

基于深度学习的行人重识别方法综述

李 擎^{1,2,3)}, 胡伟阳^{1,2)}, 李江昀^{1,2,3)}✉, 刘 艳^{1,2)}, 李梦璇^{1,2)}

1) 北京科技大学自动化学院, 北京 100083 2) 工业过程知识自动化教育部重点实验室, 北京 100083 3) 北京科技大学顺德研究生院, 佛山 528000

✉通信作者, E-mail: leejy@ustb.edu.cn

摘 要 对深度学习在行人重识别领域的应用现状进行总结与评价。首先, 对行人重识别进行介绍, 包括行人重识别的应用场景、数据集与评价指标, 并对基于深度学习的行人重识别的基本方法进行总结。之后, 针对行人重识别的研究现状, 将近年来国内外学者的研究工作归纳为基于局部特征、基于生成对抗网络、基于视频以及基于重排序 4 个方向, 并对每个方向所使用的方法分别进行梳理、性能对比以及总结。最后, 对行人重识别领域现存的问题进行了分析与讨论, 并探讨了行人重识别未来的发展方向。

关键词 深度学习; 行人重识别; 局部特征; 生成对抗网络; 视频数据; 重排序

分类号 TG183

A survey of person re-identification based on deep learning

LI Qing^{1,2,3)}, HU Wei-yang^{1,2)}, LI Jiang-yun^{1,2,3)}✉, LIU Yan^{1,2)}, LI Meng-xuan^{1,2)}

1) School of Automation & Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China

2) Key Laboratory of Knowledge Automation for Industrial Processes, Ministry of Education, Beijing 100083, China

3) Shunde Graduate School, University of Science and Technology Beijing, Foshan 528000, China

✉ Corresponding author, E-mail: leejy@ustb.edu.cn

ABSTRACT Person re-identification is an important part of multi-target tracking across cameras; its aim is to identify the same person across different cameras. Given a query image, the purpose of person re-identification is to find the best match for the query image in an image set. Person re-identification is a key component in an intelligent security system; it is beneficial for building a smart bank or smart factory and plays a crucial role in the construction of a smart city. Nowadays, with the development of artificial intelligence and increasing demand for precise identification in practical scenarios, deep learning-based person re-identification technology has become a popular research topic; this technology has achieved state-of-the-art results in comparison with conventional approaches. Although there are many recently proposed networks with stronger representation ability and a high level of accuracy for person re-identification, there also exist some problems that should be considered and solved. These include the insufficient generalization ability of various poses, the inability to fully utilize the temporal information, and the ineffective identification of occluded objects. As a result, many scholars have researched this field and have pointed out some promising solutions to cope with the aforementioned problems. This paper aims to summarize the application of deep learning in the field of person re-identification along with its advantages and shortcomings. First, the background of person re-identification is introduced, including the application scenarios, datasets, and evaluation indicators. Additionally, some basic methods of person re-identification based on deep learning are summarized. According to the existing research on person re-identification, the main approaches proposed by scholars worldwide can be summarized into four aspects, which are based

收稿日期: 2020–12–22

基金项目: 中央高校基本科研业务费专项资金资助项目 (FRF-DF-19-002); 北京科技大学顺德研究生院科技创新专项资金资助项目 (BK20BE014)

on local features, generative adversarial networks, video data, and re-ranking. A detailed comparative study of these four methods is then conducted. Finally, the existing problems and future studies that can be done in the field of person re-identification are analyzed and discussed.

KEY WORDS deep learning; person re-identification; local feature; generating adversarial networks; video data; reranking

行人重识别(Person re-identification, Person ReID), 是利用计算机视觉技术对跨摄像头的人员图像进行特征提取, 从而进行人员匹配的技术, 广泛应用于智慧安防、智慧零售等领域, 是人工智能技术与产业结合的重要支点. 在早期的行人重识别研究中, 大量的工作是基于传统计算机视觉方法, 利用手工提取的图像特征, 进行行人重识别的研究. 近年来, 随着深度学习在许多领域的成功应用, 研究人员们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, 将图像之间的差异量化, 并使用大量标注数据对模型进行训练迭代. 本文首先对行人重识别的概念、数据集以及两个基本方法进行介绍. 之后, 根据研究思路的不同, 将当前的方法分为基于局部特征、基于生成对抗网络、基于视频以及基于重排序 4 个方向并分别对其进行分析和讨论. 最后, 对行人重识别现存问题进行归纳整理, 并对未来的研究方向进行展望.

1 行人重识别

行人重识别是长时、跨域多目标跟踪的核心技术^[1], 其主要目标是对同一个行人进行跨摄像头的再识别. 在研究中, 通常给定一个目标行人图像或者视频片段, 在待匹配的图像序列或者视频片段中, 识别出目标行人, 并给出目标图像与匹配图像的相似度. 如图 1 所示, 两个视野不重合的监控摄像头, 在不同的时段分别采集了一些行人图像, 从摄像头 1 采集的图像中选择一张作为目标图像. 行人重识别的目标就是在摄像头 2 所采集的图像中, 通过寻找和目标图像匹配度最高的图像, 再次将该人物识别出来, 实现对行人的跨摄像头跟踪.

行人重识别是一项具有挑战性的任务, 它的数据一般来自于安防系统. 在系统中, 不同摄像设备之间的背景、风格具有较大差异, 同时行人外观易受遮挡、穿着、视角和姿态等多种因素的影响, 这就使得行人重识别成为一个具有一定挑战性的研究课题, 同时具有较高的研究价值. 图 2 对当前基于深度学习的行人重识别存在的问题与解决方

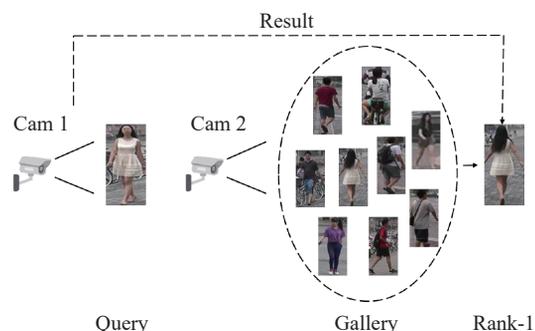


图 1 行人重识别的应用场景示例

Fig.1 An example of person re-identification application scenarios

法进行了梳理.

1.1 行人重识别数据集介绍

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, 离不开大规模数据集的发展^[2]. 目前常用的行人重识别数据集主要有以下几种: CUHK03 数据集^[3], Market1501^[4] 数据集以及 DukeMTMC-reID^[5], 其中 CUHK03 数据集的 ID 划分分为新旧两种协议, 即 1367/100 与 767/700^[6], 具体信息如表 1 所示. 在基于视频的行人重识别领域广泛应用的数据集主要有 PRID2011^[7] 以及 iLIDS-VID^[8], 分别收录了 984 与 300 个行人的视频信息, 其他视频数据集信息如表 2 所示.

