

控制与决策

Control and Decision

基于多尺度特征表示的行人再识别

卢健, 王航英, 陈旭, 张凯兵, 刘薇

引用本文:

卢健, 王航英, 陈旭, 等. 基于多尺度特征表示的行人再识别[J]. *控制与决策*, 2021, 36(12): 3015–3022.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2020.0952>

您可能感兴趣的其他文章

Articles you may be interested in

融合稀疏编码与深度学习的草图特征表示

A feature representation of sketch based on fusion of sparse coding and deep learning

控制与决策. 2021, 36(3): 699–704 <https://doi.org/10.13195/j.kzyjc.2019.0941>

Anchor-free的尺度自适应行人检测算法

Anchor-free scale adaptive pedestrian detection algorithm

控制与决策. 2021, 36(2): 295–302 <https://doi.org/10.13195/j.kzyjc.2020.0124>

一种基于多层语义特征的图像理解方法

An image understanding method based on multi-level semantic features

控制与决策. 2021, 36(12): 2881–2890 <https://doi.org/10.13195/j.kzyjc.2020.0927>

改进YOLOv2的端到端自然场景中文字符检测

End-to-end Chinese character detection in natural scene based on improved YOLOv2

控制与决策. 2021, 36(10): 2483–2489 <https://doi.org/10.13195/j.kzyjc.2020.0270>

结合注意力机制的循环神经网络复述识别模型

Recurrent neural networks based paraphrase identification model combined with attention mechanism

控制与决策. 2021, 36(1): 152–158 <https://doi.org/10.13195/j.kzyjc.2019.0638>

基于多尺度特征表示的行人再识别

卢健^{1†}, 王航英¹, 陈旭², 张凯兵¹, 刘薇¹

(1. 西安工程大学电子信息学院, 西安 710600; 2. 南通师范高等专科学校信息技术学院, 江苏南通 226010)

摘要: 应用于复杂场景下的行人再识别方法, 常采用结合全局特征和局部特征的行人表示策略来提升模型的判别能力. 但是, 提取局部特征往往需要针对特定的语义区域设计专门的模型, 增加了算法的复杂性. 为解决上述问题, 提出一种基于多尺度特征表示的行人再识别模型. 该模型通过对不同细粒度局部特征与全局特征的联合表示, 得到多层次具有互补性的判别信息, 端对端地完成行人再识别任务. 为了在获取高分度信息的同时保留更多的细节信息, 采用最大池化加平均池化的方式对特征进行下采样; 此外, 通过引入 TriHard loss 约束全局特征并采用随机擦除方法增强数据来进一步提升模型对复杂场景的适应性. 在 Market-1501 和 DukeMTMC-reID 数据集上进行对比实验, 实验结果表明, rank-1 的准确率分别达到 94.9% 和 87.1%, 从而验证了所提出方法的有效性.

关键词: 行人再识别; 全局特征; 局部特征; TriHard loss; Market-1501; DukeMTMC-reID

中图分类号: TP183

文献标志码: A

DOI: 10.13195/j.kzyjc.2020.0952

开放科学(资源服务)标识码(OSID):



引用格式: 卢健, 王航英, 陈旭, 等. 基于多尺度特征表示的行人再识别[J]. 控制与决策, 2021, 36(12): 3015-3022.

Multi-scale feature representation for person re-identification

LU Jian^{1†}, WANG Hang-ying¹, CHEN Xu², ZHANG Kai-bing¹, LIU Wei¹

(1. School of Electronics and Information, Xi'an Polytechnic University, Xi'an 710600, China; 2. Institute of Information Technology, Nantong Normal College, Nantong 226010, China)

Abstract: The strategy of pedestrian representation merging global features with local features is frequently utilized to improve the discriminability of the model for person re-identification (re-ID) in complex scenes. However, extracting local features generally necessitates specialized models for specific semantic regions, which increases the complexity of the algorithm. Therefore, a re-ID model is proposed based on multi-scale feature learning. The model acquires discrimination information of multi-level complementary with the combination of different fine-grained local features and global features, and then to realize end-to-end person re-identification. In order to obtain high-resolution information while retaining more detailed information, max pooling and average pooling are employed to downsample the features. In addition, this paper introduces TriHard loss to constrain global features and uses random erasure to enhance the data, which further ameliorates the adaptability of the model in complex scenes. Comparative experiments on Market-1501 and DukeMTMC-reID datasets show that the accuracy of rank-1 has reached 94.9% and 87.1%, respectively, which verifies the effectiveness of the proposed method.

Keywords: person re-identification; global feature; local feature; TriHard loss; Market-1501; DukeMTMC-reID

0 引言

行人再识别任务需要从其他监控设备所采集到的行人图像中检索出目标行人图像, 已在智能监控、交通管理等领域得到了广泛的应用. 由于行人图像大多是在实际场景中拍摄的, 行人姿态多变、行人被遮挡、背景复杂等问题使得如何获取更具区分性的行人特征成为该领域研究的难点和热点问题. 近年来, 利用卷积神经网络(convolutional neural networks, CNN)^[1]自主地提取行人深度特征成为解

决该问题的主流方法, 相较于根据经验提取手工设计特征^[2]的传统方法, 该方法操作简单, 易于实现, 并使行人再识别准确率得到较大的提升^[3].

