

文章编号:1671-4229(2022)02-0030-12

基于深度学习的行人重识别综述

黄业文, 许智聪, 单纯*

(广东技术师范大学 电子与信息学院, 广东 广州 510665)

摘要: 行人重识别是指从不同的摄像机采集的大量行人图像中检索特定的人员,是计算机视觉领域继人脸识别之后的一个重要的研究子课题。近年来,将深度学习应用在行人重识别方面已经取得了很大的成就。为了让读者更加了解国内外行人重识别研究进展,文章介绍了近几年国内外的相关文献并进行分析总结。首先,对行人重识别进行介绍;然后,对基于深度学习的行人重识别的基本方法归纳为有监督学习、弱监督学习和跨模态学习等3个方向,并讨论它们的现状,列举了常用的性能指标,以及常用的数据集和目前在这些数据集上最优的性能,并对行人重识别现存的问题进行阐述;最后,对未来研究热门方向进行了展望。

关键词: 行人重识别; 有监督学习; 弱监督学习; 跨模态

中图分类号: TP 391 **文献标志码:** A

A survey of person Re-ID based on deep learning

HUANG Ye-wen, XU Zhi-cong, SHAN Chun*

(School of Electronics and Information, Guangdong Polytechnic Normal University, Guangzhou 510665, China)

Abstract: Person re-identification (Re-ID) refers to retrieving a specific person from a large number of pedestrian images collected by different cameras, which is an important sub-topic in the field of computer vision after face recognition. In recent years, great achievements have been made in applying deep learning to person re-identification. In order to let readers know more about the research progress of pedestrian re-identification at home and abroad, this paper introduces the relevant literature at home and abroad in recent years and makes analysis and summary. Firstly, person re-identification is introduced, and then the basic methods of person re-identification based on deep learning are summarized into three directions: supervised learning, weakly supervised learning, and cross-modal learning, and their status quo is discussed. Common performance indicators are listed, as well as commonly used datasets and the current best performance on these datasets. After that, the existing problems of person re-identification are expounded, and finally, the future research direction has prospected.

Key words: person re-identification; supervised learning; weakly supervised learning; cross-modality

行人重识别(Person re-identification, Re-ID)是将被不同摄像头捕捉到的特定的人进行检索,由于其在监控系统中的实际重要性,重新识别技术引起了广泛的研究与关注,这对安全监控和行人行为分析具有重要意义。

当前行人重识别被认为是一个图像检索的子问题。行人重识别的主要作用是给定一个监控行人图像,然后通过检索其他监控拍的行人图像,找出该行人在其他监控下的图像。在监控视频中,由于相机分辨率和拍摄角度

基金项目: 国家重点研发计划资助项目(2018YFB1802402); 南方海洋科学与工程广东省实验室(珠海)创新团队建设资助项目(311021011); 广东省普通高校青年创新人才类资助项目(2019KQNCX069); 广州市基础研究计划基础与应用基础研究资助项目(202102020307)

作者简介: 黄业文(1979—),男,博士研究生. E-mail: yewenh@gpnu.edu.cn

* 通信作者. E-mail: shanchun@gpnu.edu.cn

引文格式: 黄业文, 许智聪, 单纯. 基于深度学习的行人重识别综述[J]. 广州大学学报(自然科学版), 2022, 21(2): 30-41.

的原因,人脸有可能被遮挡或者比较模糊,通常情况下无法得到质量比较高的人脸图片。所以,在当人脸识别失效的情况下,行人重识别这一技术就成了非常重要的替代技术。因为行人重识别的关键技术是跨摄像头追踪同一行人的行走轨迹,所以如何从不同摄像头监控图像中准确识别同一行人是行人重识别研究的主要问题。

然而,由于一个人在不同摄像机之间的类内变化很大,如遮挡、光照、视点和姿态,因此,在计算机视觉领域,重新识别仍然是一个具有挑战性的任务。如何提取特征和匹配特征是行人重识别的关键问题。2016年之前的行人重识别大量的工作是基于传统计算机视觉方法,利用手工提取一些低级视觉特征^[1-5],进行行人重识别的研究。

行人重识别虽然已经在学术界研究多年,但是自2016年以来,随着深度学习在许多领域的成功应用,研究人员开始尝试将深度学习应用到行人重识别上,这项研究才取得比较好的成果。深度学习不仅应用于提取高级特征,也为度量学习的研究带来了新的思路。虽然当前深度学习在规模较小的数据集上的结果没有明显的提升,但是随着越来越多新的性能更好的模型的提出以及大规模的数据集的出现,许多研究者投入到深度学习应用在行人重识别这一研究上。目前,这些深度学习的方法主要研究思路为采用深度学习方法进行特征向量提取,再利用度量学习对特征向量进行判别,将图像之间的差异量化,并使用大量标注数据对模型进行训练迭代。主要的流程是从待查询(Quary)的行人图像中利用神经网络提取特征,然后将提取的特征和其他摄像头拍摄的图像(Gallery)进行相似度对比,从中得出最为相似的图像并输出结果,如图1所示。将基于深度学习的方法分为有监督学习、弱监督学习和跨模态学习3个部分,并且就行人重识别目前的常用数据集进行了一些比较及其最优的模型分析。

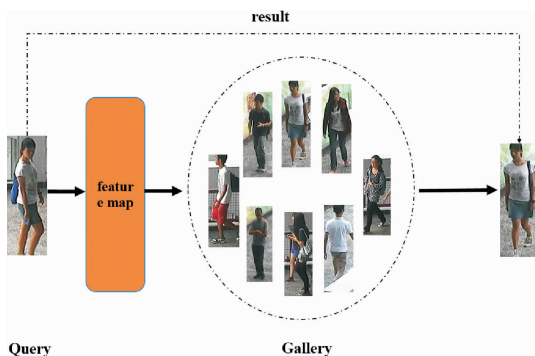


图1 行人重识别模型原理

Fig. 1 Principle of the person re-identification model

1 基于深度学习的行人重识别方法

近年来,随着深度学习在许多领域的成功应用,研究人员开始尝试将深度学习应用到行人重识别上,并已取得一定的成果。其主要研究思路为采用深度学习方法进行特征向量提取,再利用度量学习对特征向量进行判别,将图像之间的差异进行量化,并使用大量标注数据对模型进行训练迭代。现有的依赖于深度学习的行人重识别模型绝大部分采用的是有监督方式,即训练数据需要人工打上标签,然后进行训练。目前在这种有监督的方式下的识别率可以做得很高。但是,这样的方式需要消耗大量的人力物力,而且大多数的图像数据都比较少或者没有标签,因此,弱监督的方式被提出来。这种方式不需要大量的标签数据,由于其目前识别率较低且贴近现实生活,现在也是热门的研究领域。弱监督学习包括在只有少量标签的数据集中进行特征提取的半监督学习,以及在无任何标签的数据集中特征提取的无监督学习。

