

基于生成对抗网络的行人重识别方法研究综述

朱敏, 明章强*, 闫建荣, 杨勇, 朱佳旻

(四川大学计算机学院 成都 610065)
(mingzhangqiang@stu.scu.edu.cn)

摘要: 近年来, 随着公共安全需求的不断增长以及智能监控网络的快速发展, 行人重识别已成为计算机视觉领域的热门研究课题之一, 其目标是在不同摄像头中检索具有相同身份的行人。首先, 介绍目前经典的行人重识别数据集; 然后, 重点梳理了近年来基于生成对抗网络的行人重识别方法, 根据生成对抗网络的特点和应用场景将这些方法归纳为风格转换、数据增强和不变性特征学习3类, 并总结每类方法的基本原理和优缺点; 再在主流数据集上对经典算法进行比较; 最后, 总结现阶段行人重识别面临的挑战, 并对未来的研究方向进行展望。

关键词: 行人重识别; 生成对抗网络; 风格转换; 数据增强; 特征学习
中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2022.18852

A Survey on Generative Adversarial Network Based Person Re-Identification Method

Zhu Min, Ming Zhangqiang*, Yan Jianrong, Yang Yong, and Zhu Jiamin

(College of Computer Science, Sichuan University, Chengdu 610065)

Abstract: In recent years, with the growing demand for public safety and the rapid development of intelligent surveillance networks, person re-identification (Re-ID) has become one of the hottest research topics in the field of computer vision, which aims at retrieving a person of interest through multiple cameras with non-overlapping views. First, the current classical person Re-ID datasets are introduced. Second, the latest GAN-based person Re-ID methods are focused on sorting out. These methods are classified into three categories: image-image style transfer, data enhancement and invariant feature learning according to the characteristics and application scenarios of GAN. And basic principles, advantages, and disadvantages are summarized for each category. Then, the states of the art of the classic algorithms mentioned are compared on the mainstream datasets. Finally, the present challenging problems of person Re-ID and directions for future research are discussed.

Key words: person re-identification; generative adversarial network; image-image style transfer; data enhancement; feature learning

近年来, 随着智能监控设备的快速发展和对公共安全需求的不断增大, 大量摄像头被部署在机场、社区、街道和校园等公共场所^[1]。这些摄像

头所产生的海量监控视频中, 包含了行人的活动规律与行为特征, 常用于目标检测、多摄像头多目标跟踪及人群行为分析等^[2]。行人重识别最早可追

收稿日期: 2020-12-18; 修回日期: 2021-05-06. 基金项目: 国家重点研发计划(JG2018190). 朱敏(1971—), 女, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究方向为计算机视觉、行人重识别、图像处理与模式识别; 明章强(1995—), 男, 硕士研究生, CCF 学生会员, 论文通信作者, 主要研究方向为计算机视觉、行人重识别、多目标跟踪; 闫建荣(1993—), 女, 硕士研究生, CCF 学生会员, 主要研究方向为行人重识别、图像处理; 杨勇(1998—), 男, 硕士研究生, CCF 学生会员, 主要研究方向为图像融合、跨模态行人重识别; 朱佳旻(1997—), 男, 硕士研究生, CCF 学生会员, 主要研究方向为计算机视觉、目标检测。

溯到多摄像头多目标跟踪的问题上^[3], 旨在确定不同摄像头拍摄的行人或同一摄像头不同视频片段的行人图像是否为同一个行人。完整的行人重识别系统包含行人检测和行人重识别^[4]。对于行人检测, 已有多种精度较高的算法, 如 YOLO^[5], SSD^[6] 和 Fast R-CNN^[7] 等, 均可获得高质量的检测结果。行人重识别利用检测得到的行人图像构建成大型图像数据集, 并用探针图像从中检索与之匹配的行人^[8]。但现实场景中, 不同摄像头视角、行人姿态、光照和分辨率的变化较大, 增大了行人重识别的难度。

传统的行人重识别方法主要集中于如何手工提取固定的判别性特征^[9-13], 或者学习更好的相似性度量^[14-18], 不仅容易出错, 而且十分耗时, 极大地影响行人重识别的准确性和实时性。近年来, 出现了大量基于深度学习的方法, 如一些研究者通过设计不同的损失函数以约束网络模型对行人判别性特征学习^[19-24], 还有一些是引入局部特征学习提取行人更具鲁棒性的特征^[25-29], 或者使用注意力机制关注行人身体部位的关键信息^[27,30-36], 以及融合行人的全局和局部特征以增强最终的特征表示^[37-43]。为了弥补单帧图像信息不足的缺陷, 一些研究工作利用视频序列的空间和时间互补线索, 有效地融合视频序列中的更多信息^[44-49]。近年来, 出现了基于图神经网络的方法^[48,50-52], 通过对行人图像建立图关系模型学习更具判别性和鲁棒性的特征。

虽然上述方法针对行人重识别问题已经有了突破性的进展, 但行人重识别在真实场景中仍然面临诸多挑战。一方面, 深度学习的方法依赖大量的训练数据, 目前公开的数据集标注的行人数据规模有限, 并且不同的行人重识别数据集之间存在域差, 即在不同数据集上分别训练和测试时会导致性能的严重下降, 使模型很难泛化到其他应用场景^[53]; 另一方面, 同一数据集内不同摄像头拍摄背景、分辨率和光照的变化会导致图像风格差异, 相同的行人在不同摄像头中的姿态和外观变化大, 并且不同行人可能具有相同的外观和姿态, 这些差异性导致很难捕获到行人判别性的特征^[18,54]。

解决上述挑战的重要思路是: 利用生成对抗网络(generative adversarial networks, GAN)进行图像风格转换或统一不同图像风格, 缓解不同数据集之间或同一数据集内的图像风格差异^[53,55-62], 或合成不同姿态、外观、光照和分辨率的行人图像用于扩充数据集, 以提高模型的泛化能力^[63-71];

或学习与噪声无关而与身份相关的行人特征, 以提高特征匹配的准确度^[54,72-74]。相比一些基于深度学习的行人重识别方法的综述文献^[1,75-76], 以及综述总结基于局部特征学习的行人重识别方法^[77]和基于弱监督的行人重识别方法^[8], 本文更聚焦 GAN 这一特定技术在行人重识别任务中的发展和应用。这类方法属于目前研究的热点, 并且可以反映基于 GAN 的行人重识别方法的研究趋势。

本文将基于 GAN 的方法归纳为基于风格转换的方法、数据增强的方法和不变性特征学习的方法, 并结合应用场景划分子类。其中, 基于风格转换的方法分为跨域风格转换的方法^[53,55-59]和域内风格转换的方法^[60-62]; 基于数据增强的方法分为无辅助信息引导的图像合成方法^[63-66]和辅助信息引导的图像合成方法^[67-71]; 基于不变性特征学习的方法分为姿态不变性特征学习的方法^[72]和降级不变性特征学习的方法^[54,73-74]。

1 数据集介绍

近年来出现了多种行人重识别数据集, 它们在类型、行人标注方式和应用场景等方面有很大的区别。本节将选取部分经典数据集分别进行介绍。

(1) VIPeR 数据集^[78]是最早提出的小型行人重识别数据集。由 2 个摄像头在校园采集得到, 且每个视角的摄像头只拍摄行人的一幅图像, 总共包含 632 个不同行人的 1264 幅图像, 每幅图像都被裁剪并缩放为 128×48 的大小, 采用手工标注行人检测框。

(2) CUHK03 数据集^[15]属于大规模行人重识别数据集。其中的行人图像是在香港中文大学校园由 10 个摄像头采集得到的, 包含 1360 个不同的行人总共 13164 幅图像, 每幅图像的大小不定。CUHK03 数据集在 CUHK01 数据集^[79]和 CUHK02 数据集^[80]的基础上进行改进, 增加了摄像头和采集图像的数量, 可以采集更多视角下的行人图像。该数据集采用手工标记和自动检测器(deformable part model, DPM)^[81]这 2 种方式标注行人检测框, 相比单一的手工标注, DPM 标注的行人更符合真实场景。

(3) Market-1501 数据集^[82]是在 2015 年公布的大规模行人重识别数据集。其中的行人图像是由 5 个高分辨率摄像头和一个低分辨率摄像头在清华大学超市前采集得到的, 包含 1501 个不同行人的总共 32668 幅图像, 每幅图像的大小为 128×64,

采用手工标记和自动检测器 DPM 这 2 种方式标注行人检测框。相比 CUHK03 数据集, Market-1501 数据集包含 502 793 个干扰因素和更多的标注图像。

(4) DukeMTMC-reID 数据集^[63]是多摄像头、多目标行人跟踪数据集 DukeMTMC^[83]的子集, 同样属于大规模行人重识别数据集。该数据集是在杜克大学校园内由 8 个静态高清摄像头采集得到的, 包含 1812 个不同行人的总共 36 441 幅图像, 每幅图像的大小不定, 采用手工标注行人检测框。

(5) MSMT17 数据集^[9]是在 2018 年公布的一个大规模行人重识别数据集。其中的行人图像是由校园内的 15 个摄像头采集得到的, 包含 4 101 个不同行人的总共 126 441 幅图像, 每幅图像的大小不定, 采用行人自动检测器 Faster R-CNN^[84]标注行人检测框。该数据集是目前行人重识别数据集中行人和标注图像最多的数据集, 相比早期单一场景和没有显著光照变化的数据集, 该数据集覆盖了更多的场景, 并且能够捕获多个时间段光照变化的行人图像。

以上列举的数据集都属于基于图像的行人重识别数据集。其中, ViPeR 属于早期小规模数据集, 图像分辨率较低, 行人识别效果较差; CUHK03, Market-1501 和 DukeMTMC-reID 数据集均在校园内采集, 图像的质量较高, 本文基于 GAN 的行人重识别方法大部分都在这 3 个数据集上进行实验; MSMT17 数据集包含更多的标注行人, 且包含室内和室外的多个场景, 现有的大部分行人重识别方法在该数据集上依然不能得到很好的识别效果。

除基于图像的数据集, 还有针对大规模真实场景的视频数据集, 如 PRID2011^[85], iLIDS-VID^[86],

MARS^[87], DukeMTMC-VideoReID^[19]和 LPW^[88]等。表 1 和表 2 从数据集行人 ID 数、标注框数、标注方式、摄像头数、应用场景和评估指标等角度, 分别对基于图像和视频的行人重识别数据集进行总结和对比。可以看到, 随着大规模行人重识别数据集的发展, 数据集的行人 ID 数和标注框数(或轨迹数)在不断增加, 数据集覆盖的场景也越来越丰富, 并且开始采用基于深度学习的自动检测器和手工标注结合的方式来检测行人边界框, 使最新的数据集越来越接近真实环境。另外, 大多数主流数据集均采用累计匹配特征(cumulative matching characteristics, CMC)曲线和平均均值精度(mean average precision, mAP)进行性能评估。