在基于深度学习的行人再识别模型中, 常通过 CNN 直接提取行人显著的外观特征来区分不同的行人^[4]. 但这种利用全局特征表示的方法更关注人体躯干部分, 容易忽略人体不显著的细节信息. 使用局部特征可以很好地补充全局特征中未能观测到的信息, 且由于各部件之间相关性小, 局部特征具有很强的互

收稿日期: 2020-07-12; 修回日期: 2020-10-04.

基金项目: 国家自然科学基金项目(61971339, 61471161); 陕西省自然科学基金重点项目(2018JZ6002).

[†]通讯作者. E-mail: lujian@xpu.edu.cn.

补性,将二者结合使用能够有效地解决姿态多变和行人遮挡等对特征可靠性的不利影响.近年来,采用基于全局-局部特征表示的方法^[5]已成为行人再识别研究的主流方法,是提升模型性能的重要途径之一.

相较于单一的从行人图像中直接获取全局特征的方法,局部特征的获取途径主要有3个:1)利用人体骨架关键点^[6-7]或候选区域^[8]定位出具有特殊语义信息的人体部件,再从各部件中获取局部信息;2)由于目前常用的公开数据集中每张图像仅包含一个行人,可以通过水平划分行人图像或特征图的方式,获取人体不同区域内的信息^[9-10];3)引入注意力机制^[11]进行特征增强,突出所感兴趣的行人信息^[12-13].其中:定位行人身体部件的方法和特征增强的方法均需引入额外的网络,增加了模型的复杂度;而水平分割方法受限于行人的对齐程度.

本文提出一种多尺度特征表示方法,端对端地学习全局信息和多粒度局部信息,获取更具判别性的行人特征,提升模型的识别能力.本文主要贡献如下:

1)本文采用多分支学习策略,在不同分支中对特征图进行不同数量的等分操作以获取多尺度特征,且各分支采用单独训练的方式,在一定程度上可以补偿行人不对齐所造成的不利影响.

2)平均池化是取当前邻域内所有特征点的均值进行下采样,可以保留更多局部细节信息,但对于背景占比较大的区域,平均操作会弱化具有判别性的信息.而最大池化可以获取当前邻域内最大响应特征,故本文采用最大池化与平均池化相结合的方式,既可获取最大响应特征,又保留了局部细节信息,有效提升模型的识别准确率.

3)在训练过程中,本文使用TriHard loss对全局特征进行约束,增强正样本之间的聚集性,使学习到的行人特征更具区分性.并采用随机擦除^[14]方法进行数据增强,提升模型在行人遮挡等实际中的适应性.

1 相关工作

文献[15-16]利用CNN提取行人特征,判断输入行人图像对是否为同一ID的孪生网络(siamese network),将深度学习引入行人再识别任务中,并使用基于部件匹配的方法缓解了行人被遮挡所造成的不利影响.局部特征仅包含人体一小部分信息,不受其他区域信息的干扰,并且局部特征能够更加集中于行人的身份信息,与全局特征形成功能上的互补,能够有效地提升模型的准确率.文献[6]利用14个人体关键点,定位出7部分感兴趣的行人局部区域(头、上半身、下半身以及四肢),从中提取行人局部特征并

融合全局特征来表示行人,可以有效地解决由于行人姿态多变而导致识别效果差的问题,但需要额外增加一个姿态估计模型.文献[17]将行人图像划分成多个水平条纹,从每个条纹中学习局部特征,但这类方法需要引入行人对齐策略以解决行人不对齐所造成的不利影响,并且模型的识别效果严重依赖于对齐策略^[18-19].

针对上述问题,采用独立对分割后的水平特征条进行分类的方法在一定程度上缓解了行人不对齐造成的不利影响^[9],并且随着水平分割条数的增多,所获取到的局部特征更加精细.文献[20]引入金字塔池化^[21-22]思想,提出了水平金字塔匹配(horizontal pyramid matching, HPM)模型,将特征图分别水平等分成2、4、8个不同大小的局部特征,与全局特征分开学习,避免了融合噪声带来的不利影响.文献[23]所提出的多粒度网络(multiple granularity network, MGN)使用3个不同的分支(1个全局分支和2个局部分支)学习全局特征表示,并将特征图水平均匀划分成2、3份,分别输入不同的局部分支中独立地完成分类任务.上述两种模型对特征图的划分方式是固定不变的,在数据集发生改变或其他复杂的应用场景中,并不能保证其划分方式一定适用.

行人再识别可以被看作一种图像检索问题,按照相似度由大到小的顺序对行人图像进行排序.相似度量结果取决于行人特征之间的距离,在训练阶段采用度量损失可以拉近正样本之间的距离,增大负样本之间的距离.常用的度量损失函数有:对比损失(contrastive loss)^[24]、三元组损失(triplet loss)^[25]、以及由三元组损失衍生的度量函数,如四元组损失(quadruplet loss)^[26]等.对比损失常用于孪生网络;三元组损失广泛应用于多分类模型,但传统的三元组损失随机地从训练数据中抽取3张图像组成三元组,这样获取的样本典型性不够,不利于网络更好地学习行人特征.本文采用难样本采样三元组损失(triplet loss with batch hard mining, TriHard loss)^[27],选取训练数据中最难样本组成三元组训练网络,以提升模型的泛化能力.

