

基于姿态估计与特征相似度的行人重识别算法

李杲, 蒋敏*

江南大学人工智能与计算机学院模式识别与计算智能工程实验室, 江苏 无锡 214122

摘要 针对行人重识别场景复杂引起的局部特征不对齐,以及在背景杂乱情况下难以提取出具有不变性行人特征的问题,提出一种基于人体姿态估计算法和相似度矩阵引导的多尺度融合网络。网络引入姿态估计算法构造对齐的行人特征,通过多分支结构将低层局部特征和高层全局特征进行融合提升网络的表征能力;此外特征相似度矩阵将全局特征分割出相似度引导的背景、前景分支,再利用区域级的三元损失函数,提取出对复杂背景鲁棒的行人特征。在 Market-1501、DukeMTMC-ReID、CUHK03 和 MSMT17 四个主流数据集上的实验结果表明,本文提出的方法均能达到甚至超过当前主流算法的水平。在最具挑战的 MSMT17 数据集中,与目前精度领先的算法相比,首次命中精度提高了 1.4 个百分点,平均精度均值提高了 3.4 个百分点。

关键词 图像处理; 行人重识别; 局部特征; 姿态估计; 特征相似度

中图分类号 TP391.4 文献标志码 A

DOI: 10.3788/LOP212869

Person Re-Identification Based on Pose Estimation with Feature Similarity

Li Rui, Jiang Min*

Engineering Laboratory of Pattern Recognition and Computational Intelligence, School of Artificial Intelligence and Computer, Jiangnan University, Wuxi 214122, Jiangsu, China

Abstract This paper proposes a part aligned multiscale fusion network, based on a human pose estimation algorithm and similarity matrix to address the misalignment problem of local features caused by complex person re-identification scenes and difficulty in extracting invariant person features in cluttered backgrounds. The proposed network introduces pose-estimation algorithms to construct aligned local features and integrates low-level local features and high-level global features through a multibranch structure. In addition, the feature similarity matrix divides the global features into the similarity-guided background and foreground branches and uses the regional-level triplet loss to extract person features robust to complex backgrounds. Extensive experiments are conducted for four datasets (Market-1501, DukeMTMC-ReID, CUHK03, and MSMT17). The proposed method achieves state-of-the-art performance. In particular, it improves the accuracy of first hit accuracy by 1.4 percentage points and mean average precision by 3.4 percentage points on the most challenging MSMT17 dataset.

Key words image processing; person re-identification; local feature; pose estimation; feature similarity

1 引言

行人重识别通过匹配行人图像,实现行人的跨摄像头追踪,在公共安全和刑事侦查领域有广阔的应用场景,随着深度学习方法的不断发展,其在行人重识别任务上的表现也明显优于传统方法^[1]。Ahmed 等^[2]使用 Resnet-50 作为主干网络,只使用全局特征便使得网络的精度达到一个较高的水平。张涛等^[3]通过引入类

内距离,改进原始三元损失函数,提取出判别性良好的全局特征。但行人重识别场景下经常出现行人遮挡、背景杂乱等问题,仅使用全局特征通常无法获得很好的效果,因此越来越多的研究人员开始关注全局特征和局部特征结合的行人重识别算法。

引入局部特征可以与全局特征形成互补,但是局部特征与全局特征关注点并不相同,局部特征应该更聚焦于细节特征。深度神经网络提取的高层特征包含

收稿日期: 2021-11-03; 修回日期: 2021-12-16; 录用日期: 2022-01-11; 网络首发日期: 2022-01-21

基金项目: 国家自然科学基金(61362030,61201429)、中国博士后科学基金(2015M581720,2016M600360)、中央高校基本科研基金(JUSRP41908)

通信作者: *minjiang@jiangnan.edu.cn

丰富的语义信息,低层特征细节更加丰富,可以将低层和高层进行特征融合,充分发挥深度网络的优势。Sun等^[4]将特征图水平分割成6个固定大小的区域作为局部特征,并引入精细局部池化缓解特征不对齐的问题;李聪等^[5]通过多分支网络结构,实现全局特征和局部特征的有效融合;Zhao等^[6]提出了行人对齐网络,综合全卷积神经网络和区域提取网络,提取图像中 k 个最有判别力的区域,拼接成最终的行人特征。Chen等^[7]通过引导级联抑制机制,挖掘不同阶段的最有效特征并将其集成到最终特征中,达到特征互补的目的。张正一等^[8]将不同深度的特征进行级联,充分利用低层的细节特征,获得了很好的效果。这些方法通常不能完全解决行人特征不对齐的问题,同时针对每个局部特征均需要引入单独分支,提升了网络的复杂度。