面对复杂的实际场景,有研究人员开始进入到跨模态相关的研究中。跨模态学习是指待处理的行人图像不同于在正常太阳光下的摄像头拍摄的RGB图像,而是一些其他模态的图像。将现有的跨模态学习分为可见光-红外行人重识别、深度图像行人重识别、文本-图像行人重识别和跨分辨率行人重识别4种类别。

1.1 有监督学习

有监督学习是一直以来的研究热点,依据充分标签的实验数据,提取具有判别力的特征数据,依据提取的特征来区别行人,从而达到识别的目的。现有的行人重识别的模型,将它们分为特征学习、度量学习和排序学习。

1.1.1 特征学习

特征学习,即表示特征学习,特征学习是对复杂的原始数据化繁为简,把原始数据的无效信息剔除,把有效信息更有效地进行提炼,形成特征,它的目的是提取图像特征,便于图像之间差异的量化。随着深度学习的发展,广泛使用卷积神经网络作为特征提取器,提取图像或者视频的特征向量。

在图像分类领域中,已有一些较为有效的特征提取方法^[6-8],因此在行人重识别研究的初期,一些学者将行人重识别转换为分类问题进行研究,全局特征学习是主要的研究方法。全局特征学习为每一个行人图像提取全局特征向量。全局特征学习比较简单,Wang等^[9]

提出了一种结合单张图像特征的方法,对输入的图像对,分别使用独立的卷积神经网络提取各自的特征,并抽取特征提取过程中的特征图进行融合,这种方法结合了提取单张图像特征的高效率以及 CIR 方法对图像间信息提取的优势。Chen 等^[10]将注意力机制和全局特征学习结合起来,提出了一种自我批判的注意学习方法进行行人重识别。全局特征提取方法虽然能较好地提取图像特征,但是由于行人重识别的数据集存在着遮挡、部位不齐等问题,直接提取全局特征并不能达到较好的效果,由此,有学者把局部特征学习用于行人重识别。

局部特征学习是从图像局部区域中抽取的特征,包括边缘、角点、线、曲线和特别属性的区域等。早期的一些方法,将图像按照预设的划分方式进行划分,但当使用从不同角度拍摄的数据时,固定的分割方式并不能很好地将图像中身体各部位分开,图像间对应区域包含的部位也不能很好地对应。Sun 等^[11]提出了一个基于零件的卷积基线(PCB)来学习零件信息的特征。PCB 采用简单的统一划分策略,将部分信息特征组装成卷积描述符。Li 等^[12]提出一个多尺度上下文感知网络(MSCAN)来捕获行人特征学习的上下文知识,通过在每一层叠加多尺度卷积,可以较好地捕获局部上下文知识。近年来,有学者丢弃以往基于卷积神经网络的框架,使用别的框架来实现算法的提升。He 等^[13]首次提出基于视觉 Transformer(ViT)的行人重识别框架,能简单自然地将相机和视点等非可视化信息编码为向量嵌入到训练过程,并在多个数据集上表现出优秀性能,显示此框架具有较大的开发潜能。

1.1.2 度量学习

对于行人重识别来说,特征学习是把行人图像特征提取,但是还需要应用度量学习把提取的特征进行量化对比、分析相似度,从而识别行人。以前的度量学习是构造矩阵,现在深度学习框架下的度量学习主要体现在损失函数上,用来更新网络参数,从而提高图像的识别度。目前主流研究的损失函数主要是对比损失函数、分类损失函数和三元组损失函数。

对比损失函数:对比损失函数用于训练孪生网络(Siamese network)^[14],孪生网络的输入是一组共 2 张图片(I_a 和 I_b),设定这一组照片可以为同一行人,也可以为不同行人。每一组训练图片都有一个标签 y ,其中, $y=1$ 表示 2 张图片属于同一个行人,称之为正样本组,反之 $y=0$ 表示它们属于不同行人,称之为负样本组,因此,变量 y 值为 +1 或 -1,作为图像组的标签,引入阈值 margin,对图像对距离加以约束。最后对比损失函数:

$$l_{con} = (1 - \delta_{ij}) \{ \max(0, \rho - d_{ij}) \}^2 + \delta_{ij} d_{ij}^2 \quad (1)$$

其中, d_{ij} 表示 2 个输入样本 x_i 和 x_j 的嵌入特征之间的欧几里德距离。 δ_{ij} 是一个二元标签指示符(当 x_i 和 x_j 属于同一身份时, $\delta_{ij}=1$, 否则 $\delta_{ij}=0$), ρ 是训练阈值参数。

分类损失函数^[15]:分类损失函数又称 ID 损失函数,训练集中行人的 ID 数为网络的类别数,特征层后接一个分类全连接层,经过 Softmax 激活函数计算交叉熵损失。测试阶段使用倒数第二层的特征向量进行检索,分类全连接层丢弃。每批中的训练样本数为 n , 给定一个标签为 y_i 的输入图像 x_i , 通过交叉熵计算 ID 损失函数。

$$l_{id} = -\frac{1}{n} \sum_{i=1}^n q(k) \lg(p(y_i | x_i)) \quad (2)$$

三元组损失函数^[16]:三元组损失函数基本思想是正样本对之间的距离应该比负样本对之间的距离小于预定义的余量。通常,一个三元组包含一个固定样本 x_i 、一个相同身份的正样本 x_j 和一个不同身份的负样本 x_k 。带有阈值参数的三元组损失表示为

$$l_{tr}(i, j, k) = \max(\rho + d_{ij} - d_{ik}, 0) \quad (3)$$

其中, ρ 是训练阈值参数, d_{ij} 表示正样本的欧几里得距离, d_{ik} 表示负样本的欧几里得距离。三元组损失函数结构简单,如果直接使用,会造成网络鉴别能力下降。由此 Hermans 等^[17]指出,让网络一直学习简单的样本组合会限制网络的泛化能力,针对这一问题,提出三元组损失函数的改进方法,通过选择一个批次中较难区分的正样本和负样本,来对三元组损失函数进行训练,加强三元组损失函数对于困难样本对的挖掘能力,从而提升三元组损失函数的性能。图 2 为 3 种损失函数的示意图。