本文采用多分支学习策略分别获取全局特征和局部特征,局部分支采用递增的水平等分策略来学习多粒度局部特征(局部分支数量可依据不同应用场景进行调整),全局分支中加入TriHard loss,以学习更具区分性的行人特征.为增强模型在行人被遮挡等非理想化场景中的适应能力,本文使用随机擦除的方法进行数据增强.

2 模型提出

2.1 网络结构

IDE (ID-discriminative embedding)^[4]是应用于行人再识别任务中的一个基准模型,该模型采用预训练好的ResNet50网络(不包含分类层)提取行人特征,已取得了不错的成绩.在之后的行人再识别研究方法中,行人特征的提取大多是基于该网络框架的^[1].本文所提出的多尺度特征表示模型同样采用ResNet50作为主网络,并在网络结构功能上做出如下设计:1)ResNet50网络含有5个卷积模块(记作backbone),每一个卷积模块中包含一个步长为2的卷积层来实现降维.为了获取更多细节信息,本文将网络Conv5_1的卷积步长改变为1,使得到的特征图大小为原来的两倍;2)在本文所提出的模型中,设计学

习全局特征的全局分支和学习多粒度局部特征的多个局部分支,每部分特征设有独立的池化层;3)在每一池化层后添加一个BN层和全连接层. BN层用于参数归一化,并缓解过拟合;全连接层用于分类,其输出维度为训练集中行人ID的类别数.依据文献[28],本文在训练阶段,将归一化之前的全局特征用于度量学习;在测试阶段,将归一化之后的所有特征进行拼接(12 288维),用于计算图像相似度.

本文所提出的多尺度特征表示模型结构如图1所示,大小为 $384 \times 128 \times 3$ 的行人图像经backbone后得到 $24 \times 8 \times 2048$ 的特征图(feature map).将特征图作为全局特征直接输入全局分支;将特征图递增地等分成 j 份($j = 2, 3, \dots, t$)作为局部特征再对应地输入分支 j 中.各分支中特征尺寸如表1所示.

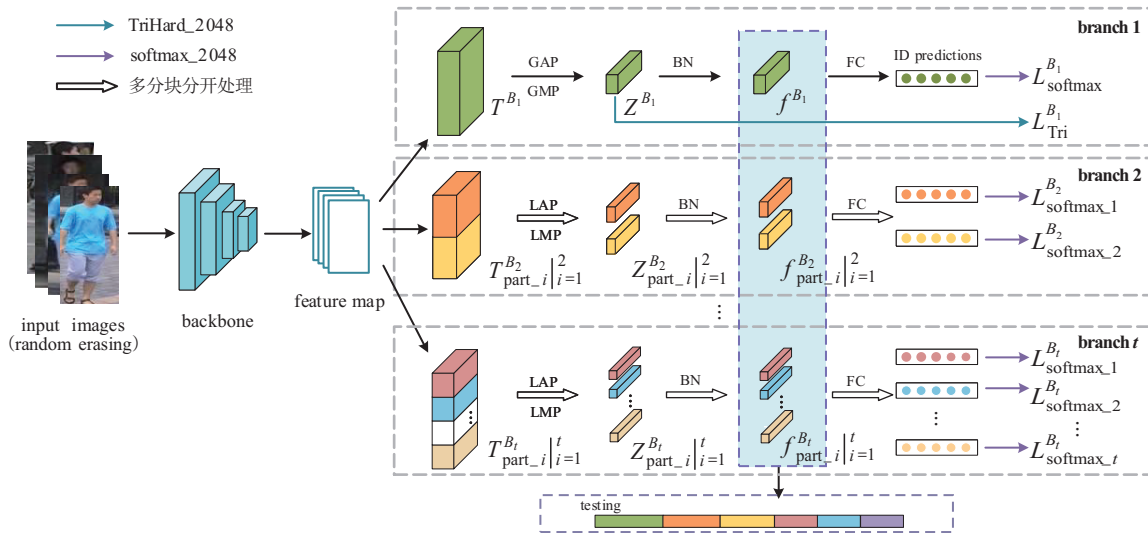


图1 尺度特征表示模型

表1 不同分支中的特征尺寸

分支	特征尺寸	特征数量
branch 1	$24 \times 8 \times 2048$	1
branch 2	$12 \times 8 \times 2048$	2
\vdots	\vdots	\vdots
branch t	$24/t \times 8 \times 2048$	t

平均池化通过获取行人区域信息和背景区域信息的平均值,能够获取特征上下文之间的联系.最大池化通过获取采样区域内的最大响应特征,能够获得具有区分性的信息.在本文提出的多尺度特征表示模型中,对全局特征和各局部特征分别采用平均池化和全局池化进行下采样,在获取上下文信息关系的同时去除干扰信息的影响.本文中,将全局特征的池化操作称为全局平均池化(global average pooling, GAP)/全局最大池化(global max pooling, GMP),相应地,将局部特征的池化操作称为局部平均池化

(local average pooling, LAP)/局部最大池化(local max pooling, LMP). 本文将平均池化结果和最大池化结果相加,作为最终的池化结果.全局分支的池化操作可表示为

$$Z^{B_1} = \text{GAP}(T^{B_1}) + \text{GMP}(T^{B_1}), \quad (1)$$

其中 T^{B_1} 表示由backbone网络提取的全局特征.局部分支的池化操作表示为

$$Z_{\text{part}_i}^{B_j} = \text{LAP}(T_{\text{part}_i}^{B_j}) + \text{LMP}(T_{\text{part}_i}^{B_j}), \quad (2)$$

其中 $T_{\text{part}_i}^{B_j}$ 表示第 j 分支中第 i ($i = 1, 2, \dots, j$)个局部特征块.

