·开发研究与工程应用·　　文章编号：1000-3428(2022)12-0281-07　　文献标志码：A　　中图分类号：TP391

面向行人重识别的通道与空间双重注意力网络

曾　涛，薛　峰，杨　添

（合肥工业大学 计算机与信息学院，合肥 230601）

摘　要：针对现实场景下因受到摄像机视角变化、行人姿态变化、物体遮挡、图像低分辨率、行人图片未对齐等因素影响导致行人判别性特征难以获取的问题，设计混合池通道注意模块（HPCAM）和全像素空间注意力模块（FPSAM），并基于这两种注意力模块提出一种通道与空间双重注意力网络（CSDA-Net）。HPCAM模块能够在通道维度上抑制无用信息的干扰，增强显著性特征的表达，以提取得到判别性强的行人特征。FPSAM模块在空间维度上增强行人特征的判别能力，从而提高行人重识别的准确率。通过在传统行人重识别深度模型框架中分阶段融入HPCAM模块和FPSAM模块，获得由粗糙到细粒度的注意力特征。实验结果表明，CSDA-Net网络在行人重识别主流数据集CUHK03、DukeMTMC-ReID和Market1501上的Rank-1准确率分别为78.3%、91.3%和96.0%，平均精度均值(mAP)分别为80.0%、82.1%和90.4%，与MGN网络相比，Rank-1准确率分别提升14.9、2.6和0.3个百分点，mAP分别提升13.7、3.7和3.5个百分点，能够提取更具鲁棒性和判别性的表达特征。

关键词：行人重识别；双重注意力机制；行人特征；深度学习；平均精度均值

开放科学（资源服务）标志码（OSID）：

中文引用格式：曾涛，薛峰，杨添. 面向行人重识别的通道与空间双重注意力网络[J]. 计算机工程，2022，48(12)：281-287，295.

英文引用格式：ZENG T，XUE F，YANG T，et al. Channel and spatial dual-attention network for person re-identification[J]. Computer Engineering，2022，48(12)：281-287，295.

Channel and Spatial Dual-Attention Network for Person Re-Identification

ZENG Tao，XUE Feng，YANG Tian

（School of Computer and Information，Hefei University of Technology，Hefei 230601，China）

【Abstract】To address the challenge in obtaining the discriminative features of pedestrians in actual scenes due to changes in camera angle，pedestrian postures，object occlusions，low image resolutions，and misaligned pedestrian images，a Hybrid Pooling Channel Attention Module（HPCAM）and a Full Pixel Spatial Attention Module（FPSAM）are designed. Based on these two attention modules，a Channel and Spatial Dual-Attention Network（CSDA-Net）is proposed. The HPCAM module suppresses the interference of meaningless information and enhances the expression of salient features in the channel dimension to extract highly discriminative pedestrian features.The FPSAM module enhances the discrimination ability of pedestrian features in the spatial dimension and then improves the accuracy of person Re-Identification（ReID）. By integrating the HPCAM and FPSAM modules into the traditional person ReID depth model framework in stages，attention features ranging from rough to fine ones are obtained. Experimental results show that the Rank-1 accuracies of CSDA-Net on mainstream datasets CUHK03，DukeMTMC-ReID，and Market1501 in the field of person ReID are 78.3%，91.3%，and 96.0%，respectively，and its mean Average Precision（mAP）values are 80.0%，82.1%，and 90.4%，respectively. Compared with MGN networks，the three networks mentioned above show higher Rank-1 accuracies by 14.9，2.6，and 0.3 percentage points，respectively，and higher mAP values by 13.7，3.7，and 3.5 percentage points，respectively. This indicates that the CSDA-Net can extract more robust and discriminative expression features.

【Key words】person Re-Identification（ReID）；dual-attention mechanism；pedestrian features；deep learning；mean Average Precision（mAP）

DOI：10. 19678/j. issn. 1000-3428. 0063136

基金项目：国家自然科学基金（61772170）。

作者简介：曾　涛（1997—），男，硕士研究生，主研方向为行人重识别；薛　峰（通信作者），教授、博士；杨　添，硕士研究生。

收稿日期：2021-11-04　　**修回日期：**2021-12-14　　**E-mail：**feng.xue@hfut.edu.cn

0 概述

随着智能安防和视频监控领域的需求与日俱增,行人重识别(Re-Identification,ReID)受到了越来越多研究人员的关注[1-2]。行人重识别可以看成一个图片检索任务,利用计算机视觉技术判断给定的图片或视频序列中是否存在特定行人,即给定一张待识别的行人图片,在其它摄像头拍摄到的视频中检索出与待识别行人具有相同身份的行人图片。在现实场景中,视角变化、行人姿态变化、物体遮挡、图像低分辨率等不利因素[3]导致行人重识别算法难以提取充分、有效的行人特征,造成行人重识别精度较低。因此,如何提取判别性强的行人特征是行人重识别研究的重点。

