其图像生成过程通常需要额外的计算成本, 并受到不可避免的噪声影响。文[44]进一步提出了通道增强联合学习策略。与 HAT 中直接采用灰度变换作为基线方法将三通道 RGB 图像转换为单通道灰度图像不同, 文[44]采用通过随机交换颜色通道来生成与颜色无关图像的增广方法, 并将随机灰度变换作为增强补充操作, 增强了可见光红外识别对颜色变化的鲁棒性, 计算成本更低且避免了噪声问题。由于仅基于可见图像生成的过渡图像很难建立三模态之间的分布相关性, 影响了度量学习能力, 因此文[45]提出了一种新颖的同化模态协同学习(SMCL)模型, 通过轻量级网络生成融合了可见光和红外图像特征的同化模态, 保留了两种图像的共同表示, 促进模态共享特征的生成, 在性能上与前面提到的方法相比有很大的优势。基于图片级方法的本质是在像素上将可见光模态和红外模态统一到中间模态, 对模态间的一致性特征提取较粗糙, 在巨大模态差异下进行模态统一

对噪声敏感且转换过程中会产生新的噪声。

基于特征级的方法是在经过骨干网络提取的特征层面通过细微纹理提取^[46-51]、对齐和互补融合^[52-55]等方式实现模态间的匹配和转换。文[49]通过动态双注意力聚合学习(DDAG)方法挖掘每个模态内的局部信息和跨模态的图结构信息来增强特征表达, 文[54]提出了联合模态与模式对齐网络(MPANet), 在发现跨模态细微差别的同时减轻模态差异。与文[49]和文[54]主要依赖于传统的特征提取和对齐方法不同, 文[51]提出的 PSFLNet 通过参数共享直接在网络结构上进行创新, 使得模型在特征提取阶段就开始处理模态差异, 更有效地整合了可见光和红外模态的信息。尽管这些方法均能有效处理模态间差异, 提高特征的分歧性和模态鲁棒性, 但模态间物理成像的巨大差异导致模态间的共享外观特征有限, 只能在干净的数据集环境下提取到可辩别的行人特征, 对于复杂环境(如遮挡等)特征的可辨别性不足。一种新的基于对比学习的方法也被用于可见光红外跨模态行人重识别任务, 例如文[56]通过利用增加的模态感知和模态不可知的质心代理来进行集群对比和相互信息学习, 通过增加内存库中的负特征有效地减小了模态间的差异, 从而增强了模型的学习能力和泛化能力, 能够在光照、遮挡和视角变化等实际应用场景中表现得更好。

基于 Transformer 的跨模态行人重识别方法是采用 Transformer 架构来提取每个行人图像的特征, 这类方法通常利用 Transformer 的强大全局注意力机制来捕捉不同模态间的复杂关系和特征, 有助于模型在面对遮挡、不同角度和照明条件变化时保持稳定的性能。由于自注意力层的权重是根据输入数据动态调整的, 这使得 Transformer 能够根据不同的输入灵活调整其处理方式, 适合用于处理模态间的差异。文[57]提出模态嵌入和模态感知增强损失来让模型学习模态不变特征。文[58]使用灰度图作为辅助模态并采用渐进式学习策略减少模态差异。这两种方法都提高了跨模态特征的分歧能力和鲁棒性, 并有效缩小了模态间的差距。虽然 Vision Transformer 在许多任务包括单模态的行人重识别任务上取得了比 CNN 更好的性能效果, 但由于其更擅长捕捉全局依赖, 在处理细粒度局部特征方面可能不如 CNN, 并且 Transformer 通常需要大量的标注数据来有效训练, 而跨模态行人重识别相关数据集的数据量不足以使 Transformer 对跨模态巨大的

模态差异敏感,因此相关研究的性能也低于同期的CNN方法。

1.2 辅助模型方法

行人重识别关键技术是提取同一行人身份不同图片之间的不变特征和不同行人身份之间的可鉴别特征,一般的特征提取方法对行人身份的属性、姿态、身体结构等可鉴别信息关注度不够且性能受姿势变化、遮挡和视角差异的影响,为此引入姿态估计^[59-63]、身体解析^[64]、分割^[65]等辅助模型到行人

重识别模型以对齐行人身体的结构区域,抑制噪声干扰,对行人重识别任务的性能有极大提升。如图4所示,4(a)表示一般方法模型注意力可视化图,4(b)表示利用姿态估计模型作为辅助模型在行人重识别模型中模型注意力的可视化图,颜色越红表示网络该区域关注度越高。对比发现引入姿态估计辅助模型后,网络能从全局关注到更多可判别特征,同时对于遮挡等复杂情况也能引导网络从身体区域进行特征学习,避免被遮挡等噪声信息错误引导。

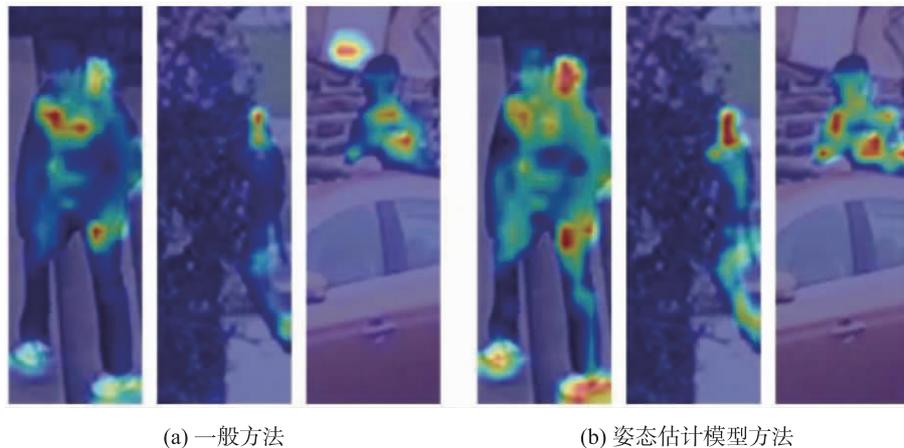


图4 姿态估计模型方法与一般方法的注意力对比

Fig.4 Attention comparison of pose estimation modeling methods with general methods

可见光相机和红外相机成像原理的不同使可见光红外跨模态行人重识别任务存在巨大的模态间差异(类内距离),使得可见光红外跨模态行人重识别模型需要集中于模态之间的一致性信息(如身体结构、步态等)提取,忽视颜色等由成像原理导致的差异信息。所以在可见光红外跨模态行人重识别任务中引入辅助模型能有效地提升性能。

对于辅助模型的使用可以分为两种方法:1)辅助模型作为行人重识别模型的一部分,在训练和测试过程中都需要额外的辅助模型。例如,文[60]提出了一个联合训练姿态估计和行人重识别任务的端到端框架,将姿态估计模块作为辅助模型整合到主模型中,不仅在训练过程中用于提供额外的监督信息,还在测试过程中继续发挥作用,确保了模型在所有阶段都能利用到姿态信息,以增强特征的辨别力和模态的一致性。2)辅助模型只在训练过程中使用。在行人重识别模型的训练过程中辅助模型起监督作用,使模型具有可鉴别区域特征提取能力。文[64]在训练过程中使用了形状先验作为指导,通过抹除与身体形状相关的语义概念来强制重

识别模型提取更多其他的模态共享特征,在实际推理或测试阶段则不需要这些辅助模型,有效减少了推理阶段的计算负担。两种方法相比,前者由于在测试阶段也需要额外的辅助模型,使得在推理过程中的模型复杂度和计算时间都显著增加,后者则可能由于训练和推理阶段的不一致而导致模型泛化能力的下降,两者在训练过程中模型复杂度和训练时间都有明显增加。通过采用类激活映射等自我鉴别区域检测技术以及理论约束方法,可以在仅少量增加模型参数的情况下,实现与额外辅助模型相当的效果。这种策略能有效降低模型在训练过程中的复杂度大和训练时间长的问题。通过这种方法,模型能够自动识别和强化对识别任务关键的图像区域,从而提升模型的学习效率和性能。同时,理论约束方法如知识蒸馏或正则化技术,可以进一步引导模型学习更加鲁棒的特征表示,确保模型在实际应用中的有效性和可靠性。

