

DOI:10.19651/j.cnki.emt.2210307

视听资料中的目标人物重识别方法*

李萌 徐麟 宋伟宁
(东华理工大学 南昌 330013)

摘要: 视频监控系统实时采集视频数据,可以作为有效的第三方目击者,为案件侦破提供有利线索与信息。但由于数据量巨大,检索工作量很高,给案件取证带来了不便。针对这一问题,本文以视听资料中的目标人物检索为目标,对现有的跨摄像头人体重识别方法进行改进,实现目标人物的快速重识别。首先对目标人物图像进行切分,获得分块特征图;其次引入局部融合模块,充分保留局部特征信息及局部关联信息;然后引入全局融合模块,在去除背景噪声的同时,全面表征图像全局特征;最后,综合交叉熵损失与三元组损失函数,加速模型收敛并有效防止过拟合。仿真实验结果表明,与现有人体重识别方法相比,本文方法的准确度更高;应用软件结果表明,本文方法可以实现跨摄像头目标人物快速定位,满足视听资料目标人物的快速检索需求。

关键词: 重识别;卷积神经网络;特征融合

中图分类号: TP37 **文献标识码:** B **国家标准学科分类代码:** 520.2

Re-identification method of target person in video materials

Li Meng Xu Lin Song Weining
(East China University of Technology, Nanchang 330013, China)

Abstract: The video surveillance system collects video data in real time, which can serve as an effective third-party witness and provide favorable clues and information for case detection. However, due to the huge amount of data and the high retrieval workload, it brings inconvenience to the case collection. Aiming at this problem, this paper aims to retrieve the target person in the video evidence, and improves the existing cross-camera person weight recognition method to realize the rapid re-identification of the target person. Firstly, the target person image is segmented to obtain the block feature map. Secondly, a local fusion module is introduced to fully retain local feature information and local correlation information. Then a global fusion module is introduced to fully characterize the global image features while removing background noise. Finally, the cross-entropy loss and triplet loss function are integrated to accelerate the model convergence and effectively prevent over-fitting. The simulation experiment results show that compared with the existing methods of human weight recognition, the method in this paper has higher accuracy. The application software results show that the method in this paper can quickly locate the target person across cameras and meet the fast retrieval requirements of the target person in the video evidence.

Keywords: re-identify; convolutional neural network; feature fusion

0 引言

当前,随着我国网络技术的广泛应用,视频监控系统已经遍布城市的各个关键角落,为保卫我国人民的安全起到了重要作用。作为有效的“第三方目击者”,视频监控系统提供了7×24小时的视频监控资料,当案件发生时,可以成为案件侦破取证的重要线索与信息来源。

当前,对于视频系统的取证检索方式一般为人工检索

方式,即,公安人员从某一摄像头中发现目标人物后,则需从与该摄像头相关联的摄像头视频中快速播放关联影像,以多人同时观看的方式,完成对视听资料中的目标人物检索。该方式虽然有效,但由于所涉及的视频量巨大,给公安人员办案带来了极大的工作量增加^[1]。因此,需要一种能够对跨摄像头视听资料进行目标人物重识别的快速搜索方法,以降低公安人员工作量,提高办案效率。

当前,对于跨摄像头视频流中的人体重识别技术,大

收稿日期:2022-06-13

* 基金项目:国家重点研究计划(2018YFB1702702)项目资助

多采用基于深度神经网络进行图像信息提取,进而提取出目标人物的全局信息或局部信息,以完成对目标人物的特征表达。Xiong 等^[2]通过计算各个区域图像的直方图,实现人体纹理特征及颜色信息的特征提取;Dong 等^[3]融合距离视图与双线性视图,实现局部最大出现增强,并综合双线性相似度与马氏距离,实现广义相似度特征应用;张博^[4]对图像仿射变换,实现多帧图像特征目标跟踪;Sathish 等^[5]通过人体外观特征构建了标识符,以此进行人体重识别中的差异信息计算;范彩霞等^[6]应用空间直方图,从 3 个局部角度进行颜色及形状特征提取,有效提升了重识别的鲁棒性。在此基础上,为进一步提升识别准确率,Xie 等^[7]综合人体的全局与局部特征,构建了融合特征学习模型,实现多特征联合识别;贾旭等^[8]基于矩阵特征匹配方法,在车脸重识别中得到了很好应用;张勃兴等^[9]采用全局特征与每个局部特征分别融合,实现了多尺度特征提取。但是,上述特征融合方法,未能充分考虑到人体不同身体部位之间的关联关系,降低了特征的关联性,识别准确率仍有待进一步提高。