2.2 损失函数

一般情况下,同一行人不同图像之间的相似度大于不同行人图像之间的相似度.通过计算查询集(query)中行人特征与候选集(gallery)中行人特征的

相似度大小,并按照由大到小的顺序进行排序,从候选集中检索出非同一设备下与查询行人ID一致的行人图像.相似度大小取决于行人特征之间距离的远近.为缩小同一行人图像之间的距离,增大不同行人图像之间的距离,本文在全局分支中加入TriHard loss同时学习类内相似性和类间差异性,增强了行人特征的可区分性.

如图2所示,从每一个训练批次(batch)内随机挑选 p 个行人,每个行人有 k 张图像.将批次中的每一张行人图像 x_a 作为锚点(anchor),分别挑选最难正样本 x_p 和最难负样本 x_n 组成三元组.最难正负样本表示为:每一批次中选取所有正样本(与锚点标签一致)中距锚点最远的一个作为最难正样本;选取所有负样本(与锚点标签不一致)中距锚点最近的一个作为最难负样本.TriHard loss可表示为

$$L_{\text{Tri}} = (\max d(x_{a,p}) - \min d(x_{a,n}) + \alpha)_+. \quad (3)$$

其中: $d(x_{a,p})$ 表示正样本与锚点之间的距离, $d(x_{a,n})$ 表示负样本与锚点之间的距离, α 是人为设定的阈值参数.

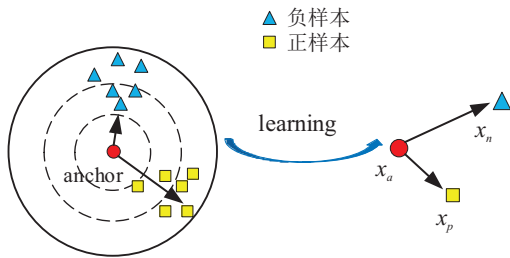


图2 TriHard loss

行人再识别是一项多分类任务,本文首先通过softmax函数计算输入图像 I 属于每一类别的概率,对应概率值最大的类别就是模型预测的结果 S_j ,即

$$S_j = \operatorname{argmax}_{j \in C, q \neq j} \frac{e^{\mathbf{w}_j^T f(I)}}{\sum_{q=1}^C e^{\mathbf{w}_q^T f(I)}}, \quad (4)$$

其中 f 表示行人图像 I 的特征,是全连接层的权重.然后,将预测结果 S_j 作为cross entropy loss的输入来计算与所对应真实标签 y_j 的距离,通过反向传播更新网络参数.本文采用的分类损失函数为

$$L_{\text{softmax}} = - \sum_j^C y_j \log(S_j). \quad (5)$$

2.3 多尺度特征表示模型算法

本文所提出的多尺度特征表示模型训练算法如算法1所示.

算法1 多尺度特征表示模型训练算法.

input: 行人图像 I (图像面积 $S = H \times W$);擦除区

域 I_e 面积为 $0.02S \leq S_e = h \times w \leq 0.4S$,长宽比 r 的上下限 $r_1 = 1, r_2 = 0.3$.

output: 收敛的多尺度特征表示模型.

//采用随机擦除对图像进行预处理

1) for each I :

2) if $w < W$ and $h < H$:

3) $(x_1, y_1)(0 : H - h, 0 : W - w)$ //随机找出一像素位置

4) $I_e \leftarrow 0$ //对所选区域进行像素擦除

5) end if

6) end for

//训练模型

7) 初始化backbone模型参数

8) for each epoch:

9) 从训练集中选取 $p \times k$ 张预处理后的行人图像,经backbone网络得到行人特征图

10) 特征图直接作为全局特征 T^{B_1}

11) 根据式(1)进行下采样得到 Z^{B_1}

12) 对 Z^{B_1} 归一化后得到 f^{B_1}

13) 根据式(3)计算 L_{Tri}

14) for j in range(2, t):

15) 将特征图水平均分,获取 j 个局部特征

$$T_{\text{part}_i}^{B_j} \Big|_{i=1}^j$$

16) 根据式(2)进行下采样得到 $Z_{\text{part}_i}^{B_j} \Big|_{i=1}^j$

17) 对 $Z_{\text{part}_i}^{B_j} \Big|_{i=1}^j$ 归一化后得到 $f_{\text{part}_i}^{B_j} \Big|_{i=1}^j$

18) end for

19) 根据式(5)计算各分支的分类损失 L_{softmax}

20) end for

3 实验分析

3.1 实验环境

本实验的环境配置为:64位Ubuntu16.04操作系统;NVIDIA GTX 1080Ti GPU;深度学习框架Pytorch1.1.0;编程语言Python3.5.