尽管现有的行人重识别数据集包含了更多的行人标注和摄像头, 在一定程度上缓解了早期数据集数据规模不足的问题, 但相对于复杂的真实世界, 数据量依然很小, 缺乏不同拍摄视角、遮挡、行人姿态、行人外观、分辨率和光照变化等场景下的行人样本数据。针对上述问题, 目前最常用的方法是利用 GAN 合成多样性的行人样本, 以扩充训练数据。

2 基于 GAN 的行人重识别方法分类

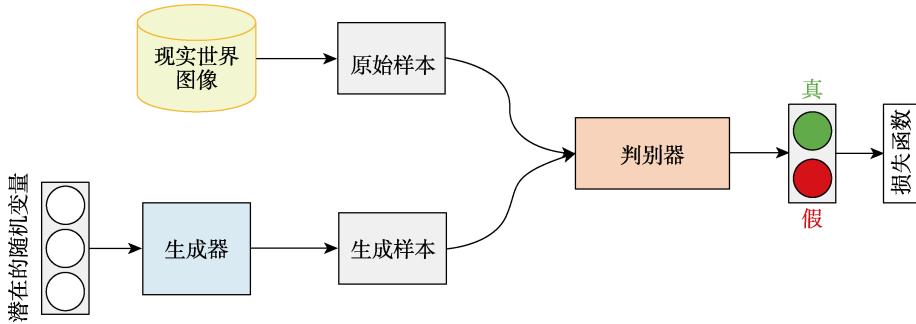
2014 年, GAN 由 Goodfellow 等^[89]首次提出, 并在近几年得到快速发展。之后出现了大量 GAN 的变体和应用^[53,55,60,67,71-73,90-91]。图 1 显示了 GAN 的基本框架, 其可描述为 2 个神经网络的竞争, 分别是生成器和判别器。这 2 个网络在训练过程中不断进行博弈, 生成器尝试生成让判别器无法区分的样

表 1 基于图像的行人重识别数据集

| 数据集 | 发布年份 | ID 数 | 标注框数 | 标注方式 | 摄像头 | 场景 | 评估 |
|-------------------------------|------|-------|---------|-------------|-----|-----------|---------|
| ViPeR ^[78] | 2007 | 632 | 1 264 | 手工 | 2 | 室外+小场景 | CMC |
| CUHK03 ^[15] | 2014 | 1 467 | 13 164 | DPM+手工 | 10 | 室内+小场景 | CMC+mAP |
| Market-1501 ^[82] | 2015 | 1 501 | 32 217 | DPM+手工 | 6 | 室外+大场景 | CMC+mAP |
| DukeMTMC-reID ^[63] | 2017 | 1 812 | 36 441 | 手工 | 8 | 室外+大场景 | CMC+mAP |
| MSMT17 ^[9] | 2018 | 4 101 | 126 441 | Faster RCNN | 15 | 室内+室外+大场景 | CMC+mAP |

表 2 基于视频的行人重识别数据集

| 数据集 | 发布年份 | ID 数 | 轨迹数 | 标注方式 | 摄像头 | 场景 | 评估 |
|----------------------------|------|-------|--------|-----------|-----|--------|---------|
| PRID-2011 ^[85] | 2011 | 934 | 400 | 手工 | 2 | 室内+小场景 | CMC |
| iLIDS-VID ^[86] | 2014 | 300 | 600 | 手工 | 2 | 室内+小场景 | CMC |
| MARS ^[87] | 2016 | 1 261 | 20 715 | DPM+GMMCP | 6 | 室外+大场景 | CMC+mAP |
| DukeMTMC-V ^[19] | 2018 | 1 812 | 4 832 | DPM | 8 | 室外+大场景 | CMC+mAP |
| LPW ^[88] | 2018 | 2 731 | 7 694 | DPM+NN+手工 | 11 | 室外+大场景 | CMC+mAP |

图 1 GAN 的基本框架^[89]

本，判别器尝试将真实样本和生成器生成的样本区分开。在训练过程中生成器会生成更为逼真的样本，而判别器会不断地提高判别能力。理想条件下，生成器和判别器会达到一个纳什均衡的状态^[89]。

生成图像作为 GAN 的重要应用之一，被广泛应用于计算机视觉领域。其工作原理可以概括为，利用 GAN 的生成器将真实图像特定的像素迁移到生成图像上，并使用判别器判断生成图像的真伪，然后反馈给生成器以生成更相似的图像。目前，利用 GAN 转换图像的风格减小相同数据集或不同数据集行人图像风格的差异，或生成多样化的图像扩充训练样本，或学习行人不变性的特征表示，已

经成为行人重识别领域的热门研究方向^[54]。

本文依据 GAN 的特点和应用场景，将基于 GAN 的行人重识别方法归纳为风格转换^[53,55-62,92]、数据增强^[63-71]和不变性特征学习^[54,72-74,93-95] 3 类，并在此基础上进行细分。对于风格转换的方法，根据数据域的不同划分为跨域风格转换的方法和域内风格转换的方法；对于数据增强的方法，根据其技术类型划分为无辅助信息引导图像合成的方法和辅助信息引导图像合成的方法；对于不变性特征学习的方法，根据不同视觉变化划分为姿态不变性特征学习和降级不变性特征学习的方法。表 3 总结了上述方法在不同应用场景下的优势和局限性。

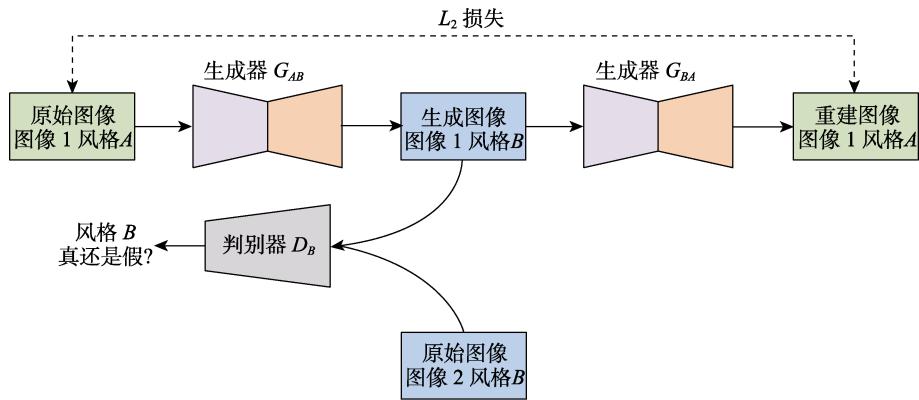
表 3 依据 GAN 的应用场景总结各类方法的优势和局限性

| 一级分类 | 二级分类 | 应用场景 | 优点 | 局限性 |
|--------------|------------------------------------|--------------------|----------------------------------|---|
| 基于风格转换的方法 | 跨域风格转换的方法 ^[53,55-59] | 不同数据集之间存在域差 | 合成大量具有源域行人 ID 和目标域的样本图像 | 合成的样本图像不稳定，可能和原风格的行人图像，在一定程度上可能导致图像存在冲突，进而影响模型的性能 |
| | 域内风格转换的方法 ^[60-62] | 同一数据集摄像头采集的图像风格不一致 | 统一或转换同一数据集内不同摄像头采集的图像风格，缓解摄像头的偏差 | 容易生成噪声样本，且随着摄像头的增加，会消耗大量的计算资源 |
| 基于数据增强的方法 | 无辅助信息引导图像合成的方法 ^[63-66] | 数据集规模小 | 合成具有不同外观、姿态的行人图像 | 缺乏大规模姿态和外观引导，合成数据存在大量的噪声样本 |
| | 辅助信息引导图像合成的方法 ^[67-71] | 姿态或外观变化大 | 合成不同姿态或外观的样本，提高模型泛化能力 | 需要额外的辅助信息如姿态检测器，增加模型的计算量 |
| 基于不变性特征学习的方法 | 姿态不变性特征学习的方法 ^[72] | 姿态变化大 | 学习行人身份有关而与姿态无关的特征 | 采用随机姿态或简单姿态生成图像，会不可避免地引入噪声伪像，需要额外的判别器和损失函数进行筛选 |
| | 降级不变性特征学习的方法 ^[54,73-74] | 光照、分辨的变化大 | 学习行人身份有关而与降级无关的特征 | 降级图像容易丢失某些判别性特征，使不变性特征提取困难 |

2.1 基于风格转换的方法

在行人重识别任务中经常会出现数据的偏差，主要表现在 2 个方面：一方面，由于地域的差异，不同数据集(域)的采集对象、背景和时间段的不同，使不同数据集(域)之前存在偏差；另一方面，由于同一数据集(域)采集视频的相机规格、视角和光线的不同，相同数据域中相机与相机之前存在偏差。因此，文献[53,55-56,60,64]使用 CycleGAN^[90]实现

不同行人图像的转换减小图像间的风格差异。虽然各算法实现的细节不同，但图像风格转换的方法可统一为图 2 所示流程。在训练阶段，生成器 G_{AB} 将图像 1 从风格 A 转换成风格 B，生成器 G_{BA} 将图像从风格 B 转换成风格 A；判别器 D_B 判断生成的图像是否逼近原始图片 2 的风格 B。通过最小化判别损失和 L_2 损失，使生成器和判别器不断对抗直至收敛^[75]，最终实现 $A \rightarrow B$ 的风格转换。本文

图 2 CycleGAN 进行图像风格转换 ($A \rightarrow B$) 流程图^[90]

根据风格转换数据域的不同, 将风格转换方法归纳为2类: 基于跨域风格转换的方法和基于域内风格转换的方法。

2.1.1 基于跨域风格转换的方法

行人重识别任务中的不同数据集之间通常存在域差^[53], 即在不同数据集上分别进行训练和测试时, 模型的性能会出现严重下降, 从而无法有效地将模型泛化到新的测试集^[96-97]. 为了解决域差, 传统有监督图像到图像的转换方法使用对齐的图像对训练模型, 以学习输入图像到输出图像之间的映射, 但大规模的标注图像对在现实任务中很难实现. 因此, 当目标域未进行行人标注时, 有监督的单域行人重识别方法会受限^[55].