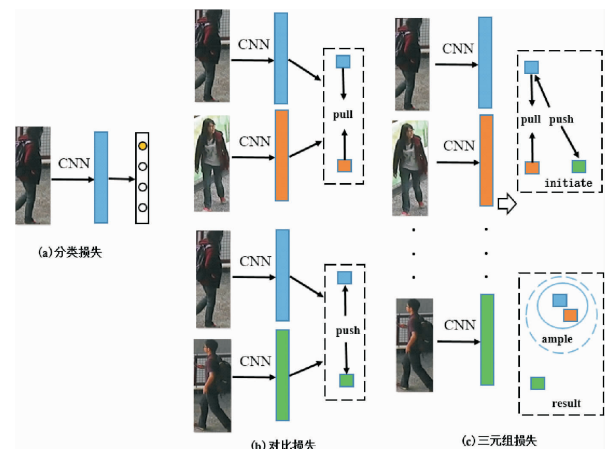


图 2 3 种损失函数对比

Fig. 2 Comparison of three loss functions

1.1.3 排序学习

除了对于特征学习与度量学习的探索,近年来,一些研究学者将主要精力集中于对行人重识别结果排序

的优化上,即结果重排序(Re-ranking)。一般情况下,给定一张目标图像,通过提取图像特征并计算目标图像与待匹配图像(Gallery)的距离,得到匹配图像的排序结果,根据该结果计算 Rank-n Accuracy,并绘制 CMC 曲线。重排序就是利用相关方法,对当前的排序结果进行优化,从而提升行人重识别的精度,如图 3。

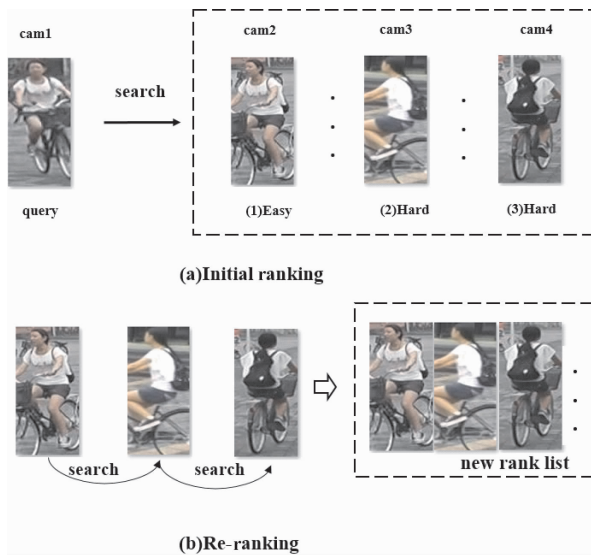


图 3 重排序示例

Fig. 3 Example of re-ranking

Luo 等^[18] 推导出了一一种名为 Local Blurring Re-ranking 的轻量级重排序方法,采用聚类结构来改进邻域相似性度量。考虑到 Query 的差异,一些方法设计了 Query 自适应检索策略来代替统一搜索引擎以提高性能。Zhou 等^[19] 提出了一种新颖的在线局部度量自适应算法,用于在测试阶段为每个探针学习专用的马氏度量。这种方法仅使用负样本进行度量适应,这在实际情况是实用的。它在很大程度上减少了现有 Re-ID 方法对大量正训练数据的需求,并且只产生最小的计算成本来执行在线训练。Ye 等^[20] 提出了一种基于 KNN 的排序聚合方法,分别提取图像的全局特征与局部特征,分别根据两者得到结果的排序,通过对 2 种特征的结合,利用最近邻算法对排序结果进行优化。Li 等^[21] 率先提出了利用最近邻算法对行人重识别的排序结果进行优化的方法。通过相关方法对排序结果进行优化,在没有大幅增加计算量与参数数量的前提下,有效地提升了行人重识别算法的精度,同时重排序方法可以直接应用到其他高精度的行人重识别算法中,是一种提升模型性能的有效方法,同时也是未来的一个热门研究方向。

1.2 弱监督学习

尽管现在在有监督的行人重识别的识别率已经很高

了,但是人工打标签是一件费时费力的事。因此,越来越多的研究人员投入到更接近实际情况的少标签或者无标签的弱监督学习中。此类学习方法是对于输入数据少量标记或者无标记,也没有确定的结果。因为输入的样本数据类型未知,所以需要根据样本间的相似性来对样本集进行分类,然后用相关方法使得样本数据类内差距最小化,类间差距最大化。弱监督学习分为半监督学习和无监督学习 2 种情况。

1.2.1 半监督学习

近年来,研究者开始关注标签少的半监督学习中,研究如何利用少量的标签应用在行人重识别中。研究者面临的问题是:如何学习训练中的少量的标签的特征数据,并且应用在测试集中的大量无标签的数据,并且建立优化模型。由于标注数据有限,Bak 等^[22] 提出一种新颖学习方法来学习度量,假设一个度量可以分成独立的颜色和纹理分量而不会损失性能。基于此,对于纹理,该方法只学习深度颜色不变的特征,另外学习色块的颜色度量来捕获特定相机对的颜色变化。Wang 等^[23] 提出了一种新的网络由一个共享的特征提取主干和 2 个分支组成,分别用于摄像机内和摄像机间的学习。根据 ICS 的逐相机标记特性,提出了联合学习相机特有的非参数分类器和混合挖掘五元组损失的相机内学习方法。设计的组件充分利用了每个相机的标签,因此,该方法的相机内学习部分只比大多数现有的 ICS 方法性能更好。

1.2.2 无监督学习

无监督学习使用的数据是无标签的,因此更具有适应性和鲁棒性。Fu 等^[24] 提出了一种自相似分组(Self-similarity Grouping, SSG)方法,该方法利用未标记样本的潜在相似性(从整体到局部),从不同的视图自动构建多个聚类。然后给这些独立的集群分配标签,这些标签作为伪身份来监督培训过程。Ye 等^[25] 设计了一个动态图匹配(DGM)框架,通过从中间估计的标签学习更好的相似性度量,迭代地细化图结构,从而改进标签估计过程。此外,还设计了一种正加权策略来细化中间标签,提高了对不准确匹配输出和噪声初始训练数据的鲁棒性。为了进一步提高性能,Liao 等^[26] 提出的 QAConv 方法在不需要进一步迁移学习的情况下,具有较好的匹配效果,且具有比现有基线更好的泛化能力。一些研究还尝试了局部特征学习,Yang 等^[27] 提出了一个基于 patch 的无监督学习框架(PAUL),该框架设计了 Patch-Net 来从人脸图像的特征图中抽取 patch,并在一个没有标记的重新标识数据集上有区别地学习 patch 特征。为此,还开发了一种补丁鉴别特征学习损耗,为在无标记