在深度学习技术普及之前,与大部分图像分类识别一样,行人重识别主要通过手工设计图像特征来实现,计算过程繁琐,识别效果较差,难以满足复杂环境变化下行人重识别任务的要求。近年来,基于深度学习的行人识别技术取得了较大的进展和突破,其识别准确率较基于人工特征的方法有了大幅提高,成为行人重识别领域的主流方法[4]。按照行人特征表达方式的不同,可以将基于深度学习的行人重识别方法概括为基于全局特征、基于局部特征和基于注意力机制3种方法。

基于全局特征的行人重识别将整张行人图像表示成一个不包含任何空间信息的特征向量,并对图像进行相似性度量。文献[5]提出一个身份判别性编码方法,将行人重识别问题看成分类问题,即将同一个行人的图片看成同一类的图片,在训练过程中利用行人的ID标签计算分类损失。文献[6]使用三元组损失来训练深度模型,通过拉近同一类样本之间的距离,拉开不同类样本之间的距离,从而获得判别性强的行人特征。文献[7]提出一个融合分类损失和验证损失的孪生网络,在模型训练过程中学习行人特征表示和相似度度量。然而,基于全局特征的方法忽略了图像局部细节信息,在物体遮挡、行人图片未对齐、视角变化等复杂场景下,辨别能力较差。

基于局部特征的行人重识别则先将整张图像表示成若干局部特征的集合,再进行度量学习,这类方法可通过图像切片、分割、人体骨架关键点检测等方式实现。文献[8]提出一种基于分块卷积基线方法提取局部特征,使用分块修正池化方法对齐分块信息,并针对每个分块信息分别采用分类损失进行训练。文献[9]提出一种水平金字塔匹配方法,通过在水平方向提取多个尺度的局部特征并分别进行训练,从而增强行人局部特征的鲁棒性和判别能力。文献[10]提出一种以人体区域划分的多阶段特征提取和融合方法,从而对齐不同图像中人体区域特征。但这类方法通常仅考虑粗粒度的局部特征,缺乏对行人全局特征的统筹考虑,因此难以获得辨别力强的行人特征。

基于注意力机制的行人重识别方法是近年来行人重识别领域出现的新方法。文献[11]提出一个端到端的比较性注意力网络,以长短时记忆网络为基础设计了注意力机制,并加入时空信息模拟人类的感知过程来比较行人之间的显著性区域,判断两幅照片是否属于同一个行人。文献[12]通过将注意力机制融入孪生网络中,发现具有相同身份的行人图像中的一致性注意力区域,在跨视角的匹配中有较强的鲁棒性。文献[13]提出一个轻量级的注意力网络,能够联合学习行人图片中的硬区域注意力和软像素注意力,优化未对准图像中的行人识别。文献[14]提出一种批丢弃块网络,通过在一个批次中随机丢弃所有输入特征图的相同区域,加强局部区域的注意力特征学习,从而获得更加鲁棒和判别性高的行人特征。文献[15]提出一个类激活图增强模型,通过在主干模型后面连接一系列有序分支进行扩展,并通过引入一个重叠激活惩罚的新损失函数,使当前分支更多地关注那些被先前分支较少激活的图像区域,从而获得多种辨别力的细粒度行人特征。然而,目前这些基于注意力机制的方法主要考虑行人局部区域的注意力特征,缺乏对行人图片不同区域细粒度注意力特征的关注,特征粒度不够精细,特征表达的辨别能力有待进一步增强。

本文提出一种通道与空间双重注意力网络(Channel and Spatial Dual-Attention Network,CSDA-Net),通过改进设计一种混合池通道注意力模块(Hybrid Pooling Channel Attention Module,HPCAM),在通道维度上获得更加具有判别性的特征,并在HPCAM模块之后设计一种新型像素级细粒度的全像素空间注意力模块(Full Pixel Spatial Attention Module,FPSAM),从而在空间维度上增强行人特征的判别能力。