解决此类问题的常用策略是利用GAN进行跨数据域的风格转换. 由于CycleGAN^[90]可以实现任意2幅图像风格的转换, 研究者考虑在其基础上改进, 以实现不同数据集之间的自适应行人样式转换, 从而缩小或消除域差. 受到CycleGAN的启发, Wei等^[53]提出人员转移GAN(person transfer GAN, PTGAN), 在保留源域中行人身份的前提下, 将源域中的行人转移到目标数据集, 使源域中的行人具有目标域的背景和光照射样. 但PTGAN是随机生成行人图像, 生成样本中可能存在未知的噪声. Deng等^[55]利用孪生网络和CycleGAN组成相似性保留GAN(similarity preserving GAN, SPGAN), 以无监督的方式将有标签的行人从源域迁移到目标域.

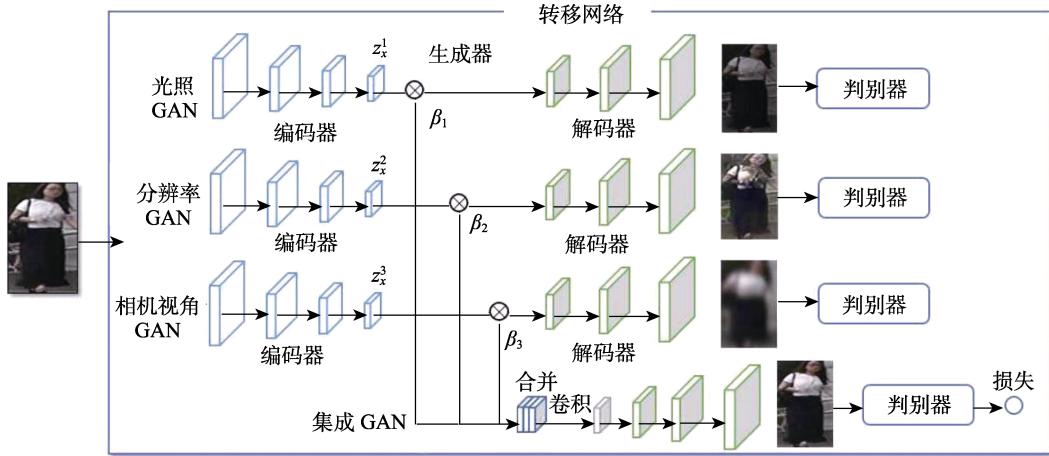
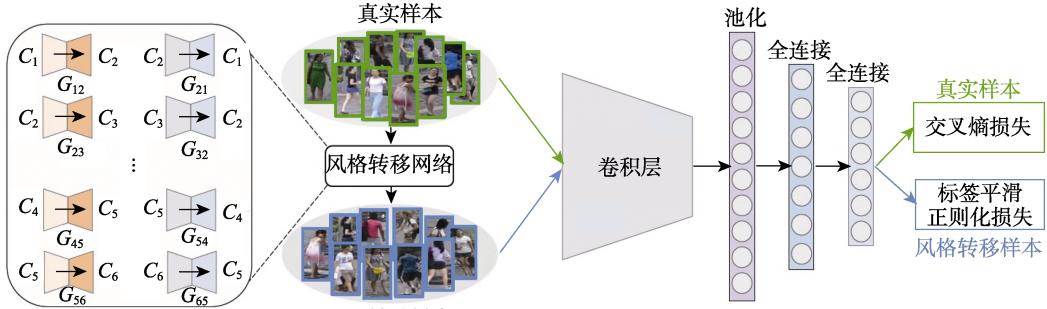
上述方法都是基于单一因素的风格转换. 实际上, 域差可能由多个因素(如光照、分辨率和摄像头视角)的变化引起^[98], 并且不同图像受到这些因素影响的程度不同, 使单一因素样式转换的方法识别性能不佳. 因此, Liu等^[56]提出一种自适应转移网络(adaptive transfer network, ATNet), 其网络结构如图3所示. ATNet使用3个CycleGAN实现摄像头

视角、光照和分辨率的风格转换, 并根据不同因素影响的程度自适应地为每个CycleGAN分配权重, 从而进行精确的风格转换. Zhong等^[58]不仅考虑不同数据集之间的域差, 同时考虑目标域内摄像头的风格差异对跨域自适应行人重识别性能的影响, 提出一种异构同质学习(hetero-homogeneous learning, HHL)方法. HHL利用StarGAN^[91]以无监督的方式生成具有不同风格的行人图像, 并将源域图像和目标域图像作为负样本对训练跨域连通性. 在Market-1501, DukeMTMC-reID和CUHK03数据集上的实验结果证明, HHL方法能够得到较高的跨域识别精度. 同样地, 梁文琦等^[59]在跨域风格转换的同时考虑域内相机各视角的差异, 提出一种基于多对多GAN的非对称跨域迁移的方法, 通过增加视角分类器用于鉴别不同视角的分布, 并针对不同源域和目标域的视角采取不同的迁移方式. 该方法能够平滑目标域摄像头视角差异, 进而使模型学习到视角不变性的特征, 有效地提高跨域行人重识别的准确性.

2.1.2 基于域内风格转换的方法

同一数据集通常包含多个摄像头, 由于光照、分辨率和背景的不同, 同一个行人图像在不同摄像头中会出现不同的风格. 另外, 同一个摄像头由于时间跨度不同也会产生不同风格的图像. 现有方法大多数通过学习行人判别性特性或稳定不变性特征来解决此类问题, 但当不同摄像头行人图像差异较大时, 会出现大量图像伪像, 从而干扰特征的学习^[61]. 因此, 同一数据集内行人图像风格的差异依然是行人重识别的主要挑战之一.

Zhong等^[60]引入相机样式(camera style, CamStyle)解决同一数据集中不同摄像头风格变化的问题, 其工作流程如图4所示. CamStyle首先使用CycleGAN将标注的训练数据迁移到不同的摄像

图 3 ATNet 结构示意图^[56]图 4 CamStyle 工作流程示意图^[60]

头, 使合成的样本在保留行人标签的同时具有不同摄像头的风格, 以此平滑同一数据集内不同摄像头之间的风格差异; 然后, 将生成图像加入训练集增加多样性, 以防止过拟合, 并引入标签平滑正则化(label smooth regularization, LSR)减少噪声的影响。虽然 CamStyle 通过平滑同一数据集内不同摄像头风格在一定程度上能够提高模型的性能, 但依然存在以下问题。

(1) 当同一数据集内不同摄像头拍摄的图像之间存在较大的风格差异时, CycleGAN 生成的转移样本会产生更多的失真图像, 使模型引入噪声, 并需要额外的 LSR 调整网络的性能。

(2) 生成的不同风格的图像只能作为扩展训练集, 当出现无效的生成样本时会严重影响训练模型的鲁棒性。

(3) CamStyle 为每个摄像头生成不同风格的行人图像, 随着摄像头的数量增加, 需要训练的模型数量越来越多, 将消耗大量的计算资源, 在实际场景中并不适用。

针对上述存在的问题, Liu 等^[61]提出统一风格自适应的方法, 称为 UnityStyle, 其工作流程如图 5 所示。与生成同一数据集内不同摄像头风格行人

样的方法不同, 该方法可以平滑同一摄像头以及同一数据集内不同摄像头之间的样式差异。通过为每个摄像头生成形状稳定的风格图像, 消除不同图像之间的风格差异, 并将真实图像和合成图像共同训练以增强训练集, 提高行人匹配的准确度。UnityStyle 合成的样本近乎原始图像, 无须引入 LSR 消除噪声的影响, 并且 UnityStyle 为所有的摄像头统一风格, 无须额外训练大量的模型。除了可见光图像之间的转换, GAN 的摄像头风格的转换也被应用于红外和可见光之间的相互转换。Wang 等^[62]提出一种双级差异减少学习(dual-level discrepancy reduction learning, D2RL)方案, 通过分解形态和外观差异, 实现红外图像和可见光图像的相互转换。

2.1.3 基于风格转换的方法总结

表 4 所示为从基于跨域风格转换和基于域内风格转换 2 个方面归纳总结相关方法。可以看到, 以 Resnet50^[99]作为基础网络和使用分类交叉熵损失约束网络模型训练是目前风格转换方法的常用策略。表 4 的第 1 部分为以 DukeMTMC-reID 作为源域数据集, Market-1501 作为目标域数据集(Duke→Market)的跨域风格转换的方法。从实验结果来看,

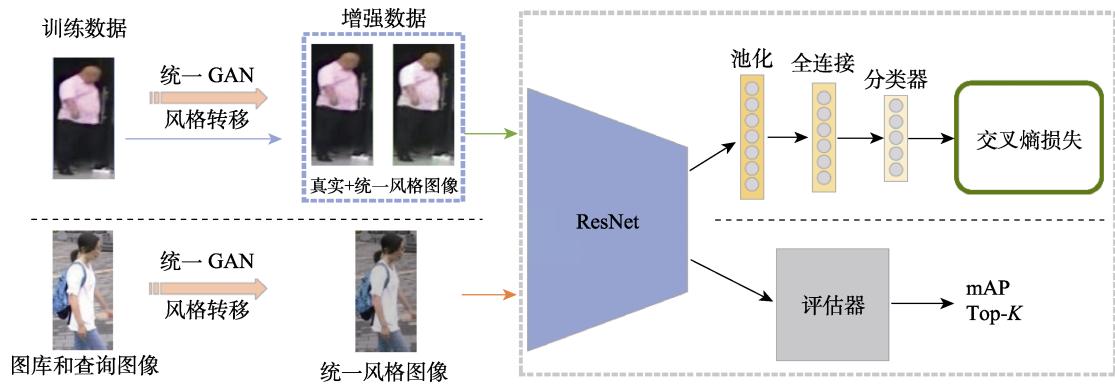
图5 UnityStyle统一摄像头风格的工作流程^[61]