复杂背景也是影响行人重识别算法精度的一个重要因素,Tian等^[9]研究表明,背景偏置会影响模型的训练效果,模型仅学习背景特征,就会得到比随机猜测要高得多的正确率,这明显很不合理。Song等^[10]利用图像分割算法将行人图形从背景中识别出来,但图像分割算法的引入会为模型引入新的误差,算法精度会受到图像分割准确性的影响;Qi等^[11]在主干网络的基础上引入分支网络,实现对行人图像前景和背景的分割。这种方法会增加模型结构的复杂度,增加训练的难度。

针对行人遮挡、复杂背景等一系列问题,本文提出一种基于人体姿态估计算法和特征相似度引导的多尺度融合的行人重识别算法。针对局部特征不对齐的问题,算法基于人体姿态估计引入区域对齐池化,并使各尺度特征进行互补;针对复杂背景问题,算法提出一种基于相似度矩阵的人物前景特征分割方法,并引入区域级三元损失函数对前景特征进行区分,有效抑

制了背景杂乱的影响。算法在Market-1501^[12],DukeMTMC-ReID^[13]、CUHK03^[14]和MSMT7^[15]数据集上进行了测试评估,实验结果表明本文提出的算法具有很高的性能,可以有效提高行人重识别的精度。

2 基于人体姿态估计算法和相似度矩阵的融合网络

2.1 网络结构

以Resnet-50^[16]作为骨架,其在行人重识别任务中表现出较好的性能;移除特征提取后的平均池化层和全连接层,用独立的分类器代替,这样可以获得良好的性能。同时,研究结果表明,Resnet-50第5层的下采样操作会降低特征的细粒度特征,因此也给予去除^[4]。网络从尺度、粒度多方面构建多分支网络结构,提取具有不变性的行人特征,完成行人重识别任务。其中,第1个分支是行人局部特征提取分支,利用Resnet-50第4层提取的特征图,聚焦更加注重细节的低层特征;第2个分支是行人全局特征提取分支,在Resnet-50第5层提取的特征图基础上,关注更加丰富的语义信息;第3个分支是相似度引导的前景和背景特征提取分支,引入基于特征相似度的分割模块(Similarity-based separate method, SSM),通过相似度矩阵分割出行人的前景和背景,利用区域级的三元损失函数降低杂乱背景对模型的影响。所有分支均引入局部特征区域对齐池化(Part region aligned pooling, PRAP)机制,来解决特征不对齐的问题,对池化后的特征向量,通过 1×1 的卷积层进行特征整合,形成256维的特征向量,最后利用全连接层和Softmax获得最终的ID Loss。