1.3 无监督方法

行人重识别是一个有实际应用价值的课题,所以保证其实际性是有意义的。虽然有监督的行人重

识别在性能上表现更好,但由于标签成本过高,有监督方法无法扩展到大型无标签数据集和新领域。相反无监督行人重识别则不需要标签即可完成匹配任务,所以其使用价值更高。近年来在可见光单模态行人重识别领域,发表了大量基于无监督的行人重识别方法,且其性能已经接近基于监督的方法^[66-68]。可见光单模态无监督行人重识别可分为域自适应方法^[69-74]和纯无监督方法^[75-77]。基于域自适应的方法是在有标签的源域进行预训练,然后迁移到无标签的目标域,虽然一些研究^[78-81]已经证明将知识从源域转移到目标域是有效的,但它假设源域数据和目标域数据分布差异小,当源域数据和目标域数据分布差异大时,此方法可能无法达到理想效果。纯无监督方法只在无标签数据集上训练模型,其难度也更大。本文只对纯无监督方法(后文简称无监督方法)在可见光红外跨模态行人重识

别中研究现状和挑战进行分析。

目前单模态无监督行人重识别的主流方法是基于聚类的方法^[82-83],其结构如图5所示。基于聚类的无监督行人重识别可分为两个过程:聚类过程和训练过程。在聚类过程中 ReID 模型将整个数据集图像映射到特征空间,在 DBSCAN 和 K-means 等聚类算法作用下为每个类别生成伪标签,并将伪标签分配给每张图片以重构一个新的有伪标签的数据集;在训练过程,ReID 模型在新的伪标签数据集上进行监督形式的训练。由于视角、遮挡、光照等原因无监督的聚类结果存在相同行人被分成多个不同类别(多分)和不同行人被分成同一类别(错分)问题,这也是当前大部分研究^[84-86]去解决的核心问题,并取得了显著的效果,最新的单模态无监督行人重识别方法^[87]已经接近 SOTA 的监督行人重识别方法^[88-89]。

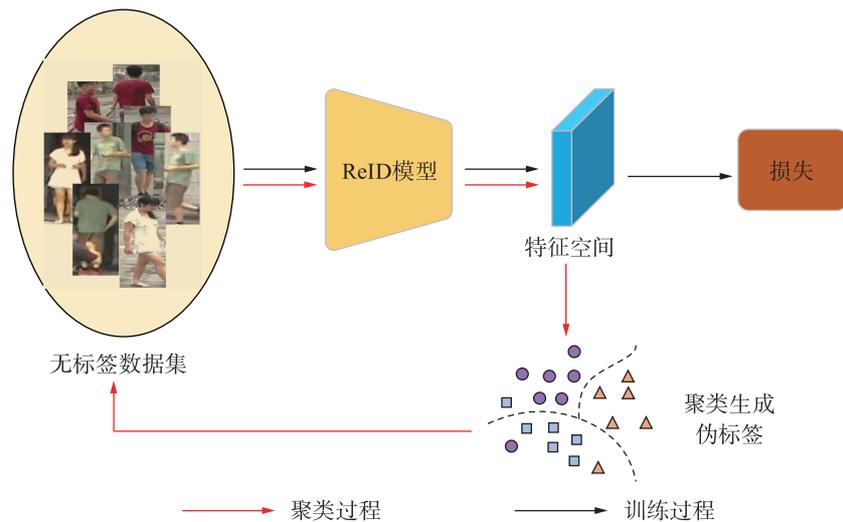


图5 单模态无监督行人重识别方法

Fig.5 Unimodal unsupervised person re-identification method

相比于单模态的无监督行人重识别,可见光红外跨模态无监督行人重识别^[90-95]面临更大的挑战。如图6所示,可见光红外跨模态无监督行人重识别方法在可见光红外两个模态分别进行单模态形式聚类生成伪标签得到新的伪标签数据集的聚类过程,之后也是在新的伪标签数据集上进行监督形式的训练过程。文[90]提出了第一个无监督跨模态行人重识别方法,即“同质到异质”两阶段学习(H2H)方法,虽然H2H在可见光到红外域不需要身份标注,但在源域仍然需要进行预训练,因此H2H并不是一种完全无监督的方法,而是一种跨域方法。后续研究中ADCA(Augmented Dual-Contrastive Aggre-

gation)^[92]和DFC(Diversity Feature Constraint)^[93]不再依赖标记源域,采用完全无监督的方案,进一步提升了模型性能。OTAL(Optimal Transport for Label-efficient)^[94]利用标准的无监督域自适应技术,在有良好标注的RGB数据集的帮助下,为可见子集生成伪标签,并尝试将可见模态的伪标签分配到红外模态。然而这些方法往往依赖于模态内初始化和跨模态实例选择。与上述方法不同,文[95]提出了跨模态层次聚类和伪标签细化(CHCR)方法,不需要任何预训练和模态内初始化,完全在无监督的框架下工作,直接在模态间进行聚类,有效减少了计算复杂度和训练时间。然而可见光红外跨模态的

无监督行人重识别不仅有单模态中的聚类多分和错分问题,还因为红外模态的类间距离小导致聚类结果更容易出现错分情况,使得模态间聚类结果分布和数量上都有较大差异,所以可见光红外跨模态无监督行人重识别出现一个聚类结果不平衡的新问题。文[96]提出了渐进式图匹配的方法来完成模态间的聚类结果配对,以解决聚类结果不平衡问

题,通过交替对比学习降低模态差异。虽然这些工作取得了一定效果,但由于模态间的差异巨大使得聚类 and 匹配结果都未能达到理想效果,导致最终的行人匹配结果不佳。从单模态无监督行人重识别结果接近监督行人重识别结果可以发现,可见光红外跨模态无监督行人重识别缺少一个能消除模态间巨大差异的鲁棒模型,从而提取模态间的不变特征。