表4 基于风格转换的方法总结

| 域间/内 | 方法 | 基础网络 | 损失函数 | mAP/% | Rank-1/% | 方法思想 |
|-------------|--|-------------------|------------|-------|----------|--|
| Duke→Market | PTGAN ^[53] | PTGAN+GoogleNet | 对抗损失+分类损失 | | 38.60 | 将源域中的行人转移到目标域, 使生成样本同时具有源域的行人ID和目标域的背景、光照样式 |
| | SPGAN ^[55] | CycleGAN+Resnet50 | 对比损失+分类损失 | 22.80 | 51.50 | 转换后的行人图像保持源域行人ID信息不变, 并具有目标域行人的风格 |
| | ATNet ^[56] | CycleGAN+Resnet50 | 对抗损失+分类损失 | 25.60 | 55.70 | 利用CycleGAN自适应地合成跨域不同光照、摄像头视角和分辨率的风格迁移图像 |
| | CR-GAN ^[57] | CR-GAN+Resnet50 | 对抗损失+分类损失 | 29.60 | 59.60 | 利用大量目标域的行人图像的背景信息和源域行人图像的身份信息生成具有丰富上下文信息的跨域图像 |
| | HHL ^[58] | StarGAN+Resnet50 | 三元组损失+分类损失 | 31.40 | 62.20 | 行人图像的摄像头风格转移的同时保持原图像ID不变, 并且源域和目标域图像组成负样本对以学习跨域连通性 |
| Market-1501 | IDE ^[4] +CamStyle ^[60] | CycleGAN+Resnet50 | 分类损失+LSR损失 | 68.72 | 88.12 | 合成具有不同摄像头风格的行人样本, 并引入标签平滑正则化以减少噪声的影响 |
| | IDE ^[4] +UnityStyle ^[61] | UnityGAN+Resnet50 | 分类损失 | 89.30 | 93.20 | 为每个相机生成形状稳定的风格图像, 以消除不同图像之间的样式差异, 并将真实图像和合成图像共同作用增强训练集 |

增加三元组损失约束可以提高跨域风格转换行人重识别的准确度。表4的第2部分为域内图像风格转换的方法, 转换的策略主要是以合成同一数据集内不同摄像头风格的样本或统一不同摄像头风格为主。从Market-1501数据集上的实验结果可以看出, 统一摄像头风格的方法能够得到更好的识别性能。

综上所述, 跨域风格转换的方法通过转换不同数据集行人图像的风格, 可以平滑域间行人图像的样式差异, 这种方式可以获得大量自动标注和具有目标域风格的合成图像, 并且可以和原始图像共同增强训练集, 缓解在不同数据集之间的域差。这类方法存在的问题在于合成的图像包含噪声, 当与源域图像共同用于模型训练时可能出现冲突, 影响模型对判别性特征的学习。对于域内图像风格转换的方法, 通过合成同一数据集内不同摄像头风格或者统一摄像头风格的行人样本, 可以最小化同一摄像头内和不同摄像头之间的图

像风格差异。但随着现实场景数据规模的扩大, 需要训练更多的模型用于生成不同摄像头风格的行人样本, 这将消耗大量的计算机资源, 严重制约了模型的泛化能力。在未来的研究中, 采用GAN生成更高质量的多样化行人样本, 设计更优的损失函数, 以约束模型对判别性特征和不变性特征的学习, 并提出更高效的训练策略, 是解决上述问题的研究重点。

2.2 基于数据增强的方法

与利用GAN进行风格转换以减小数据偏差的方法不同, 基于数据增强的方法从模型的训练出发, 通过增加多样性的训练数据提高模型的泛化能力。目前公开的行人重识别数据集相对现实世界场景的数据规模依然较小^[75], 并且缺乏大量姿态和外观变化的跨视角行人数据, 使模型很难学到行人判别性特征和不变性特征。本节将重点论述基于无辅助信息引导的图像合成方法和基于辅

助信息引导的图像合成方法在行人重识别任务中的影响与应用.

2.2.1 基于无辅助信息引导的图像合成方法

深度神经网络需要大量的数据用于模型训练, 手工标注行人图像需要耗费大量的人力成本, 极大地限制了训练数据集的规模. 为了得到更多数据, 通常使用无辅助信息引导合成未标注的行人图像, 并分配伪标签扩充训练数据.

Zheng 等^[63]最先使用深度卷积 GAN (deep convolutional GAN, DCGAN)^[94]生成样本数据, 其大致流程如图 6 所示. 首先, 通过无监督学习的方式, 使用原始训练数据生成未标注的行人样本; 然后, 通过半监督学习的方式, 利用离散标签正则化(label smoothing regularization for outliers, LSRO)为未标注图像均匀分配标签. 实验表明, 添加 DCGAN 生成的数据可使 CNN 学习到更具判别能力的特征. 由于 LSRO 为未标注的样本分配相同的虚拟标签, 且生成图像和原始图像之间存在偏差, 相同的标签可能分配给不同类别的生成图像, 使训练模型产生错误的预测结果. 因此, Huang 等^[64]提出一种多伪正则化标签(multi-pseudo regularized label, MpRL), 为每个生成的样本分配一个适当的虚拟标签, 以建立真实图像和生成图像的对应关系, 可以有效地区分不同的生成数据, 并且在 Market-1501, DukeMTMC-reID 和 CUHK03 等数据集上均能取得较好的识别效果.

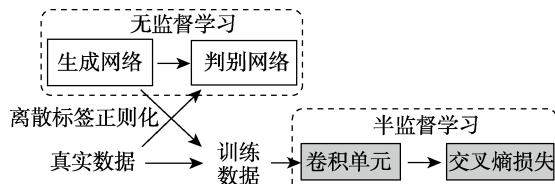


图 6 GAN 生成无标签样本用于模型训练^[63]

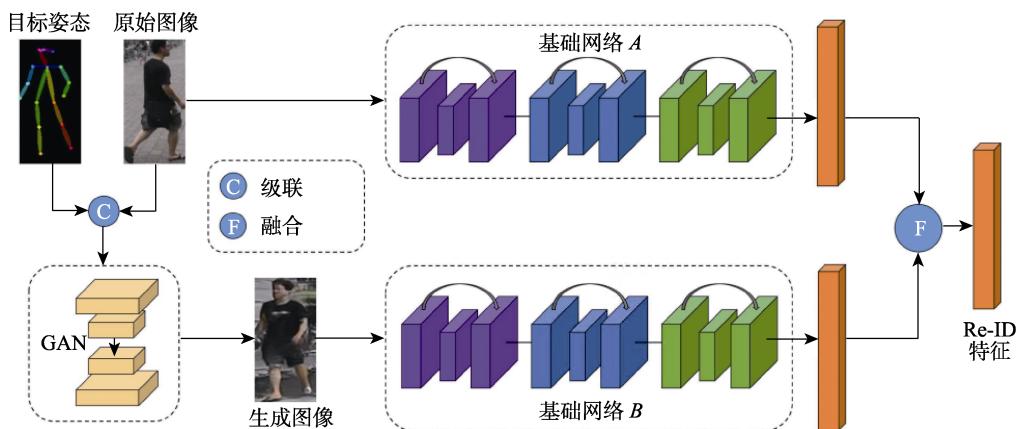


图 7 PN-GAN 的网络流程图^[68]

但是, 这类传统的无辅助信息引导的 GAN 存在以下问题: (1) 这类方法只关注生成图像的视觉效果, 生成样本的质量得不到保证, 容易生成一些严重扭曲的噪声图像, 从而影响训练模型的性能; (2) 传统无辅助信息引导的方法没有充分利用行人的结构和外观信息, 生成的图像不能很好地指导模型对判别性特征的学习, 极大地限制了生成样本的有效使用^[67].

2.2.2 基于辅助信息引导的图像合成方法

针对无辅助信息引导的图像合成方法存在的问题, Liu 等^[67]引入行人姿态信息辅助 GAN 生成样本. 具体地, 首先, 利用 MARS^[87]数据集的丰富行人信息引入大量行人骨架结构特征, 以构建姿态引导模型; 然后, 将得到的姿态信息和现有数据集行人的外观信息进行配准; 最后, 利用 GAN 生成同时具有 MARS 数据行人的姿态结构和现有数据集行人外观的样本图像. 这种方式生成的新样本能够保留现有数据集行人目标的身份信息, 无须重新标注.

姿态的多变性是影响网络模型对有效身份特征学习的重要原因. Qian 等^[68]利用姿态归一化 GAN(pose-normalization GAN, PN-GAN)生成具有统一姿态的行人图像, 其网络框架如图 7 所示. 首先, 利用 PN-GAN 合成具有不同姿势的辅助图像; 然后, 通过 2 个基础网络分别学习原始图像和合成图像的互补特征; 最后, 将 2 种类型的特征进行融合用作最终行人重识别任务的特征匹配. 早期的方法生成行人图像的姿态容易产生很大的偏差. Zhu 等^[69]采用多层次级联注意力网络训练判别器, 可以有效地利用姿态和外观特征优化行人的姿态转换, 使生成的行人图像与输入图像具有更好的姿态和外观一致性. 为了生成更丰富多样的行人

图像, Ma 等^[70]利用 GAN 将输入图像的 3 种变化因素(前景、背景和姿态)解构并编码为嵌入特征, 采用对抗性方式学习特征映射, 生成具有新的前景、背景和姿势的真实人物图像。

传统的基于无辅助信息引导的 GAN 合成图像的质量较差, 而基于姿态指导的 GAN 需要引入额外的辅助信息(如行人姿态检测器等), 增加了模型的训练成本和网络结构复杂程度。同时, 大多数方法生成器和判别器相互分离, 仅采用独立的生成模型生成图像, 可能使生成模块的优化无法与行

人重识别任务匹配, 限制了生成数据的可扩展性。因此, Zheng 等^[71]提出将判别器和生成器结合的学习框架, 其结构如图 8 所示。该框架能够将数据生成和行人重识别进行端到端的结合。生成模块利用现有数据集内部的行人外观和结构编码生成新的图像; 判别模块与生成模块共享外观编码器, 并作为行人重识别的主干网络。通过切换外观和结构编码, 可以生成高质量的交叉外观和结构的行人图像。该框架仅利用现有数据集的行人姿态和外观就能够合成多样性的行人样本。