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: 北京大学采集的 MSMT17^[9] 数据集, 包含室内室外 15 个相机的 12 万多张行人图像, 有 4 千多个行人 ID, 是目前最大的单帧 ReID 数据集; 北京大学和微软研究院联合采集的 LvreID 数据集, 包含室内室外 15 个相机的 3 千多个行人 ID 的序列图像, 总共 14943 个序列的 3 百多万张图像; 北京航空航天大学等采集的 LPW^[10] 数据集, 包含 2731 个行人的 7694 个轨迹序列, 总共有 56 万多张图像, 该数据集的特点是有多个独立的场景, 每个场景都可以作为一个独立的数据集, 训练集和测试集按照场景分开, 因此更加接近真实使用情况; Yu 等^[11] 提出一种衣着变换的数据集 COCAS, 含有 5266 个行人 ID, 共 62382 个行人检测框, 由手工标注得到. 2020 年以来, 一些优秀的行人重识别数据集也被相继提出; 由中山大学提出的 SYSU-30^[12]

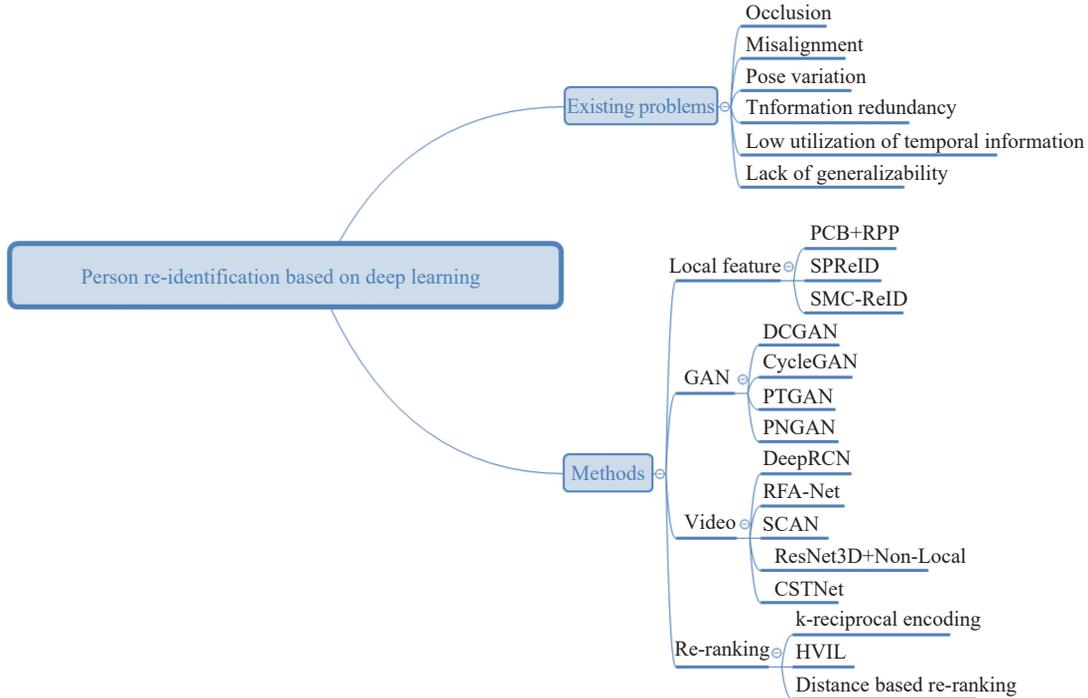


图 2 基于深度学习的行人重识别研究问题与方法归纳

Fig.2 Research problems and methods of person re-identification based on deep learning

表 1 部分行人重识别公开数据集

Table 1 Part of person re-identification public datasets

Dataset	Camera numbers	ID numbers	Image numbers	Body images
Market-1501 ^[4]	6	1501	32668	DPM
DukeMTMC-reID ^[5]	8	1812	36411	Hand
MSMT17 ^[9]	15	4101	126441	Faster RCNN
CUHK03 ^[3]	2	1467	14096	Hand
LPW ^[10]	11	2731	562438	DPM+Hand
COCAS ^[11]	30	5266	62382	Hand

表 2 部分行人重识别视频数据集

Table 2 Part of person re-identification video datasets

Dataset	Camera numbers	ID numbers	Sequence length	Body images
PRID2011 ^[7]	2	200	400	Hand
DukeMTMC VideoReID ^[15]	6	702	4832	—
iLIDS-VID ^[8]	2	300	600	Hand
MARS ^[16]	6	1261	20715	DPM+GMMCP
EgoReID ^[17]	3	900	10200	YOLO9000+ FSDSC
LS-VID ^[18]	15	3772	14943	Faster R-CNN

数据集包含了 30000 个行人 ID, 同时数据集涵盖了剧烈光照变化、遮挡、低像素、俯视拍摄以及真实场景下的复杂背景等多种复杂情况; 由起源人工智能研究院提出的 RandPerson^[13] 行人数据集,

利用一种随机自动生成大量三维人物模型的方法, 并在游戏环境中进行渲染而得到, 包含 8000 个行人 ID, 38 段视频以及 1801816 张行人切片图像. 为了研究黑人问题而建立的首个黑人数据集 Black-reID^[14], 该数据集图像来自于 Market-1501、DukeMTMC-reID、Partial 等几个数据集, 建立者从中选取对应的图像建立黑人数据集, 该数据集训练集包含 688 个行人 ID, 测试集包含 1644 个行人 ID.

1.2 行人重识别评价指标

行人重识别的评测指标主要有以下几种.

(1) Rank-*n* Accuracy^[19]: 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 Rank-1 Accuracy 与 Rank-5 Accuracy 2 种评价指标, 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^[19](Cumulative match characteristic): 累积匹配曲线, 通过绘制 Rank-*n* accuracy 随着参数 *n* 变化的曲线, 直观地了解 top-*n* 击中概率的变化, 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、识别任务中常用的一项评价指标, 表示随着分类阈值的变化, 行人重识别模型对正样本与负样本的区分情况.

(4) mAP^[11](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.

行人重识别模型最为常用的评价指标为 Rank-*n* Accuracy 与 mAP, 表 3 整理了目前各个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, 其中 OSNet 与 DG-Net 的精度数据测试于按照新版协议划分的 CUHK03 数据集.

表 3 各数据集的性能最优模型以及精度数据

Dataset	SOTA	Rank-1 accuracy	mAP
Market-1501 ^[4]	St-ReID(RE, RK) ^[20]	97.20	86.70
	Viewpoint-Aware Loss ^[21]	96.79	95.43
	DG-Net ^[22]	94.80	84.00
DukeMTMC-reID ^[5]	St-ReID(RE, RK, Cam) ^[20]	94.50	92.70
	ABD-Net(ResNet-50) ^[23]	89.00	78.95
	Viewpoint-Aware Loss ^[21]	93.90	91.80
CUHK03 ^[3,6]	FD-GAN ^[19]	92.60	91.30
	OSNet ^[24]	67.80	—
	DG-Net ^[22]	61.10	—
MSMT17 ^[9]	ABD-Net(ResNet-50) ^[23]	82.30	60.80
	OSNet ^[24]	78.70	52.90
	DG-Net ^[22]	77.20	52.30

2 行人重识别方法研究框架

行人重识别框架与大多数基于深度学习的图像处理问题相似, 首先使用卷积神经网络(CNN), 利用表征学习对数据进行特征提取得到特征向量, 之后再利用度量学习的方法进行类别判断. 行人重识别的研究框架如图 3 所示, 本章对表征学习与度量学习的定义以及基本方法进行简要介绍.