局部对齐多尺度融合网络(Part aligned multiscale fusion network, PAMFNet)的总体结构如图1所示。

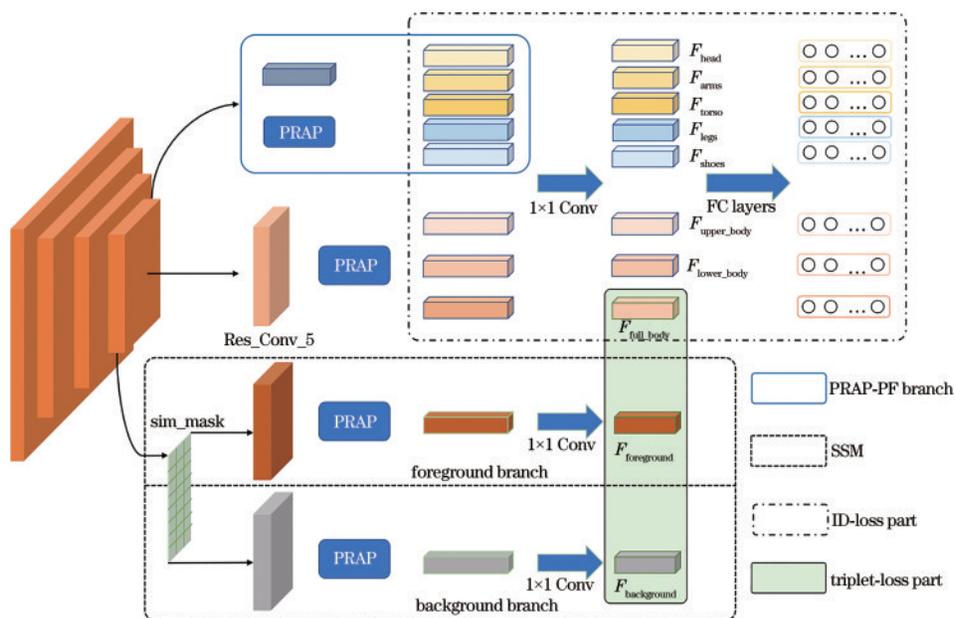


图1 PAMFNet结构

Fig. 1 Structure of PAMFNet

2.2 PRAP 机制

之前有研究人员通过对行人特征图进行水平切分提取局部特征,取得了较好的效果,但是平均切分的特征存在明显的特征无法对齐的问题。为解决此问题,引入行人姿态估计算法,提出了PRAP机制,该机制利用行人姿态估计算法提取行人身体关键点,通过关键点的定位能力结合图像的长宽比例将行人图像分割成具有语义特征的不同部分,再对特征图中具有相同语义的区域进行池化,最终得到对齐的行人特征。

对于局部特征分支,针对特征图中语义相同的部

分划分出5个局部特征,分别是头、臂膀、上半身躯干、腿、鞋,构成5个特征对齐的局部特征;对于全局特征分支,仅利用人体胯部关键点位置信息,将特征图划成仅上下两部分而非更细小的对齐特征,再加上完整图片的特征,构成3个全局特征。行人重识别数据图片中存在比较明显的遮挡问题,为解决此问题,引入可见向量 v ,描述不同分支特征是否存在遮挡。 v 根据关键点置信度确定,当特征点置信度小于0.2时, $v=0$;当置信度大于0.2时, $v=1$ 。

图2为引入行人姿态估计之后,局部特征池化区域的可视化结果。



图2 局部特征区域对齐池化结果。(a)原始行人图像;(b)姿态估计算法提取出的行人关键点;(c)基于姿态估计得出的局部特征池化区域;(d)对齐的上半身和下半身全局特征

Fig. 2 Results of PRAP. (a) Original pedestrian images; (b) pedestrian key points extracted by pose estimation algorithm; (c) local feature pooling region based on attitude estimation; (d) aligned upper and lower body global features

由图2可知,通过姿态估计模型,可以使得局部特征池化覆盖人体语义相近的部分,实现局部特征的对齐。

为了弥补高层特征对细节的忽视,将局部特征在神经网络的低层作为单独的分支引入。分支的位置对模型性能有比较明显的影响,基于以下原因考虑:1)对于Resnet-50的低层特征,网络的感受野较小,特征主要关注空间细节,包含更多的细粒度特征,但是背景的杂乱信息更容易造成干扰;2)网络的高层特征感受野更大,更加关注语义特征,不太关注细节信息。综

合以上两点,选择在Resnet-50网络的第4层引入局部特征分支。同时,通过PRAP机制提取局部特征,并不与高层特征进行直接融合,而是作为独立特征计算ID-loss,实现低层和高层特征的互相补充。

2.3 SSM

背景杂乱是影响行人重识别模型精度的重要因素,提出一种基于特征相似度矩阵引导的掩模分割,同时引入区域尺度的三元损失函数来去除杂乱背景的干扰。

在Resnet-50主干网络的第4层,计算像素级别特

征和特征向量的相似度,形成像素级别的相似度矩阵,利用相似度矩阵作为掩模,对特征图进行软分割。相似度矩阵的计算原理如图 3 所示。

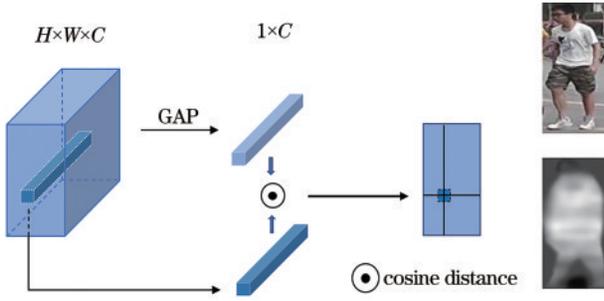


图 3 相似度矩阵结构

Fig. 3 Structure of similarity matrix

将 $H \times W \times C$ 的特征图经过全局平均池化 (GAP) 之后,得到 $1 \times 1 \times C$ 的基准特征向量,计算其与特征图每个像素点的 $1 \times 1 \times C$ 特征向量之间的距离,该距离可以视作基准特征向量与像素点特征向量之间的相似度,最终得到 $H \times W \times C$ 的相似度矩阵。相似度矩阵 $M_{\text{sim}(i,j)}$ 的计算过程如下:

$$M_{\text{sim}(i,j)} = \text{SIM}(\mathbf{F}_{(i,j)}, \mathbf{F}_{\text{GAP}}), \quad (1)$$