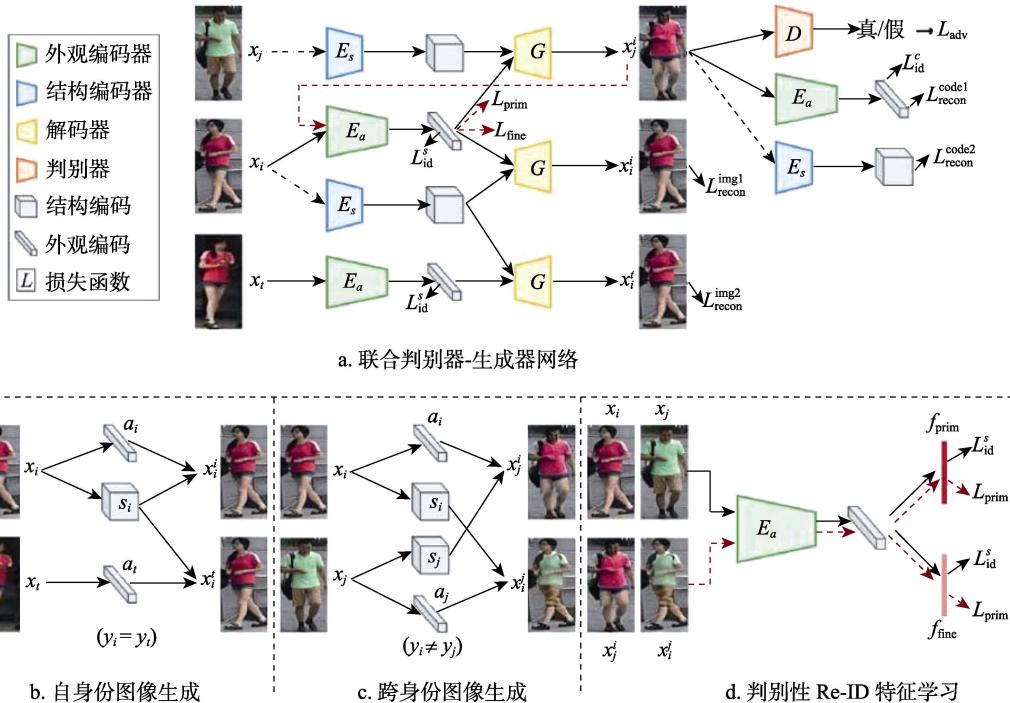


图 8 联合生成模块和判别模块的行人重识别框架^[71]

2.2.3 基于数据增强的方法总结

表 5 归纳比较了基于数据增强方法用到的基础网络、实验性能和损失函数等, 并简单地概述了其基本思想。这些方法包括基于无辅助-辅助信息引导的图像合成方法, 它们通常都采用具有简洁结构和竞争性的 Resnet50 和 DenseNet169^[100]作为特征学习的基础网络, 并且大部分方法都使用在图像像素对齐方面表现优异的 CGAN^[101]来生成行人样本。从 Market-1501 数据集上的实验结果可以看到, 采用辅助信息引导的方法性能优于传统无辅助信息引导的方法, 结合行人外观和姿态可以得到更好的识别性能。

综上所述, 采用 GAN 生成多样的行人图像在一定程度上缓解了现有训练数据不足的问题。对于无辅助信息引导的图像合成方法, 仅使用现有

训练数据即可合成大量的未标注行人图像。以半监督学习的方式为这些未标注样本分配标签, 可节省大量的人工标注成本, 并且生成图像和原始图像共同用于模型训练, 有助于提高神经网络对判别性特征的学习。虽然这类方法容易实现, 但生成图像的质量不高, 不具有足够的区别性信息。相比上述无辅助信息引导的图像合成方法, 有辅助信息引导的图像合成方法可以生成更多姿态和外观的行人图像, 能够有效地解决行人重识别任务中由跨视角配对训练数据缺乏、姿态变化引起的判别性特征和不变性特征学习困难的问题。但此类方法大多需要设计复杂的网络结构学习多样的行人姿态, 增加了额外的训练成本, 且通常情况下生成模型和行人重识别任务分离, 很难扩展到实际应用中。因此, 设计简单、高效的姿态提取方法和端到

表 5 基于数据增强的方法总结

| 方法 | 基础网络 | 损失函数 | mAP/% | Rank-1/% | 方法思想 |
|-------------------------------|-------------------|----------------|-------|----------|---|
| LSRO ^[63] | DCGAN + Resnet50 | 验证损失+分类损失 | 66.07 | 83.97 | 在不引入辅助信息的条件下生成未标注的行人样本, 利用半监督 LSRO 均匀分配标签 |
| MpRL ^[64] | GAN + Resnet50 | 分类损失 | 67.53 | 85.75 | 在不引入辅助信息的条件下生成未标注的行人样本, 采用半监督 MpRL 动态分配虚拟标签 |
| Post-transfer ^[67] | CGAN+ DenseNet169 | 分类损失+三元组损失 | 68.92 | 87.65 | 利用 MARS 数据集行人丰富的结构信息和现有数据集行人的外观信息生成多样化的样本 |
| PN-GAN ^[68] | GAN + Resnet50 | 分类损失 | 72.58 | 89.43 | 利用姿态归一化 GAN 生成统一姿态的行人样本, 并融合原始图像的互补信息生成更具判别性的特征 |
| PATN ^[69] | CGAN + Resnet50 | 对抗损失+ L_1 损失 | | | 采用多层级联判别器分别优化行人图像的外观和结构, 以提升生成图像的质量 |
| DPIG ^[70] | CGAN + Resnet50 | 对抗损失+ L_2 损失 | | | 利用 GAN 合成具有新的前景、背景和姿态的真实行人图像 |
| DG-Net ^[71] | GAN + Resnet50 | 对抗损失+分类损失 | 86.00 | 94.80 | 利用现有数据集内部行人的外观代码和结构代码生成新的、具有不同外观和姿态的行人图像 |

注. mAP 和 Rank-1 为在 Market-1501 数据集上的实验结果.

端的行人重识别任务依然是未来研究的重点工作.

2.3 基于不变性特征学习的方法

GAN 除了合成图像外, 还可用于特征学习. 通常现实场景下的行人重识别任务包含高级视觉变化和低级视觉变化^[54]. 前者主要包括行人的遮挡、姿态和摄像头视角的变化等; 后者主要包括分辨率、光照和天气变化等, 通常将低级视觉变化得到的图像称为降级图像. 这些视觉变化可能导致行人区别性特征信息丢失, 极大地阻碍行人身份特征的学习, 并可能引起特征不匹配, 使检索性能大大降低^[102]. 目前处理这些视觉变化的主要方法之一是利用 GAN 学习与身份相关而与视觉变化无关的特征^[54]. 本节根据视觉变化的不同, 将不变性特征学习的方法分为基于姿态不变性特征学习和基于降级不变性特征学习的方法.

2.3.1 基于姿态不变性特征学习的方法

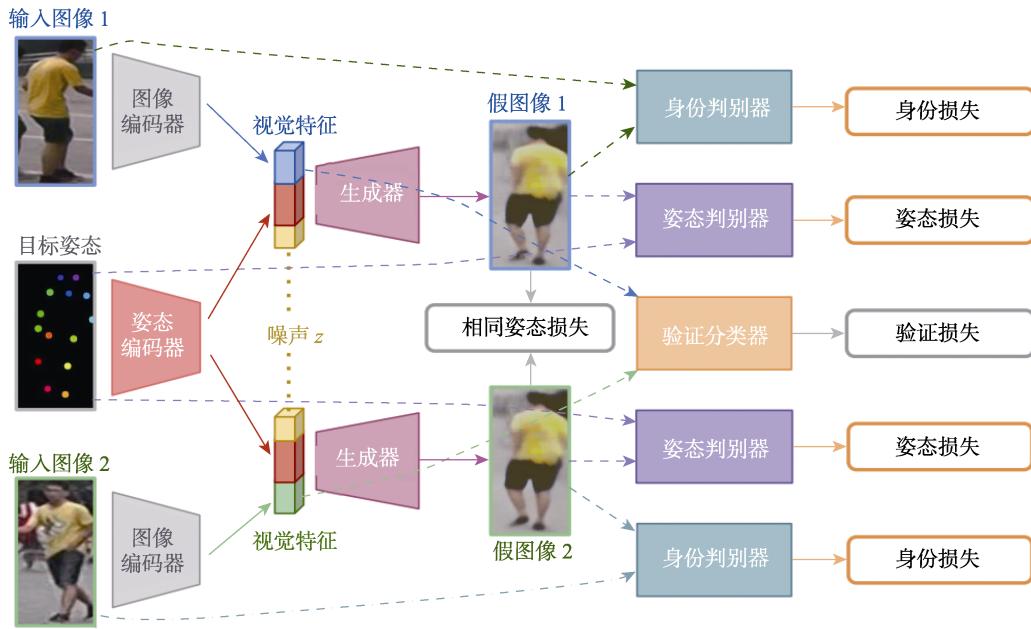
高级视觉行人姿态的变化, 主要影响模型对行人判别性特性的学习. 解决此类问题的常用方法是学习行人身体部位的局部特征或生成多姿态的行人用于模型训练. 但这些方法需要引入辅助的姿态估计信息和增加额外计算成本才能得到较好的效果. 针对这些问题, Ge 等^[72]提出特征提取 GAN(feature distilling GAN, FD-GAN), 学习与行人身份相关而与姿态无关的特征, 用以进行具有姿势变化的行人重识别, 其网络结构如图 9 所示. FD-GAN 采用孪生网络结构进行特征学习, 每个分支由一个图像编码器和图像生成器组成. 其中前者根据输入图像学习行人鲁棒性的与身份相关而与姿势无关的嵌入视觉特征, 后者根据编码器的姿态信息和输入行人特征合成新的行人样本. 身份判别器、姿态判别器、验证分类器和相同姿势

损失共同作用确保学习更多姿态不变性特征. 该方法无需额外的计算成本或辅助姿态信息, 并且在 Market-1501, CUHK03 和 DukeMTMC-reID 数据集上具有先进性的实验结果.

2.3.2 基于降级不变性特征学习的方法

现实场景中, 不同摄像头拍摄得到的行人图像分辨率大小差异很大, 查询图像和待查询图像的分辨率差异会导致特征很难对齐. 解决该问题的常用方法是利用图像的超分辨率(super resolution, SR)对低分辨率的行人图像进行上采样, 合成高分辨率的图像^[103-104], 从而减小图像之间的分辨率差异. 但这类方法的高分辨率图像合成和行人重识别任务相互分离, 无法保证合成的高分辨率图像满足行人匹配的要求^[54]. 因此, 有研究者考虑利用 GAN 学习低分辨率和高分辨率行人图像的共有不变性特征. Chen 等^[73]提出一种端到端的自适应分辨率行人重识别网络(resolution adaptation and re-identification network, RAIN), 通过在低分辨率图像和高分辨率图像特征上增加对抗损失, 学习和对齐不同分辨率行人图像的不变特征. Li 等^[74]提出的模型除了可以学习不同分辨率行人图像的不变性特征, 还能够利用 SR 恢复低分辨率图像丢失的细粒度细节信息, 有助于判别性特征的学习.