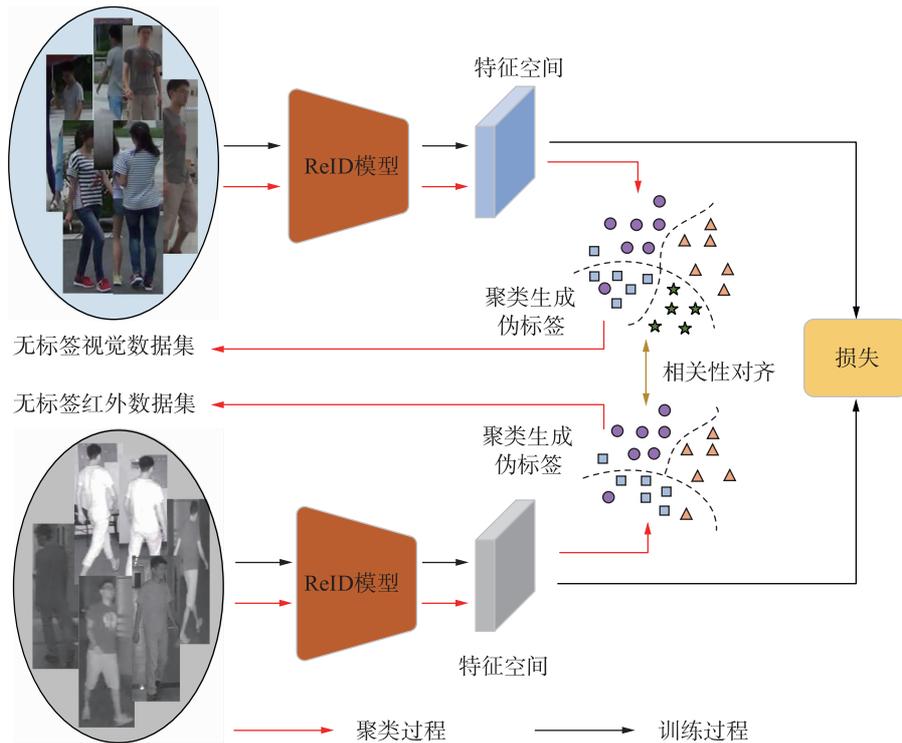


图6 可见光红外跨模态无监督行人重识别方法

Fig.6 Visual-infrared cross-modal unsupervised person re-identification method

1.4 基于视频的方法

与基于图像的行人重识别不同,基于视频的行人重识别通过一系列图像来表示一个人,提供了时间信息和更丰富的外观^[11],它可以通过多帧图像来缓解由于姿态变化、遮挡和光照变化等引起的问题,更具有鲁棒性。3维卷积^[97-101]在视频行人重识别中提供了一种强大的方法,它可以用于对视频序列进行时空建模,行人在视频中的移动和动作可以通过卷积核捕捉,从而建模行人的外观和行为特征,区分不同行人。然而在视频行人重识别中经常会出现行人错位问题,导致图片同一个位置在时间轴上可能不是同一行人部位,采用普通3维卷积提取的特征并不好,因此可以通过加入重构模块或结合注意力机制等进行特征对齐。为了弥补CNN处理时间维度信息的缺陷,还可以将RNN^[102-105]用于

视频行人重识别,RNN可以有效捕捉视频中每一帧和前后帧的时序关系,从之前的帧中存储时间上下文,帮助模型理解行人在视频序列中的动作和行为变化。为了去除噪声的影响,部分方法利用时空信息^[106-112]进行视频行人重识别,文[106]提出了时空注意模型,文[107]提出了空间和时间记忆网络,文[108]提出了金字塔时空聚合架构,将帧级特征分层整合为最终的视频级表示,上述这些方法都聚焦于提高视频行人重识别的准确性和鲁棒性,但都面临不同程度的计算复杂度和计算资源需求问题。

在跨模态行人重识别中,用视频代替静止图像更为重要。在某些情况下,只根据一张图片可能很难识别两个外表相似的人,而视频是包含空间和时间信息的图像序列,可以利用运动信息进行判别。

如图 7 所示, 对两个不同的人拍摄的两幅红外图像具有相似性, 然而与穿蓝色 T 恤的人相比, 穿黑色 T 恤的人具有特定的手臂姿势。这些运动特征可以提供更多的判别信息, 从而实现更鲁棒和准确的识别模型。与现有的基于视频的可见光行人重识别方法相比, 跨模态视频行人重识别需要提取可见光和红外模态之间的一致性。文[113]首次将跨模态和

基于视频的行人重识别结合在一起, 通过对抗学习和时间信息处理来增强特征的模态不变性和时间连续性, 并构建了基于视频的可见光红外数据集 HITSZ-VCM。文[114]则是通过对抗性扰动和时空关系挖掘相结合, 不仅处理模态间的差异, 还强化了对复杂场景变化的适应能力, 增强了模型的整体鲁棒性。

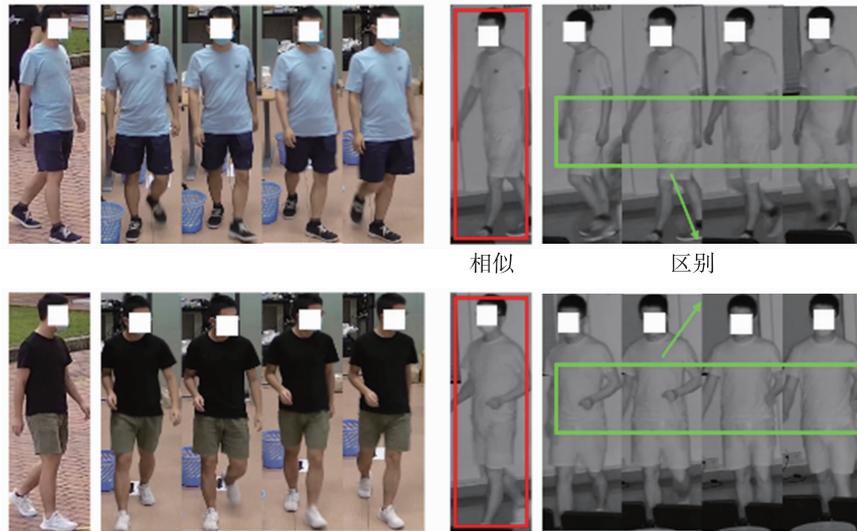


图 7 基于视频的跨模态行人重识别的优点^[113]

Fig.7 Advantages of video-based cross-modal person re-identification^[113]

1.5 现有方法总结

基本方法通过在图片级或者特征级进行模态转换、生成、融合或者利用 Transformer 骨干集中于提取模态具体特征和模态共享特征去消除模态间差异。这类方法没有考虑实际场景的复杂性, 如光照、遮挡、服装改变等因素的影响, 所以一般方法在单一、且摄像头采集的图像干净情况下效果较好, 在面对实际场景中的复杂问题, 性能会迅速下降。图片级方法无法保证模态转换或者模态生成实现模态统一过程的转换和生成的质量, 进而引入额外的噪声限制了模型对于行人可辨别特征的学习。未来可在模态转换的特征层面设计损失函数来动态控制模态转换和生成, 减少噪声和冗余, 从而实现更好的性能。对于特征级方法, 现有的可见光和红外跨模态行人重识别方法都集中行人外观特征提取模态共享特征, 但由于跨模态物理成像的限制导致模态间的模态共享外观特征比单模态外观特征更少, 使得跨模态行人重识别的模态共享特征不足以表示类间差异, 在性能上不如单模态。未来需要从其他角度, 如人体结构、类内实例关系等, 挖掘更

多的模态共享特征来提高特征的代表能力。Transformer 在很多任务包括单模态行人重识别任务上都体现了强大的特征提取能力, 但在可见光红外跨模态行人重识别效果并不理想, 普遍低于同期的 CNN 方法。相比于 CNN, Transformer 更加注重于全局特征的学习, 不同模态间图像频域差异大, 导致高频部分信息被稀释^[82], 进而反映出更低的性能。未来需要在结合跨模态巨大的模态差异设计合理的 Transformer 范式, 而不是直接利用 Transformer 骨干在跨模态数据集上进行训练。

辅助模型方法利用其它模型提取行人身份的属性、姿态、身体结构等信息对齐行人的特征表示。这种方法应对实际场景中的一些特定问题。例如, 采用的人体语解析模型或者关键点检测模型对具有遮挡或者采集不完整的图像有一定的鲁棒性; 采用光照增强模型则对光线变化具有更好的鲁棒性。并且现有的辅助模型方法都主要集中在对齐特征学习或者说是消除背景干扰, 并不能促进模型提取更加丰富的个人信息, 并且模型的性能存在上限, 因此未来需要有效利用这些辅助信息去丰富行人的信息

表示。

由于人员的流动性和隐私等问题,导致开放世界下的行人身份标签是不可得的。无监督方法主要集中在解决标签不可得问题,所以无监督跨模态行人重识别更符合开放世界下的全天候监控。但目前这方面的研究成果并不多,大都还处于结合单模态无监督方法来处理跨模态中的无标签问题,而模态不一致使得无监督任务更加具有挑战。本文认为未来主要有3个问题需要解决:1)模态间的差异导致聚类不平衡,使得单模态中基于聚类的无监督方法不能直接用在跨模态;2)无监督跨模态任务仍然会面临遮挡、光照等实际问题的挑战;3)在模型部署后需要随着时间的变化持续学习。例如在室外随着季节的变化,环境在改变,同时人也会发生变化,模型应该需要在过去采集到的行人图像库上进行持续学习。