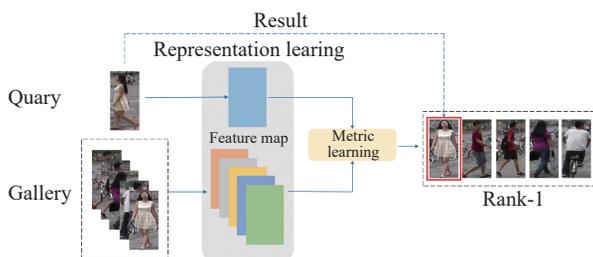


图 3 行人重识别研究方法框架

Fig.3 Research method framework of person re-identification methods

2.1 表征学习

表征学习, 即特征表示学习, 它的目的是提取图像特征, 即将图像数据抽象为高维特征向量, 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, 已有一些较为有效的特征提取方法^[25-27], 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. 将行人的 ID 作为标签对数据进行标注, 使用图像分类方法进行行人的分类. 在之后的研究中, 一些学者受到图像检索方法的启发, 提出了 CIR(Cross-image representation)方法, 使用图像对作为输入, 最终输出两者的相似度. 该方法基于孪生卷积神经网络(Siamese CNN)^[28], 将 2 张图像分别输入到 2 个独立的子网络中, 得到图像的特征图, 再通过度量学习的方法判断图像对是否属于同一个行人. Yi 等^[29]首次将孪生卷积神经网络应用到行人重识别领域, 考虑到行人重识别图像数据的背景、光照等特征的不同, 舍弃了原网络共享权值的做法, 使 2 个子网络相互独立. Ahmed 等^[30]基于孪生卷积神经网络, 针对行人重识别提出了一种深度网络, 将图像对作为网络输入, 计算图像特征图差异, 最终判断图像对是否属于同一类别. 在此基础上, Wu 等^[31]提出了 PersonNet, 在原有的结构上, 使用多个 3×3 卷积层加深网络, 使得效果得到显著提升. CIR 方法强调输入的两张图像的相互关系, 在网络结构中可以很好地体现出来. Wang 等^[32]提出一种名为 Wconv 的卷积结构, 让每张输入图像都经过两个独立的卷积层, 生成两张特征图并进行融合, 从而得到两张图像各自的特征图. Wang 等^[33]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. 熊炜等^[34]提出了一种深层特征融合行人重识别模型, 利用卷积层与池化层多次提取网络深层特征, 使用融合后的深层特征作为行人的特征属性, 提升了网络提取全局特征的能力. 同样的思想, 该作者在文献 [35] 中提出, 将特征提取网络最后三个卷积层提取的特征图进行拼接, 再使用卷积神经网络提取深层特征, 提升了网络的深层特征提取能力. Wang 等^[20]提出的 St-ReID 方法, 利用跨模态融合方法, 将图像的视觉特征以及时空

特征进行融合,该方法在多个数据集上取得了较好的性能表现。

评价与总结:表征学习是提取图像特征的基本方法,在行人重识别上的应用也取得了一定效果。但是,由于行人重识别数据的特殊性,存在遮挡、部位不对齐以及姿态差异等问题,直接提取图像的全局特征并不能达到最好的识别精度。针对这些问题,学者们相继提出了提取局部特征以及引入注意力机制等方法,在后文中予以归纳总结。

2.2 度量学习

对于行人重识别,首先需要利用特征提取器得到行人图像的特征向量,然后利用度量学习方法判断目标图像与待匹配图像的相似度。度量学习是机器学习的常用方法,通过定量方法,确定特征向量的相似程度。在网络的训练过程中,度量学习方法主要体现在损失函数上,用以更新深度网络的参数,从而提取更具判别力的图像特征。行人重识别使用的损失函数主要有以下几种。

对照损失:在最初的行人重识别研究中,通过提取图像对的特征并进行距离的计算^[3,29],根据特征向量 I_a 与 I_b 的距离 d 判断图像对是否属于同一个行人。通常给定网络输入为图像对,变量 y 值为 +1 或 -1,作为图像对的标签,引入阈值 margin ,对图像对距离加以约束。

$$L_{\text{con}} = yd_{I_a, I_b}^2 + (1 - y)\max(\text{margin} - d_{I_a, I_b}, 0)^2 \quad (1)$$

Yi 等^[29] 将对照损失应用在行人重识别中,网络输入为图像对,分别提取图像的特征向量,计算特征向量的余弦(Cosine)距离,输出一个介于 +1、-1 之间的相似度。这种方法的优点在于对于一张目标图像,相同类别图像的相似度往往会高于不同类别图像的相似度。在之后的研究中,学者们提出对每一张待匹配的图像,根据相似度进行排名,将排名最高的图像作为行人重识别的结果,这也是现在广泛使用的确定识别结果的方法。但是一些学者发现,分类方法和排名方法都有各自的优缺点,分类方法难以确定一个固定的区分阈值,而排名方法得出的 Top1 图像相似度最高,却很容易因为丢失细节信息导致匹配错误。针对这一问题,Chen 等^[36] 提出了一种多任务深度网络 MTDNet,文章指出应用排名方法往往更为关注全局特征,而分类方法则更关注图像对之间的联系,比较注重局部特征。MTDNet 的输入为 3 张图像,首先通过卷积神经网络提取低维特征后,执行排名任务,之后通过三元组损失函数,判断正负样本的距离,从

而完成分类任务。该方法整合了排名方法和分类方法的优点,并使两者的缺点得到互补。同样的思想,Chen 等^[37] 提出了整合分类与排名的训练方法,使用 GoogleNet^[25] 进行特征提取,便于迁移训练。

三元组损失:三元组损失是图像检索领域广泛应用的损失函数^[38-40]。它来源于人脸识别的方法 FaceNet^[41],应用三元组损失函数的网络架构输入为 3 张图像,分别是目标图像、正样本和负样本。三元组损失函数通过调整系数,拉近正样本特征向量 I_p 与目标图像特征向量 I_a 的距离,扩大负样本特征向量 I_n 与目标图像特征向量 I_a 的距离。最终相同 ID 的图像在特征空间内聚类,完成行人重识别任务,同时三元组损失函数可以将行人重识别作为一个端对端的问题进行研究。

$$L_{\text{tri}} = \max(d_{I_a, I_p} - d_{I_a, I_n} + \text{margin}, 0) \quad (2)$$

三元组损失函数的缺点在于构造简单,如果在网络中直接优化三元组损失函数,会导致网络的鉴别能力降低,一些学者通过加入更丰富的信息,设计出性能更强的三元组损失函数,Hermans 等^[42] 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。Cheng 等^[43] 提出一种改进的三元组损失函数,原始的三元组损失函数旨在拉近正样本与目标图像的距离,扩大负样本与目标图像的距离^[41],但是损失函数并没有对样本距离进行约束,导致正样本与目标图像的绝对距离很远,作者通过在原始三元组损失函数中加入了一个新的阈值,用于限制类内图像的最大距离,保证网络不仅能够在特征空间把正负样本推开,也能保证正样本对的距离处于一定的阈值之下。

四元组损失:为了解决三元组损失函数类内距离较大的问题,一些学者在三元组损失函数的基础上,提出了四元组损失函数。Chen 在文章^[44] 中提出了一种应用四元组损失函数的行人重识别网络架构。网络输入为 4 张图像,相对于三元组损失,又增加了一张不同 ID 的负样本 n_2 。四元组损失不仅可以学习正负样本的相对距离,同时限制最小的类间距离高于最大的类内距离,同时引入阈值 margin_1 与 margin_2 对损失函数的值域加以约束,使网络的性能得到提升。

$$L_{\text{qua}} = \max(d_{I_a, I_p} - d_{I_a, I_n} + \text{margin}_1, 0) + \max(d_{I_a, I_p} - d_{I_n, I_{n_2}} + \text{margin}_2, 0) \quad (3)$$

现有的行人重识别方法在训练过程中, 广泛使用交叉熵损失函数与三元组损失函数联合对网络的训练过程进行约束, 在训练过程中, 使用网络层最后输出的特征向量进行三元组损失函数的计算, 之后连接全连接层进行交叉熵损失函数的计算, 在这一过程中, 由于2种损失函数的收敛方向不同, 在训练过程中会产生一定冲突. Luo等^[45]提出BNneck分类器, 在最后一个网络层与全连接层之间插入一个批次正则化层, 从而实现两种损失函数的解耦, 统一网络训练的收敛方向, 使网络模型的性能得到有效提升.

评价与总结: 利用度量学习对卷积神经网络提取的特征向量进行处理, 从而得到行人重识别的结果. 对照损失注重图像对之间的联系, 但是在类别较多的问题中, 又会损失类间的相互关系. 三元组损失函数则通过拉近类内图像距离、扩大类间距离的方法, 使行人重识别的性能得到有效提升. 它的问题在于类内距离与类间距离没有一个明确的阈值, 导致类内距离有时会大于类间距离. 针对这个问题, 学者们提出改进的三元组损失函数以及四元组损失函数, 对图像对距离加以约束, 从而取得了识别精度的提升.