Huang 等^[54]提出一种降级不变性学习框架, 借助自我监督和对抗性训练策略, 可以保留与身份相关的鲁棒性特征并删除与降级相关的特征. 通过重组内容和降级特征, 可以生成具有降级提供者的降级特征和内容提供者的内容特征的新样本. 该框架提取的与降级无关的特征可作为行人重识别任务的身份特征表示, 并且生成的新样本可以增强训练数据, 进一步提高模型的性能.

图9 FD-GAN 的结构示意图^[72]

2.3.3 基于不变性特征学习的方法总结

表6归纳总结了GAN用于不变性特征学习的相关工作,包括基于姿态-降级不变性特征学习的方法。可以看出,以Resnet50作为基础网络,使用分类损失和三元组损失训练模型依然是目前大部

分研究者的选择。从各类方法在CUHK03数据集上的实验结果(mAP和Rank-1)来看,姿态不变性学习的方法FD-GAN能够得到最优的实验性能,同时关注光照和分辨率的降级不变性特征能够得到更好的实验结果。

表6 基于不变性特征学习的方法总结

| 方法 | 基础网络 | 损失函数 | mAP/% | Rank-1/% | 方法思想 |
|-------------------------|-----------------|------------|-------|----------|---|
| FD-GAN ^[72] | FD-GAN+Resnet50 | 分类损失+验证损失 | 91.3 | 92.6 | 能够学习与身份相关而与姿态无关的特征,测试阶段无需额外的计算成本或姿态信息 |
| RAIN ^[73] | GAN+Resnet50 | 分类损失+三元组损失 | 75.9 | 78.9 | 通过在低分辨率图像和高分辨率图像特征上增加对抗损失,用以学习和对齐不同分辨率行人图像的不变特征 |
| CAD-Net ^[74] | CRGAN+Resnet50 | 分类损失+三元组损失 | | 82.1 | 可以学习不同分辨率行人图像的不变性特征表示,同时能够恢复低分辨率图像丢失的细粒度细节信息 |
| DI-ReID ^[54] | DDGAN+Resnet50 | 分类损失+三元组损失 | | 85.7 | 借助自我监督和对抗性训练策略,可以保留与身份相关的鲁棒性特征并删除与降级相关的特征 |

综上所述,不变性特征学习的方法通过学习与行人身份相关,而与姿态、分辨率和光照无关的特征能够缓解行人特征未对齐的问题,提高行人身份匹配的准确度。但这类方法依然存在不足:(1)该方法采用随机姿态或简单姿态生成图像,因此会不可避免地引入噪声伪像,需要额外的判别器和损失函数对生成的图像进行筛选,增加了网络的复杂程度;(2)低分辨率图像容易丢失细粒度的区分信息,使学习高分辨率图像和低分辨率图像的不变性特征变得困难。在未来的研究中,设计更加简单、高效的模型,挖掘姿态、分辨率和光照不变性特征,依然是值得关注的研究内容。

3 基于GAN的方法性能对比分析

针对基于GAN的行人重识别方法,本文从近年来计算机视觉顶级会议(如CVPR, ICCV, ECCV和AAAI等)中选取具有代表性且准确度较高的一系列方法,对比分析了它们在CUHK03^[15],Market-1501^[82]和DukeMTMC-reID^[63]这3个经典数据集上的实验结果。这些实验结果均从原文献中得到且为无重排序^[9]的实验结果。

表7分别将Market-1501和DukeMTMC-reID数据集作为源域和目标域,对比了传统的手工特征提取的方法(LOMO^[105]和Bow^[82])、传统的无监督的

方法(如 UMDL^[106], CAMEL^[107]和 PUL^[96])以及基于GAN 的跨域风格转换的方法(如 CycleGAN(base)^[90], PTGAN^[53], SPGAN^[55], HHL^[58], ATNet^[56]和 CR-GAN^[57]). 从实验结果可以看出, 基于跨域风格转换的方法明显优于传统的手工特征提取的方法和传统无监督的方法. CycleGAN(base), PTGAN, SPGAN 和 ATNet 的实验结果明显劣于 CR-GAN 和 HHL 的方法. 前 4 种方法均使用 CycleGAN 生成无

标签的行人图像, CycleGAN 生成的转移样本会产生更多的失真图像, 使模型引入噪声, 并且需要额外的标签平滑正则化来调整网络的性能. CR-GAN 考虑将源域的行人身份转移到目标域的上下文中, 以保证合成更直观、更合理的多样化行人图像. HHL 在转换风格的过程中保留源域行人 ID 不变, 并且将源域和目标域图像组成负样本对以学习跨域连通性, 提高了模型在目标测试集上的泛化能力.

表 7 跨域风格转换的方法在 Market-1501 和 DukeMTMC-reID 上的实验结果

| 方法 | DukeMTMC-reID → Market-1501/% | | | | Market-1501 → DukeMTMC-reID/% | | | | 发表会议 |
|--------------------------------|-------------------------------|-------------|-------------|-------------|-------------------------------|-------------|-------------|---------|-----------|
| | mAP | Rank-1 | Rank-5 | Rank-10 | mAP | Rank-1 | Rank-5 | Rank-10 | |
| LOMO ^[105] | 8.0 | 27.2 | 41.6 | 49.1 | 4.8 | 12.3 | 21.3 | 26.6 | CVPR2015 |
| Bow ^[82] | 14.8 | 35.8 | 52.4 | 60.3 | 8.3 | 17.1 | 28.8 | 34.9 | ICCV2015 |
| UMDL ^[106] | 12.4 | 34.5 | 52.6 | 59.6 | 7.3 | 18.5 | 31.4 | 37.6 | CVPR2016 |
| CAMEL ^[107] | 26.3 | 54.5 | | | | | | | ICCV2017 |
| PUL ^[96] | 20.5 | 45.5 | 60.7 | 66.7 | 16.4 | 30.0 | 43.4 | 48.5 | ACMMM2018 |
| CycleGAN(base) ^[90] | 19.1 | 45.6 | 63.8 | 71.3 | 19.6 | 38.1 | 54.4 | 60.5 | CVPR2017 |
| PTGAN ^[53] | | 38.6 | | 66.1 | <u>27.4</u> | | 50.7 | | CVPR2018 |
| SPGAN ^[55] | 22.8 | 51.5 | 70.1 | 76.8 | 22.3 | 41.4 | 56.6 | 63.0 | CVPR2018 |
| HHL ^[58] | 31.4 | 62.2 | 78.8 | 84.0 | 27.2 | <u>46.9</u> | 61.0 | 66.7 | ECCV2018 |
| ATNet ^[56] | 25.6 | 55.7 | <u>73.2</u> | <u>79.4</u> | 24.9 | 45.1 | <u>59.5</u> | 63.0 | CVPR2019 |
| CR-GAN ^[57] | <u>29.6</u> | <u>59.6</u> | | | 30.0 | 52.2 | | | ICCV2019 |

注: 粗体和带下划线的数字分别表示排名最优和次优结果.

表 8 展示了基于域内风格转换、数据增强和不变性特征学习的方法在上述 3 个经典数据集上的实验结果. 对于基于域内风格转换的方法, IDE^[4]+ CameraStyle^[60] 和 IDE^[4]+UnityGAN^[61] 的方法均利用 GAN 生成同一数据集内不同摄像头风格的行人图像, 在 Market-1501 和 DukeMTMC-reID 数据集中测试均能得到较好的效果. 两者的区别在于, 前

者为同一数据集内的不同摄像头拍摄的行人图像相互转换风格, 随着摄像头数量的增加, 需要训练的模型数量越来越多, 消耗的计算资源越多, 在实际场景中并不适用; 后者通过为同一数据集内的每个摄像头生成形状稳定的风格图像, 消除不同图像之间的样式差异, 并将真实图像和合成图像共同训练以增强训练集, 从而提高行人匹配的准确度.

表 8 基于域内风格转换、数据增强和不变性特征学习的方法在 3 种图像数据集上的实验结果

| 分类 | 方法 | CUHK03/% | | Market-1501/% | | DukeMTMC-reID/% | | 发表会议 |
|-------------|---|-------------|-------------|---------------|-------------|-----------------|-------------|-------------|
| | | mAP | Rank-1 | mAP | Rank-1 | mAP | Rank-1 | |
| 域内风格 转换 | IDE ^[4] +CameraStyle ^[60] | | | 68.7 | 88.1 | 57.6 | 78.3 | CVPR2018 |
| | IDE ^[4] +UnityStyle ^[61] | | | 89.3 | <u>93.2</u> | <u>65.2</u> | <u>82.1</u> | CVPR2020 |
| 数据增强 | LSRO ^[63] | 87.4 | 84.6 | 66.1 | 84.0 | 47.1 | 67.7 | ICCV2017 |
| | PNGAN ^[68] | | | 79.7 | 72.6 | 89.4 | 53.2 | ECCV2018 |
| | PT ^[67] | | 42.0 | 45.1 | 68.9 | 87.6 | 56.9 | CVPR2018 |
| | MpRL ^[64] | | <u>87.5</u> | 85.4 | 67.5 | 85.7 | 58.5 | TIP2019 |
| | DG-Net ^[71] | | | | <u>86.0</u> | 94.8 | 74.8 | 86.6 |
| 不变性 特征学习 | FD-GAN ^[72] | 91.3 | 92.6 | 77.7 | 90.5 | 64.5 | 80.0 | NIPS2018 |
| | RAIN ^[73] | | | 78.9 | | | | AAAI2019 |
| | CAD-Net ^[74] | | | 82.1 | | 83.7 | | ICCV2019 |
| | DI-REID ^[54] | | | 85.7 | | | | CVPR2020 |

注: 粗体和带下划线的数字分别表示排名最优和次优结果.