的 RE-ID 数据集上学习具有鉴别性的补丁特征提供了有效的指导。

由于数据标签的缺失,有些研究者进行将有标记的源域的判别性信息迁移到目标域^[28],由于源域数据集强大的监督学习,它是另一种流行的没有目标数据集标签的无监督行人重识别方法。学者们利用生成对抗网络(Generative Adversarial Networks, GAN)^[29],来进行数据集的扩充。使用生成的图像,可以实现在未标记的目标域中实现有监督的行人重识别模型学习,从而加强行人重识别模型的泛化能力。Zheng 等^[30]第一次尝试将 GAN 技术应用到行人重识别,它使用生成的行人图像改进监督特征表示学习,然而生成图像质量低,对 Re-ID 模型性能提高有限。Wei 等^[31]提出 PTGAN,这是为弥补数据集之间的领域差距而提出的一项关于人员转移的原创工作。PTGAN 能有效地减小间隙。不同的相机可能呈现不同的风格,这使得一个映射功能很难实现多种风格的转换。面对缺乏交叉视图配对训练数据,以及在存在大姿态变化的情况下学习区分身份敏感和视图不变特征这 2 个问题,Qian 等^[32]提出了一种新的基于位姿条件的真实感人物图像生成模型来解决这 2 个问题。该模型专门为 Re-ID 中的姿态归一化而设计,称为姿态归一化 GAN(PN-GAN)。利用合成的图像,该模型可以学习一种新的不受姿态变化影响的深度识别特征。

有些方法直接在源域训练好模型,然后直接应用在目标域上。Liu 等^[33]提出了一种新的自适应转移网络(ATNet)的跨域行人重识别。该网络提出了一种自适应集成策略,通过感知各因子对图像的影响程度来融合因子方向的传输。这种“分解与集成”的策略使 ATNet 能够在要素水平上进行精确的风格迁移,并最终实现跨域的有效迁移。Chen 等^[34]提出了一种新的基于实例的上下文呈现方案,用于跨领域的 Re-ID 模型学习。通过精心设计的双重条件映射,丰富的目标实例被用作图像生成的上下文指导。由于领域差距和不令人满意的聚类性能,上述的方法没有充分利用所有有价值的信息。为了解决这些问题,Ge 等^[35]提出了一种基于混合记忆的自定步长对比学习框架。混合内存动态地生成源域类级、目标域集群级和非集群实例级监督信号,用于学习特征表示。与传统的对比学习策略不同,该框架将源域类、目标域集群和非集群实例结合起来进行区分。最重要的是,所提出的自定步长方法逐渐创建更可靠的集群来细化混合记忆和学习目标,这是该模型取得优异性能的关键。

1.3 跨模态学习

跨模态行人重识别是指不同类型行人数据相互匹

配的问题。大多数行人重识别主要集中在同一模态下的行人重识别上,如监控下的 RGB 图像,这些图像大部分采集于光源比较充足的场景。但是随着现实生活中出现的弱光甚至是无光情况下,因此,为了克服可见光摄像头无法全天候进行图像采集的缺点,越来越多的研究者关注从其他模态对行人进行识别。

跨模态行人重识别是近年来计算机视觉领域的热点问题,在实际生活除了一般行人 RGB 图像还有许多其他模态的图像,如红外图像、深度图像、文本信息和跨分辨率图像等。因此,跨模态行人重识别与一般行人重识别相比更具有挑战性和实用性。

1.3.1 可见光-红外行人重识别

可见光-红外行人再识别(VI-ReID)的目标是实现跨模态行人图像的匹配,突破了单模态行人重识别在黑暗环境下的局限性。可见光-红外行人再识别(VI-ReID)的问题,即在交叉模式设置中检索一组由可见或红外摄像机捕获的人员图像。VI-ReID 的 2 个主要挑战是人物图像的类内变化,以及可见光和红外图像之间的跨模态差异。Wu 等^[36]首次尝试解决这个问题,提出了像素级对齐和联合判别策略,以自适应地学习模态共享特性。Fu 等^[37]发现适当地分离 BN 层可以产生更好的性能,由此提出了一种新的 CM-NAS 来解决具有挑战性的 VI-ReID。该方法开发一种面向 BN 的 NAS 算法,能够自动决定 BN 层的分离,搜索最优的架构。Ye 等^[38]提出了一个层次化的交叉模态匹配模型,通过联合优化模态特定和模态共享的度量。特定于模态的度量将 2 种不同的模态转化为一致的空间,从而学习模态共享度量。不同光谱相机拍摄的异构行人图像在图像风格上存在显著差异,导致特征表征的可分辨性较差。为了解决这个问题,Wei 等^[39]提出了一种新的融合模态协同学习(SMCL)模型来学习 VI-ReID 的模态不变标识-判别表示。融合模态的自生成特征保留了可见光和红外图像的重要信息,可以引导网络将异构图像投射到公共空间,挑战增强同质性学习和辅助分布相似性学习。利用跨模态人物图像之间的密集对应关系,Park 等^[40]提出了一种新颖的特征学习框架双流 CNN 框架,从相应的人物图像中提取 RGB 和 IR 特征,然后将特征与 CMAlign 模块对齐。它在 RGB 和 IR 特征之间建立密集的跨模态对应关系,并使用相应的匹配概率将这些特征相互扭曲。还提出了利用像素级关联的 ID 一致性和密集三元组损失,使得该模型能够学习更具辨别力的人表示。

1.3.2 深度图像行人重识别

深度图像捕捉行人身体和骨骼信息,这为低光照和换衣场景下行人重识别提供了可能性,对个性化的人机

交互应用也很重要。Haque 等^[41]提出了一个基于注意力的模型,在没有 RGB 信息的情况下,从人体形状和运动动力学的角度来识别个体。提出的模型是基于强化学习和回归神经网络的结合,目标是识别小的区别性区域,表明人类身份。当人们出现在极端的照明或更换衣服,RGB 外观为基础的重现方法往往失败。为了克服这个问题,Wu 等^[42]提出利用深度信息来提供更多的不变的身体形状和骨架信息,而不管光照和颜色的变化。为了解决数据稀缺的问题,Karianakis 等^[43]提出了分离率 RGB 深度转移,以此有效地利用预训练模型的 RGB 数据,并学习强大的帧级特性。为了提高对视频序列的重新识别能力,该方法提出了加强时间注意单元,它位于帧级特征之上,不依赖于网络结构。