3 基于深度学习的行人重识别方法

3.1 基于局部特征的行人重识别方法

在行人重识别的研究中, 针对行人遮挡与部位不对齐等问题, 学者们通过提取局部特征来代替全局特征, 提出了固定分块、基于注意力机制以及其他融合局部特征的方法, 对这些固有问题进行研究与探讨. 下面对基于局部特征的行人重识别方法的研究进展进行总结与讨论.

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^[46]使用PCB(Part-based convolutional baseline)方法, 如图4所示, 在提取特征的过程中, 将特征图分为多个水平特征块后进行池化操作, 再通过小尺寸卷积核进行特征降维生成多个特征向量输入到分类器中, 进行ID的区分. 同时为了解决在划分过程中伴随的部位不一致问题, 提出了改良的局部池化方法, 进一步地提升了PCB的

性能. Li等^[47]提出MSCAN, 利用空洞卷积完成多个尺度下的特征提取, 充分利用了图像的上下文信息. 引入STN(Spatial transformer networks)用于图像的分割, 利用MSCAN网络提取局部特征, 并将提取的全局特征与局部特征向量进行融合. Kalayeh等^[48]提出了一种SPReID架构, 为了避免背景信息的干扰, 使用语义分割方法替换常用的检测方法来确定目标区域, 从而进行局部特征的提取. Xu等^[49]利用多任务金字塔重叠匹配的方法, 将全局信息与多尺度的局部信息进行融合, 并结合各分块的边缘信息, 提升了网络的识别性能. Zhu等^[50]提出一种ISP(Identity-guided human semantic parsing)模型, 通过人体语义解析的方法, 仅通过人的身份标签, 进行人体部位的定位与像素级别的分类.

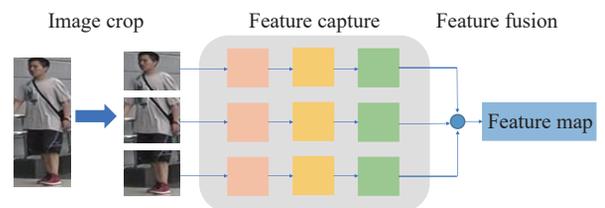


图4 采取固定分块方式的局部特征提取方法^[46]

Fig.4 Local feature extraction method based on fixed blocks^[46]

一些学者引入注意力机制进行局部特征的提取. Zhao等^[51]提出一种方法, 没有使用固定的区域分割方法, 而是在注意力机制的启发下, 在网络中根据不同权值分为了 K 个分支, 用于提取不同区域的特征, 解决了关键点未对齐的问题. 董亚超等^[52]利用注意力机制提取注意力特征图, 并与卷积神经网络提取的全局特征进行融合, 之后将特征图进行切块, 相邻块之间进行协作式融合, 保持了特征块之间的联系. Wang等^[53]提取一种有效地解决遮挡问题的行人重识别方法, 将图像的局部特征作为图的节点, 通过一种自适应方向的图卷积(ADGC)来聚合节点间的信息.

此外, Yang等^[54]使用LSTM(Long short-term memory)^[55]对局部特征进行融合, 有效地改善了遮挡问题. Varior等^[56]在孪生卷积神经网络的中间层加入了匹配门(Matching gate)机制, 通过对比图像对在卷积中间层的特征, 有选择地增强局部细节, 提升了网络的判别能力. 杨婉香等^[57]提出了一种多尺度生成对抗网络, 使用低分辨率图像作为输入重构高分辨率的行人图像, 有效融合不同尺度的图像细节信息, 提升了行人图像去遮挡的性能, 表4给出了不同方法在多个数据集上的

精度表现, 其中 PCB 方法的识别精度测试于按照新版协议划分的 CUHK03 数据集.

表 4 基于局部特征的行人重识别方法的性能表现

Table 4 Performance of person re-identification method based on local feature

Methods	Year	Market-1501		DukeMTMC-reID		CUHK03	
		Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
PCB+RPP ^[46]	2018	93.8	81.6	83.3	69.2	63.7	57.5
SPReID+re-ranking ^[48]	2018	94.6	90.9	88.9	84.9	—	—
RNLSTM _A ^[54]	2019	90.3	76.4	77.0	62.1	86.1	83.6
mGD+RNLSTM _A ^[56]	2020	91.3	77.9	80.8	63.9	88.0	84.2
SMC-ReID ^[49]	2020	95.3	93.0	—	—	—	—
HOReID ^[53]	2020	94.2	84.9	86.9	75.6	—	—
ISP ^[50]	2020	95.3	88.6	89.6	80.0	—	—

评价与总结: 为了解决行人图像的遮挡问题, 学者们提出了基于局部特征的方法. 提取局部特征可以使网络更加关注图像的细节信息. 但由于固定的分块方式, 并不能解决部位不对齐问题, 因此学者们引入了注意力机制增强重要部位的显著性, 能够实现部位对齐. 对于行人重识别中存在的遮挡与部位不对齐问题, 基于注意力机制是目前主流的方法, 也是未来的研究热点.

3.2 基于生成对抗网络的行人重识别方法

在行人重识别任务的研究中, 对于训练数据的处理具有较高的要求, 首先需要对从监控系统中采集的图像进行切割, 得到行人的切片图像. 之后还要对图像进行标注, 标明行人 ID 以及图像采集时间、摄像头编号等信息. 所以, 行人重识别的一些主流数据集, 训练数据都相对较少, 在数据集 Market-1501 上, 平均每一个行人类别只有 17.2 张图像, CUHK03 数据集平均为 9.6 张, DukeMTMC-reID 数据集平均为 23.5 张. 由于数据集之间背景差异较大, 导致模型的泛化能力较差. 针对这些问题, 学者们利用生成对抗网络(Generative adversarial networks, GAN)^[58], 来进行数据集的扩充, 从而加强行人重识别模型的泛化能力.

Zheng 等^[5]首次将生成对抗网络用于行人重识别的问题上来, 他们使用行人重识别数据集对 DCGAN^[59]进行训练, 输入随机向量生成新的行人图像, 对数据集进行了有效扩充, 同时提出一种标签平滑正则化(Label Smoothing Regularization for Outliers, LSRO)的方法, 进行半监督学习. 作为生成对抗网络在行人重识别领域的首次应用, 该方法在精度上取得了小幅提升, 为行人重识别的数

据扩充提供了新思路. 在这个方法的基础上, Zhong 等^[60]使用 CycleGAN^[61]对数据集进行训练, 对不同摄像机拍下的图像数据进行迁移, 生成一个行人在不同摄像机风格下的新图像, 保留了行人的标签信息. 同时文章对前文提出的 LSRO 进行了推广, 提出了一种 LSR (Label smoothing regularization) 方法, 减轻了生成数据噪声对模型训练的影响. Deng 等^[62]将 CycleGAN 与孪生网络结合, 提出了一种图像在数据集之间迁移的方法, 在将图像由原数据集向目标数据集迁移的过程中, 保留了图像的标签信息. Wei 等^[9]提出 PTGAN, 完成了行人图像在不同数据集之间的风格迁移, 并提出了 MSMT17 数据集. Li 等^[63]提出了一种 ReadNet 用于解决数据不平衡问题. Zhai 等^[64]提出增广判别聚类方法实现了数据的迁移. Zhang 等^[65]根据源域图像与目标与图像的语义一致性, 基于弱标注的源域行人图像, 将源域图像的语义信息迁移到目标域上, 增强了跨域的行人重识别特征表达能力. Liu 等^[66]提出 UnityGAN, 学习不同摄像机之间的背景风格差异, 并生成基于这些背景风格差异的平均风格图像, 提升了行人重识别模型对于相机背景风格的泛化能力. 表 5 给出了部分基于生成对抗网络的行人重识别方法的性能表现.