式中:SIM 为向量距离计算函数,选取向量余弦距离作为衡量距离的计算函数; $\mathbf{F}_{(i,j)}$ 为像素点 (i,j) 处的特征向量; \mathbf{F}_{GAP} 为 GAP 后得到的特征向量。

将相似度矩阵进行归一化和阈值化后,得到可以区分行人前景和背景的特征掩模,特征图经过分割后形成前景、背景两个分支。这两个分支分别通过主干网络的最后一层提取高层语义特征,再经过池化和 1×1 的卷积层进行特征整合,形成可以表征背景和人物前景的特征向量。将背景特征作为负样本,前景特征作为正样本,全局特征作为固定样本,建立单个图像区域级别的三元损失函数,约束背景杂乱对模型的影响。

通过区域级别的三元损失函数约束,在特征空间中,图像全局和人物前景提出的特征将会相互靠近,而与背景特征远离。借助这一机制,模型可以更加关注于图像中的人物前景特征而忽视图像中杂乱的背景,提高行人重识别模型的精度。

2.4 损失函数设计

研究表明,对于不同的特征分支,训练独立的全连接层分类器,得到多个交叉熵损失函数并进行组合,这种方式可以提高模型的精度。在模型中,基于行人身份标签 (ID) 的交叉熵损失函数的计算如下:

$$L_{\text{ID}} = \sum_{i=1}^p L_{\text{ID}}^{(i)} \times v_i, \quad (2)$$

式中: $v_i \in \{0, 1\}$ 为第 i 个分支特征的可见性,表示第 i 个局部特征是否被遮挡; $L_{\text{ID}}^{(i)}$ 为第 i 个局部特征损失函数计算结果。

具体第 i 个分支的损失函数计算如下:

$$L_{\text{ID}}^{(i)} = - \sum_{j=1}^N q_i \ln \frac{\exp(\mathbf{w}_j^T \mathbf{f})}{\sum_{k=1}^N \exp(\mathbf{w}_k^T \mathbf{f})}, \begin{cases} q_i = 0, y \neq i \\ q_i = 1, y = i \end{cases}, \quad (3)$$

式中: \mathbf{w}_j 为全连接层分类器对应类别的权重; \mathbf{w}_k^T 为第 k 类的权重; \mathbf{f} 为特征向量; y 为样本的标签; N 为不同行人 ID 的数量。

由于行人重识别任务的特殊性,训练集和测试集的行人 ID 是不同的。因此,直接采用交叉熵损失会造成一定程度的过拟合。研究表明,采用 Label smoothing 可以缓解过拟合现象, q_i 定义为

$$q_i = \begin{cases} \epsilon/N, y \neq i \\ 1 - (N-1) \cdot \epsilon/N, y = i \end{cases}, \quad (4)$$

式中: ϵ 为模型训练中值为 $[0, 1]$ 的超参数,用于降低训练集样本的置信度,设置为 0.1。

为解决背景杂乱问题,对同一张图片引入区域级别的三元损失函数,计算公式如下:

$$L_{\text{trip}} = \|h_{\text{raw}} - h_{\text{fgf}}\|_2^2 + \max\{m - \|h_{\text{raw}} - h_{\text{bgf}}\|_2^2, 0\}, \quad (5)$$

式中: h_{raw} 为经处理的原始图片特征; h_{fgf} 为图像的前景特征(主要包含行人信息); h_{bgf} 为图像的背景特征(主要包含背景信息)。

在三元损失函数中, h_{raw} 为 anchor 样本, h_{fgf} 为正样本, h_{bgf} 为负样本,通过区域级别的三元损失函数,模型可以将行人前景和复杂背景区分开,有效提高模型的特征能力和泛化性能。

3 实验结果与分析

在行人重识别的主流数据集上对提出的模型进行实验与分析,选取的数据集为 Market-1501^[12]、DukeMTMC-ReID^[13]、CUHK03^[14] 和 MSMT7^[15]。在这 4 个数据集上均与当前最新算法进行了性能比较。同时,为验证各模块的有效性,还设计了一系列的消融实验作为佐证。

3.1 数据集与评估方法

Market-1501 数据集包含 1501 个不同行人的 36036 张图片,这些图片采集自 6 个不同的摄像头,并通过可变形的组件模型 (DPM) 检测器^[17] 自动框定行人边界。训练集由 12936 张图片组成,包含 750 个不同行人身份;测试集由 19732 张图库图像和 3368 张查询图像组成,包含 751 个不同行人身份。数据集中还包含 6617 张干扰图像,主要是编注不合格等原因造成。