针对这一问题,本文研究了一种融合全局特征与局部特征的特征提取方法,通过对人体图像分块,应用残差网络

与平均池化层进行局部特征提取,实现人体各部位之间的关联关系对应;进一步,研究了全局融合方法,并通过使用交叉熵和三元组联合损失函数,优化人体特征度量,可以更好地实现视听资料中的目标人物重识别需求。

1 系统模型

1.1 系统模型

本文所设计的人体重识别融合特征提取方法如图 1 所示,包括:主体网络、局部特征融合模块以及全局特征融合模块等 3 部分。其中,主体网络应用 ResNet50 网络实现,网络输出为 6 段待检图像;通过融合平均池化层,输出人体的分块特征图,实现图像基本信息的保留;然后,应用局部融合模块(partial fusion module,PFM),实现局部特征关联,从而有效避免图像局部信息互相独立,并输出含有全部局部信息的局部特征;其次,应用全局融合模块(global fusion module,GFM),将局部特征与全局特征进行融合,有效滤除背景噪声影响以提升信噪比,实现了全局信息与局部信息的有机融合,保留了局部特征关联性的同时,得到了分辨力更强的全局特征信息;最后,应用三元组与交叉熵联合损失函数实现网络优化。

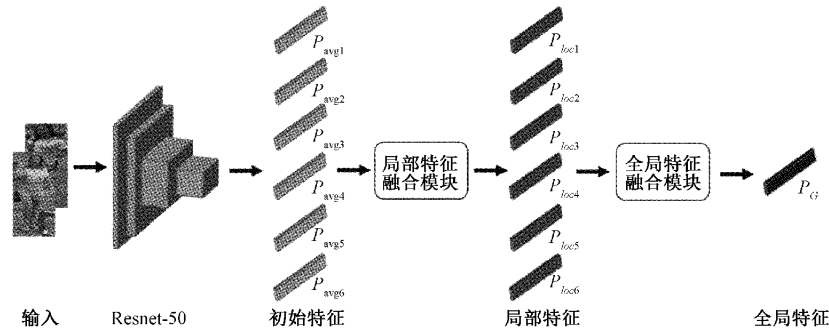


图 1 主体网络框图

1.2 初始特征提取

由于直接应用全局特征提取方法会导致人体身体各部位之间没有对准的问题,为更好提取图像局部信息,本文采用分块卷积方法,对待追踪的目标人体图像进行全局特征提取,即以垂直方向入手,将图像平均分割成 6 个子块,然后进入后续网络中分别提取其局部特征;具体实现时,以 ResNet50 网络作为基础网络,去除最后的全连接层,将引入平均池化层,以尽可能地保留每块图像的信息,最终输出各子块的初始特征图 P_{avg} 。

1.3 局部特征融合模块

经过初始特征提取模块提取的图像子块,可以表征人体的不同部位,从而进入人体局部特征识别。为克服当前研究孤立利用局部信息特征导致部分特征失效的问题,本文应用局部融合模块(partial fusion module,PFM),构建人体不同部位之间的关联特征,进而将各局部特征实现彼此关联融合,可以有效表征人体的几何特征,并对遮挡部

位特征进行推演反应,从而提升总体判决水平。具体实现如下:

将均分后的人体部位特征输出表示为 P_{avg} , 假定第 i 个初始特征为 P_{avg_i} , 并以其进行当前部位与其他部位信息进行融合。计算第 i 个初始特征的融合特征为 \overline{P}_i , 计算方法如下:

$$\overline{P}_i = \frac{1}{5} \sum P_{avg_j} (j \neq i, i = 1, L, 6) \quad (1)$$