表 5 基于生成对抗网络的行人重识别方法的性能表现

Table 5 Performance of person re-identification method based on GAN

Methods	Year	Market-1501		DukeMTMC		CUHK03	
		Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
DCGAN+LSRO ^[59]	2017	—	—	—	—	84.6	87.4
IDE+CamStyle+RE ^[62]	2018	89.5	71.6	78.3	58.6	—	—
Pose-Transfer ^[9]	2018	87.7	68.9	78.5	56.9	45.1	42.0
PNGAN ^[67]	2018	89.4	72.6	73.6	53.2	79.8	—
PCB+UnityGAN ^[66]	2020	95.8	93.6	89.3	96.2	—	—
st-ReID+UnityGAN ^[66]	2020	98.5	95.8	95.1	93.6	—	—

为了解决数据集之间姿态的差异, 利用生成对抗网络进行姿态的迁移, 成为了行人重识别数据集扩充的重要方向. 一些优秀的神经网络模型可以在一定程度上弥补数据集中行人姿态较少的缺点, 但是扩充姿态多样性的数据仍然是提升模型性能的关键. Liu 等^[68]提出了一种姿态迁移的方法, 选择姿态丰富的数据集, 利用姿态检测算法^[69]提取图像中行人的姿态信息, 并使用 RGB 图像进行表征, 将姿态信息与待迁移数据集中的行人图像作为输入数据, 训练 CGAN^[70]生成新的行人图像, 完成姿态信息的迁移. 同时提出了一种指导模

型,对生成的行人图像进行选择,之后使用LSR^[59]方法对生成数据进行标注,完成了对数据集的扩充。Qian等^[66]提出一种一般性的生成对抗网络模型PNGAN,预设8种行人姿势,以一张原始图像为基准,利用生成对抗网络生成八张预设姿态的图像,将原始数据集扩大了8倍,姿态的多样性与统一性使网络有能力学习一些更具判别力的特征,该方法使行人重识别模型的泛化能力得到了有效地提升,不再需要重新训练就可以进行跨数据集的应用。

评价与总结:为了扩充数据集与实现数据集的域适应,学者们引入了生成对抗网络。但是,生成对抗网络产生了大量未标注的数据,因此需要结合半监督学习方法进行训练。以上方法的应用使基于图像的行人重识别方法的精度达到了峰值,但

图像数据在时序信息上的缺失使得精度达到了瓶颈,这就引发了基于视频的行人重识别的研究热潮。

3.3 基于视频的行人重识别方法

基于图像的行人重识别方法虽然在大部分数据集上取得了较好的效果,但是由于数据集较小,导致基于单帧图像的行人重识别方法没有很好的泛化能力。同时,单帧图像很难获得行人的动作以及姿态变化等信息,导致当行人被遮挡或者光线变化较大时,模型的判别能力大大下降。针对这些问题,学者们提出了基于视频的行人重识别方法,在提取单帧特征的基础上,能有效地考虑到帧间关系。基于视频的行人重识别方法主要由3个部分组成:单帧图像特征提取、视频序列时序特征提取以及度量学习,表6给出了部分基于视频的行人重识别方法的性能表现。

表6 基于视频的行人重识别方法的性能表现

Table 6 Performance of video-based person re-identification method

Methods	Year	Rank-1		MARS		DukeMTMC VideoReID ^[15]	
		PRID2011 ^[7]	iLIDS-VID ^[8]	Rank-1	mAP	Rank-1	mAP
CNN+RNN+Temporal Pooling ^[77]	2016	70.0	58.0	—	—		
Deep RCN ^[78] +KISSME	2016	69.0	46.1				
RFA-Net ^[80]	2016	58.2	49.3				
SCAN+ResNet50 ^[81]	2018	92.0	81.3	86.6	76.7		
ResNet3D-50+Non-Local ^[82]	2018	91.2	81.3	84.3	77.0		
Spatial Attention+Temporal Attention ^[83]	2018	93.2	80.2	82.3	65.8		
AP3D ^[84]	2020	—	86.7	90.1	85.1	96.3	95.6
STCNet ^[86]	2020	—	83.4	88.5	82.3	95.0	93.5
MGH ^[88]	2020	94.8	85.6	90.0	85.8		

对于单帧图像的特征提取与损失函数,学者仍然沿用基于图像的行人重识别领域的方法^[71-75],基于视频的行人重识别方法的重点在于对时序信息的处理,目前最为广泛应用的方法是递归神经网络(Recurrent neural network, RNN)^[76]。McLaughlin等^[77]提出了一种基于CNN与RNN的行人重识别方法,如图5所示,基于孪生网络,网络以视频对作为输入,首先通过CNN提取视频里每一帧图像的特征,按照时间顺序送入RNN中。之后,RNN结合时序信息得到特征向量序列。最后,为了避免对前一时间信息的过度依赖,加入了时序池化层,对所有时间的信息进行聚合,得到视频序列的特征向量。该方法的优点在于可以使用任意长度的视频序列作为输入,同时使得每一时刻的信息都能占有一定的比重。在这个方法的基础上,Wu等^[78]

对网络进行了修改,使用GRU模块^[79]代替LSTM模块,在效果基本不变的情况下,降低了参数量。

上述方法中,都采用了平均池化层用于时序信息的融合,这种方法的优点在于每一时刻的信息都占有相同的比重。但针对行人重识别,一段视频中只有行人出现的段落具有特征提取的价值,同时,某一异常值会对结果造成影响。针对这一问题,Yan等^[80]提出RFA-net,提取单帧图像的特征,表示行人外观信息,使用LSTM提取图像序列特征,表示行人动态信息。融合外观信息与动态信息,获得更好的识别精度。Zhang等^[81]提出了一种SCAN行人重识别架构,利用无参数注意力机制,在视频序列中选取具有辨别力的关键帧,并最终通过计算视频对的相似度得出识别结果。Liao等^[82]将3D卷积应用到行人重识别领域进行视频特征的

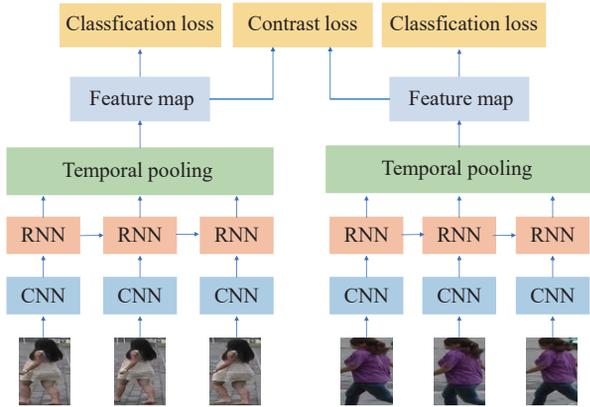


图 5 视频帧序列的时序信息融合方法^[77]

Fig.5 Temporal information fusion of video frames sequence^[77]

提取, 并使用非局部块(Non-local blocks)来解决部位失调问题, 并捕获时空远程依赖关系. 表 6 给出了部分基于视频的行人重识别方法的性能表现.