DukeMTMC-ReID 数据集是 DukeMTMC 数据集在行人重识别领域的子集,其包含 36411 张图像,1404 个有效的行人身份,这些图像采集自 8 个不同的摄像头,采用手工标注的方法确定行人边框。702 个行人身份组成 16522 张图片用于训练,702 个行人身份组成 17661 张图片形成图库图像,2228 张图像组成查询图像用于测试。同时,DukeMTMC-ReID 还有

408个只出现在单一摄像头中的干扰行人身份。由于数据集中行人图像相似度较高,DukeMTMC-ReID是个很有挑战的数据集。

CUHK03数据集是一个规模相对较小的数据集,包含1467个行人身份和14097张行人图像,图像来自一对(2个)摄像机。数据集提供2种类型的行人边框检测方法,分别是DPM检测和手工检测。训练集和测试集分别包含767和700个行人身份。

MSMT17是目前规模最大,也是最具挑战性的数据集。数据集的行人图像采自15个不同摄像头,为了更加贴近真实场景,数据集采集上午、中午、下午3个时间段的行人图像,利用Faster RCNN^[18]裁剪出行人框。数据集包含4101个不同的行人身份共126441张图像,其中,训练集包含1401个行人身份、32621张图像,测试集包含2700个行人身份、82161张图库图像和11659张查询图像。从数据集中可知,MSMT17的图片规模和场景数量非常大,同时也是目前最接近真实场景的行人重识别数据集。

使用累积匹配曲线中的首次命中精度(Rank-1)和平均精度均值(mAP)作为衡量模型性能的指标。Rank-1为搜索结果第1张图片可以匹配正确行人身份的概率,mAP为查询图像在精度召回曲线下方面积的均值,mAP衡量匹配正确的图像在排序列表前列的程度,与Rank-1相比,可以更加全面地评价ReID算法的性能。此外,为确保评价的公正性,没有使用重排序算法,单个行人身份只用特定摄像头中的一张行人图像作为查询图像。

3.2 实验设置

采用ImageNet上预训练的Resnet-50模型来初始化主干网络权重,图像输入模型之前,大小处理为 256×128 。特征经过池化后生成512维特征向量,经过 1×1 的卷积层变为256维特征向量,再经过全连接层分类器生成最终的ID损失函数。在训练阶段,仅使用图像翻转进行数据增强。使用随机梯度下降(SGD)进行训练,动量值为0.9,权重衰减系数为0.0005。初始学习率设置为0.02,每过25个训练周期,学习率下降为之前的10%,总共训练60个周期。训练时的批量大小为32,采用PyTorch框架搭建网络模型,使用NVIDIA 2080Ti GPU进行加速训练,训练时为单GPU模式。行人姿态估计模型在COCO数据集中进行训练,在COCO 2017中的平均精度(AP)为73.5%。

3.3 与最新方法的比较

Market1501数据集上的测试比较结果如表1所示。实验结果表明,本文算法的Rank-1为95.4%、mAP为88.7%,二者均已同时达到目前最高的精度范围。本文算法在Rank-1方面,仅略低于目前精度最高的方法MGN、Pyramid和DSAP;在mAP方面,本文算法高于现有方法,比目前最高的方法DSAP(87.6%)

表1 与Market-1501数据集上最新方法的比较
Table 1 Comparison to state-of-the-art methods on Market-1501 dataset

Method	Market1501	
	Rank-1 / %	mAP / %
PCB+RPP[4]	93.8	81.6
AANet[19]	93.9	83.4
Pyramid[20]	95.7	88.2
MGN[21]	95.7	86.9
DSAP[22]	95.7	87.6
CAMA[23]	94.7	84.5
OSNet[24]	94.8	84.9
MHN[25]	95.1	85.0
FPR[26]	95.4	86.6
Ours	95.4	88.7

提高了1.1个百分点,虽然Rank-1低于MGN和Pyramid,但是本文算法拥有更高的mAP性能,表明总体的准确程度相比之下并不逊色。

DukeMTMC-ReID数据集上的测试比较结果如表2所示。如前文所述,DukeMTMC-ReID是一个很有挑战的数据集,本文算法的Rank-1为88.7%、mAP为78.9%,性能依然达到了目前的最高水平。其中,Pyramid算法的Rank-1为89.0%、mAP为79.0%与本文算法精度相近,但是Pyramid算法采用更加复杂的主干网络Resnet-101,同时Pyramid算法使用了21个局部特征数量,而本文只用了8个局部特征数量。同样与以Resnet-50作为主干网络的算法相比,FPR是目前mAP最好的算法,本文算法与之相比提升了0.5个百分点。