将融合特征 \overline{P}_i 与初始特征 P_{avg_i} 相叠加, 求出人体当前部位与其他部位的关联特征 \overline{W}_i , 进而完成第 i 个初始特征融合后的局部特征 P_{loc} , 如图 2 所示。

图 2 中所示为局部特征 P_{loc1} 的提取过程。其中, \boxtimes 表示从图像子块中提取的聚合初始特征, \oplus 表示多子块的聚合特征与初始特征相叠加。

同理, 对于其他图像子块的初始特征也实现如图 2 所示的提取过程, 最终完成 6 个局部特征 P_{loc} 输出。

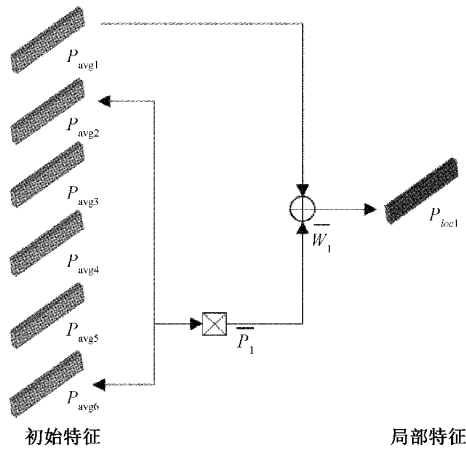


图2 局部特征融合图

1.4 全局特征融合模块

1.3 节中所提取的局部融合特征 P_{loc} 由于采用了平均池化方法,其特征图容易受到背景噪声影响,导致难以准确定位到人体部位,使得图像的局部特征难以准确反映图像信息。最大池化技术能降低高频噪声影响,实现图像特征锐化,从而凸显图像的边缘特征,可以有效弥补平均池化方法的不足。为此,引入了全局特征融合模块(global feature module, GFM),基于人体图像的最大池化特征 P_{max} 与前述所提取的局部特征 P_{loc} 进行特征融合,从而进行图像全局特征 P_G 提取,不仅可以有效消除背景噪声对特征提取影响,也可以有效实现人体各部位之间关联关系表征,使得人体的特征表达更加精准可靠。其实现过程如图3所示。

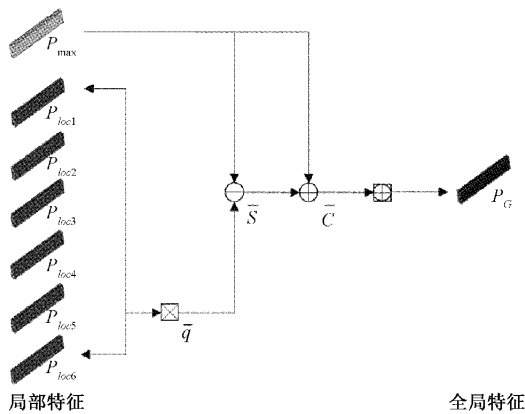


图3 GFM 模块特征融合示意图

其中, \ominus 表示特征相减操作, \oplus 表示特征相加操作, \boxtimes 表示局部特征聚合操作, \boxdot 表示对特征使用大小 1×1 的卷积核进行卷积操作。具体实现如下:

1) 对输入的人体图像进行最大池化操作,提取出图像的最大池化特征 P_{max} ;

2) 将 6 个局部特征 P_{loc} 聚合为局部整合特征 \bar{q} , 计算方法如下:

$$\bar{q} = \frac{1}{6} \sum P_{loci} (i = 1, \dots, 6) \quad (2)$$

3) 将最大池化特征 P_{max} 与局部整合特征 \bar{q} 相减并卷积, 得出 \bar{q} 与 P_{max} 的差异特征 \bar{S} ;

4) 将差异特征 \bar{S} 与 P_{max} 求和, 并卷积, 从而得到图像的全局特征 P_G 为:

$$P_G = \delta_{1 \times 1} (P_{max} + \delta_{1 \times 1} (P_{max} - \bar{q})) \quad (3)$$