视频方法的可见光红外跨模态行人重识别方法不仅能提供个人的外观信息,还包含图像序列的时间和空间信息。这种方法更加准确的检索到目标人物,并且能直接检索出监控系统下的目标人物视频。但相比于单图片检索,根据检索结果再调取对应视频,视频行人重识别对数据集要求更高,需要跟踪相机视角下的人,并对其当前轨迹视频存入数据库中,增加了部署的难度。

2 评估方法和数据集

2.1 评估方法

行人重识别评估指标能够客观衡量行人重识别算法在识别和匹配行人图像时的性能和准确度,对于评估和比较不同算法的效果至关重要。它们帮助研究人员了解不同算法的优劣,并推动行人重识别算法的发展和改进。同时,这些指标也对实际应用中的行人重识别系统的性能和鲁棒性具有指导意义,帮助选择最适合的算法或系统。以下是几个常见的行人重识别评估指标。

2.1.1 Rank- n 准确率

Rank- n 准确率 (Rank- n Accuracy) 用于衡量在给定 Probe 时在 Gallery 匹配的前 n 个最相似的图像中是否有正确的匹配项,如果有,则 Rank- n 为 1,否则为 0。例如 Rank-1 是首位命中率,即排在第一位的图像是否与 Probe 是同一个人,如果是则 Rank-1 为 1,反之为 0。最后,计算 Query 中所有 Probe 查询图像的 Rank- n 取平均得到最终的 Rank- n 准确率值。Rank- n 准确率常用于评估行人重识别

算法在实际应用中的实用性,例如在视频监控中通过查询图像来识别和追踪行人。较高的 Rank- n 准确率表示算法能够在前 n 个最相似的图像中更容易找到正确的匹配,因此具有更好的性能。

2.1.2 CMC 曲线

显示在不同排名 (Rank) 下的准确率,形成 CMC (cumulative match characteristic) 曲线,该曲线展示了算法在不同排名下的性能表现。在绘制 CMC 曲线时,横轴表示排名 (Rank),从 1 到 N , N 为 Gallery 图像库中的总图像数量。纵轴表示准确率值,即在给定排名下的正确匹配的比例,曲线向上的趋势表示算法能够在更高的排名下找到正确的匹配。通过观察 CMC 曲线,可以了解算法在不同排名下的准确率,比较不同算法之间的性能差异,以及评估算法在实际应用中的可行性和实用性。通常,随着排名的增加,准确率会逐渐升高,而好的算法在较低的排名下能够保持较高的准确率。

2.1.3 mAP 平均精度均值

mAP (mean Average Precision) 综合考虑了检索结果的准确性和排名的精确度,反映了检索的人在 Gallery 图像库中的所有正确匹配图片排在结果队列前面的程度,而不止首位命中。具体操作是分别求出每个类别的平均准确值后再次取平均值,而平均准确值是求准确率-召回率曲线下的面积。较高的 mAP 值表示算法具有更好的性能,能够在检索过程中产生更准确和稳定的结果。mAP 在行人重识别任务中被广泛使用,特别是在评估算法的整体性能和比较不同算法之间的差异时,是一项重要的指标。

2.2 数据集介绍

目前,行人重识别领域的数据集类型比较广泛,针对可见光红外跨模态的行人重识别数据集^[18, 113, 115-117]主要有 5 个,如表 1 所示。

表 1 数据集统计
Tab.1 Dataset statistics

数据集	ID 数	图片数	相机数	类型
RegDB	412	8 240	2	图片
SYSU-MM01	491	303 420	6	图片
LLCM	1 064	46 767	9	图片
RGBNT201	201	14 361	4	图片
HITSZ-VCM	927	463 259	12	视频

RegDB 数据集是 2017 年用于可见光红外跨模态行人重识别任务公开数据集,数据集的采集摄像

头由一个可见光摄像头、一个红外摄像头共 2 个摄像头采集。数据集包括 412 个行人, 其中 254 个女性和 158 个男性, 正面捕获了 156 个人, 背面捕获 256 个人, 每个人 10 张图片, 每张图片在身体姿势、捕获距离和照明条件间存在差异, 数据集总共 8 240 张图片。RegDB 数据集采集摄像头数量较少, 图片总数也较少, 在同年另一个可见光红外的跨模态行人重识别数据集 SYSU-MM01 也被公开。

SYSU-MM01 数据集总共由 6 个摄像头采集(2 个红外摄像头和 4 个可见光摄像头), 包括一个室内红外摄像头、一个过道红外摄像头、2 个室内可见光摄像头、2 个室外可见光摄像头, 所以大部分模型实验是分为室内检索和全检索两种模式。数据集包括 491 个行人, 287 628 张可见光图片和 15 792 张红外图片。与 RegDB 数据集相比, SYSU-MM01 数据集更加复杂, 图片数量也更多。

LLCM 数据集由 9 个可见光红外自动切换的相机在低光照的复杂环境下获得。数据集包括 1 064 个行人, 25 626 张可见光图片和 21 141 张红外图片。与其它可见光红外跨模态数据集相比采集环境更复杂、行人更多, 所以也更具挑战。

RGBNT201 数据集用于多模态行人重识别的综合评价, 包含了在校园内以 4 个非重叠视图采集的 201 个身份, 每个视图由 3 个摄像头组成, 同时记录 RGB、NI 和 TI 数据, 其中 141 个身份用于训练, 30 个身份用于验证, 30 个身份用于测试。为了高效评估, 在每 5 ~ 10 张相邻图像中自动选择人物图像, 然后手动检查以避免数据冗余。每个记录由至少 20 个不相邻的 3 组图像以三模态的方式组成, 以不同的姿势捕获, 形成每种模态 4 787 张图像用于实验评价。与仅在白天光照良好的情况下捕获的大多数 RGB 行人重识别数据集不同, 该数据集进一步在恶劣的光照条件下捕获大量具有挑战性的图像, 如夜间黑暗或烟雾和雾等低能见度天气。具体来说, 使用分辨率为 700 × 580, 帧率为 15 帧/s 的 HIKVISION RGB-NI 相机来捕捉 RGB 和 NI 模态图像。TI 图像由 FLIR T610 同时捕获, 分辨率 640 × 480, 帧率 20 帧/s。此数据集 3 种模态相机是同时拍摄, 使得每张行人图片都同时具有 3 种模态, 所以此数据集的研究重点也从跨模态变成了多模态即利用多模态之间的信息互补来完成行人重识别任务。

在行人重识别任务中, 图片数据集存在着视角有限、时间信息缺失、遮挡问题等不足之处, 而视频数据集则可以提供多个角度下的行人图像, 更全

面地捕捉行人的外貌特征, 有助于提高重识别的准确性; 可以提供行人的运动轨迹和动态特征, 有助于更好地理解行人的行为模式和空间关系; 视频数据集中的连续帧可以提供丰富的上下文信息, 有助于识别和验证行人身份。尽管图片数据集在行人重识别中有其应用和研究的重要性, 但视频数据集相对具有更多的优点, 可以提供更全面、准确和丰富的信息, 有助于提升行人重识别算法的性能和鲁棒性。