由于所使用的数据集中每个行人ID所对应的样本数量不等,为进行均衡采样,在每一批次中随机地选择 $p = 16$ 个不同ID的行人,每个ID行人随机选择 $k = 4$ 张图像,保证每一ID行人都有相同数量的训练样本.在训练过程中,使用Adam优化器,初始学习率设置为 3.5×10^{-4} ,并在第40 epoch和第70 epoch,学习率进行系数为0.1的指数衰减,训练总次数为120 epoch.

3.2 性能评估

本文所提出的方法采用多尺度特征来表征行人,并且各特征之间单独训练的方式有效提升了模型在

行人不对齐等场景下的适用性;在全局分支中加入了TriHard loss以增强同一行人图像之间的聚集性,提升模型的识别能力;结合最大池化和平均池化在下采样中的优势,提取具有区分性的行人特征又防止干扰信息的影响;加入随机擦除增强数据,使模型更适应于行人遮挡等实际场景.为验证上述方法的有效性,本文进行了以下实验.

1) 局部特征划分数量对模型精度的影响.

增加对特征图水平划分条纹的数量,可以获取更加细致的局部特征.为研究不同细粒度特征对行人再识别模型的影响,本文将特征图分别进行1~7等分,在Market-1501和DukeMTMC-reID数据集上进行实验(在下采样过程中使用了最大池化和平均池化两种方式).不同分块数量对rank-1准确度的影响如图3所示,从图中可以看出,增加对特征图的等分数量可以明显提升识别精度,但在达到一定的识别准确率后,继续增加等分数量并不能相应提升模型性能,

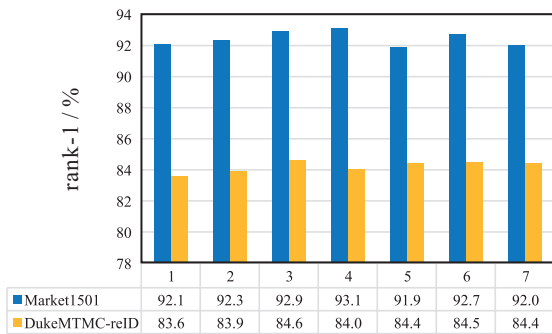


图3 不同分块数量对rank-1准确度的影响

甚至出现下降的趋势.在Market-1501数据集的测试中,等分数量为4时模型取得了93.1%的最优成绩;在DukeMTMC-reID数据集的测试中,等分数量为3时模型取得了84.6%的最优成绩.

不同分块数量得到不同细粒度的局部特征与全局特征相互补充,可以学习更多具有判别性的行人特征.从上述实验中可以看出,细粒程度与模型准确率不完全成正比.为探究全局特征与不同细粒度局部特征组合方式对模型的影响,本文选取等分数量逐渐递增的组合方式进行实验.特别地,由于上述实验中不同数据集所取得的最优rank-1准确率所对应的划分数量不一致,针对本文所使用的Market-1501和DukeMTMC-reID数据集,当组合分支数达到3时,将图1所示的模型结构分别设置为 $j = 2, 3$ (模型记作branch123)和 $j = 2, 4$ (模型记作branch124)进行实验.

表2中Market-1501和DukeMTMC-reID数据集的测试结果显示:rank-1准确率的最优结果所对应的组合方式为branch1234;mAP的最优结果所对应的组合方式为branch123.在组合方式为branch12345时,模型的准确率骤降:在Market-1501和DukeMTMC-reID数据集的测试中,相较于branch1234组合方式,rank-1的准确率分别下降了0.7%和1.3%;mAP分别下降了1.9%和1.4%.由此可见,特征数量过多时易出现过拟合,导致模型准确率下降.

表2 不同分块数目以及不同组合方式下度量损失对准确度的影响

模型	Market-1501				DukeMTMC-reID			
	rank-1	rank-5	rank-10	mAP	rank-1	rank-5	rank-10	mAP
branch 1	92.1	96.7	97.8	79.2	83.6	92.1	94.0	68.9
branch 2	92.3	96.9	98.1	77.9	83.9	92.1	94.0	69.8
branch 3	92.9	97.4	98.2	80.3	84.6	92.0	94.3	70.9
branch 4	93.1	97.1	98.0	79.8	84.0	91.3	94.0	70.5
branch 5	91.9	96.9	97.9	75.5	84.4	91.6	94.2	70.2
branch12	92.8	97.0	98.2	78.9	84.6	92.7	94.6	70.1
branch123	93.4	97.5	98.7	81.6	85.4	93.1	95.1	72.7
branch124	93.2	97.4	98.5	80.7	85.2	93.0	95.0	72.1
branch1234	93.6	97.4	98.5	81.0	85.9	93.1	94.8	71.7
branch12345	92.9	97.2	98.3	79.1	84.6	92.1	94.2	70.3
branch123+triplet	94.2	97.9	98.9	82.2	86.4	93.2	94.9	73.2
branch123+TriHard	94.3	98.0	98.8	82.8	87.0	93.6	95.3	74.4
branch124+triplet	93.5	97.4	98.5	80.9	85.9	93.3	94.9	72.2
branch124+TriHard	93.8	97.7	98.5	81.7	86.3	93.0	95.1	72.4
branch1234+triplet	93.9	98.0	98.6	82.1	86.2	93.1	94.7	73.5
branch1234+TriHard	94.2	98.0	98.7	82.2	87.0	93.3	95.3	74.2

当分支数量为3时,branch123组合方式在两个数据集上均取得了较优成绩.在Market-1501数据集和DukeMTMC-reID数据集的测试中,相较于单独

使用全局特征(branch1),rank-1准确率分别提升了1.3%和1.8%.在Market-1501数据集中,branch123组合方式相较于单独使用等分数量为4的局部特征

(branch4), rank-1 准确率和 mAP 分别提升了 0.3% 和 1.8%; 相较于 branch124 组合方式, rank-1 准确率和 mAP 分别提升了 0.2% 和 0.9%。

2) 度量学习对模型精度提升的影响.