对于数据增强的方法, LSRO^[63], MpRL^[64], PNGAN^[68], PT^[67]和 DG-Net^[71]利用 GAN 生成不同外观和姿态的行人, 用于扩充训练集以提高模型的泛化能力。从表 8 所示的实验结果可以看到, DG-Net 的方法明显优于前 4 种方法, 原因是 DG-Net 生成不同姿态和外观的多样化行人图像, 而 PNGAN 仅生成不同姿态的行人, 并且 LSRO, MpRL 和 PT 属于无监督图像生成的方法, 生成样本的质量得不到保证, 容易生成一些严重扭曲的噪声图像。

对于不变性特征学习的方法, FD-GAN^[72], RAIN^[73], CAD-Net^[74]和 DI-REID^[54]利用 GAN 学习与行人姿态、分辨率和光照无关而与行人身份相关的特征。从表 8 中的实验结果可以看到, FD-GAN 在 3 种数据集测试性能最高, 其学习与行人身份相关和姿态无关的特征, 减小姿态对行人重识别准确度的影响; 后 3 个方法通过学习不同光照、分辨率下的行人不变性特征, 也能够得到较好的实验结果, 但相比 RAIN 和 CAD-Net 仅关注分辨率不变性特征, DI-REID 同时考虑光照和分辨率的降级不变性特征学习, 能得到更好的实验结果。

4 结语

行人重识别是计算机视觉领域的热门研究话题, 具有广泛的应用前景。针对基于 GAN 的行人重识别方法, 本文首先介绍目前行人重识别主流的数据集; 然后, 重点归纳了现有的基于 GAN 的行人重识别方法, 从风格转换、数据增强和不变性特征学习 3 类方法对现有研究工作进行综述, 并总结各类方法的优缺点; 最后, 对比分析这些方法在几种主流行人重识别数据集上的实验性能。

尽管基于 GAN 的方法对解决行人重识别任务中多样性训练数据缺乏、域差和摄像头风格变化等问题具有较好的性能, 但仍然在以下几个方面存在一些问题。

(1) 高质量数据集的获取。现有的行人重识别数据集采集的区域和时间跨度相对较小, 缺乏大量姿态和外观变化的跨视角行人数据, 使模型很难学到行人的判别性特征或不变性特征。同时, 受光照、分辨率和摄像头视角等因素的影响, 传统无监督 GAN 生成的行人样本包含失真图像, 使模型引入噪声, 需要增加额外的标签平滑正则化来调整网络的性能。未来基于 GAN 的数据集扩充应考虑结合更多现实场景下的行人外观、姿态和摄像头

样式信息, 设计更优的生成器、判别器和损失函数来约束训练模型, 从而获得更高质量的行人图像或视频样本。

(2) 模型的跨域自适应。由于不同数据集的数据分布存在很大差异, 当在不同数据集上分别进行训练和测试时, 模型的性能会出现严重下降, 因此无法有效地将模型泛化到新测试集。传统基于 GAN 的跨域风格转换的方法需要针对每个场景都训练一个模型, 可扩展性差。未来可考虑引入迁移学习的方法, 使一个场景的训练模型能够自适应地迁移到其他场景。

(3) 特征的对齐。虽然不变性特征学习的方法能缓解因行人姿态、分辨率和光照变化引起的特征未对齐问题, 但这类方法需要设计更多的判别器和损失函数对模型进行约束, 增加了训练模型的复杂度, 并且低分辨率图像容易丢失细粒度的区分信息, 使学习高分辨率图像和低分辨率图像的不变性特征变得困难。未来可考虑引入具有强大自动关系建模能力的图卷积神经网络, 对行人图像的时间和空间关系进行建模, 以挖掘更多的区分性和鲁棒性的特征。

(4) 端到端生成模型与行人重识别模型的设计。目前基于 GAN 的大部分行人重识别任务, 生成模型和行人重识别模型相互独立, 生成模型的优化无法与行人重识别任务匹配, 因此, 很难推广到实际应用当中。如何有效地结合生成模型和行人重识别模型, 设计端到端的行人重识别任务, 也是未来需要研究者关注的方向。

参考文献(References):

- [1] Wu D, Zheng S J, Zhang X P, et al. Deep learning-based methods for person re-identification: a comprehensive review[J]. Neurocomputing, 2019, 337(14): 354-371
- [2] Bedagkar-Gala A, Shah S K. A survey of approaches and trends in person re-identification[J]. Image & Vision Computing, 2014, 32(4): 270-286
- [3] Wang X G. Intelligent multi-camera video surveillance: a review[J]. Pattern Recognition Letters, 2013, 34(1): 3-19
- [4] Zheng L, Yang Y, Hauptmann A G. Person re-identification: past, present and future[OL]. [2020-12-18]. <https://arxiv.org/abs/1610.02984>
- [5] Redmon J, Farhadi A. YOLOv3: an incremental improvement[OL]. [2020-12-18]. <https://arxiv.org/abs/1804.02767>
- [6] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[C] //Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2016: 21-37
- [7] Girshick R. Fast R-CNN[C] //Proceedings of the IEEE Interna-

- tional Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2015: 1440-448
- [8] Qi Lei, Yu Peize, Gao Yang. Research on weak-supervised person re-identification[J]. Journal of Software, 2020, 31(9): 2883-2902(in Chinese)
(祁磊, 于沛泽, 高阳. 弱监督场景下的行人重识别研究综述[J]. 软件学报, 2020, 31(9): 2883-2902)
- [9] Zhong Z, Zheng L, Cao D, et al. Re-ranking person re-identification with k -reciprocal encoding[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 1318-1327
- [10] Zhao R, Ouyang W, Wang X. Person re-identification by salience matching[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2013: 2528-2535
- [11] Martinel N, Micheloni C, Foresti G L. Saliency weighted features for person re-identification[C] //Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2014: 191-208
- [12] An L, Chen X J, Liu S, et al. Integrating appearance features and soft biometrics for person re-identification[J]. Multimedia Tools and Applications, 2017, 76(9): 12117-12131
- [13] Hu H M, Fang W, Zeng G D, et al. A person re-identification algorithm based on pyramid color topology feature[J]. Multimedia Tools & Applications, 2017, 76(24): 26633-26646
- [14] Dikmen M, Akbas E, Huang T S, et al. Pedestrian recognition with a learned metric[C] //Proceedings of the Asian Conference on Computer Vision. Heidelberg: Springer, 2011: 501-512
- [15] Li W, Zhao R, Wang X G. Human reidentification with transferred metric learning[C] //Proceedings of the Asian Conference on Computer Vision. Heidelberg: Springer, 2012: 31-44
- [16] Chen Y C, Zheng W S, Lai J H. Mirror representation for modeling view-specific transform in person re-identification[C] //Proceedings of the 24th International Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2015: 3402-3408
- [17] Wang X J, Zheng W S, Li X, et al. Cross-scenario transfer person on reidentification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 26(8): 1447-1460
- [18] Lu Ping, Dong Husheng, Zhong Shan, et al. Person re-identification by cross-view discriminative dictionary learning with metric embedding[J]. Journal of Computer Research and Development, 2019, 56(11): 2424-2437(in Chinese)
(陆萍, 董虎胜, 钟珊, 等. 基于跨视角判别词典嵌入的行人再识别[J]. 计算机研究与发展, 2019, 56(11): 2424-2437)
- [19] Wu Y, Lin Y T, Dong X Y, 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. Los Alamitos: IEEE Computer Society Press, 2018: 5177-5186
- [20] Zheng M, Karanam S, Wu Z Y, et al. Re-identification with consistent attentive Siamese networks[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 5735-5744
- [21] Varior R R, Haloi M, Wang G. Gated Siamese convolutional neural network architecture for human re-identification[C] //Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2016: 791-808
- [22] Hermans A, Beyer L, Leibe B. In defense of the triplet loss for person re-identification[OL]. [2020-12-18]. <https://arxiv.org/abs/1703.07737>
- [23] Chen W H, Chne X T, Zhang J G, et al. Beyond triplet loss: a deep quadruplet network for person re-identification[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 403-412
- [24] Chen Qiaoyuan, Chen Ying. TriHard loss based multi-task person re-identification[J]. Journal of Computer-Aided Design & Computer Graphics, 2019, 31(7): 1156-1165(in Chinese)
(陈巧媛, 陈莹. 基于困难样本三元组损失的多任务行人再识别[J]. 计算机辅助设计与图形学学报, 2019, 31(7): 1156-1165)
- [25] Sun Y F, 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. Heidelberg: Springer, 2018: 480-496
- [26] Sun Y F, Xu Q, Li Y L, et al. Perceive where to focus: learning visibility-aware part-level features for partial person re-identification[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 393-402
- [27] Chen T L, Ding S J, Xie J Y, et al. ABD-Net: attentive but diverse person re-identification[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2019: 8351-8361
- [28] Miao J X, Wu Y, Liu P, et al. Pose-guided feature alignment for occluded person re-identification[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2019: 542-551
- [29] Zhao H Y, Tian M Q, Sun S Y, et al. Spindle net: person re-identification with human body region guided feature decomposition and fusion[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 1077-1085
- [30] Liu X H, Zhao H Y, Tian M Q, et al. HydraPlus-Net: attentive deep features for pedestrian analysis[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 350-359
- [31] Li W, Zhu X T, Gong S G. Harmonious attention network for person re-identification[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 2285-2294
- [32] Zhao L M, Li X, Zhuang Y T, et al. Deeply-learned part-aligned representations for person re-identification[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 3219-3228
- [33] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks[C] //Proceedings of the IEEE conference on computer vision and pattern recognition. Los Alamitos: IEEE Computer Society Press, 2018: 7132-7141
- [34] Chen B H, Deng W H, Hu J N. Mixed high-order attention network for person re-identification[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2019: 371-381
- [35] Tay C P, Roy S, Yap K H. AANet: attribute attention network