HITSZ-VCM 是一个基于视频的可见光红外跨模态行人重识别任务的公开数据集, 数据由 12 个能自动切换可见光红外模态的高清摄像机采集, 包括 7 个室外、3 个室内和 2 个通道场景。涵盖的场所比较多, 如办公室、咖啡馆、通道、游乐场和花园都在内。此外, 还考虑了每个人在每个摄像机下的多个角度捕获, 构建了丰富外观的数据集。摄像机拍摄频率是 25 帧/s, 每隔 5 帧选取一帧图像作为数据集一个视频序列的一帧, 每选取的 24 个连续图像作为一个行人的一个视频序列。最终整个数据集包括 927 个行人, 251 452 张可见光图像和 211 807 张红外图像, 分别被分成 11 785 个可见光视频序列和 10 078 个红外视频序列。虽然 HITSZ-VCM 数据集采集摄像机有 12 个, 但大部分行人都是由其中的 3 个视觉摄像机和 3 个红外摄像机采集。

3 总结与展望

本文从基本方法、辅助模型方法、无监督方法和基于视频的方法对可见光红外跨模态行人重识别进行总结分析。基本方法是在图片层面和特征层面减少模态间差异, 由于跨模态物理成像的限制导致模态间的模态共享外观特征少, 只能应用于干净环境。辅助模型方法通过引入额外模型指导 ReID 模型提取更具判别性特征减少无关信息干扰, 对一些复杂环境(如遮挡)具有更好的适应性, 也因此引入了额外的模型参数和计算时间。无监督方法因其不需要人工标注, 解决了实际场景中跨相机标注难题, 使其备受欢迎, 其难度也更大, 需要在没有标签情况下, 应对模态间差异、遮挡、光照等问题。基于视频的方法能同时利用时空信息, 但现有方法无法自动处理动态长度轨迹视频。最后, 总结可见光红外跨模态行人重识别未来的发展方向:

1) 文本引导的跨模态检索。通过大语言模型对文本的强大编码能力, 实现图文复合表征强化行人表示的互补性。例如, 利用文本提示引导模型关注行人身份特征, 减少模态关注; 行人图像被遮挡

时,也可利用文本提示避开遮挡;行人有突出特征时,利用文本提示进行着重学习。在跨模态中,“一图+千言”对行人进行图文互补表示提高行人跨模态跨相机检索的准确性。同时由于缺乏成对的行人图像文本数据集,并且进行文本描述的标注工作量庞大,所以利用固定句式加提示学习来生成图像的文本描述方法来实现文本图像跨模态行人重识别具有很高的研究价值。

2) 多模态数据的统一。目前在可见光-红外,可见光-素描,可见光-文本,可见光-事件等跨模态行人重识别任务中取得了许多进展,但缺少能同时处理不同模态的统一 ReID 模型。例如,在真实世界中有时可能只能获取文本或者根据描述得到的素描图等单一模态的信息,统一模型能利用一个模态信息检索出目标行人;有时也能同时具有目标行人多种模态信息,也能同时利用多种模态信息检索目标行人。多模态的统一能提供更全面和多样的检索方式,可以单一模态输入,也可以多模态输入,具有更好的灵活性和交互性。

3) 复杂场景的适应性。加强模型对复杂场景、

极端光照和天气状况下的适应性,如拥挤场所,雨、雾、雪等恶劣天气条件下,以及在不同的光照条件(如黎明、黄昏、全光照)下的跨模态行人重识别。这涉及到更高级的场景分析和语义理解技能,是未来跨模态行人重识别在面对实际场景中的巨大挑战。

4) 无监督的持续学习。应用部署中,模型应该根据环境、人员变化进行持续学习,学习中既要学习环境和人员的新知识也要保留以前学习的旧知识。所以有必要研究持续学习在跨模态行人重识别中的应用。但在应用部署的持续学习过程中采集的新的行人图像数据集并没有标签,所以结合无监督技术实现无监督持续学习的跨模态行人重识别,在未来具有很高的研究意义和价值。

5) 识别过程的隐私保护。随着行人重识别技术在社会安防、智能监控系统等领域的应用,如何在确保个人隐私的同时实现高效的行人重识别成为一个重要的议题。未来研究可以关注隐私保护行人重识别技术的发展,以确保识别过程中个人信息不被泄露。

参考文献

- [1] ZHENG L, YANG Y, HAUPTMANN A G. Person re-identification: Past, present and future[EB/OL]. [2023 - 06 - 02]. <https://arxiv.org/abs/1610.02984>. DOI: 10.48550/arXiv.1610.02984.
- [2] HUANG T, RUSSELL S. Object identification in a Bayesian context[C]//International Joint Conference on Artificial Intelligence. New York, USA: ACM, 1997: 1276 - 1282.
- [3] ZAJDEL W, ZIVKOVIC Z, KROSE B J A. Keeping track of humans: Have I seen this person before? [C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2005: 2081 - 2086.
- [4] YI D, LEI Z, LIAO S, et al. Deep metric learning for person re-identification[C]//22nd International Conference on Pattern Recognition. Piscataway, USA: IEEE, 2014: 34 - 39.
- [5] BAI X, YANG M, HUANG T, et al. Deep-person: Learning discriminative deep features for person re-identification[J/OL]. Pattern Recognition, 2020[2023 - 11 - 02]. <https://www.sciencedirect.com/science/article/pii/S0031320319303395>. DOI: 10.1016/j.patcog.2019.107036.
- [6] LIN Y, ZHENG L, ZHENG Z, et al. Improving person re-identification by attribute and identity learning[J]. Pattern Recognition, 2019, 95: 151 - 161.
- [7] WEI P, ZHANG C, TANG Y, et al. Reinforced domain adaptation with attention and adversarial learning for unsupervised person ReID[J]. Applied Intelligence, 2023, 53(4): 4109 - 4123.
- [8] GU X, CHANG H, MA B, et al. Clothes-changing person re-identification with rgb modality only[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2022: 1060 - 1069.
- [9] WANG G, YANG S, LIU H, et al. High-order information matters: Learning relation and topology for occluded person re-identification[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2020: 6449 - 6458.
- [10] 叶钰, 王正, 梁超, 等. 多源数据行人重识别研究综述[J]. 自动化学报, 2020, 46(9): 1869 - 1884.
YE Y, WANG Z, LIANG C, et al. A survey on multi-source person re-identification[J]. Acta Automatica Sinica, 2020, 46(9): 1869 - 1884.
- [11] YE M, SHEN J, LIN G, et al. Deep learning for person re-identification: A survey and outlook[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(6): 2872 - 2893.
- [12] 管宇, 陈希爱, 田建东, 等. 基于多曝光图像生成的低照度图像增强[J]. 机器人, 2023, 45(4): 422 - 430.