1 通道和空间双重注意力网络

1.1 网络结构

受文献[16]中bagTricks模型的启发,本文设计了一种通道和空间双重注意力网络CSDA-Net,其结构如图1所示。可以看到,CSDA-Net网络在骨干网络的第1个、第2个、第3个残差块后引入HPCAM模块,在骨干网络的第4个残差块后引入FPSAM模块。相比于批归一化(Batch Normalization,BN)模块,实例批归一化(Instance Batch Normalization,IBN)模块[17]对一个批次里的每一个样本均进行正则化处理,而不是对整个批次样本进行正则化处理。研究表明IBN泛化能力较强[18],更加适合处理在复杂环境下拍摄的行人图片。因此,CSDA-Net骨干网络采用ResNet50-IBN网络。

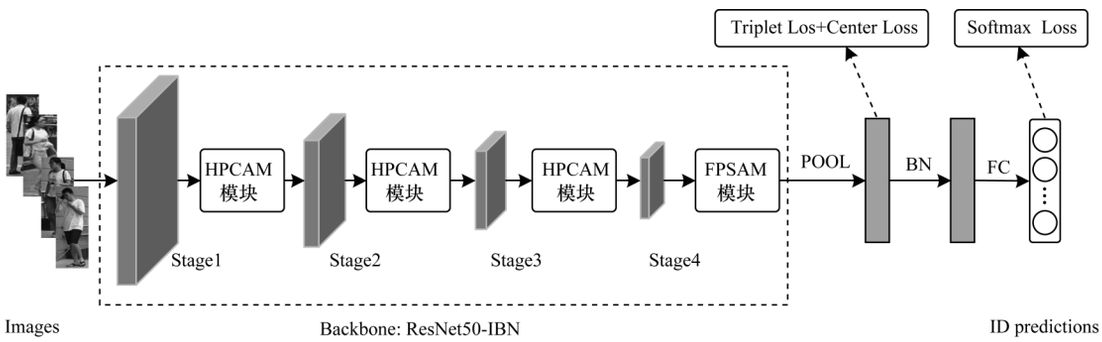

图1　CSDA-Net结构

Fig.1　Structure of CSDA-Net

1.2　混合池通道注意模块

通道注意力机制可以在引入少量参数的情况下,在通道维度上抑制无用信息的干扰及增强显著性特征的表达,从而提高行人重识别的准确度和精度。为了在通道维度上获得更加具有判别性的行人特征,本文在传统通道注意力网络机构上进行改进,设计了混合池通道注意力模块HPCAM,其结构如图2所示。本文的HPCAM模块在通道域注意力(Squeeze and Excitation SE)模块[19]的全局均值池化(Global Average Pooling,GAP)单分支结构设计的基础上,加入全局最大池化(Global Max Pooling,GMP)分支结构,即通过全局均值池化操作混合全局最大池化操作,进一步挖掘通道维度上的显著性特征,从而提取到更加具有判别性的特征。

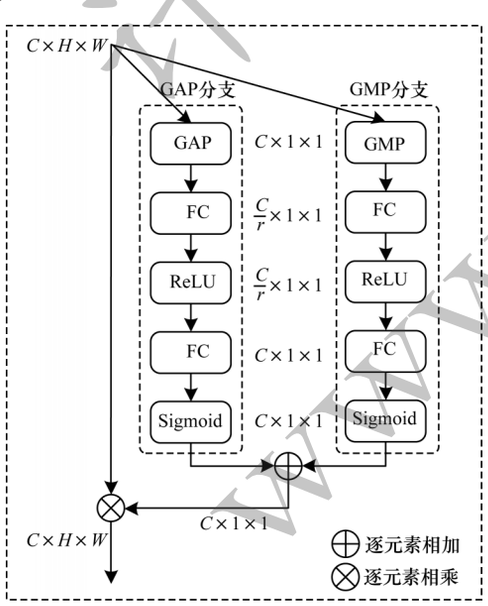

图2　HPCAM模块

Fig.2　HPCAM module

如图2所示,HPCAM模块的输入是一个三维向量$\boldsymbol{X}^{CA}\in\mathbb{R}^{C\times H\times W}$,其中$W$、$H$、$C$分别表示特征图的宽度、高度和通道个数。HPCAM模块旨在学习得到一个在通道维度上的权重偏好向量$\boldsymbol{A}_c\in\mathbb{R}^{C\times 1\times 1}$,用于捕获特征在通道维度上的显著性信息。HPCAM模块的输出计算过程如式(1)所示:

$$\tilde{\boldsymbol{X}}^{CA}=\boldsymbol{X}^{CA}\times\left(1+\boldsymbol{A}_c\right) \tag{1}$$