- for person re-identifications[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 7134-7143
- [36] Chen Qiaoyuan, Chen Ying. Correlation channel-wise based part aligned representations for person re-identification[J]. Journal of Computer-Aided Design & Computer Graphics, 2020, 32(8): 1258-1266(in Chinese)
(陈巧媛, 陈莹. 通道互注意机制下的部位对齐行人再识别[J]. 计算机辅助设计与图形学学报, 2020, 32(8): 1258-1266)
- [37] Wang G S, Yuan Y F, Chen X, et al. Learning discriminative features with multiple granularities for person re-identification [C] //Proceedings of the 26th ACM International Conference on Multimedia. New York: ACM Press, 2018: 274-282
- [38] Wei L H, Zhang S L, Yao H T, et al. GLAD: global-local-alignment descriptor for pedestrian retrieval[C] //Proceedings of the 25th ACM International Conference on Multimedia. New York: ACM Press, 2017: 420-428
- [39] Su C, Li J N, Zhang S L, et al. Pose-driven deep convolutional model for person re-identification[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 3960-3969
- [40] Chen X S, Fu C M, Zhao Y, et al. Salience-guided cascaded suppression network for person re-identification[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 3300-3310
- [41] Zhang Z Z, Lan C L, Zeng W J, et al. Relation-aware global attention for person re-identification[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 3186-3195
- [42] Zheng F, Deng C, Sun X, et al. Pyramidal person re-identification via multi-loss dynamic training[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 8514-8522
- [43] Yao H T, Zhang S L, Hong R C, et al. Deep representation learning with part loss for person re-identification[J]. IEEE Transactions on Image Processing, 2019, 28(6): 2860-2871
- [44] Chung D, Tahboub K, Delp E J. A two stream Siamese convolutional neural network for person re-identification[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 1983-1991
- [45] Liao X Y, He L X, Yang Z W, et al. Video-based person re-identification via 3D convolutional networks and non-local attention[C] //Proceedings of the Asian Conference on Computer Vision. Heidelberg: Springer, 2018: 620-634
- [46] Fu Y, Wang X Y, Wei Y C, et al. STA: Spatial-temporal attention for large-scale video-based person re-identification[C] //Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2019: 8287-8294
- [47] Hou R B, Ma B P, Chang H, et al. VRSTC: occlusion-free video person re-identification[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 7183-7192
- [48] Yang J R, Zheng W S, Yang Q, et al. Spatial-temporal graph convolutional network for video-based person re-identification [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 3289-3299
- [49] Yan Y C, Qin J, Chen J X, et al. Learning multi-granular hypergraphs for video-based person re-identification[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 2899-2908
- [50] Wu Y M, Bourahla O E F, Li X, et al. Adaptive graph representation learning for video person re-identification[J]. IEEE Transactions on Image Processing, 2020, 29: 8821-8830
- [51] Yan Y C, Zhang Q, Ni B B, et al. Learning context graph for person search[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 2158-2167
- [52] Shen Y T, Li H S, Yi S A, et al. Person re-identification with deep similarity-guided graph neural network[C] //Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2018: 508-526
- [53] Wei L H, Zhang S L, 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. Los Alamitos: IEEE Computer Society Press, 2018: 79-88
- [54] Huang Y K, Zha Z J, Fu X Y, et al. Real-world person re-identification via degradation invariance learning[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 14084-14094
- [55] Deng W J, Zheng L, Ye Q X, et al. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 994-1003
- [56] Liu J W, Zha Z J, Chen D, et al. Adaptive transfer network for cross-domain person re-identification[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 7202-7211
- [57] Chen Y B, Zhu X T, Gong S G. Instance-guided context rendering for cross-domain person re-identification[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2019: 232-242
- [58] Zhong Z, Zheng L, Li S Z, et al. Generalizing a person retrieval model hetero and homogeneously[C] //Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2018: 172-188
- [59] Liang Wenqi, Wang Guangcong, Lai Jianhuang. Asymmetric cross-domain transfer learning of person re-identification based on the many-to-many generative adversarial network[J]. Acta Automatica Sinica, 2022, 48(1): 1001-1018(in Chinese)
(梁文琦, 王广聰, 赖剑煌. 基于多对多生成对抗网络的非对称跨域迁移行人再识别[J]. 自动化学报, 2022, 48(1): 1001-1018)
- [60] Zhong Z, Zheng L, Zheng Z D, et al. Camera style adaptation for person re-identification[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 5157-5166
- [61] Liu C, Chang X J, Shen Y D. Unity style transfer for person

- re-identification[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 6887-6896
- [62] Wang Z X, Wang Z, Zheng Y Q, et al. Learning to reduce dual-level discrepancy for infrared-visible person re-identification[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 618-626
- [63] Zheng Z D, 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. Los Alamitos: IEEE Computer Society Press, 2017: 3754-3762
- [64] Huang Y, Xu J S, Wu Q, et al. Multi-pseudo regularized label for generated data in person re-identification[J]. IEEE Transactions on Image Processing, 2019, 28(3): 1391-1403
- [65] Dai Chenchao, Wang Hongyuan, Ni Tongguang, et al. Person re-identification based on deep convolutional generative adversarial network and expanded neighbor reranking[J]. Journal of Computer Research and Development, 2019, 56(8): 1632-1641(in Chinese)
(戴臣超, 王洪元, 倪彤光, 等. 基于深度卷积生成对抗网络和拓展邻近排序的行人重识别[J]. 计算机研究与发展, 2019, 56(8): 1632-1641)
- [66] Yang Wanxiang, Yan Yan, Chen Si, et al. Multi-scale generative adversarial network for person reidentification under occlusion[J]. Journal of Software, 2020, 31(7): 1943-1958(in Chinese)
(杨婉香, 严严, 陈思, 等. 基于多尺度生成对抗网络的遮挡行人重识别方法[J]. 软件学报, 2020, 31(7): 1943-1958)
- [67] Liu J X, Ni B B, Yan Y C, et al. Pose transferrable person re-identification[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 4099-4108
- [68] Qian X L, Fu Y W, Xiang T, et al. Pose-normalized image generation for person re-identification[C] //Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2018: 661-678
- [69] Zhu Z, Huang T T, Shi B G, et al. Progressive pose attention transfer for person image generation[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 2347-2356
- [70] Ma L Q, Sun Q R, Georgoulis S, et al. Disentangled person image generation[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 99-108
- [71] Zheng Z D, Yang X D, Yu Z D, et al. Joint discriminative and generative learning for person re-identification[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 2138-2147
- [72] Ge Y X, Li Z W, Zhao H Y, et al. FD-GAN: pose-guided feature distilling gan for robust person re-identification[C] //Proceedings of the Advances in Neural Information Processing Systems. New York: ACM Press, 2018: 1230-1241
- [73] Chen Y C, Li Y J, Du X F, et al. Learning resolution-invariant deep representations for person re-identification[C] //Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2019: 8215-8222
- [74] 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. Los Alamitos: IEEE Computer Society Press, 2019: 8090-8099
- [75] Luo Hao, Jiang Wei, Fan Xing, et al. A survey on deep learning based person re-identification[J]. Acta Automatica Sinica, 2019, 45(11): 2032-2049(in Chinese)
(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)
- [76] Ye M, Shen J B, Lin G J, et al. Deep learning for person re-identification: A survey and outlook[OL]. [2020-12-18]. <https://arxiv.org/pdf/2001.04193v1.pdf>
- [77] Yao Zu, Gong Xun, Chen Rui, et al. Research progress, challenge and prospect of local features for person re-identification[J]. Acta Automatica Sinica, 2021, 47(12): 1-19(in Chinese)
(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): 1-19)
- [78] Gray D, Brennan S, Tao H. Evaluating appearance models for recognition, reacquisition, and tracking[C] //Proceedings of the IEEE International Workshop on Performance Evaluation for Tracking and Surveillance. Los Alamitos: IEEE Computer Society Press, 2007: 41-47
- [79] Li W, Wang X G. Locally aligned feature transforms across views[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2013: 3594-3601
- [80] 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. Los Alamitos: IEEE Computer Society Press, 2014: 152-159
- [81] Felzenszwalb P F, Girshick R B, McAllester D, et al. Object detection with discriminatively trained part-based models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627-1645
- [82] Zheng L, Shen L Y, Tian L, et al. Scalable person re-identification: A benchmark[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2015: 1116-1124
- [83] Ristani E, Solera F, Zou R, et al. Performance measures and a data set for multi-target, multi-camera tracking[C] //Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2016: 17-35
- [84] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149
- [85] Hirzer M, Beleznai C, Roth P M, et al. Person re-identification by descriptive and discriminative classification[C] // Proceedings of the Scandinavian Conference on Image Analysis. Heidelberg: Springer, 2011: 91-102
- [86] Wang T Q, Gong S G, Zhu X T, et al. Person re-identification by video ranking[C] //Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2014: 688-703
- [87] Zheng L, Bie Z, Sun Y F, et al. MARS: A video benchmark for large-scale person re-identification[C] //Proceedings of the

- European Conference on Computer Vision. Heidelberg: Springer, 2016: 868-884
- [88] Song G L, Leng B, Liu Y, et al. Region-based quality estimation network for large-scale person re-identification[OL]. [2020-12-18]. <https://arxiv.org/abs/1711.08766v2>
- [89] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[C] //Proceedings of the Advances in Neural Information Processing Systems. New York: ACM Press, 2014: 2672-2680
- [90] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 2223-2232
- [91] Choi Y, Choi M, Kim M, et al. StarGAN: unified generative adversarial networks for multi-domain image-to-image translation[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 8789-8797
- [92] Bak S, Carr P, Lalonde J F. Domain adaptation through synthesis for unsupervised person re-identification[C] //Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2018: 193-209
- [93] Cheng Z Y, Dong Q, Gong S G, et al. Inter-task association critic for cross-resolution person re-identification[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 2605-2615
- [94] Liu J W, Zha Z J, Hong R C, 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: ACM Press, 2019: 665-673
- [95] Zha Z J, Liu J W, Chen D, et al. Adversarial attribute-text embedding for person search with natural language query[J]. IEEE Transactions on Multimedia, 2020, 22(7): 1836-1846
- [96] Fan H H, Zheng L, Yan C G, et al. Unsupervised person re-identification: clustering and fine-tuning[J]. ACM Transactions on Multimedia Computing Communications and Applications, 2018, 14(4): Article No.83
- [97] Torralba A, Efros A A. Unbiased look at dataset bias[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2011: 1521-1528
- [98] Patel V M, Gopalan R, Li R N, et al. Visual domain adaptation: A survey of recent advances[J]. IEEE Signal Processing Magazine, 2015, 32(3): 53-69
- [99] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 770-778
- [100] 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. Los Alamitos: IEEE Computer Society Press, 2017: 4700-4708
- [101] Mirza M, Osindero S. Conditional generative adversarial nets [OL]. [2020-12-18]. <https://arxiv.org/abs/1411.1784>
- [102] Mao S N, Zhang S L, Yang M. Resolution-invariant person re-identification[OL]. [2020-12-18]. <https://arxiv.org/abs/1906.09748>
- [103] Jiao J N, Zheng W S, Wu A C, et al. Deep low-resolution person re-identification[C] //Proceedings of the 32nd AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2018: 6967-6974
- [104] Wang Z, Ye M, Yang F, et al. Cascaded SR-GAN for scale-adaptive low resolution person re-identification[C] //Proceedings of the 27th International Joint Conference on Artificial Intelligence. New York: ACM Press, 2018: 3891-3897
- [105] Liao S C, Hu Y, Zhu X Y, et al. Person re-identification by local maximal occurrence representation and metric learning[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2015: 2197-2206
- [106] Peng P X, Xiang T, Wang Y W, et al. Unsupervised cross-dataset transfer learning for person re-identification[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 1306-1315
- [107] Yu H X, Wu A C, Zheng W S. Cross-view asymmetric metric learning for unsupervised person re-identification[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 994-1002