表2 与DukeMTMC-ReID数据集上最新方法的比较
Table 2 Comparison to state-of-the-art methods on DukeMTMC-ReID dataset

Method	DukeMTMC-ReID	
	Rank-1 / %	mAP / %
PCB+RPP[4]	83.3	69.2
AANet[19]	87.7	74.3
Pyramid[20]	89.0	79.0
MGN[21]	88.7	78.4
DSAP[22]	86.2	74.3
CAMA[23]	85.8	72.9
OSNet[24]	88.6	73.5
MHN[25]	89.1	77.2
FPR[26]	88.6	78.4
Ours	88.7	78.9

CUHK03数据集上的测试比较结果如表3所示。CUHK03分为手工标注和行人检测器标注2种模式,所以数据验证也分为2种情况。在手工标注的情况下,本

表 3 与 CUHK03 数据集上最新方法的比较

Table 3 Comparison to state-of-the-art methods on CUHK03 dataset

Method	CUHK03			
	Labeled		Detected	
	Rank-1 / %	mAP / %	Rank-1 / %	mAP / %
PCB+RPP[4]	-	-	63.7	57.5
MGN[21]	68.0	67.4	68.0	66.0
CAMA[23]	70.1	66.5	66.6	64.2
Auto-ReID[27]	77.9	73.0	73.3	69.3
MHN[25]	77.2	72.4	71.7	65.4
FPR[26]	-	-	76.1	72.3
Ours	78.5	75.6	76.2	73.7

文算法的 Rank-1、mAP 分别为 78.5%、75.6%，均为目前最好的性能，其中 Rank-1 比目前精度最高的算法提升了 0.6 个百分点，mAP 精度提升了 2.6 个百分点。在行人边框为行人检测器标注的情况下，本文算法的 Rank-1 为 76.2%、mAP 为 73.7%，也为目前最高的结果。所有算法在手工标注的情况下，精度均优于行人检测器标注，这说明行人边界框的准确性对算法精度有着较大的影响，行人边界框越精准，行人重识别算法的精度也会随之提升。

MSMT17 数据集发布的时间较短，所以在该数据集上验证的算法并不多。算法在 MSMT17 数据集上的验证结果如表 4 所示。结果表明，本文算法在 MSMT17 数据集上的结果可以达到最优精度，Rank-1 为 80.1%、mAP 为 56.3%，比目前精度最高的算法分别提高了 1.4 个百分点和 3.4 个百分点。

3.4 消融实验

在 Market1501 和 DukeMTMC-ReID 数据集上设计了一系列实验来评估每个模块的影响。其中，使用 PCB 作为基线模型，PCB 采用水平平均分割的方法将

表 4 与 MSMT17 数据集上最新方法的比较

Table 4 Comparison to state-of-the-art methods on MSMT17 dataset

Method	MSMT17	
	Rank-1 / %	mAP / %
IANet[28]	75.5	45.8
DGNet[29]	77.2	52.3
Auto-ReID[27]	78.2	52.5
OSNet[24]	78.7	52.9
Ours	80.1	56.3

特征图分割成 6 个局部特征，再通过 6 个独立的分类器配合 Softmax 损失进行训练。

3.4.1 PRAP 机制对实验结果的影响

在 PCB 基线模型的基础上，利用行人姿态估计算法，构建 PRAP 机制。表 5 为 PRAP 机制加入基线模型之后的实验结果。从实验结果可知，PRAP 机制可以有效解决行人重识别中特征无法对齐的问题。首先，本文使用了头部、臂膀、躯干、腿、鞋 5 个局部特征 (PRAP-5P)，结果显示，PRAP-5P 与 PCB 相比，虽然特征数量减少，但是性能尤其是 mAP 却有大幅度提高 (提高 9.1 个百分点)，有效证明 PRAP 机制与直接水平切分池化相比，可以有效提升模型准确性。同时，为解决局部特征对行人整体的学习能力不足问题，再引入上半身、下半身、全身 3 个全局特征组成 PRAP-8P 混合特征，结果显示模型的性能有较大提升，这表明局部特征可以和全局特征形成互补。为了进一步确定最优的局部特征数量，再次将局部特征进行细化，臂膀和腿特征进行拆分，形成 PRAP-10P。结果显示，模型的性能有所下降。这其中有两方面原因：一是局部特征划分太细，局部特征之间的相似度较高，不具有很好的区分度；二是行人重识别任务遮挡比较严重，局部特征过细可能会造成部分局部特征无法有效匹配，影响最终精度。