1.3.3 文本-图像行人重识别

文本-图像的行人重识别解决了文本描述和 RGB 图像之间的匹配问题。当无法获得查询人的可视图像时,必须提供自然语言文本描述。图文匹配的关键是如何准确地度量视觉输入和文本输入之间的相似性。Zhang 等^[44]提出了一种新的跨模态投影匹配损失算法和跨模态投影分类损失算法,用于深度识别图文嵌入。Cmpm 损失利用 kl 发散最小化匹配图文对的兼容性得分,同时最大化匹配图文对之间的相关性。该算法在不同批量下对图像和文本进行关联具有很好的稳定性和优越性,不存在传统的双向排序损失的三重抽样和边缘选择。Liu 等^[45]设计了一种基于图关系挖掘的深度对抗性图注意力卷积网络(A-GANet)。利用模态鉴别器和特征变换器开发了一个对抗式学习模块,用于学习匹配跨模态的联合文本-视觉特征空间,图形注意力卷积层有效地利用了图形结构来学习视觉和文本具有判别性的特征。传统的基于文本的人员再识别方法严重依赖于身份标注。然而,这种标签过程是昂贵和耗时的。为了解决这种问题,Zhao 等^[46]提出了一个跨模式相互培训(CMMT)框架。具体来说,为了缓解类内的变化,使用了一种聚类方法来为可视和文本实例生成伪标签。为了进一步细化聚类结果,CMMT 提供了标签细化模块,该模块利用一种模式的聚类结果来细化受文本-图像成对关系约束的另一种模式的聚类结果。

1.3.4 跨分辨率行人重识别

跨分辨率(Cross-resolution)行人重识别将低分辨率和高分辨率图像进行匹配,解决了较大的行人图像分辨率变化难匹配的问题。面对低分辨率的问题,为了提高图像的尺度自适应提升和图像特征提取能力,Wang 等^[47]提出了一种新的结构 CSR-GAN。设计了 common-human 损失,使得生成的行人图像更加真实,设计了 u-

nique-human 损失,使得行人图像特征更具有判别力。为了克服由于相机和相关人员之间的距离不同,可能会出现分辨率不匹配的问题,Li 等^[48]提出了一种新的生成对抗网络来解决交叉分辨率的人身份验证,允许不同分辨率的查询图像。通过改进对抗学习技术,提出的模型学习分辨率不变的图像表示,同时能够恢复低分辨率输入图像中缺失的细节。由于保持了分辨率不变性和恢复了面向重定向的鉴别细节,所得特征可以联合应用于提高人体识别性能。从低分辨率到高分辨率中,大多数只关注行人特征的提取,忽略了原始行人图像的有效信息。为了克服这个问题,Zhang 等^[49]提出了一种基于分辨率的分辨率重构网络(RNN),无论输入分辨率如何,都可以同时生成 HR 和 LR 2 个版本。此外,还设计了双特征融合网络(DFFN)来提取鉴别性的多分辨率表示。

2 行人重识别的评价指标

行人重识别的评价指标有以下 4 种:

(1) Rank-n Accuracy^[50]:这是一种在图像分类、检索领域广泛使用的评价指标。给定一张目标图像,训练好的行人重识别模型会给出一个按照置信度排序的识别结果序列。最常用的 2 种评价指标是 Rank-1 Accuracy 与 Rank-5 Accuracy,分别代表在测试中,第 1 位即为正确匹配的比例,以及前 5 位存在正确匹配的比例。

(2) CMC 曲线^[50](Cumulative Match Characteristic):累积匹配曲线,在候选行人库(gallery)中检索待查询(probe)的行人,前 r 个检索结果中包含正确匹配结果的比率。其中,第 1 匹配率 $r=1$ 指标 rank-1 反映了最匹配候选目标刚好为待查询图片目标的概率,即该指标为真正的识别能力,因此很重要。但是当 r 值很小但大于 1 时,由于可以通过人眼进行辅助识别查找目标,因此也很有现实意义,如第 5 匹配率 $r=5$ 指标反映前 5 个匹配候选目标中存在待查询图片目标的概率,主要用来评估闭集中排序结果的正确率。

(3) ROC 曲线(Receiver Operating Characteristic Curve):ROC 曲线是检测、分类、识别任务中很常用的一项评价指标。ROC 曲线上的每一点反映的是不同的阈值对应的 FP(False Positive)和 TP(True Positive)之间的关系。

(4) mAP(mean Average Precision):平均精度均值,是多目标检测与多标签分类的常用评价指标,绘制精确率与召回率的关系曲线(P-R 曲线),该曲线与坐标轴的面积即为平均精确率(Average Precision),对所有类别测试得到各自的平均精确率,取均值得到平均精度均值 mAP。平均精度指标不仅体现了模型的精确率,还能对

模型给出的排名顺序做出评价。

3 数据集及实验比较

行人重识别的研究方法从基于传统的人工特征发展为基于深度学习方法,这离不开大规模数据集的发

展。当前比较常用的行人重识别数据集主要有以下几种:CUHK03 数据集^[51]、Market-1501 数据集^[52]、DukeMTMC-ReID 数据集^[53]和 MSMT17 数据集^[54],以及 4 个视频数据集:PRID-2011 数据集^[55]、iLIDS-VID 数据集^[56]、MARS 数据集^[57]和 DukeMTMC-VideoReID 数据集^[58]。具体常用的数据集参数如表 1 和表 2 所示。

表 1 行人重识别常用图像类型数据集

Table 1 Datasets of common image types for person re-identification

类型	数据集	发表时间	人数	相机数	图片数	标注方式
图像	CUHK03	2014	1 467	10	13 164	手工 + DPM
	Market-1501	2015	1 501	6	32 217	手工 + DPM
	DukeMTMC-ReID	2017	1 812	8	36 441	手工
	MSMT17	2018	4 101	15	126 441	Faster RCNN

表 2 行人重识别常用视频类型数据集

Table 2 Datasets of common video types for person re-identification

类型	数据集	发表时间	人数	相机数	视频序列(边界框)	标注方式
视频	PRID-2011	2011	200	2	400(40K)	手工
	iLIDS-VID	2014	300	2	600(44K)	手工
	MARS	2016	1 261	6	20 715(1M)	DPM + GMMCP
	DukeMTMC-VideoReID	2018	1 812	8	4 832(-)	手工

下面简单介绍几个常用开源的数据集:

(1) CUHK03

CUHK03 数据集示例,如图 4 所示。



图 4 CUHK03 数据集示例

Fig. 4 Example of CUHK03 dataset

CUHK03 数据集是香港中文大学于 2014 年提出来的数据集,其数据集图像来源于 6 个监控摄像头捕获的。每个身份由 2 个不相交的相机视图观察,每个视图中平均有 4.8 张图像。

(2) Market-1501

Market-1501 数据集示例,如图 5 所示。



图 5 Market-1501 数据集示例

Fig. 5 Example of Market-1501 dataset

Market-1501 数据集是 2015 年发布的,由 6 个摄像头收集行人重新识别大规模数据集。它包含 19 732 张用于测试的图像和 12 936 张用于训练的图像。