以上对单帧特征融合的几种方法进行了介绍. 在基于视频的行人重识别中, 主要存在的问题有行人姿态失调与遮挡、视频数据冗余信息过多等. 为了解决视频序列中存在的遮挡问题以及部位不对齐问题, Li 等^[83] 在行人重识别的模型中引入注意力机制, 在连续的图像序列中, 某一帧的信息并不会影响全局的信息, 所以使用限定随机抽样法在视频序列中提取部分数据作为网络的输入, 提取图像特征后, 应用空间注意力模型自动发现突出的图像区域. 这种方法很好地解决了遮挡问题, 同时有效提取了行人的一些非常规特征, 使得网络的辨别力得到显著提升. 针对 3D 卷积在行人部位不对齐时, 存在的特征抓取性能不佳的问题, Gu 等^[84] 提出一种 AP3D 方法, 在卷积层之前, 通过像素级的特征对齐, 解决由于行人部位失调带来的问题. Feng 等^[85] 提出一种门控机制, 判断哪些区域有助于人的重新识别, 让这些区域通过门, 不相关的背景区域被门过滤掉. Hou 等^[86] 提出 STCnet, 利用时空补全网络, 基于未被遮挡的部分精确地恢复遮挡部分的信息, 增强了网络的鲁棒性, 为解决行人重识别中的遮挡问题提供了新的思路. Gao 等^[87] 提出了一种姿态匹配方法用于学习局部特征, 并判别该部分是否被遮挡. Yan 等^[88] 提出一种多粒度超图网络框架, 在空间角度, 通过不同细粒度的图像分割, 获取不同尺度的信息; 时序角度, 则通过连接不同视频帧的相同部位构建超图, 实现时序信息的捕捉.

对于视频数据的过多冗余信息, Zhang 等^[89] 提出多细粒度参考辅助注意力特征聚合方法(Multi-granularity reference-aided attentive feature aggregation,

MG-RAFA)用于空间信息与时序信息的融合. Liu 等^[90] 为了抑制背景信息的干扰, 提出了 CSTNet, 捕获视频帧之间常见的突出前景区域, 并从这些区域探索了时空长程文本的相互依赖关系, 以学习更具判别力的行人特征.

评价与总结: 与图像数据相比, 视频数据具有较强的时序信息, 所以如何将这些时序信息进行融合是基于视频的重识别问题的重要研究方向. 现有的方法主要包括利用 RNN 结合平均池化进行时序信息的融合, 以及利用 3D 卷积的方法进行融合. 学者们在此基础上对细节部分加以创新, 使得识别精度不断提升. 基于视频的行人重识别方法更加接近实际应用, 需要大量的数据标注工作, 同时模型的参数量与计算量较大, 因此结合半监督学习与无监督学习方法以及对网络轻量化的研究会成为今后的热点研究方向.

3.4 基于重排序的行人重识别

除了对于表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.

在图像检索领域, Shen^[91] 使用 K 最近邻算法(*k*-Nearest Neighbor, KNN)对模型输出排序结果进行二次优化. 在行人重识别领域, Li 等^[92] 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. Ye 等^[93] 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. Zhong^[6] 等利用 K 最近邻互编码(*k*-reciprocal encoding)的方法对识别结果进行重排序, 使用该编码方法, 对匹配图像与目标图像在特征空间内求取最近邻, 利用该距离并联合度量学习中的图像特征距离, 对匹配图像进行重排序, 从而提升行人重识别的精度. Wang^[94] 等提出一种在线优化排序结果的方法, 即人体验证在线机器学习(Human verification incremental learning, HVIL), 在预设的周期内, 人工地从排序结果中找出正确匹配, 并给出 Rank-1 结果的相似度(True match/similar/dissimilar), 根据以上结果进行误差计算并更新网

络参数,实现在线人工标注,有效地提升识别的精度. Sarfraz 等^[95]提出了一种基于交叉最近邻距离的无监督自动重排序框架,是当时性能最佳的重排序方法.

评价与总结:针对常规方法在排序结果上存在的一些问题,学者们另辟蹊径,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的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.

4 当前问题与未来展望

4.1 当前存在问题

(1)现有的研究方法大多基于切割的行人图像,而实际的工业场景应用则基于全局图像,这就需要人工或者基于目标检测方法对图像进行切割.

(2)全监督学习方法需要对数据进行完整的标注,而行人重识别的数据标注较为复杂,且需要标注人员 ID、摄像头的编号、视频编号等属性,导致行人重识别数据集相对较小.

(3)行人重识别模型在不同数据集的表现差异很大,泛化能力较差.

(4)行人重识别算法一般需要部署在边缘计算设备上,如智能监控摄像头、嵌入式设备等. 现有的算法参数量与计算量较大,无法满足边缘设备的算力要求.

4.2 未来研究方向

(1)端到端学习:结合目标检测算法与行人重识别方法,进行端到端行人重识别模型的研究,实现直接使用采集的图像或者视频进行训练和实际的应用.

(2)半监督学习的性能优化:如何在保证模型表征学习能力的基础上,提出更加有效的半监督学习方法,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.

(3)数据集的迁移:模型的泛化能力是深度学习领域所面对的一个共同问题,因此跨数据集的行人重识别模型研究仍然是未来的研究热点.

(4)轻量化模型:对行人重识别进行轻量化模型研究,降低模型参数量,提升模型运行速度,为算法在工业领域的部署应用提供可能.

4.3 总结

本文针对行人重识别领域的研究现状,从基于局部特征、基于生成对抗网络、基于视频以及

基于重排序 4 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,针对近一年的成果做出分析与讨论,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,仍然存在诸多未解决的问题. 本文针对当前现有的行人重识别方法进行综述,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帮助.

参 考 文 献

- [1] Ye M, Shen J B, Lin G J, et al. Deep learning for person re-identification: a survey and outlook [J/OL]. *Arxiv Preprint* (2020-01-13) [2020-12-22]. <https://arxiv.org/abs/2001.04193>
- [2] Karanam S, Gou M R, Wu Z Y, et al. A systematic evaluation and benchmark for person re-identification: Features, metrics, and datasets. *IEEE Trans Pattern Anal Mach Intell*, 2019, 41(3): 523
- [3] Li W, Zhao R, Xiao T, et al. DeepReID: Deep filter pairing neural network for person re-identification // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, 2014: 152
- [4] Zheng L, Shen L, Tian L, et al. Scalable person re-identification: a Benchmark // *2015 IEEE International Conference on Computer Vision*. Santiago, 2015: 1116
- [5] Zheng Z D, Zheng L, Yang Y. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro // *2017 IEEE International Conference on Computer Vision (ICCV)*. Venice, 2017: 3774
- [6] Zhong Z, Zheng L, Cao D L, et al. Re-ranking person re-identification with k -reciprocal encoding // *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, 2017: 3652
- [7] Hirzer M, Beleznai C, Roth P M, et al. Person re-identification by descriptive and discriminative classification // *Scandinavian Conference on Image Analysis*. Ystad Sal, 2011: 91
- [8] Wang T Q, Gong S G, Zhu X T, et al. Person re-identification by discriminative selection in video ranking. *IEEE Trans Pattern Anal Mach Intell*, 2016, 38(12): 2501
- [9] Wei L H, Zhang S L, Gao W, et al. Person transfer GAN to bridge domain gap for person re-identification // *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, 2018: 79
- [10] Song G L, Leng B, Liu Y, et al. Region-based quality estimation network for large-scale person re-identification [J/OL]. *Arxiv Preprint* (2017-11-23) [2020-12-22]. <https://arxiv.org/abs/1711.08766>
- [11] Yu S J, Li S H, Chen D P, et al. COCAS: A large-scale clothes changing person dataset for Re-identification // *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, 2020: 3397
- [12] Wang G R, Wang G C, Zhang X J, et al. Weakly supervised person