- GUAN Y, CHEN X A, TIAN J D, et al. Low-light image enhancement based on multi-exposure images generation[J]. *Robot*, 2023, 45(4): 422–430.
- [13] 彭锦佳, 宋鹏鹏, 王辉兵. 面向隐私保护的联邦域锐化行人重识别方法[J]. *模式识别与人工智能*, 2023, 36(7): 634–646.
- PENG J J, SONG P P, WANG H B. Federated domain generalization person re-identification with privacy preserving[J]. *Pattern Recognition and Artificial Intelligence*, 2023, 36(7): 634–646.
- [14] 姚足, 龚勋, 陈锐, 等. 面向行人重识别的局部特征研究进展, 挑战与展望[J]. *自动化学报*, 2021, 47(12): 2742–2760.
- YAO Z, GONG X, CHEN R, et al. Research progress, challenge and prospect of local features for person re-identification[J]. *Acta Automatica Sinica*, 2021, 47(12): 2742–2760.
- [15] LIU X, ZHANG P, YU C, et al. Watching you: Global-guided reciprocal learning for video-based person re-identification [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2021: 13334–13343.
- [16] 周勇, 王瀚正, 赵佳琦, 等. 基于可解释注意力部件模型的行人重识别方法[J]. *自动化学报*, 2023, 49(10): 2159–2171.
- ZHOU Y, WANG H Z, ZHAO J Q, et al. Interpretable attention part model for person re-identification[J]. *Acta Automatica Sinica*, 2023, 49(10): 2159–2171.
- [17] REN M, HE L, LIAO X, et al. Learning instance-level spatial-temporal patterns for person re-identification[C]//IEEE/CVF International Conference on Computer Vision. Piscataway, USA: IEEE, 2021: 14930–14939.
- [18] WU A, ZHENG W S, YU H X, et al. RGB-infrared cross-modality person re-identification[C]//IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2017: 5380–5389.
- [19] BAK S, ZAIDENBERG S, BOULAY B, et al. Improving person re-identification by viewpoint cues[C]//11th IEEE International Conference on Advanced Video and Signal Based Surveillance. Piscataway, USA: IEEE, 2014: 175–180.
- [20] LI X, ZHENG W S, WANG X, et al. Multi-scale learning for low-resolution person re-identification[C]//IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2015: 3765–3773.
- [21] CHO Y J, YOON K J. Improving person re-identification via pose-aware multi-shot matching[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2016: 1354–1362.
- [22] HUANG H, LI D, ZHANG Z, et al. Adversarially occluded samples for person re-identification[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2018: 5098–5107.
- [23] LIU M, ZHANG Y, LI H. Survey of cross-modal person re-identification from a mathematical perspective[J]. *Mathematics*, 2023, 11(3): 654–658.
- [24] 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.
- [25] 万磊, 李华锋, 张亚飞. 多模态特征融合和自蒸馏的红外-可见光行人重识别[J/OL]. *计算机辅助设计与图形学学报*, 2024[2024-04-26]. <https://www.jcad.cn/cn/article/id/01ea269a-2c53-4c9e-8e24-dca9ce886115>.
- WAN L, LI H F, ZHANG Y F. Infrared-visible person re-identification via multi-modality feature fusion and self-distillation[J/OL]. *Journal of Computer-Aided Design & Computer Graphics*, 2024[2024-04-26]. <https://www.jcad.cn/cn/article/id/01ea269a-2c53-4c9e-8e24-dca9ce886115>.
- [26] 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.
- [27] 冯素坤, 王志全. 公共安全治理中可疑目标变形 Transformer 跟踪器[J]. *信息与控制*, 2024, 53(1): 98–107.
- FENG S K, WANG Z Q. Suspicious target Transformer tracker in public security governance[J]. *Information and Control*, 2024, 53(1): 98–107.
- [28] 闫於虎, 王永雄, 潘志群. 结合 Transformer 的显著性目标检测[J]. *信息与控制*, 2023, 52(3): 382–390.
- YAN Y H, WANG Y X, PAN Z Q. Salient object detection with Transformer[J]. *Information and Control*, 2023, 52(3): 382–390.
- [29] 王晋溪, 鲁鸣鸣. 基于场景图知识的文本到图像行人重识别[J]. *模式识别与人工智能*, 2024, 37(11): 947–959.
- WANG J X, LU M M. Scene graph knowledge based text-to-image person re-identification[J]. *Pattern Recognition and Artificial Intelligence*, 2024, 37(11): 947–959.
- [30] 葛斌, 卢洋, 夏晨辰, 等. 基于文本图像互学习的换衣行人重识别方法[J]. *模式识别与人工智能*, 2024, 37(11): 960–973.
- GE B, LU Y, XIA C C, et al. Clothes-changing person re-identification method based on text-image mutual learning[J]. *Pat-*

- tern Recognition and Artificial Intelligence, 2024, 37(11): 960–973.
- [31] ZHANG Y, ZHAO S, KANG Y, et al. Modality synergy complement learning with cascaded aggregation for visible-infrared person re-identification[C]//17th European Conference on Computer Vision. New York, USA: the ACM, 2022: 462–479.
- [32] LIU J, SUN Y, ZHU F, et al. Learning memory-augmented unidirectional metrics for cross-modality person re-identification [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: 2022: 19366–19375.
- [33] ZHANG Q, LAI C, LIU J, et al. Fmcnet: Feature-level modality compensation for visible-infrared person re-identification [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2022: 7349–7358.
- [34] HUANG Z, LIU J, LI L, et al. Modality-adaptive mixup and invariant decomposition for RGB-infrared person re-identification [C]//AAAI Conference on Artificial Intelligence. Keystone, USA: AIAA, 2022: 1034–1042.
- [35] JIANG K, ZHANG T, LIU X, et al. Cross-modality transformer for visible-infrared person re-identification[C]//17th European Conference on Computer Vision. Berlin, Germany: Springer, 2022: 480–496.
- [36] SUN Z, ZHAO F. Counterfactual attention alignment for visible-infrared cross-modality person re-identification[J]. Pattern Recognition Letters, 2023, 168: 79–85.
- [37] MUKHTAR H, KHAN M U G. CMOT: A cross-modality transformer for RGB-D fusion in person re-identification with online learning capabilities[J/OL]. Knowledge-Based Systems, 2024[2024–05–12]. <https://www.sciencedirect.com/science/article/pii/S095070512300905X>. DOI: 10.1016/j.knosys.2023.111155.
- [38] ZHENG X, HUANG X, JI C, et al. Multi-modal person re-identification based on transformer relational regularization[J/OL]. Information Fusion, 2024[2024–05–20]. <https://www.sciencedirect.com/science/article/pii/S156625352300444X>. DOI: 10.1016/j.inffus.2023.102128.
- [39] CHAI Z, LING Y, LUO Z, 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. .
- [40] 杜鹏, 宋永红, 张鑫瑶. 基于自注意力模态融合网络的跨模态行人再识别方法研究[J]. 自动化学报, 2022, 48(6): 1457–1468.
- DU P, SONG Y H, ZHANG X Y. Self-attention cross-modality fusion network for cross-modality person re-identification[J]. Acta Automatica Sinica, 2022, 48(6): 1457–1468.
- [41] YE M, SHEN J, SHAO L. Visible-infrared person re-identification via homogeneous augmented tri-modal learning[J]. IEEE Transactions on Information Forensics and Security, 2020, 16: 728–739.
- [42] WANG Z, WANG Z, ZHENG Y, et al. Learning to reduce dual-level discrepancy for infrared-visible person re-identification [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2019: 618–626.
- [43] YE M, RUAN W, DU B, et al. Channel augmented joint learning for visible-infrared recognition[C]//IEEE/CVF International Conference on Computer Vision. Piscataway, USA: IEEE, 2021: 13567–13576.
- [44] LI D, WEI X, HONG X, et al. Infrared-visible cross-modal person re-identification with an x modality[C]//AAAI Conference on Artificial Intelligence. Keystone, USA: AIAA, 2020: 4610–4617.
- [45] WEI Z, YANG X, WANG N, et al. Syncretic modality collaborative learning for visible infrared person re-identification[C]//IEEE/CVF International Conference on Computer Vision. Piscataway, USA: IEEE, 2021: 225–234.
- [46] WANG G A, ZHANG T, YANG Y, et al. Cross-modality paired-images generation for RGB-infrared person re-identification [C]//AAAI Conference on Artificial Intelligence. Keystone, USA: AIAA, 2020: 12144–12151.
- [47] 霍东东, 杜海顺. 基于通道重组和注意力机制的跨模态行人重识别[J]. 激光与光电子学进展, 2023, 60(14): 1417–1429.
- HUO D D, DU H S. Cross-modal person re-identification based on channel reorganization and attention mechanism[J]. Laser & Optoelectronics Progress, 2023, 60(14): 1417–1429.
- [48] FU C, HU Y, WU X, et al. CM-NAS: Cross-modality neural architecture search for visible-infrared person re-identification [C]//IEEE/CVF International Conference on Computer Vision. Piscataway, USA: IEEE, 2021: 11823–11832.
- [49] YE M, SHEN J, J. CRANDALL D, et al. Dynamic dual-attentive aggregation learning for visible-infrared person re-identification[C]//16th European Conference on Computer Vision. Berlin, Germany: Springer, 2020: 229–247.
- [50] CHOI S, LEE S, KIM Y, et al. Hi-CMD: Hierarchical cross-modality disentanglement for visible-infrared person re-identification[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2020: 10257–10266.
- [51] CHAN S, DU F, TANG T, et al. Parameter sharing and multi-granularity feature learning for cross-modality person re-identification[J]. Complex & Intelligent Systems, 2024, 10(1): 949–962.
- [52] HAO X, ZHAO S, YE M, et al. Cross-modality person re-identification via modality confusion and center aggregation[C]//

- IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2021: 16403 – 16412.
- [53] PARK H, LEE S, LEE J, et al. Learning by aligning: Visible-infrared person re-identification using cross-modal correspondences[C]//IEEE/CVF international conference on computer vision. Piscataway, USA: IEEE, 2021: 12046 – 12055.
- [54] WU Q, DAI P, CHEN J, et al. Discover cross-modality nuances for visible-infrared person re-identification[C]//The IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2021: 4330 – 4339.
- [55] HUANG N, LIU J, LUO Y, et al. Exploring modality-shared appearance features and modality-invariant relation features for cross-modality person re-identification[J/OL]. Pattern Recognition, 2023[2023 – 11 – 10]. <https://www.sciencedirect.com/science/article/pii/S0031320322006240>. DOI: 10.1016/j.patcog.2022.109145.
- [56] CHENG D, WANG X, WANG N, et al. Cross-modality person re-identification with memory-based contrastive embedding [C]//AAAI Conference on Artificial Intelligence. Keystone, USA: AIAA, 2023: 425 – 432.
- [57] LIANG T, JIN Y, LIU W, et al. Cross-modality Transformer with modality mining for visible-infrared person re-identification [J]. IEEE Transactions on Multimedia, 2023, 25: 8432 – 8444.
- [58] LU H, ZOU X, ZHANG P. Learning progressive modality-shared transformers for effective visible-infrared person re-identification[C]//AAAI Conference on Artificial Intelligence. Keystone, USA: AIAA, 2023, 37(2): 1835 – 1843.
- [59] ZHENG L, HUANG Y, LU H, et al. Pose-invariant embedding for deep person re-identification[J]. IEEE Transactions on Image Processing, 2019, 28(9): 4500 – 4509.
- [60] MIAO Y, HUANG N, MA X, et al. On exploring pose estimation as an auxiliary learning task for visible-infrared person re-identification[J/OL]. Neurocomputing, 2023[2023 – 11 – 09]. <https://www.sciencedirect.com/science/article/pii/S0925231223007750>. DOI: 10.1016/j.neucom.2023.126652.
- [61] WAN L, SUN Z, JING Q, et al. G2DA: Geometry-guided dual-alignment learning for RGB-infrared person re-identification[J/OL]. Pattern Recognition, 2023[2023 – 11 – 10]. <https://www.sciencedirect.com/science/article/pii/S003132032200629X>. DOI: 10.1016/j.patcog.2022.109150.
- [62] CUI J, CHAN S, MU P, et al. Pure detail feature extraction network for visible-infrared re-identification[J/OL]. Intelligent Automation & Soft Computing[2024 – 01 – 21]. <https://www.techscience.com/iasc/v37n2/53259/html>. DOI: 10.32604/iasc.2023.039894.
- [63] CHEN C, YE M, QI M, et al. Structure-aware positional transformer for visible-infrared person re-identification[J]. IEEE Transactions on Image Processing, 2022, 31: 2352 – 2364.
- [64] FENG J, WU A, ZHENG W. Shape-erased feature learning for visible-infrared person re-identification[C]//The IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2023: 22752 – 22761.
- [65] HUANG H, YANG W, LIN J, et al. Improve person re-identification with part awareness learning[J]. IEEE Transactions on Image Processing, 2020, 29: 7468 – 7481.
- [66] WANG X, LIU M, WANG F, et al. Relation-preserving feature embedding for unsupervised person re-identification[J]. IEEE Transactions on Multimedia, 2023, 26: 714 – 723.
- [67] YANG L. Continuous epoch distance integration for unsupervised person re-identification[C]//5th International Conference on Communications, Information System and Computer Engineering. Piscataway, USA: IEEE, 2023: 464 – 469.
- [68] LAN L, TENG X, ZHANG J, et al. Learning to purification for unsupervised person re-identification[J]. IEEE Transactions on Image Processing, 2023, 32: 3338 – 3353.
- [69] ZHU X, MORERIO P, MURINO V. Unsupervised domain-adaptive person re-identification based on attributes[C]//IEEE International Conference on Image Processing. Piscataway, USA: IEEE, 2019: 4110 – 4114.
- [70] GUO J, LIU H, SHI W, et al. Unsupervised domain adaptation person re-identification by camera-aware style decoupling and uncertainty modeling[C]//IEEE International Conference on Image Processing. Piscataway, USA: IEEE, 2022: 761 – 765.
- [71] BAI Y, WANG C, LOU Y, et al. Hierarchical connectivity-centered clustering for unsupervised domain adaptation on person re-identification[J]. IEEE Transactions on Image Processing, 2021, 30: 6715 – 6729.
- [72] NI H, LI Y, GAO L, et al. Part-aware transformer for feneralizable person re-identification[C]//IEEE/CVF International Conference on Computer Vision. Piscataway, USA: IEEE, 2023: 11280 – 11289.
- [73] 张栩嘉, 白淼源, 郭继峰. 结合域间一致性与抗噪学习的跨域行人重识别方法[J]. 小型微型计算机系统, 2024, 45(6): 1444 – 1450.
- ZHANG X J, BAI M Y, GUO J F. Cross-domain person re-identification method combining inter-domain consistency and anti-noise learning[J]. Journal of Chinese Computer Systems, 2024, 45(6): 1444 – 1450.