由式(1)可以看出,当权重偏好向量$\boldsymbol{A}_c$为0时,式(1)变成恒等映射,能有效避免随着深度网络层数的加深而出现梯度消失和网络退化的情况,有助于深度模型的训练学习,挖掘出在通道维度上的行人显著性特征。其中权重偏好向量$\boldsymbol{A}_c$是通过GAP分支和GMP分支共同学习得到,其表达式如式(2)所示:

$$\boldsymbol{A}_c=\boldsymbol{A}_{ca}+\boldsymbol{A}_{cm} \tag{2}$$

在GAP分支中,$\boldsymbol{A}_{ca}$的计算式如式(3)所示:

$$\boldsymbol{A}_{ca}=\sigma\left(\boldsymbol{W}_{a2}\times\delta\left(\boldsymbol{W}_{a1}\times\mathrm{GAP}\left(\boldsymbol{X}^{CA}\right)\right)\right) \tag{3}$$

其中:$\sigma(\cdot)$表示Sigmoid函数;$\delta(\cdot)$表示ReLU激活函数;$\boldsymbol{W}_{a1}$表示GAP分支中第1个全连接层中权重向量,其特征维度为$(C/r)\times1\times1$,其中r表示降维比例因子,本文中r取16;$\boldsymbol{W}_{a2}$表示第2个全连接层中权重向量,其特征维度为$C\times1\times1$;GAP$(\cdot)$表示全局均值池化操作。

在GMP分支中,$\boldsymbol{A}_{cm}$的计算式如式(4)所示:

$$\boldsymbol{A}_{cm}=\sigma\left(\boldsymbol{W}_{m2}\times\delta\left(\boldsymbol{W}_{m1}\times\mathrm{GMP}\left(\boldsymbol{X}^{CA}\right)\right)\right) \tag{4}$$

其中:$\boldsymbol{W}_{m1}$表示GMP分支中第1个全连接层的权重向量;$\boldsymbol{W}_{m2}$表示第2个全连接层的权重向量;GMP$(\cdot)$表示全局最大池化操作。

1.3　全像素空间注意力模块

为进一步在空间维度上增强行人特征的判别能力,本文在HPCAM模块之后加入了全像素空间注意力模块FPSAM,其结构组成如图3所示。FPSAM模块学习到的是一个和输入特征图维度一致的全像素细粒度的注意力权重向量,能够在HPCAM模块获得粗粒度的注意力特征基础上,进一步获得更加具有判别性的行人特征,提高行人重识别的准确率和精度。

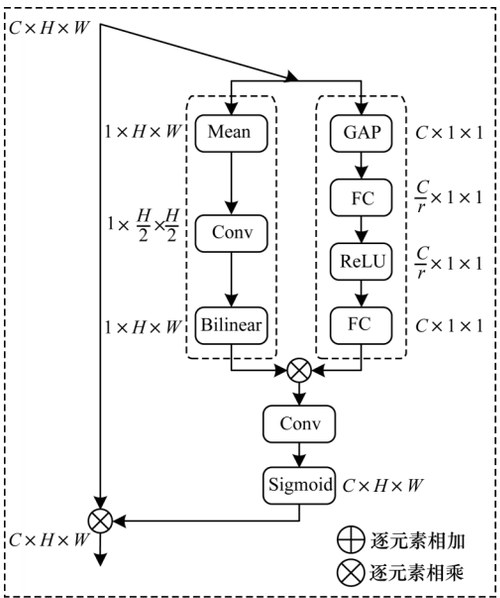

图3 FPSAM模块

Fig.3 FPSAM module

FPSAM模块的输入是一个三维向量$\boldsymbol{X}^{PA}\in\mathbb{R}^{C\times H\times W}$，其中$W$、$H$、$C$分别表示特征图的宽度、高度和通道个数。FPSAM模块旨在学习得到一个和输入维度相同的权重向量$\boldsymbol{A}_s\in\mathbb{R}^{C\times H\times W}$。FPSAM模块的输出为输入向量$\boldsymbol{X}^{PA}$和权重偏好向量$\boldsymbol{A}_s$的乘积，其计算式如式(5)所示：

$$\tilde{\boldsymbol{X}}^{PA}=\boldsymbol{X}^{PA}\times\boldsymbol{A}_s \tag{5}$$