行人再识别任务可以采用三元组损失从细节上区分差异较小的行人特征. 本文在 branch123、branch124 和 branch1234 三种分块组合方法的全局分支中均加入了 triplet loss 和 TriHard loss 进行实验, 与 triplet loss 不同的是, TriHard loss 在样本的选取过程中选择最难样本组成三元组来计算三元组损失. 从表2中可以看出, 加入三元组损失的训练策略对以上3种组合方式均有不同程度提升. 在 Market-1501 数据集的测试中使用 triplet loss 和 TriHard loss 后, branch124 组合方式的 mAP 分别提升了 0.2% 和 1.0%; Branch123 组合方式的 mAP 提升了 0.6% 和

1.2%; branch1234 组合方式的 mAP 提升了 1.1% 和 1.2%. 相比于 triplet loss, TriHard loss 能更好地优化模型, 使模型更好地区分不同类样本之间的差异性, 提升识别的准确率, 并且实验表明, branch123 组合方式结合 TriHard loss 的识别结果最优, 故后续实验均在该网络结构中实现.

3) 不同池化方式性能比较.

本文以多尺度特征表示模型为基础, 进行不同池化方式效果的消融实验. 在表3中, 相较于单独使用平均池化或最大池化, 采用将二者结合起来使用的方式, 在 Market-1501 数据集的测试中, mAP 分别提升了 1.1% 和 0.4%; 在 DukeMTMC-reID 数据集的测试中, mAP 分别提升了 1.3% 和 1.2%. 从而验证了结合使用最大池化和平均池化方法的有效性.

表3 不同方法在 Market-1501 和 DukeMTMC-reID 数据集上的准确度

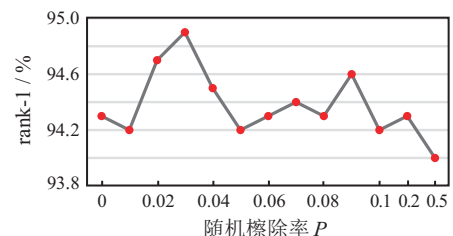
模型	Market-1501				DukeMTMC-reID			
	rank-1	rank-5	rank-10	mAP	rank-1	rank-5	rank-10	mAP
Re-rank ^[29]	77.11	—	—	63.63	—	—	—	—
SVDNet ^[30]	82.3	92.3	95.2	62.1	76.7	—	—	56.8
PAN ^[24]	82.8	—	—	63.4	71.67	—	—	51.5
GAN ^[31]	84.0	—	—	66.1	67.7	—	—	47.1
TriNet ^[20]	84.9	94.2	—	69.1	72.4	—	—	53.5
KPM ^[32]	90.1	96.7	97.9	75.3	80.3	89.5	91.9	63.2
HA-CNN ^[33]	91.2	—	—	75.7	80.5	—	—	63.8
AlignedReID ^[5]	91.8	97.1	—	79.3	—	—	—	—
PABR ^[34]	91.7	96.9	98.1	79.6	84.4	92.2	93.8	69.3
DNN-CRF ^[35]	93.5	97.7	—	81.6	84.9	92.3	—	69.5
PCB+RPP ^[9]	93.8	97.5	98.5	81.6	83.3	90.5	92.5	69.2
SCPNet ^[36]	94.1	97.7	—	81.8	84.8	91.9	—	68.5
HPM ^[25]	94.2	97.5	98.5	82.7	86.6	93.0	95.1	74.3
ours(avgpool+TriHard)	93.9	97.8	98.8	81.7	86.4	93.5	95.2	73.1
ours(maxpool+TriHard)	94.4	97.7	98.5	82.4	86.2	93.1	94.9	73.2
ours(avg+maxpool+TriHard)	94.3	98.0	98.8	82.8	87.0	93.6	95.3	74.4
ours (avg+maxpool+TriHard)+ 随机擦除	94.9	98.2	98.8	83.7	87.1	93.9	95.7	75.4

4) 采用随机擦除增强数据.

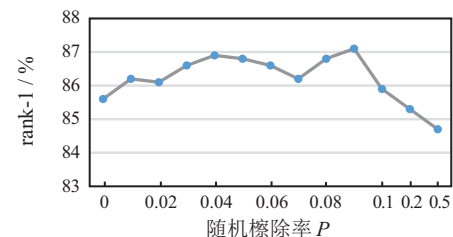
为增强本文所提出模型在图像存在遮挡下的鲁棒性, 采用随机擦除的方法增强数据, 其中随机擦除率(擦除面积与图像总面积的比值)是一个重要的参数. 当擦除率选择过低时, 对解决行人遮挡问题没有实质性意义; 当擦除率选择过高时会丢失重要的行人信息. 本文选用不同的擦除率 P 进行实验. 从图4中观察到, 在 Market-1501 数据集中, 当 $P = 0.03$ 时, rank-1 的准确率取得了 94.9% 的最优结果; 在 DukeMTMC-reID 数据集中, 当 $P = 0.09$ 时, rank-1 的准确率取得了 87.1% 的最优结果.