其中, $\delta_{1 \times 1}(\cdot)$ 表示使用 1×1 卷积核。

由式(3)中可以看出, 全局特征 P_G 中, 包含了连接了局部特征信息和最大池化特征信息, 在覆盖人体部位的同时, 对背景杂波具有良好的鲁棒性, 可以很好表征人体特征。

1.5 损失函数

在人体识别模型中, 常用的损失函数主要有三元组损失(triplet loss)、对比损失(contrastive loss)以及交叉熵损失函数(cross entropy loss)等三类。其中, 三元组损失函数能够有效进行特征空间分布控制, 可以更好学习和表达差异性较小的特征; 但其由于采用局部优化方法, 在模型训练中会产生收敛困难与过拟合问题; 为此, 本文引入交叉熵损失, 构建三元组损失函数与交叉熵损失联合优化模型^[7], 在增强特征区分度的同时, 加速模型收敛。联合优化模型的损失函数如式(4)所示。

$$\mathcal{L}_{combine} = \lambda_1 \mathcal{L}_{ce} + \lambda_2 \mathcal{L}_{triplet} \quad (4)$$

其中, $\mathcal{L}_{combine}$ 为联合模型损失函数, 由加权后的交叉熵损失函数 \mathcal{L}_{ce} 与三元组损失函数 $\mathcal{L}_{triplet}$ 求和得到, λ_1 和 λ_2 为相应的加权参数。

交叉熵损失函数 \mathcal{L}_{ce} 为:

$$\mathcal{L}_{ce} = - \sum_{n=1}^N \sum y^n \log \hat{y}^n \quad (5)$$

其中, N 表示为批量图像的数量, y^n 为图像的真值标签, \hat{y}^n 表示为对全局特征 P_G 的预测识别标签, 定义为:

$$\hat{y}^n = \operatorname{argmax}_{c \in K} \frac{\exp((W^c)^T P_G)}{\sum_{k=1}^K \exp((W^k)^T P_G)} \quad (6)$$

其中, K 为识别标签的数量, $(W^k)^T$ 为特征 P_G 和标签 k 的分类器。

为减少正样本对距离, 扩大负样本对距离, 本文选用批量硬三元组损失^[7], 选取与当前样本度量距离最近的正样本对以及度量距离最近的负样本对, 构建正负样本对进行模型训练, 如式(7)所示:

$$\mathcal{L}_{triplet} = \sum_{k=1}^p \sum_{m=1}^N [\alpha + \max_{p-1, \dots, m} \|f_{k,m}^A - f_{k,p}^p\|_2 - \min_{t=1, \dots, k} \|f_{k,m}^A - f_{t,p}^N\|_2] \quad (7)$$

其中, p 为小批量中标识的数量, m 为小批量中每个标识标签的图像数量, α 为三元组损失的边际参数, 用于控制特征空间中正负对之间的距离, $f_{k,m}^A$, $f_{k,p}^p$ 和 $f_{t,m}^N$ 分别为

当前图像、正图像和负图像的人体特征表示。

2 仿真结果与分析

2.1 预训练

为提高模型初始准确率,本文采用 ImageNet 对 ResNet50 基础网络进行预训练;训练时,采用 Market1501 数据集,该数据集包含 1 501 位行人共计 32 217 张图片;实验批次大小设置为 64,使用随机梯度下降法(stochastic gradient descent,SDG)作为优化器对训练过程进行优化,动量衰减因子为 0.9,权值衰减为 5×10^{-4} ,模型优化共 80 个 epoch,初始学习率为 1×10^{-2} ,在训练的 40 和 60 个 epoch 将学习率衰减到 1×10^{-3} 和 1×10^{-4} 。采用三元组损失和交叉熵损失联合优化模型,其权重参数 λ_1 和 λ_2 分别为 2 和 1。

2.2 消融实验

为验证本文方法的可行性与有效性,应用 Market1501 数据集,对模型各模块开展消融实验。

1)分块对比消融实验

将人体图像,在竖直方向分割为大小均匀的 6 个子块,以提取图像的局部特征。由于子块分割的数量会影响网络的识别率,为此,以子块分割数量为参数,对人体图像的分割方法进行消融分析,结果如表 1 与图 4 所示。其中,part 为子块分割数量。