(3) DukeMTMC-ReID

DukeMTMC-ReID 数据集示例,如图 6 所示。

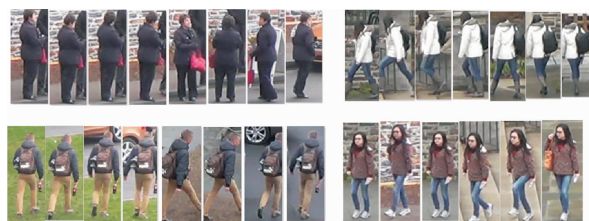


图 6 DukeMTMC-ReID 数据集示例

Fig. 6 Example of DukeMTMC-ReID dataset

DukeMTMC-ReID 数据集是 2017 年发布的行人重识别数据集。它包括来自 8 个高分辨率相机的 1 812 个身份的 36 411 张照片。其中,训练集包含从整体图像中随机选取的 702 个身份的 16 522 张照片,测试集包含其他 2 228 张查询照片和 17 661 张图库照片。

(4) MSMT17

MSMT17 数据集示例,如图 7 所示。

MSMT17 数据集是北京大学采集的行人重识别数据集,该数据集于 2017 年发布,有 15 个网络摄像头拍摄的 4 101 个身份的 126 441 张照片,包含 12 个室外摄像机和 3 个室内摄像机。



图7 MSMT17数据集示例

Fig.7 Example of MSMT17 dataset

除了以上几个已经开源的常用数据集以外,近几年内还有一些优秀的数据集被提出,其中比较典型的有:①2017年由中山大学发布的SYSU-MM01数据集^[36],包括来自6个摄像机(其中2个红外摄像机和4个可见光摄像机)的491个身份的RGB和IR图像,共得到287 628张RGB图像和15 792张IR图像。②2017年3月发布的RegDB数据集^[59]包含了412个行人身份,每个行人收集了10张RGB图像和10张热图像,其中有254个女性和158个男性,并且412个人中有156个人是从正面拍摄,256个人从背面拍摄。由于该数据集图像小,清晰度较差,每个身份的RGB图像和热图的姿态都是一一对应的,并且同一个身份在姿态上变化很小,这些因素都降低了该数据集RegDB上的跨模态行人重识别任务的难度。③2018年发布的由北京航空航天大学采集的LPW数据集^[60],包含2 731个行人的7 694个轨迹序列,总共有56万多张图像。该数据集尽管规模相对较大,但注释也具有较高清洁度。而且,更具挑战性的还有以下几个方面:人物的年龄从小到大不一样,人们的姿势是多种多样的,除了正常的步行状态外,还包括跑步和骑自行车。④对于跨分辨率的数据集MLR-ViPeR是从ViPeR数据集^[61]构建的。ViPeR由632个身份组成,每个身份包含从2个不同的相机捕获的2个图像。所有图像都被归一化为128×48像素。ViPeR被随机分成两等份,一份用于训练,另一份用于测试。每一半包含316个身份。对于每个标识,都是从一个相机中获取一个图像作为查询,并执行跨相机搜索;⑤2017年发布的CUHK-PEDES数据集^[62]包含13 003个身份的40 206幅行人图像。每个行人图像由2个不同的文本描述。总共收集了80 412个句子。

除了上述的常用数据集之外,许多研究人员也在开发最新的数据集用来丰富数据集类型。包括:①2021年发布的Person30K数据集^[63]是一个大规模数据集,其中包含30K个不同身份的138万张图像,都是由部署在89个不同站点的6 497个摄像机中收集到的。该数据集丰富的样本多样性,包括不同的背景和不同的人物姿势。②2021年发布的LUPerson数据集^[64],它是一个未

标记的数据集,包含从46K个场景中收集的超过200K个身份的图像。由于目前的数据集都是基于平行摄像头拍摄的,有研究学者针对此种情况,从俯瞰的角度出发,在2021年发布了名为BV-Person数据集^[65],包含18K个身份的114K张图像,其中近7.4K个身份的20K张的图像是从鸟瞰视图中获取的。

数据集对行人重识别的发展起着重要的作用。目前的行人重识别模型最为常用的评价指标为Rank-n Accuracy与mAP,表3整理了目前常用数据集上表现最优的模型(State of the Art, SOTA)以及其他一些优秀模型的精度数据。

表3 各数据集的性能最优模型以及精度数据

Table 3 State-of-the-art methods and accuracy data for each dataset

数据集	SOTA	mAP	Rank-1 accuracy
CUHK03	SCSN ^[66] (2020)	84.0	86.8
	HLGAT ^[67] (2021)	80.6	83.5
	RGA-SC ^[68] (2020)	77.4	81.1
Market-1501	HLGAT ^[67] (2021)	94.3	97.5
	SGR ^[69] (2021)	89.3	96.1
	B-VNet ^[65] (2021)	89.2	96.0
DukeMTMC-ReID	HLGAT ^[67] (2021)	87.3	92.7
	SGR ^[69] (2021)	81.3	91.1
	B-VNet ^[65] (2021)	80.6	90.5
MSMT17	SCSN ^[66] (2020)	58.5	83.8
	RGA-SC ^[68] (2020)	57.5	80.3
	CDNet ^[70] (2021)	54.7	78.9
iLIDS-VID	CTL ^[71] (2021)	-	89.7
	BiCnet-TKS ^[72] (2021)	75.1	84.6
	STMN ^[73] (2021)	66.6	80.6
MARS	CTL ^[71] (2021)	86.7	91.4
	BiCnet-TKS ^[72] (2021)	86.0	90.2
	STMN ^[73] (2021)	84.5	90.5
DukeMTMC-VideoReID	BiCnet-TKS ^[72] (2021)	96.1	96.3
	STMN ^[73] (2021)	95.9	97.0
	PSTA ^[74] (2021)	97.4	98.3

4 当前问题与未来展望

随着当前计算机算力的提升,以及大规模数据集的出现,为深度学习提供了大量的训练数据和验证数据,把深度学习应用在行人重识别的研究成果也得到了很大的突破。越来越多的研究者投入到行人重识别的研究中。在此,本文也简要提出行人重识别当前存在的问

题以及未来研究方向。

4.1 当前存在问题

(1) 现有的全监督模型整体性能虽然接近饱和状态,但是弱监督的整体性能依旧难以取得理想的成绩,而且不同数据集之间关联度不高,模型的泛化能力明显不足。

(2) 跨模态模型中度量不同模态数据之间的内容相似性的能力不足,整体性能不够理想,模态与模态之间泛化能力不高。

(3) 行人重识别算法实现一般需要布署在边缘计算设备上,如智能监控摄像头、嵌入式设备等。现有的算法参数数量与计算量较大,边缘设备的算力根本无法满足要求。

4.2 未来研究方向

(1) 弱监督学习的性能优化:应该怎样在保证模型表征学习能力的基础上,提出更为有效的弱监督学习方法,降低数据标注的工作量,是行人重识别未来研究的热点问题。

(2) 数据集的迁移:减少人工标注依赖,提高模型的泛化能力是深度学习领域所要面对的核心问题,从虚拟数据进行学习(Learning from virtual data),如何才能解决虚拟数据中的 domain gap,提高模型泛化能力,这也是一个研究热点。