表 5 PRAP 的影响

Table 5 Effect of PRAP

Method	Market1501		Duke	
	Rank-1 / %	mAP / %	Rank-1 / %	mAP / %
Baseline	92.3	77.4	81.7	66.1
Baseline+PRAP_5P	93.9	86.5	84.3	72.4
Baseline+PRAP_8P(Ours)	94.3	86.8	85.0	74.8
Baseline+PRAP_10P	94.1	86.2	84.7	72.9

3.4.2 局部特征分支对实验结果的影响

在基线网络和 PRAP 的基础上，通过局部特征模

块的位置，确定该模块的影响，实验结果如表 6 所示。

局部特征模块一方面引入局部语义，提升网络特

表 6 局部特征分支的影响

Table 6 Effect of part feature branch

Method	Market1501		Duke	
	Rank-1 / %	mAP / %	Rank-1 / %	mAP / %
Baseline	92.3	77.4	81.7	66.1
Baseline+PRAP+s3	94.4	87.8	86.9	77.8
Baseline+PRAP+s4(Ours)	94.9	87.9	87.5	78.7
Baseline+PRAP+s3+s4	94.7	87.3	87.6	78.5

征提取能力;另一方面也可以成为高层全局特征的补充,与高层特征一起组成行人图像的最终特征。结果表明,在次高层(主干网络第4层)引入局部特征分支可以有效提升基线和PRAP-8P的性能,而如果在更低层引入局部特征模块,模型的性能反而降低;这是因为低层特征存在较多的干扰信息。此外,实验结果表明

模块数量过多并不能有效提升模型的性能,同时低层局部特征与高层全局特征可以形成有效互补。

3.4.3 SSM对实验结果的影响

通过计算特征向量相似度构造相似度矩阵,借助相似度矩阵将特征的前景和背景进行区分,实验结果如表7所示。

表7 SSM的影响
Table 7 Effect of SSM

Method	Market1501		Duke	
	Rank-1 / %	mAP / %	Rank-1 / %	mAP / %
Baseline	92.3	77.4	81.7	66.1
Baseline+PRAP+fore	95.1	88.3	87.9	79.0
Baseline+PRAP+fore+triLoss	95.4	88.7	88.7	78.9

行人前景特征响应热力图如图4所示。由图4可知,经过相似度矩阵引导,前景分支的注意力集中在行人本身,且响应区域非常广泛,对背景的干扰特征没有响应。

同时,本文引入区域级别的三元损失函数能进一步降低背景干扰信息对模型精度的影响,模型也能进一步学习如何区分背景杂乱特征,从而更加关注行人前景特征。



图4 特征热力图。(a)原始行人图像;(b)响应热力图

Fig. 4 Thermodynamic diagrams of feature. (a) Original pedestrian image; (b) responded thermodynamic diagram

部分查询图像的检索结果如图5所示。虚线框是算法输出Rank-10结果中,与查询行人不属于相同ID的图

像。当不同行人类别之间相似度较高或存在遮挡和干扰时,PAMFNet依然可以有效识别,显示出较强鲁棒性。



图5 部分查询图像的Top10检索结果

Fig. 5 Top10 search results of partial query images

4 结 论

针对行人重识别局部特征存在特征不对齐和低层特征利用不足的问题,提出了基于行人姿态估计模型的 PRAP 机制,并将多尺度特征学习机制融入模型。同时,背景杂乱是影响重识别算法精度的重要因素,提出了基于相似度矩阵引导的分割机制,使模型更加关注行人前景特征。提出的单个行人图像区域级的三元损失函数,进一步降低杂乱背景对基线精度的影响。几个主流公开数据集的实验结果表明,本文算法可以有效提升行人重识别性能,后续研究将考虑引入可共同训练的区域分割分支,消除姿态估计算法精度对模型的影响,进一步提升模型的精度和鲁棒性。