表 1 人体图像子块分割结果 %

Part	Rank-1	mAP
2	90.3	79.1
4	94.6	86.9
6	95.7	87.6
8	94.8	87.5
10	91.3	80.4

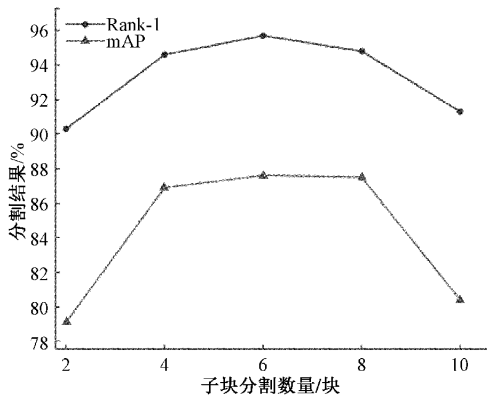


图 4 分块对比实验结果

由图 4 中可以看出,当子块分割数量为 6 时,模型的 Rank-1 和 mAP 值最高,即,模型性能最优;而子块相对较少的 part2 和 part4,由于其含有较多重叠区域,导致模型

易受背景噪声等干扰,性能较差;而子块相对较多的 part8 和 part10,由于所分割图像过于破碎,导致人体各部位之间的联系不够显著,难以充分提取局部信息特征,导致模型的识别率降低。为此,本模型设置子块分割数量为 6。

2)模块组合消融实验

为了验证算法各个模型的有效性,本节对模块组合进行了 4 组消融实验,分别为只使用基准网络、添加初始特征模型、添加初始特征模块和 PFM 模块,以及添加初始特征模块、PFM 模块和 GFM 模块,结果如表 2 所示。

表 2 各模块对比结果

	初始特征	PFM 模块	GFM 模块	Rank-1	mAP
基准网络				84.5	66.2
本文算法	✓			93.6	84.7
本文算法	✓	✓		94.9	85.8
本文算法	✓	✓	✓	95.7	87.6

由表 2 可知:

(1)仅使用 ResNet50 基准网络时,模型的识别率较低,Rank-1 和 mAP 值仅达到了 84.5%和 66.2%。

(2)在基准网络基础上进行图像分割,分别提取图像的局部特征,使得模型性能明显提升,其中,Rank-1 提高 9.1%,达到了 93.6%;mAP 提高了 18.5%,达到了 84.7%。

(3)进一步添加 PFM 模块,模型中引入了局部特征之间的联系,使得模型性能进一步提升,rank-1 提高至 94.9%,mAP 提高至 85.8%。

(4)最后,添加 GFM 模块,消除背景噪声的影响,获取更具识别力的特征表达,模型性能得到进一步提升,Rank-1 与 mAP 分别提升至 95.7%与 87.6%。

3)损失函数的权重参数对比

针对本文所使用交叉熵损失与三元组损失联合优化模型, λ_1 为交叉熵损失的权重参数, λ_2 为三元组损失的权重参数。本节对权重参数开展消融实验,结果如表 3 所示。

表 3 损失函数不同权重参数值实验结果

λ_1	λ_2	Rank-1	mAP
0	1	87.1	79.3
1	0	89.3	82.4
1	1	93.2	85.1
2	1	95.7	87.6
2	2	95.2	87.5

其中,当 $\lambda_1 = 0, \lambda_2 = 1$ 时,模型只使用三元组损失作为最终的损失函数, $\lambda_1 = 1, \lambda_2 = 0$ 时,模型仅选择交叉熵损失作为最终的损失函数。结果显示,在 λ_1 和 λ_2 分别取

到 2 和 1 时,模型的识别准确率最高,既克服了三元组在训练过程中造成收敛困难和过拟合的问题,又能关注特征实际输出与期望输出的接近程度,达到最佳性能。

2.3 对比实验

为了验证本文提出方法的有效性,应用主流的 Market1501 和 DukeMTMC-reID 数据集上开展对比实验,使用 Rank-1 和 mAP 来评估模型的准确度,与现有文献中的人体重识别算法进行对比,包括:SVDNet^[10]、HA-CNN^[11]、PCB^[12]、Manes^[13]、MGN^[14]、VPM^[15]、AligedReID^[16]、CtF^[17]等。