其中：权重向量$\boldsymbol{A}_s$是基于通道注意力权重向量$\boldsymbol{A}_{pc}\in\mathbb{R}^{C\times1\times1}$和空间注意力权重向量$\boldsymbol{A}_{ps}\in\mathbb{R}^{1\times H\times W}$计算所得，向量$\boldsymbol{A}_{pc}$和向量$\boldsymbol{A}_{ps}$的学习过程如图3所示，其中，通道注意力权重向量$\boldsymbol{A}_{pc}$的学习过程如图3中左虚线框所示。$\boldsymbol{A}_{pc}$的计算式如式(6)所示：

$$\boldsymbol{A}_{pc}=\sigma\left(\boldsymbol{W}_{pa2}\times\delta\left(\boldsymbol{W}_{pa1}\times\mathrm{GAP}\left(\boldsymbol{X}^{PA}\right)\right)\right) \tag{6}$$

其中：$\boldsymbol{W}_{pa1}$表示第1个全连接层中的权重向量；$\boldsymbol{W}_{pa2}$表示第2个全连接层中的权重向量。

空间注意力权重向量$\boldsymbol{A}_{ps}$的学习过程如图3中右虚线框所示。具体而言，首先在通道维度上求均值，然后通过一个卷积核为3×3、步长为2的卷积层，最后通过一个双线性插值层还原得到特征图$\boldsymbol{A}_{ps}$。基于向量$\boldsymbol{A}_{pc}$和向量$\boldsymbol{A}_{ps}$即可得到像素注意力权重向量$\boldsymbol{A}_s$。具体而言，向量$\boldsymbol{A}_{pc}$和向量$\boldsymbol{A}_{ps}$首先进行逐元素相乘，然后再经过一个卷积核为1×1、步长为1的卷积层，最后经过Sigmoid激活函数进行归一化处理。向量$\boldsymbol{A}_s$的计算式如式(7)所示：

$$\boldsymbol{A}_s=\sigma\left(\mathrm{Conv}\left(\boldsymbol{A}_{pc}\times\boldsymbol{A}_{ps}\right)\right) \tag{7}$$

其中：Conv(·)表示卷积操作。因为$\boldsymbol{A}_{pc}$和$\boldsymbol{A}_{ps}$是通过2个不同的分支独立学习得到的，两者直接相乘得到的权重向量存在信息冗余，所以加入1×1的卷积层后可以修正权重向量，得到更加具有判别性的特征。

1.4 损失函数

为了使模型能更加有效地提取出辨别性强的行人特征，本文采用交叉熵损失函数、三元组损失函数和中心损失函数进行联合训练。交叉熵损失结合三元组损失和中心损失进行联合训练可以获取更加有效的行人特征，这也是行人重识别研究领域常用的方式。

1.4.1 交叉熵损失函数

设进行分类识别的特征为f，交叉熵损失函数的计算式如式(8)所示：

$$\mathcal{L}_{\mathrm{softmax}}=-\sum_{i=1}^{N}q_i\log_a\frac{\mathrm{e}^{\boldsymbol{w}_i^{\mathrm{T}}f}}{\sum_{k=1}^{N}\mathrm{e}^{\boldsymbol{w}_k^{\mathrm{T}}f}} \tag{8}$$

其中：N为训练集中类别总数；$\boldsymbol{W}_i$表示全连接层中第i个类别的权重向量；y是输入图像的真实标签，当$y\neq i$时，$q_i=0$，当$y=i$时，$q_i=1$。为防止模型对训练集中行人图片过拟合，提高模型的泛化能力，本文采用了带标签平滑的交叉熵损失函数q_i，其表达式为：

$$q_i=\begin{cases}\dfrac{\varepsilon}{N}, & y\neq i\\ 1-\dfrac{N-1}{N}\varepsilon, & y=i\end{cases} \tag{9}$$