(3) 轻量化模型:轻量型快速的行人重识别算法设计,降低模型参数数量,提升模型运行速度,自适应的针对不同类型的硬件配置(小型的移动手机和大型服务器)调整和优化模型,这也是未来的研究热点之一。

(4) 基于 Transformer 的神经网络有着比 CNN 更多的优点,将其应用在行人重识别已经有了初步的成效,进一步挖掘 Transformer 在行人重识别中也是一个热门的研究方向。

5 总 结

本文针对行人重识别领域的研究现状,从有监督学习、弱监督学习和跨模态学习 3 个方向,对现有的基于深度学习的行人重识别方法进行了归纳与总结,针对近几年的最新成果作出分析与讨论,并对现存的主要问题与未来的研究方向进行阐述。当前基于深度学习的行人重识别问题研究尚未成熟,特别是针对弱监督和跨模态这 2 大热门方向的性能还尚有不足。目前行人重识别还面临着诸多问题和挑战。对于行人重识别的未来研究热门方向也亟需研究者进行深耕。本文针对行人重识别相关概念、方法和数据集进行了简单阐述,旨在使读者对基于深度学习的行人重识别方法有全面的了解,为开展行人重识别领域的研究提供些许帮助。

参考文献:

- [1] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). Piscataway: IEEE, 2005: 886-893.
- [2] Koestinger M, Hirzer M, Wohlhart P, et al. Large scale metric learning from equivalence constraints[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE, 2012: 2288-2295.
- [3] Martinel N, Micheloni C, Foresti G L. Saliency weighted features for person re-identification[C]//European Conference on Computer Vision. Berlin: Springer, 2014: 191-208.
- [4] Lowe D G. Object recognition from local scale-invariant features[C]//Proceedings of the Seventh IEEE International Conference on Computer Vision. Piscataway:IEEE, 1999: 1150-1157.
- [5] Zheng W S, Gong S, Xiang T. Reidentification by relative distance comparison[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 35(3): 653-668.
- [6] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE, 2016: 770-778.
- [7] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE, 2017: 4700-4708.
- [8] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE, 2015: 1-9.
- [9] Wang F, Zuo W, Lin L, et al. Joint learning of single-image and cross-image representations for person re-identification [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE, 2016: 1288-1296.
- [10] Chen G, Lin C, Ren L, et al. Self-critical attention learning for person re-identification[C]//Proceedings of the IEEE/

- CVF International Conference on Computer Vision. Piscataway:IEEE,2019: 9637-9646.
- [11] Sun Y, Zheng L, Yang Y, et al. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline) [C]//Proceedings of the European Conference on Computer Vision (ECCV). Berlin: Springer, 2018: 480-496.
- [12] Li D, Chen X, Zhang Z, et al. Learning deep context-aware features over body and latent parts for person re-identification [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE, 2017: 384-393.
- [13] He S, Luo H, Wang P, et al. Transreid: Transformer-based object re-identification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway:IEEE, 2021: 15013-15022.
- [14] Yi D, Lei Z, Liao S, et al. Deep metric learning for person re-identification[C]//2014 22nd International Conference on Pattern Recognition. Piscataway:IEEE, 2014: 34-39.
- [15] Zheng L, Zhang H, Sun S, et al. Person re-identification in the wild[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE,2017: 1367-1376.
- [16] Schroff F, Kalenichenko D, Philbin J. Facenet: A unified embedding for face recognition and clustering[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE,2015: 815-823.
- [17] Hermans A, Beyrer L, Leibe B. In defense of the triplet loss for person re-identification[EB/OL]. (2017-11-21)[2021-12-01]. <http://arXiv preprint arXiv:1703.07737>, 2017.
- [18] Luo C, Chen Y, Wang N, et al. Spectral feature transformation for person re-identification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway:IEEE,2019: 4976-4985.
- [19] Zhou J, Yu P, Tang W, et al. Efficient online local metric adaptation via negative samples for person re-identification[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway:IEEE,2017: 2420-2428.
- [20] Ye M, Chen J, Leng Q, et al. Coupled-view based ranking optimization for person re-identification[C]//International Conference on Multimedia Modeling. Berlin:Springer, 2015: 105-117.
- [21] Li W, Wu Y, Mukunoki M, et al. Common-neighbor analysis for person re-identification[C]//2012 19th IEEE International Conference on Image Processing. Piscataway:IEEE, 2012: 1621-1624.
- [22] Bak S, Carr P. One-shot metric learning for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE,2017: 2990-2999.
- [23] Wang M, Lai B, Chen H, et al. Towards precise intra-camera supervised person re-identification[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway:IEEE,2021: 3229-3238.
- [24] Fu Y, Wei Y, Wang G, et al. Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway:IEEE,2019: 6112-6121.
- [25] Ye M, Li J, Ma A J, et al. Dynamic graph co-matching for unsupervised video-based person re-identification[J]. IEEE Transactions on Image Processing, 2019, 28(6): 2976-2990.
- [26] Liao S, Shao L. Interpretable and generalizable person re-identification with query-adaptive convolution and temporal lifting [C]//European Conference on Computer Vision. Berlin:Springer, 2020: 456-474.
- [27] Yang Q, Yu H X, Wu A, et al. Patch-based discriminative feature learning for unsupervised person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE,2019: 3633-3642.
- [28] Ma A J, Yuen P C, Li J. Domain transfer support vector ranking for person re-identification without target camera label information[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway:IEEE,2013: 3567-3574.
- [29] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[J]. Advances in Neural Information Processing Systems, 2014, 27:2672-2680.
- [30] Zheng Z, Zheng L, Yang Y. Unlabeled samples generated by gan improve the person re-identification baseline in vitro[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway:IEEE,2017: 3754-3762.
- [31] Wei L, Zhang S, Gao W, et al. Person transfer gan to bridge domain gap for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE,2018: 79-88.