1)Market1501 数据集上对比结果

将本文方法与上述方法在 Market1501 数据集上进行对比,结果如表 4 所示。由表 4 可知,本文所研究方法的 Rank-1 和 mAP 的准确度分别为 95.7%和 87.6%,与当前文献研究的方法相比,其评估标准 Rank-1 和 mAP 值都有不同的提高。

表 4 Market1501 数据集上各方法对比结果

Methods	Publication	Rank-1	mAP
SVDNet	ICCV2017	82.3	62.1
HA-CNN	CVPR2018	91.2	75.7
PCB	ECCV2018	92.3	77.4
Manes	ECCV2018	93.1	82.3
MGN	ACM2018	95.7	86.9
VPM	CVPR2019	93	82.3
AligedReID	PR2019	91.8	79.1
CtF	ECCV2020	94.2	84.9
本文	—	95.7	87.6

其中,PCB 算法融合了水平切块的局部特征,但没有考虑图像的全局信息;本文方法通过融合全局特征和局部特征,与比 PCB 相比,Rank-1 提高了 3.4%,mAP 提高了 10.2%;与上述文献中性能最高的 MGN 算法相比,本文算法的 Rank-1 指标与之相同,但 mAP 指标提高了 0.7%;与最新的融合了人体的外观和结构属性 CtF 算法相比,本文算法的 Rank-1 指标比其提高了 1.5%,mAP 上提高了 2.7%。

2)DukeMTMC-reID 数据集上对比结果

将本文方法及上述方法在 DukeMTMC-reID 数据集上进行对比,结果如表 5 所示。

由表 5 可知,在 DukeMTMC-reID 数据集上,本文方法 Rank-1 和 mAP 的准确度分别为 89.6%和 79.1%,与现有文献的结果相比均有所提高。其中,与最经典的 SVDNet 算法相比,本文算法的 Rank-1 和 mAP 分别提高了 12.9%和 22.3%;与计算全局特征和局部特征距离之和的 AligedReID 算法相比,本文算法由于考虑了局部特征之间的联系,在 Rank-1 上提升了 7.5%,在 mAP 上提升了

表 5 DukeMTMC-reID 数据集上各方法对比结果

Methods	Publication	Rank-1	mAP
SVDNet	ICCV2017	76.7	56.8
HA-CNN	CVPR2018	80.5	63.8
PCB	ECCV2018	81.9	65.3
Manes	ECCV2018	84.9	71.8
MGN	ACM2018	88.7	78.4
VPM	CVPR2019	83.6	72.6
AligedReID	PR2019	82.1	69.7
CtF	ECCV2020	86.9	75.6
本文	—	89.6	79.1

9.4%;与基于注意力机制的 HA-CNN 算法相比,本文算法融合了图像的全局和局部信息,削弱了背景因素的影响,在 Rank-1 上提升了 9.1%,在 mAP 上提升了 15.3%。

3 模型应用

为更直观反应本文方法的有效性,基于本文方法设计了简易的应用系统,如图 5 所示。点击“开始检测”按钮,即可开始人体重识别系统。系统将自动调用所选择的网络模型提取图片的人体特征,并与查询库中人体图像进行匹配识别,并基于欧式距离对图像进行相似度排序,输出前 5 张最相似的人体图像。

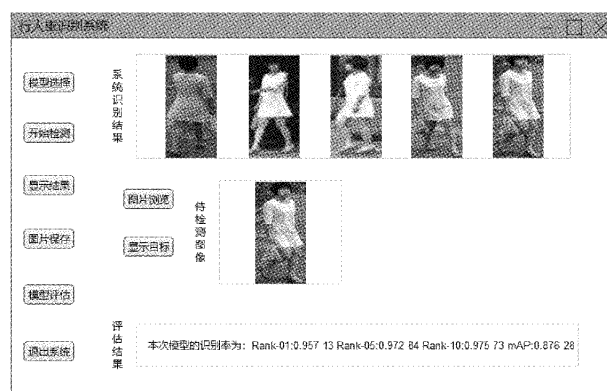


图 5 系统识别结果图

由图 5 可以看出,所设计系统通过对人体特征提取,可以有效分辨出人的背影、侧影等图像,从而有效实现目标人体追踪。

4 结 论

为了获取更好的目标人体的图像特征表达,提升算法的识别性能,本文从全局特征和局部特征的融合角度出发,提出了基于特征融合的人体重识别方法。该方法以 ResNet50 作为基准网络,将人体图像沿竖直方向分块,提取图像的局部信息;进一步通过 PFM 模块融合分块提取