- [74] 杨章静, 吴数立, 黄璞, 等. 基于相机感知的域自适应行人重识别形式[J]. 模式识别与人工智能, 2024, 37(5): 383–397. YANG Z J, WU S L, HUANG P, et al. Domain adaptive person re-identification model based on camera perception[J]. Pattern Recognition and Artificial Intelligence, 2024, 37(5): 383–397.
- [75] CHEN H, LAGADEC B, BREMOND F. Ice: Inter-instance contrastive encoding for unsupervised person re-identification[C]//IEEE/CVF International Conference on Computer Vision. 2021: 14960–14969.
- [76] DAI Z, WANG G, YUAN W, et al. Cluster contrast for unsupervised person re-identification[C]//Asian Conference on Computer Vision. Piscataway, USA: IEEE, 2022: 1142–1160.
- [77] ZHANG X, LI D, WANG Z, et al. Implicit sample extension for unsupervised person re-identification[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2022: 7369–7378.
- [78] JIANG K, ZHANG T, ZHANG Y, et al. Self-supervised agent learning for unsupervised cross-domain person re-identification[J]. IEEE Transactions on Image Processing, 2020, 29: 8549–8560.
- [79] WU X, HUI Y, HU X, et al. Unsupervised cross-domain person re-identification method combining improved squeeze and excitation block[C]//International Conference on Computer Network, Electronic and Automation. Piscataway, USA: IEEE, 2022: 238–242.
- [80] YAO L, LIN B Y, HAQ Q M U, et al. Unsupervised cross-domain adaptation through mutual mean learning and GANs for person re-identification[C]//3rd International Conference on Artificial Intelligence. Piscataway, USA: IEEE, 2023: 122–128.
- [81] HAN G, ZHANG X, LI C. One-shot unsupervised cross-domain person re-identification[J/OL]//IEEE Transactions on Circuits and Systems for Video Technology, 2023[2024–01–23]. <https://ieeexplore.ieee.org/abstract/document/10176137>. DOI: 10.1109/TCSVT.2023.3293130.
- [82] LIN X, REN P, YE H C, et al. Unsupervised person re-identification: A systematic survey of challenges and solutions[EB/OL]. [2024–01–23]. <https://arxiv.org/abs/2109.06057>. DOI: 10.48550/arXiv.2109.06057.
- [83] WANG M, LAI B, HUANG J, et al. Camera-aware proxies for unsupervised person re-identification[C]//AAAI Conference on Artificial Intelligence. Keystone, USA: AIAA, 2021: 2764–2772.
- [84] LI M, LI C G, GUO J. Cluster-guided asymmetric contrastive learning for unsupervised person re-identification[J]. IEEE Transactions on Image Processing, 2022, 31: 3606–3617.
- [85] YIN J, ZHANG X, MA Z, et al. A real-time memory updating strategy for unsupervised person re-identification[J]. IEEE Transactions on Image Processing, 2023, 32: 2309–2321.
- [86] PANG Z, WANG C, WANG J, et al. Reliability modeling and contrastive learning for unsupervised person re-identification[J/OL]. Knowledge-Based Systems, 2023[2023–11–09]. <https://www.sciencedirect.com/science/article/pii/S0950705123000138>. DOI: 10.1016/j.knsys.2023.11026.
- [87] LI J, WANG M, GONG X. Transformer based multi-grained features for unsupervised person re-identification[C]//IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway, USA: IEEE, 2023: 42–50.
- [88] ZHANG G, ZHANG Y, ZHANG T, et al. PHA: Patch-wise high-frequency augmentation for transformer-based person re-identification[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2023: 14133–14142.
- [89] WANG T, LIU H, SONG P, et al. Pose-guided feature disentangling for occluded person re-identification based on transformer[C]//AAAI Conference on Artificial Intelligence. Keystone, USA: AIAA, 2022: 2540–2549.
- [90] PANG Z, WANG C, ZHAO 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.
- [91] LIANG W, WANG G, LAI J, et al. Homogeneous-to-heterogeneous: Unsupervised learning for rgb-infrared person re-identification[J]. IEEE Transactions on Image Processing, 2021, 30: 6392–6407.
- [92] YANG B, CHEN J, YE M. Towards grand unified representation learning for unsupervised visible-infrared person re-identification[C]//IEEE/CVF International Conference on Computer Vision. Piscataway, USA: IEEE, 2023: 11069–11079.
- [93] YANG B, YE M, CHEN J, et al. Augmented dual-contrastive aggregation learning for unsupervised visible-infrared person re-identification[C]//30th ACM International Conference on Multimedia. New York, USA: ACM, 2022: 2843–2851.
- [94] SI T, HE F, LI P, et al. Diversity feature constraint based on heterogeneous data for unsupervised person re-identification[J/OL]. Information Processing & Management, 2023[2023–11–10]. <https://www.sciencedirect.com/science/article/pii/S0306457323000419>. DOI: 10.1016/j.ipm.2023.103304.
- [95] WANG J, ZHANG Z, CHEN M, et al. Optimal transport for label-efficient visible-infrared person re-identification[C]//European Conference on Computer Vision. Berlin, Germany: Springer, 2022: 93–109.
- [96] WU Z, YE M. Unsupervised visible-infrared person re-identification via progressive graph matching and alternate learning[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2023: 9548–9558.

- [97] GU X, CHANG H, MA B, et al. Appearance-preserving 3d convolution for video-based person re-identification[C]//16th European Conference on Computer Vision. Berlin, Germany: Springer, 2020: 228 – 243.
- [98] LIU Y, YUAN Z, ZHOU W, et al. Spatial and temporal mutual promotion for video-based person re-identification[C]//AAAI Conference on Artificial Intelligence. Keystone, USA: AIAA, 2019: 8786 – 8793.
- [99] JIANG X, QIAO Y, YAN J, et al. SSN3D: Self-separated network to align parts for 3D convolution in video person re-identification[C]//The AAAI Conference on Artificial Intelligence. Keystone, USA: AIAA, 2021: 1691 – 1699.
- [100] LI J, ZHANG S, HUANG T. Multi-scale 3d convolution network for video based person re-identification[C]//AAAI Conference on Artificial Intelligence. Keystone, USA: IEEE, 2019: 8618 – 8625.
- [101] KOESTINGER M, HIRZER M, WOHLHART P, et al. Large scale metric learning from equivalence constraints[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2012: 2288 – 2295.
- [102] MCLAUGHLIN N, RINCON J M, MILLER P. Recurrent convolutional network for video-based person re-identification[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2016: 1325 – 1334.
- [103] XU S, CHENG Y, GU K, et al. Jointly attentive spatial-temporal pooling networks for video-based person re-identification [C]//IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2017: 4733 – 4742.
- [104] ZHANG W, YU X, HE X. Learning bidirectional temporal cues for video-based person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 28(10): 2768 – 2776.
- [105] ZHOU Z, HUANG Y, WANG W, et al. See the forest for the trees: Joint spatial and temporal recurrent neural networks for video-based person re-identification [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2017: 4747 – 4756.
- [106] LI S, BAK S, CARR P, et al. Diversity regularized spatiotemporal attention for video-based person re-identification [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2018: 369 – 378.
- [107] EOM C, LEE G, LEE J, et al. Video-based person re-identification with spatial and temporal memory networks[C]//IEEE/CVF International Conference on Computer Vision. Piscataway, USA: IEEE, 2021: 12036 – 12045.
- [108] WANG Y, ZHANG P, GAO S, et al. Pyramid spatial-temporal aggregation for video-based person re-identification[C]//IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2021: 12026 – 12035.
- [109] PATHAK P, ESHRATIFAR A E, GORMISH M. Video person re-id: Fantastic techniques and where to find them (student abstract)[C]//AAAI Conference on Artificial Intelligence. Keystone, USA: AIAA, 2020: 13893 – 13894.
- [110] LI J, WANG J, TIAN Q, et al. Global-local temporal representations for video person re-identification[C]//IEEE/CVF International Conference on Computer Vision. Piscataway, USA: IEEE, 2019: 3958 – 3967.
- [111] ZANG X, LI G, GAO W. Multidirection and multiscale pyramid in transformer for video-based pedestrian retrieval[J]. IEEE Transactions on Industrial Informatics, 2022, 18(12): 8776 – 8785.
- [112] HOU R, CHANG H, MA B, et al. Bicnet-tks: Learning efficient spatial-temporal representation for video person re-identification[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2021: 2014 – 2023.
- [113] LIN X, LI J, MA Z, et al. Learning modal-invariant and temporal-memory for video-based visible-infrared person re-identification[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 20973 – 20982.
- [114] LI H, XU L, ZHANG Y, et al. Adversarial self-attack defense and spatial-temporal relation mining for visible-infrared video person re-identification [EB/OL]. (2023 – 06 – 13) [2023 – 11 – 20]. <https://arxiv.org/abs/2307.03903>. DOI: 10.48550/arXiv.2307.03903.
- [115] 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 – 609.
- [116] ZHANG Y, WANG H. Diverse embedding expansion network and low-light cross-modality benchmark for visible-infrared person re-identification[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2023: 2153 – 2162.
- [117] ZHENG A, WANG Z, CHEN Z, et al. Robust multi-modality person re-identification[C]//AAAI Conference on Artificial Intelligence, Keystone, USA: AAAI, 2021: 3529 – 3537.

作者简介

范慧杰(1985 –), 女, 博士, 研究员。研究领域为机器视觉, 图像处理, 模式识别。

郁航(2000 –), 男, 硕士生。研究领域为计算机视觉与图像处理。

赵颖畅(1999 –), 女, 硕士生。研究领域为计算机视觉与图像处理。