参 考 文 献

- [1] 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.
Luo H, Jiang W, Fan X, et al. A survey on deep learning based person re-identification[J]. Acta Automatica Sinica, 2019, 45(11): 2032-2049.
- [2] Ahmed E, Jones M, Marks T K. An improved deep learning architecture for person re-identification[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 3908-3916.
- [3] 张涛, 易争明, 李璇, 等. 一种基于全局特征的行人重识别改进算法[J]. 激光与光电子学进展, 2020, 57(24): 241503.
Zhang T, Yi Z M, Li X, et al. Improved algorithm for person re-identification based on global features[J]. Laser & Optoelectronics Progress, 2020, 57(24): 241503.
- [4] Sun Y F, Zheng L, Yang Y, et al. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline) [M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11208: 501-518.
- [5] 李聪, 蒋敏, 孔军. 基于多尺度注意力机制的多分支行人重识别算法[J]. 激光与光电子学进展, 2020, 57(20): 201001.
Li C, Jiang M, Kong J. Multi-branch person re-identification based on multi-scale attention[J]. Laser & Optoelectronics Progress, 2020, 57(20): 201001.
- [6] Zhao L M, Li X, Zhuang Y T, et al. Deeply-learned part-aligned representations for person re-identification [C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 3239-3248.
- [7] Chen X S, Fu C M, Zhao Y, et al. Saliency-guided cascaded suppression network for person re-identification [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 3297-3307.
- [8] 张正一, 丁建伟, 魏慧雯, 等. 基于注意力机制的多级特征级联的行人重识别方法[J]. 激光与光电子学进展, 2021, 58(22): 2215003.
Zhang Z Y, Ding J W, Wei H W, et al. Cascaded multi-level features learning for attention based person re-identification[J]. Laser & Optoelectronics Progress, 2021, 58(22): 2215003.
- [9] Tian M Q, Yi S, Li H S, et al. Eliminating background-bias for robust person re-identification[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 5794-5803.
- [10] Song C F, Huang Y, Ouyang W L, et al. Mask-guided contrastive attention model for person re-identification [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 1179-1188.
- [11] Qi L, Huo J, Wang L, et al. A mask based deep ranking neural network for person retrieval[C]//2019 IEEE International Conference on Multimedia and Expo, July 8-12, 2019, Shanghai, China. New York: IEEE Press, 2019: 496-501.
- [12] Zheng L, Shen L Y, Tian L, et al. Scalable person re-identification: a benchmark[C]//2015 IEEE International Conference on Computer Vision, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1116-1124.
- [13] Ristani E, Solera F, Zou R, et al. Performance measures and a data set for multi-target, multi-camera tracking [M]//Hua G, Jégou H. Computer vision-ECCV 2016 workshops. Lecture notes in computer science. Cham: Springer, 2016, 9914: 17-35.
- [14] Li W, Zhao R, Xiao T, et al. DeepReID: deep filter pairing neural network for person re-identification[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 152-159.
- [15] Wei L H, Zhang S L, Gao W, et al. Person transfer GAN to bridge domain gap for person re-identification [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 79-88.
- [16] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [17] Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model[C]//2008 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2008, Anchorage, AK, USA. New York: IEEE Press, 2008.
- [18] 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.
- [19] Tay C P, Roy S, Yap K H. AANet: attribute attention

- network for person re-identifications[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 7127-7136.
- [20] Zheng F, Deng C, Sun X, et al. Pyramidal person re-identification via multi-loss dynamic training[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 8506-8514.
- [21] Wang G S, Yuan Y F, Chen X, et al. Learning discriminative features with multiple granularities for person re-identification[C]//MM '18: Proceedings of the 26th ACM international conference on Multimedia, October 22-26, 2018, Seoul, Republic of Korea. New York: ACM, 2018: 274-282.
- [22] Zhang Z Z, Lan C L, Zeng W J, et al. Densely semantically aligned person re-identification[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 667-676.
- [23] Yang W J, Huang H J, Zhang Z, et al. Towards rich feature discovery with class activation maps augmentation for person re-identification[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 1389-1398.
- [24] Zhou K Y, Yang Y X, Cavallaro A, et al. Omni-scale feature learning for person re-identification[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 3701-3711.
- [25] Chen B H, Deng W H, Hu J N. Mixed high-order attention network for person re-identification[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 371-381.
- [26] He L X, Wang Y G, Liu W, et al. Foreground-aware pyramid reconstruction for alignment-free occluded person re-identification[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 8449-8458.
- [27] Quan R J, Dong X Y, Wu Y, et al. Auto-ReID: searching for a part-aware ConvNet for person re-identification[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 3749-3758.
- [28] Hou R B, Ma B P, Chang H, et al. Interaction-and-aggregation network for person re-identification[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 9309-9318.
- [29] Zheng Z D, Yang X D, Yu Z D, et al. Joint discriminative and generative learning for person re-identification[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 2133-2142.