的初始特征,建立局部信息之间的联系,获取图像的局部特征;随后,结合 GFM 模块,在局部特征相互联系的同时,消除背景噪声的影响,提取更具辨别力的人体全局特征;最后采用三元组和交叉熵联合损失优化模型,提高模型的泛化能力和特征的鲁棒性。在主流的 Market1501 和 DukeMTMC-reID 数据集上与现有文献方法开展实验对比,证明了本文所提算法的有效性。

参考文献

- [1] 孔芳楠. 基于视频监控的证据收集问题研究[J]. 北京警察学院学报, 2020, 5(3): 95-99.
- [2] XIONG F, GOU M, CAMPS O, et al. Person re-identification using kernel-based metric learning methods [C]. European Conference on Computer Vision, Springer, Cham, 2014: 1-16.
- [3] DONG H, LU P, ZHONG S, et al. Person re-identification by enhanced local maximal occurrence representation and generalized similarity metric learning[J]. Neurocomputing, 2018, 307: 25-37.
- [4] 张博. 基于局部稀疏表示的多帧图像特征目标跟踪方法[J]. 国外电子测量技术, 2022, 41(6): 25-30.
- [5] SATHISH P K, BALAJI S. A complete person re-identification model using Kernel-PCA-based Gabor-filtered hybrid descriptors[J]. International Journal of Multimedia Information Retrieval, 2018, 7(4): 221-229.
- [6] 范彩霞, 朱虹, 蒯广逢, 等. 多特征融合的人体目标再识别[J]. 中国图象图形学报, 2013, 18(6): 711-717.
- [7] XIE J, XIN W, LIU R, et al. Global co-occurrence feature and local spatial feature learning for skeleton-based action recognition[J]. Entropy, 2020, 22(10): 1135.
- [8] 贾旭, 孙福明. 基于孪生非负矩阵分解的车脸重识别算法[J]. 仪器仪表学报, 2020, 41(6): 132-139.
- [9] 张勃兴, 马敬奇, 张寿明, 等. 利用全局与局部关联特征的行人重识别方法[J]. 电子测量与仪器学报, 2022, 36(6): 1-9.
- [10] SUN Y, ZHENG L, DENG W, et al. Svdnet for pedestrian retrieval [C]. Proceedings of the IEEE International Conference on Computer vision, 2017: 3800-3808.
- [11] LI W, ZHU X, GONG S. Harmonious attention network for person re-identification[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 2285-2294.
- [12] SUN Y, ZHENG L, YANG Y, et al. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 480-496.
- [13] WANG C, ZHANG Q, HUANG C, et al. Mancs: A multi-task attentional network with curriculum sampling for person re-identification[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 365-381.
- [14] WANG G, YUAN Y, CHEN X, et al. Learning discriminative features with multiple granularities for person re-identification[C]. Proceedings of the 26th ACM International Conference on Multimedia, 2018: 274-282.
- [15] SUN Y, XU Q, LI Y, et al. Perceive where to focus: Learning visibility-aware part-level features for partial person re-identification[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 393-402.
- [16] LUO H, JIANG W, ZHANG X, et al. AlignedReID++: Dynamically matching local information for person re-identification [J]. Pattern Recognition, 2019, 94: 53-61.
- [17] WANG G, GONG S, CHENG J, et al. Faster person re-identification [C]. European Conference on Computer Vision, Springer, Cham, 2020: 275-292.

作者简介

李萌, 助理实验师, 主要研究方向为证据法学、大数据与人工智能等。

E-mail: limeng@ecut.edu.cn