3.3 与目前主流方法的对比

在表3中, 将本文所提出的方法与主流方法进行比较. 实验结果显示, 本文所提出的多尺度特征表示



(a) Market-1501



(b) DukeMTMC-reID

图4 不同擦除率 P 对模型识别准确度的影响

模型在 Market-1501 数据集和 DukeMTMC-reID 数据集中, rank-1 准确率分别达到了 94.9% 和 87.1%, mAP 分别达到了 83.7% 和 75.4%。相比于单独使用全局特征(如 SVDNet、TriNet 等)或单独使用局部特征(如 PCB 等)的方法, 本文所采用的方法有效地提升了模型识别的准确率。相较于同样使用了多尺度特征的 HPM 模型, 本文所提出的多尺度特征表示模型, 在 Market-1501 数据集和 DukeMTMC-reID 数据集的测试中, rank-1 准确率分别提升了 0.7% 和 0.5%, mAP 分别提升了 1.0% 和 1.1%, 并且使用了更少的特征数量, 节省了计算成本。

本文随机检索了 Market-1501 查询集中 3 个行人, 并将检索结果中的前 10 张图片展示在图 5 中, 绿色数字表示为检索结果正确, 红色数字则代表检索结果错误。从图 5 中可以观察到, 无论捕获的行人姿态是正面、背面还是侧面, 本文所提出的多尺度特征表示模型均展现了良好的鲁棒性。并且在第 1 查询人中, 被检索出的第 6 序位图片被栏杆遮挡的行人亦可被检索出来, 在第 3 查询人中, 即使是与查询图像不对齐的第 10 序位行人图片也可以被检索出来, 从而验证了本文所提出的方法可以有效地缓解行人遮挡和行人不对齐所产生的不利影响。

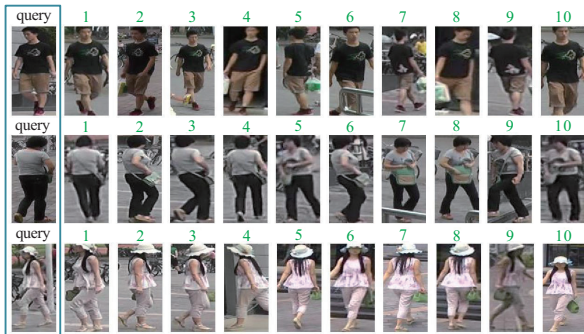


图 5 跨设备下的行人检索结果

对于图片较为模糊的查询人, 图 6 所展示的多尺度模型检索结果中, 可以将该行人在候选集中所有跨设备图片正确地检索出来(该查询行人在候选集中共有 7 张图片, 其中 1 张图片与该查询行人是在同一设备下所拍摄的, 该查询行人共有 6 张行人图片可以被正确地检索)。而 HPM 模型检测结果在第 6 序位的



图 6 与 HPM 模型行人检索结果的对比

查询结果中出现错误, 说明相比于 HPM 方法, 本文所提出的方法可以有效地提升行人再识别模型的精度。

4 结论

本文提出了一种基于多尺度特征表示的行人再识别模型, 通过实验验证了以下结论:

1) 多粒度的局部特征表示可以获取多样性的细节信息, 结合全局特征一起使用可以增强模型在图像存在遮挡和行人不对齐等场景下的适应性。

2) 相比于使用单一的池化方式, 采用平均池化与最大池化相结合的方法可以有效提升行人再识别模型的精度。

3) TriHard loss 可以增强行人样本之间的可区分性, 有效地减少误识别率; 采用随机擦除数据增强方法, 提升了模型在行人遮挡场景中的泛化能力。

参考文献(References)

- [1] Fan H H, Zheng L, Yan C G, et al. Unsupervised person re-identification[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2018, 14(4): 1-18.
- [2] Matsukawa T, Okabe T, Suzuki E, et al. Hierarchical Gaussian descriptor for person re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 1363-1372.
- [3] Chang X B, Hospedales T M, Xiang T. Multi-level factorisation net for person re-identification[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 2109-2118.
- [4] Zheng L, Yang Y, Hauptmann A G. Person re-identification: Past, present and future[J]. 2016, arXiv: 1610.02984.
- [5] Cheng D, Gong Y H, Zhou S P, et al. Person re-identification by multi-channel parts-based CNN with improved triplet loss function[C]. IEEE Computer Vision and Pattern Recognition. Las Vegas, 2016: 1335-1344.
- [6] 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]. IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 907-915.
- [7] Su C, Li J N, Zhang S L, et al. Pose-driven deep convolutional model for person re-identification[J]. IEEE International Conference on Computer Vision. Venice, 2017: 3980-3989.
- [8] 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.
- [9] Sun Y F, Zheng L, Yang Y, et al. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)[J/OL]. 2017, arXiv: 1711.09349.
- [10] Chen Y B, Zhu X T, Gong S G, et al.