其中：ε是标签平滑参数，其通过抑制真实标签在计算损失时的权重，从而抑制模型在数据集上过拟合，提高模型泛化能力。在本文中，ε设置为0.1。

1.4.2 三元组损失函数

本文采用的难样本三元组损失函数是三元组损失函数的一个改进版本。难样本三元组损失函数可以表示为式(10)所示：

$$\mathcal{L}_{\mathrm{triplet}}=-\sum_{i=1}^{P}\sum_{a=1}^{K}[\alpha+\max_{p=1,2,\cdots,K}\left\|f_a^{(i)}-f_p^{(i)}\right\|_2-\min_{\substack{n=1,2,\cdots,K\\ j=1,2,\cdots,P\\ j\neq i}}\left\|f_a^{(i)}-f_n^{(j)}\right\|_2]_+ \tag{10}$$

其中：P表示一个批次数据中行人ID个数；K表示每个行人挑选出图片的个数；$f_a^{(i)}$表示批次中一张ID为i的行人图片(anchor)的特征；$f_p^{(i)}$表示与其ID相同的正样本特征；$f_n^{(i)}$表示ID不同的负样本特征；α为预设的超参阈值，用来调整正负样本对之间的距离，本文中，α设置为0.6；$[\cdot]_+$表示max(·,0)函数。

1.4.3 中心损失函数

中心损失函数通过为每个类别学习得到一个特征中心点，并在训练过程中不断拉近深度特征和其对应的特征中心之间的距离，从而使类内特征更加紧凑，学习得到更加鲁棒的判别性特征。中心损失函数的表达式如式(11)所示：

$$\mathcal{L}_{\mathrm{center}}=\frac{1}{2}\sum_{i=1}^{B}\left\|f_i-c_{y_i}\right\|_2^2 \tag{11}$$

其中：f_i表示第i个样本经过深度网络后提取得到的特征；y_i表示第i个样本的标签；c_{yi}表示第y_i个类别对应的高维特征中心；B表示批次大小。

1.4.4　本文损失函数

综上所述,本文使用3种损失函数进行联合训练,最终的损失函数表达式如式(12)所示:

$$\mathcal{L}_{\text{total}}=\mathcal{L}_{\text{softmax}}+\mathcal{L}_{\text{triplet}}+\beta\mathcal{L}_{\text{center}} \tag{12}$$

其中:β表示中心损失对应的权重系数,在本文中,β默认设置为0.000 5。

2　实验结果与分析

2.1　数据集介绍

为证明本文网络的有效性,本文在行人重识别领域公开的3个大型数据集Market1501[20]、DukeMTMC-ReID[21]和CUHK03[22]上进行了实验,数据集的属性信息如表1所示。

表1　数据集属性信息

Table 1　Attribute information of dataset

数据集	图片数	行人数	摄像机数	年份
Market1501	32 668	1 501	6	2015
DukeMTMC-ReID	16 522	1 404	8	2017
CUHK03	13 164	1 467	10	2014

2.2　实验参数与评价指标

本文使用Pytorch深度学习框架,并采用2块GeForce GTX 1080ti显卡进行GPU加速。在训练迭代过程中,每次随机挑选16个行人,每个行人挑选4张图片来构成一个批次,且图片大小统一调整为384×192像素。在训练过程中,学习率采用预热策略,即在前10个epoch学习率由3.5×10^{-5}线性增加到3.5×10^{-4},然后在第50个、第140个和第240个epoch时学习率进行系数为0.1的指数衰减,训练总次数为400个epoch。训练过程中采用Adam优化器算法对模型参数进行优化,并使用权重衰减因子为5×10^{-4}的L_2正则化。测试时,使用BN层之后的特征作为行人检索特征,并采用余弦距离度量方式计算特征之间的距离。

本文使用累积匹配特性曲线(Cumulative Match Characteristic Curve,CMC)中的Rank-1准确率和平均精度均值(mean Average Precision,mAP)作为评估模型性能的指标。

2.3　与相关网络的比较

为证明本文网络性能,本文与近几年行人重识别领域中一些具有代表性的网络进行比较,如基于全局特征的IDE[5]、TriNet[6]、SVDNet[23]、bagTricks[16],基于局部特征的PCB[8]、PCB+RPP[8]、HPM[9]、MGN[24],以及基于注意力机制的HA-CNN[13]、CASN[11]、CAMA[15]、BDB[14]。为简化实验和直观地分析网络本身的有效性,本文所有实验均采用单帧查询模式,且未使用re-ranking[25]技术。实验结果如表2和表3所示,其中"—"表示原文献中未列出该实验结果。