- Re-ID: Differentiable graphical learning and a new benchmark [J/OL]. *arXiv preprint arXiv* (2019-04-08) [2020-12-22]. <https://arxiv.org/abs/1904.03845>
- [13] Wang Y N, Liao S C, Shao L. Surpassing real-world source training data: Random 3D characters for generalizable person re-identification // *Proceedings of the 28th ACM International Conference on Multimedia*. Seattle, 2020: 3422
- [14] Xu B Q, He L X, Liao X Y, et al. Black Re-ID: A Head-shoulder descriptor for the challenging problem of person re-identification // *Proceedings of the 28th ACM International Conference on Multimedia*. Seattle, 2020: 673
- [15] Wu Y, Lin Y T, Dong X Y, et al. Exploit the unknown gradually: One-shot video-based person re-identification by stepwise learning // *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, 2018: 5177
- [16] Zheng L, Bie Z, Sun Y, et al. Mars: A video benchmark for large-scale person re-identification // *Proceedings of the European Conference on Computer Vision*. Amsterdam, 2016: 868
- [17] Basaran E, Tesfaye Y T, Shah M. EgoReID Dataset: Person re-identification in videos acquired by mobile devices with first-person point-of-view [J/OL]. *arXiv preprint arXiv* (2018-12-22) [2020-12-22]. <https://arxiv.org/abs/1812.09570>
- [18] Li J, Wang J, Tian Q, et al. Global-local temporal representations for video person re-identification // *Proceedings of the IEEE International Conference on Computer Vision*. Seoul, 2019: 3958
- [19] Ge Y X, Li Z W, Zhao H Y, et al. FD-GAN: Pose-guided feature distilling GAN for robust person re-identification. [J/OL]. *arXiv preprint arXiv* (2018-10-06) [2021-5-5]. <https://arxiv.org/abs/1810.02936>
- [20] Wang G C, Lai J H, Huang P G, et al. Spatial-temporal person re-identification // *Proceedings of the AAAI conference on artificial intelligence*. Hawaii, 2019, 33(1): 8933
- [21] Zhu Z H, Jiang X Y, Zheng F, et al. Viewpoint-aware loss with angular regularization for person re-identification [J/OL]. *arXiv preprint arXiv* (2019-12-03) [2020-12-22]. <https://arxiv.org/abs/1912.01300v1>
- [22] Zheng Z D, Yang X D, Yu Z D, et al. Joint discriminative and generative learning for person Re-identification // *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, 2019: 2133
- [23] Chen T L, Ding S J, Xie J Y, et al. ABD-net: Attentive but diverse person Re-identification // *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, 2019: 8350
- [24] Zhou K Y, Yang Y X, Cavallaro A, et al. Omni-scale feature learning for person re-identification // *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, 2019: 3701
- [25] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions // *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, 2015: 1
- [26] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition // *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, 2016: 770
- [27] Huang G, Liu Z, Der Maaten L V, et al. Densely connected convolutional networks // *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, 2017: 2261
- [28] Filković I, Kalafatić Z, Hrkać T. Deep metric learning for person re-identification and de-identification // *2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics*. Opatija, 2016: 1360
- [29] Yi D, Lei Z, Liao S C, et al. Deep metric learning for person Re-identification // *2014 22nd International Conference on Pattern Recognition*. Stockholm, 2014: 34
- [30] Ahmed E, Jones M, Marks T K. An improved deep learning architecture for person re-identification // *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, 2015: 3908
- [31] Wu L, Shen C H, Hengel A V D. PersonNet: Person re-identification with deep convolutional neural networks [J/OL]. *Arxiv Preprint* (2016-01-27) [2020-12-22]. <https://arxiv.org/abs/1601.07255>
- [32] Wang Y C, Chen Z Z, Wu F, et al. Person re-identification with cascaded pairwise convolutions // *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, 2018: 1470
- [33] Wang F Q, Zuo W M, Lin L, et al. Joint learning of single-image and cross-image representations for person re-identification // *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, 2016: 1288
- [34] Xiong W, Xiong Z J, Yang D C, et al. Pedestrian re-identification based on deep feature fusion. *Comput Eng Sci*, 2020, 42(2): 358 (熊炜, 熊子婕, 杨荻椿, 等. 基于深层特征融合的行人重识别方法. *计算机工程与科学*, 2020, 42(2): 358)
- [35] Xiong W, Yang D C, Xiong Z J, et al. Person re-identification algorithm based on global feature stitching. *Appl Res Comput*, 2021, 38(1): 316 (熊炜, 杨荻椿, 熊子婕, 等. 基于全局特征拼接的行人重识别算法研究. *计算机应用研究*, 2021, 38(1): 316)
- [36] Chen W, Chen X, Zhang J, et al. A Multi-task deep network for person re-identification // *Proceedings of the AAAI Conference on Artificial Intelligence*. San Francisco, 2017: 3988
- [37] Chen H R, Wang Y W, Shi Y M, et al. Deep transfer learning for person Re-identification // *2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM)*. Xi'an, 2018: 1
- [38] Ding S Y, Lin L, Wang G R, et al. Deep feature learning with relative distance comparison for person re-identification. *Pattern Recognit*, 2015, 48(10): 2993
- [39] Shi H L, Yang Y, Zhu X Y, et al. Embedding deep metric for person re-identification: A study against large variations // *Proceedings of the European Conference on Computer Vision*. Amsterdam, 2016: 732
- [40] Su C, Zhang S L, Xing J L, et al. Deep attributes driven multi-camera person re-identification // *Proceedings of the European Conference on Computer Vision*. Amsterdam, 2016: 475
- [41] Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering // *2015 IEEE*

- Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, 2015: 815
- [42] Hermans A, Beyer L, Leibe B, et al. In defense of the triplet loss for person re-identification [J/OL]. *Arxiv Preprint* (2017-03-22) [2020-12-22]. <https://arxiv.org/abs/1703.07737>
- [43] Cheng D, Gong Y H, Zhou S P, et al. Person re-identification by multi-channel parts-based CNN with improved triplet loss function // *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, 2016: 1335
- [44] Chen W H, Chen X T, Zhang J G, et al. Beyond triplet loss: A deep quadruplet network for person Re-identification // *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, 2017: 1320
- [45] Luo H, Gu Y Z, Liao X Y, et al. Bag of tricks and a strong baseline for deep person Re-identification // *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Long Beach, 2019: 1487
- [46] Sun Y F, Zheng L, Yang Y, et al. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline) // *Proceedings of the European Conference on Computer Vision*. Munich, 2018: 501
- [47] Li D W, Chen X T, Zhang Z, et al. Learning deep context-aware features over body and latent parts for person Re-identification // *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, 2017: 7398
- [48] Kalayeh M M, Basaran E, Gökmen M, et al. Human semantic parsing for person Re-identification // *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, 2018: 1062
- [49] Xu L Z, Peng L, Zhu F Z. Pedestrian re-identification method based on multi-task pyramid overlapping matching. *Comput Eng*, 2021, 47(1): 239
(徐龙壮, 彭力, 朱凤增. 多任务金字塔重叠匹配的行人重识别方法. *计算机工程*, 2021, 47(1): 239)
- [50] Zhu K, Guo H Y, Liu Z W, et al. Identity-guided human semantic parsing for person re-identification [J/OL]. *ArXiv Preprint* (2020-07-27) [2020-12-22]. <https://arxiv.org/abs/2007.13467>
- [51] Zhao L M, Li X, Zhuang Y T, et al. Deeply-learned part-aligned representations for person re-identification // *2017 IEEE International Conference on Computer Vision (ICCV)*. Venice, 2017: 3239
- [52] Dong Y C, Liu H Z, Xu C. Person re-identification method based on saliency multi-scale feature collaborative fusion [J/OL]. *Computr Engineering*, <https://doi.org/10.19678/j.issn.1000-3428.0057938>
(董亚超, 刘宏哲, 徐成. 基于显著性多尺度特征协作融合的行人重识别方法[J/OL]. *计算机工程*, <https://doi.org/10.19678/j.issn.1000-3428.0057938>)
- [53] Wang G A, Yang S, Liu H Y, et al. High-order information matters: Learning relation and topology for occluded person Re-identification // *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, 2020: 6448
- [54] Yang W X, Yan Y, Chen S. Adaptive deep metric embeddings for person re-identification under occlusions. *Neurocomputing*, 2019, 340: 125
- [55] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*, 1997, 9(8): 1735
- [56] Varior R R, Haloi M, Wang G, et al. Gated Siamese convolutional neural network architecture for human re-identification // *Proceedings of the European Conference on Computer Vision*. Amsterdam, 2016: 791
- [57] Yang W X, Yan Y, Chen S, et al. Multi-scale generative adversarial network for person Re-identification under occlusion. *J Softw*, 2020, 31(7): 1943
(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. *软件学报*, 2020, 31(7): 1943)
- [58] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Nets // *Conference and Workshop on Neural Information Processing Systems*. Montreal, 2014: 2672
- [59] Radford A, Metz L, Chintala S, et al. Unsupervised representation learning with deep convolutional generative adversarial networks [J/OL]. *Arxiv Preprint* (2015-11-19) [2020-12-22]. <https://arxiv.org/abs/1511.06434>
- [60] Zhong Z, Zheng L, Zheng Z D, et al. Camera style adaptation for person Re-identification // *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, 2018: 5157
- [61] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks // *2017 IEEE International Conference on Computer Vision (ICCV)*. Venice, 2017: 2242
- [62] 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 // *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, 2018: 994
- [63] Li Y T, Xue R N, Zhu M M, et al. ReadNet: Towards accurate ReID with limited and noisy samples [J/OL]. *ArXiv Preprint* (2020-05-12) [2020-12-22]. <https://arxiv.org/abs/2005.05740>
- [64] Zhai Y P, Lu S J, Ye Q X, et al. AD-cluster: Augmented discriminative clustering for domain adaptive person re-identification // *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, 2020: 9018
- [65] Zhang X W, Lü M Q, Li H. Cross-domain person re-identification based on partial semantic feature invariance. *J Beijing Univ Aeronaut Astronaut*, 2020, 46(9): 1682
(张晓伟, 吕明强, 李慧. 基于局部语义特征不变性的跨域行人重识别. *北京航空航天大学学报*, 2020, 46(9): 1682)
- [66] Liu C, Chang X J, Shen Y D. Unity style transfer for person Re-identification // *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, 2020: 6886
- [67] Qian X L, Fu Y W, Xiang T, et al. Pose-normalized image generation for person re-identification // *Proceedings of the European Conference on Computer Vision*. Munich, 2018: 661
- [68] Liu J X, Ni B B, Yan Y C, et al. Pose transferrable person re-identification // *2018 IEEE/CVF Conference on Computer Vision*

- and Pattern Recognition. Salt Lake City, 2018: 4099
- [69] Cao Z, Simon T, Wei S H, et al. Realtime multi-person 2D pose estimation using part affinity fields // 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, 2017: 1302
- [70] Mirza M, Osindero S. Conditional generative adversarial nets [J/OL]. *ArXiv Preprint* (2014-11-06) [2020-12-22]. <https://arxiv.org/abs/1411.1784>
- [71] Farenzena M, Bazzani L, Perina A, et al. Person re-identification by symmetry-driven accumulation of local features // 2010 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Francisco, 2010: 2360
- [72] Kviatkovsky I, Adam A, Rivlin E. Color invariants for person re-identification. *IEEE Trans Pattern Anal Mach Intell*, 2013, 35(7): 1622
- [73] Ma B P, Su Y, Jurie F. Local descriptors encoded by fisher vectors for person re-identification // *International Conference on Computer Vision*. Barcelona, 2012: 413
- [74] Chen D P, Yuan Z J, Chen B D, et al. Similarity learning with spatial constraints for person re-identification // 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, 2016: 1268
- [75] Köstinger M, Hirzer M, Wohlhart P, et al. Large scale metric learning from equivalence constraints // 2012 *IEEE Conference on Computer Vision and Pattern Recognition*. Providence, 2012: 2288
- [76] Zaremba W, Sutskever I, Vinyals O, et al. Recurrent neural network regularization [J/OL]. *Arxiv Preprint* (2014-09-08) [2020-12-22]. <https://arxiv.org/abs/1409.2329>
- [77] McLaughlin N, Rincon J M, Miller P. Recurrent convolutional network for video-based person Re-identification // 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, 2016: 1325
- [78] Wu L, Shen C H, Den Hengel A, et al. Deep recurrent convolutional networks for video-based person re-identification: An end-to-end approach [J/OL]. *ArXiv Preprint* (2014-06-11) [2020-12-22]. <https://arxiv.org/abs/1606.01609v2>
- [79] Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [J/OL]. *ArXiv Preprint* (2014-06-03) [2020-12-22]. <https://arxiv.org/abs/1406.1078>
- [80] Yan Y C, Ni B B, Song Z C, et al. Person re-identification via recurrent feature aggregation // *Proceedings of the European Conference on Computer Vision*. Amsterdam, 2016: 701
- [81] Zhang R M, Li J Y, Sun H B, et al. SCAN: Self-and-collaborative attention network for video person re-identification. *IEEE Trans Image Process*, 2019, 28(10): 4870
- [82] Liao X Y, He L X, Yang Z W, et al. Video-based person re-identification via 3D convolutional networks and non-local attention // *Asian Conference on Computer Vision*. Perth, 2018: 620
- [83] Li S, Bak S, Carr P, et al. Diversity regularized spatiotemporal attention for video-based person re-identification // 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, 2018: 369
- [84] Gu X Q, Chang H, Ma B P, et al. Appearance-preserving 3D convolution for video-based person re-identification // *Proceedings of the European Conference on Computer Vision*. Glasgow, 2020: 228
- [85] Feng Y, Wang Y, Luo J B, et al. Video-based person re-identification using gated convolutional recurrent neural networks [J/OL]. *ArXiv Preprint* (2020-03-21) [2020-12-22]. <https://arxiv.org/abs/2003.09717>
- [86] Hou R B, Ma B P, Chang H, et al. VRSTC: Occlusion-free video person Re-identification // 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, 2019: 7176
- [87] Gao S, Wang J Y, Lu H C, et al. Pose-guided visible part matching for occluded person ReID // 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, 2020: 11741
- [88] Yan Y C, Qin J, Chen J X, et al. Learning multi-granular hypergraphs for video-based person re-identification // 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, 2020: 2896
- [89] Zhang Z Z, Lan C L, Zeng W J, et al. Multi-granularity reference-aided attentive feature aggregation for video-based person Re-identification // 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, 2020: 10404
- [90] Liu J, Zha Z, Zhu X, et al. Co-saliency spatio-temporal interaction network for person re-identification in videos [J/OL]. *ArXiv Preprint* (2020-05-11) [2020-12-22]. <https://arxiv.org/abs/2004.04979>
- [91] Shen X, Lin Z, Brandt J, et al. Object retrieval and localization with spatially-constrained similarity measure and k-nn re-ranking // 2012 *IEEE Conference on Computer Vision and Pattern Recognition*. Providence, 2012: 3013
- [92] Li W, Wu Y, Mukunoki M, et al. Common-near-neighbor analysis for person re-identification // 2012 *19th IEEE International Conference on Image Processing*. Orlando, , 2012: 1621
- [93] Ye M, Chen J, Leng Q M, et al. Coupled-view based ranking optimization for person re-identification // *International Conference on Multimedia Modeling*. Kaiserslautern, 2015: 105
- [94] Wang H X, Gong S G, Zhu X T, et al. Human-in-the-loop person re-identification // *Proceedings of the European Conference on Computer Vision*. Amsterdam, 2016: 405
- [95] Sarfraz M S, Schumann A, Eberle A, et al. A pose-sensitive embedding for person re-identification with expanded cross neighborhood Re-ranking // 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, 2018: 420