- [32] Qian X, Fu Y, Xiang T, et al. Pose-normalized image generation for person re-identification[C]//Proceedings of the European Conference on Computer Vision (ECCV). Berlin: Springer, 2018: 650-667.
- [33] Liu J, Zha Z J, Chen D, et al. Adaptive transfer network for cross-domain person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE,2019: 7202-7211.
- [34] Chen Y, Zhu X, Gong S. Instance-guided context rendering for cross-domain person re-identification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway:IEEE,2019: 232-242.
- [35] Ge Y, Zhu F, Chen D, et al. Self-paced contrastive learning with hybrid memory for domain adaptive object re-id[J]. Advances in Neural Information Processing Systems, 2020, 33: 11309-11321.
- [36] Wu A, Zheng W S, Yu H X, et al. RGB-infrared cross-modality person re-identification[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway:IEEE,2017: 5380-5389.
- [37] Fu C, Hu Y, Wu X, et al. Cm-nas: Cross-modality neural architecture search for visible-infrared person re-identification [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway:IEEE,2021: 11823-11832.
- [38] Ye M, Lan X, Li J, et al. Hierarchical discriminative learning for visible thermal person re-identification[C]//Proceedings of the AAAI Conference on Artificial Intelligence. New Orleans: AAAI Press,2018:7501-7508.
- [39] Wei Z, Yang X, Wang N, et al. Synergetic modality collaborative learning for visible infrared person re-identification[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway:IEEE,2021: 225-234.
- [40] Park H, Lee S, Lee J, et al. Learning by Aligning: Visible-infrared person re-identification using cross-modal correspondences[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway:IEEE,2021: 12046-12055.
- [41] Haque A, Alahi A, Li F F. Recurrent attention models for depth-based person identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE,2016: 1229-1238.
- [42] Wu A, Zheng W S, Lai J H. Robust depth-based person re-identification[J]. IEEE Transactions on Image Processing, 2017, 26(6): 2588-2603.
- [43] Karianakis N, Liu Z, Chen Y, et al. Reinforced temporal attention and split-rate transfer for depth-based person re-identification[C]//Proceedings of the European Conference on Computer Vision (ECCV). Berlin:Springer,2018: 715-733.
- [44] Zhang Y, Lu H. Deep cross-modal projection learning for image-text matching[C]//Proceedings of the European Conference on Computer Vision (ECCV). Berlin: Springer,2018: 686-701.
- [45] Liu J, Zha Z J, Hong R, et al. Deep adversarial graph attention convolution network for text-based person search[C]//Proceedings of the 27th ACM International Conference on Multimedia. New York: Association for Computing Machinery, 2019: 665-673.
- [46] Zhao S, Gao C, Shao Y, et al. Weakly supervised text-based person re-identification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway:IEEE,2021: 11395-11404.
- [47] Wang Z, Ye M, Yang F, et al. Cascaded SR-GAN for scale-adaptive low resolution person re-identification[C]//Twenty-seventh International Joint Conference on Artificial-Intelligence (IJCAI). Piscataway:IEEE,2018:3891-3897.
- [48] Li Y J, Chen Y C, Lin Y Y, et al. Recover and identify: A generative dual model for cross-resolution person re-identification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway:IEEE,2019: 8090-8099.
- [49] Zhang G, Chen Y, Lin W, et al. Low resolution information also matters: Learning multi-resolution representations for person Re-Identification[EB/OL]. (2021-05-26)[2021-12-01]. <http://arXiv:2105.12684v1>
- [50] Ge Y, Li Z, Zhao H, et al. Fd-gan: Pose-guided feature distilling gan for robust person re-identification[J]. Advances in Neural Information Processing Systems, 2018, 31:1230-1241.
- [51] Li W, Zhao R, Xiao T, et al. Deepreid: Deep filter pairing neural network for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE, 2014: 152-159.
- [52] Zheng L, Shen L, Tian L, et al. Scalable person re-identification: A benchmark[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway:IEEE,2015: 1116-1124.
- [53] Zheng Z, Zheng L, Yang Y. Unlabeled samples generated by gan improve the person re-identification baseline in vitro[C]// Proceedings of the IEEE International Conference on Computer Vision. Piscataway:IEEE,2017: 3754-3762.

- [54] Wei L, Zhang S, Gao W, et al. Person transfer gan to bridge domain gap for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE,2018: 79-88.
- [55] Hirzer M, Beleznai C, Roth P M, et al. Person re-identification by descriptive and discriminative classification[C]//Scandinavian Conference on Image Analysis. Berlin:Springer, 2011: 91-102.
- [56] Wang T, Gong S, Zhu X, et al. Person re-identification by video ranking[C]//European Conference on Computer Vision. Berlin: Springer, 2014: 688-703.
- [57] Zheng L, Bie Z, Sun Y, et al. Mars: A video benchmark for large-scale person re-identification[C]//European Conference on Computer Vision. Berlin: Springer,2016: 868-884.
- [58] Wu Y, Lin Y, Dong X, et al. Exploit the unknown gradually: One-shot video-based person re-identification by stepwise learning[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE,2018: 5177-5186.
- [59] 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, doi:103990/s/7030605.
- [60] Song G, Leng B, Liu Y, et al. Region-based quality estimation network for large-scale person re-identification[C]//Proceedings of the AAAI Conference on Artificial Intelligence. New Orleans: AAAI Press, 2018:7347-7354.
- [61] Gray D, Tao H. Viewpoint invariant pedestrian recognition with an ensemble of localized features[C]//European Conference on Computer Vision. Berlin: Springer, 2008: 262-275.
- [62] Li S, Xiao T, Li H, et al. Person search with natural language description[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE, 2017: 1970-1979.
- [63] Bai Y, Jiao J, Ce W, et al. Person30k: A dual-meta generalization network for person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE,2021: 2123-2132.
- [64] Fu D, Chen D, Bao J, et al. Unsupervised pre-training for person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE,2021: 14750-14759.
- [65] Yan C, Pang G, Wang L, et al. BV-Person: A large-scale dataset for bird-view person re-identification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway:IEEE,2021: 10943-10952.
- [66] Chen X, Fu C, Zhao Y, et al. Saliency-guided cascaded suppression network for person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE,2020: 3300-3310.
- [67] Zhang Z, Zhang H, Liu S. Person re-identification using heterogeneous local graph attention networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE,2021: 12136-12145.
- [68] Zhang Z, Lan C, Zeng W, et al. Relation-aware global attention for person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE,2020: 3186-3195.
- [69] Yan C, Pang G, Jiao J, et al. Occluded person re-identification with single-scale global representations[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway:IEEE,2021: 11875-11884.
- [70] Li H, Wu G, Zheng W S. Combined depth space based architecture search for person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE,2021: 6729-6738.
- [71] Liu J, Zha Z J, Wu W, et al. Spatial-temporal correlation and topology learning for person re-identification in videos[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE, 2021: 4370-4379.
- [72] Hou R, Chang H, Ma B, et al. Bicnet-tks: Learning efficient spatial-temporal representation for video person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE,2021: 2014-2023.
- [73] Eom C, Lee G, Lee J, et al. Video-based person re-identification with spatial and temporal memory networks[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway:IEEE,2021: 12036-12045.
- [74] Wang Y, Zhang P, Gao S, et al. Pyramid spatial-temporal aggregation for video-based person re-identification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway:IEEE,2021: 12026-12035.