- Person re-identification by deep learning multi-scale representations[C]. IEEE International Conference on Computer Vision Workshop. Venice, 2017: 2590-2600.
- [11] Liu H, Feng J S, Qi M B, et al. End-to-end comparative attention networks for person re-identification[J]. IEEE Transactions on Image Processing, 2017, 26(7): 3492-3506.
- [12] Liu X H, Zhao H Y, Tian M Q, et al. HydraPlus-Net: Attentive deep features for pedestrian analysis[J]. IEEE International Conference on Computer Vision. Venice, 2017: 350-359.
- [13] Zhou S P, Wang J J, Meng D Y, et al. Discriminative feature learning with foreground attention for person re-identification[J]. IEEE Transactions on Image Processing, 2019, 28(9): 4671-4684.
- [14] Zhong Z, Zheng L, Kang G L, et al. Random erasing data augmentation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 13001-13008.
- [15] Li W, Zhao R, Xiao T, et al. DeepReID: Deep filter pairing neural network for person re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition. Columbus, 2014: 152-159.
- [16] Yi D, Lei Z, Liao S C, et al. Deep metric learning for person re-identification[C]. International Conference on Pattern Recognition. Stockholm, 2014: 34-39.
- [17] Varior R R, Shuai B, Lu J W, et al. A siamese long short-term memory architecture for human re-identification[C]. European Conference on Computer Vision. Amsterdam, 2016: 135-153.
- [18] Zheng L, Huang Y J, Lu H C, et al. Pose invariant embedding for deep person re-identification[J]. IEEE Transactions on Image Processing, 2019, 28(9): 4500-4509.
- [19] Zheng Z D, Zheng L, Yang Y. Pedestrian alignment network for large-scale person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 29(10): 3037-3045.
- [20] Fu Y, Wei Y C, Zhou Y Q, et al. Horizontal pyramid matching for person re-identification[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33: 8295-8302.
- [21] He K M, Zhang X Y, Ren S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [22] Wu P L, He B, Hou Z G. Household tools classification recognition based on spatial pyramid pooling features[J]. Control and Decision, 2019, 34(7): 1481-1486.
- [23] Wang G S, Yuan Y F, Chen X, et al. Learning discriminative features with multiple granularities for person re-identification[J/OL]. 2018, arXiv: 1804.01438.
- [24] Hadsell R, Chopra S, LeCun Y. Dimensionality reduction by learning an invariant mapping[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, 2006: 1735-1742.
- [25] Bai X, Yang M K, Huang T T, et al. Deep-person: Learning discriminative deep features for person re-identification[J]. Pattern Recognition, 2020, 98: 107036.
- [26] Chen W H, Chen X T, Zhang J G, et al. Beyond triplet loss: A deep quadruplet network for person re-identification[J/OL]. 2017, arXiv: 1704.01719.
- [27] Hermans A, Beyer L, B L. In defense of the triplet loss for person pe-Identification[J]. 2017, arXiv: 1703.07737.
- [28] Luo H, Jiang W, Gu Y Z, et al. A strong baseline and batch normalization neck for deep person re-identification[J]. IEEE Transactions on Multimedia, 2020, 22(10): 2597-2609.
- [29] Zhong Z, Zheng L, Cao D L, et al. Re-ranking person re-identification with k -reciprocal encoding[C]. IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 3652-3661.
- [30] Sun Y F, Zheng L, Deng W J, et al. SVDNet for pedestrian retrieval[C]. IEEE International Conference on Computer Vision. Venice, 2017: 3820-3828.
- [31] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
- [32] Shen Y T, Xiao T, Li H, et al. End-to-end deep kronecker-product matching for person re-identification[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 6886-6895.
- [33] Li W, Zhu X T, Gong S G. Harmonious attention network for person re-identification[J/OL]. 2018, arXiv: 1802.08122.
- [34] Suh Y, Wang J D, Tang S Y, et al. Part-aligned bilinear representations for person re-identification[C]. Proceedings of the European Conference on Computer Vision. Munich, 2018: 402-419.
- [35] Chen D P, Xu D, Li H S, et al. Group consistent similarity learning via deep CRF for person re-identification[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 8649-8658.
- [36] Fan X, Luo H, Zhang X, et al. SCPNet: Spatial-channel parallelism network for joint holistic and partial person re-identification[J/OL]. 2018, arXiv: 1810.06996.

作者简介

卢健(1978—), 男, 副教授, 博士, 从事机器人导航与智能决策、复杂流程工业过程控制、人工智能与机器视觉等研究, E-mail: lujian@xpu.edu.cn;

王航英(1997—), 女, 硕士生, 从事深度学习、行人重识别的研究, E-mail: 271151831@qq.com;

陈旭(1995—), 女, 助教, 硕士, 从事人工智能及机器视觉的研究, E-mail: 273662018@qq.com;

张凯兵(1975—), 男, 教授, 博士, 从事影像超分辨重建及质量评价、计算机视觉检测与分析、跨模态人脸合成与识别等研究, E-mail: zhangkaibing@xpu.edu.cn;

刘薇(1982—), 女, 讲师, 博士, 从事人工智能与机器视觉、深度学习、图像超分辨等研究, E-mail: lw_xaw@foxmail.com.

(责任编辑: 李君玲)