表2　不同网络在Market1501和DukeMTMC-ReID数据集下的结果对比

Table 2　Comparison of results of different networks under Market1501 and DukeMTMC-ReID datasets　%

网络	Market1501数据集		DukeMTMC-ReID数据集	
	Rank-1	mAP	Rank-1	mAP
IDE[5]网络	72.5	46.0	65.2	44.9
TriNet[6]网络	84.9	69.1	72.4	53.5
PCB[8]网络	92.4	77.4	81.9	66.1
PCB+RPP[8]网络	93.8	81.6	83.3	69.2
HPM[9]网络	94.2	82.7	86.6	74.3
CASN[11]网络	94.4	82.8	87.7	73.7
HA-CNN[13]网络	91.2	75.7	80.5	63.8
BDB[14]网络	95.3	86.7	89.0	76.0
CAMA[15]网络	94.7	84.5	85.8	72.9
BagTricks[16]网络	94.5	85.9	86.4	76.4
SVDNet[23]网络	82.3	62.1	76.7	56.8
MGN[24]网络	95.7	86.9	88.7	78.4
本文网络	96.0	90.4	91.3	82.1

表3　不同网络在CUHK03数据集下的结果对比

Table 3　Comparison of results of different networks under CUHK03 dataset　%

网络	手工标记		检测器检测	
	Rank-1	mAP	Rank-1	mAP
PCB[8]网络	—	—	63.7	57.5
PCB+RPP[8]网络	—	—	61.3	54.2
HPM[9]网络	—	—	63.9	57.5
CASN[11]网络	73.7	68.0	71.5	64.4
HA-CNN[13]网络	44.4	41.0	41.7	38.6
CAMA[15]网络	70.1	66.5	66.6	64.2
MGN[24]网络	68.0	67.4	68.0	66.0
本文网络	82.9	81.1	78.3	80.0

由表2可知,本文CSDA-Net网络在Market1501数据集上的Rank-1和mAP分别为96.0%和90.4%,两者同时达到了对比网络中的最高精度。相比于同样基于注意力机制的CAMA[15]和BDB[14],本文CSDA-Net网络在Rank-1上分别提升了1.3和0.7个百分点,在mAP上分别提升了5.9和3.7个百分点。CAMA和BDB网络由于仅考虑的是局部区域粗粒度的注意力特征,缺乏对行人像素级细粒度注意力特征的统筹考虑,因此获取的行人判别性特征不够准确。相比于使用全局特征的bagTricks[16]和使用局部特征的MGN[24]网络,本文的CSDA-Net在Rank-1上分别提升了1.5和0.3个百分点,在mAP上分别提升了4.5和3.5个百分点。未使用注意力机制的bagTricks和MGN网络提取得到的是一个全局的特征,缺乏对判别性特征的关注,故而较难获得辨别力强的行人特征。上述

实验结果验证了本文CSDA-Net网络的有效性。

由表2可知，本文CSDA-Net网络在DukeMTMC-ReID数据集上的Rank-1和mAP分别为91.3%和82.1%，两者同时达到了对比网络中的最高精度。相比于同样基于注意力机制的CAMA[15]和BDB[14]网络，本文CSDA-Net网络在Rank-1上分别提升了5.5和2.3个百分点，在mAP上分别提升了9.2和6.1个百分点。相比于使用全局特征的bagTricks[16]网络，本文CSDA-Net网络在Rank-1和mAP分别提高了4.9和5.7个百分点。相比于使用局部特征的MGN[24]网络，本文CSDA-Net网络在Rank-1方面提高了2.6个百分点，在mAP上提高了3.7个百分点。在DukeMTMC-ReID数据集上的实验结果同样证明了本文CSDA-Net网络的有效性。

不同网络在CUHK03数据集上的对比结果如表3所示，可以看到，本文CSDA-Net网络的性能显著优于其他网络。因为该数据集中行人边界框由手工标记方式和检测器检测两种方式获得，所以分成两种情况进行验证。在使用手工标注方式获得的行人边界框下，本文CSDA-Net网络达到了81.1%的Rank-1精度和82.9%的mAP，两者同时达到了对比网络中最好的性能。相比于同样基于注意力机制的CAMA[15]和BDB[14]网络，本文CSDA-Net网络在Rank-1上分别提升了12.8和6.2个百分点，在mAP上分别提升了14.6和1.7个百分点。在使用检测器检测的方式下，本文CSDA-Net网络达到了78.3%的Rank-1和80.0%的mAP，性能同样优于其他的网络，相比于同样基于注意力机制的CAMA[15]和BDB[14]网络，本文CSDA-Net网络在Rank-1上分别提升了11.7和4.8个百分点，在mAP上分别提升了15.8和3.6个百分点。在CUHK03数据集上的实验结果进一步验证了本文CSDA-Net网络的有效性。

由表2、表3可以看出，本文CSDA-Net网络在DukeMTMC-ReID和CUHK03数据集上性能提升幅度明显高于在Market1501数据集上的性能提升幅度，这是由于Market1501数据集中的行人图片较为规整，且较少出现物体遮挡、视图变化、行人未对齐等不利情况，而DukeMTMC-ReID和CUHK03数据集具有较大挑战性，更符合真实场景下行人图片的特点，即图片来自于多个互不重叠的摄像头，普遍存在行人姿态变化、物体遮挡、行人未对齐等不利情况。实验结果论证了本文CSDA-Net网络能够有效应对上述不利情况，具有更好的鲁棒性，更适合真实场景下的行人重识别任务。

2.4 消融实验

为展现本文网络中2个主要创新改进模块（HPCAM模块和FPSAM模块）对算法性能提升的贡献，本文以Market1501数据集为例，设计了一系列消融实验。本文以未加入注意力模块作为基线（Baseline）模型，实验结果如表4所示。

表4 不同模块组合在Market1501数据集下的实验结果

Table 4 Experimental results of different module combinations on Market1501 dataset %

模块组合	Rank-1	mAP
Baseline (w/o HPCAM FPSAM)	95.34	89.63
Baseline + HPCAM	95.81	89.93
Baseline + FPSAM	95.78	90.29
Baseline + HPCAM + FPSAM	96.02	90.37

由表4实验结果可以看出，单独融入HPCAM模块或FPSAM模块均对模型效果有所提升，将HPCAM模块和FPSAM模块同时融入Baseline模型（即本文方法），Rank-1和mAP得到了进一步提升。上述实验表明，HPCAM和FPSAM两个注意力模块在“粒度”上具有互补作用，通过深度网络联合学习粗粒度和细粒度注意力特征，可以获得更具鲁棒性的行人特征，从而提高行人重识别的准确率和精度。

2.5 算法可视化结果分析

为进一步验证本文网络的先进性，本文随机检索了Market1501查询集中的3个行人，检索结果排名前10的图片如图4所示。其中，Query是待查询图像，1～10是按照相识度从大到小排列的10张检索结果正确的图像，分别表示为Rank-1~Rank-10。

图4 本文网络在Market1501数据集下排名前10的行人检索结果展示

Fig.4 Display of the top 10 pedestrian search results of network in this paper under the Market1501 dataset

从图4第1个检索示例中可以看出,当摄像机视角变化(Rank-1,Rank-4,Rank-8)和行人姿态变化(Rank-1,Rank-2,Rank-3)时,均可以被检索出来。从图4第2个检索示例中可以看出,即使是与待检索图像不对齐的行人图像(Rank-1,Rank-8)也可以被检索出来。从图4第3个检索示例中可以看出,即使是被自行车遮挡的行人图像(Rank-7)也可以被正确地检索出来。在图像低分辨率情况下,如第1个检索示例中的Rank-8和第3个检索示例中的Rank-6所示,也能被检索出来。上述实验结果验证了本文网络的有效性和鲁棒性,可以有效应对真实场景下的行人重识别问题。

3 结束语

针对在复杂环境下行人判别性特征难以获取的问题,本文提出一种面向行人重识别的通道与空间双重注意力网络CSDA-Net。通过在深度模型中分阶段融入HPCAM模块和FPSAM模块,获得不同粒度的注意力特征,并通过深度网络互补训练学习,有效挖掘行人判别性特征,提高行人重识别的准确率和精度。在CUHK03、DukeMTMC-ReID和Market1501公开数据集上的实验结果表明,本文网络能有效提高行人重识别性能。下一步将从实际应用的角度出发,在跨模态的情况下把不同模态的特征数据统一映射到一个共享的特征表征空间中,并同时约束该共享特征空间中的类内一致性和类间辨别性,使用深度模型训练得到与模态无关的行人判别性特征,提高行人重识别模型的泛化能力,从而提取具有鲁棒性和判别性的行人特征。

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]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2022,44(6):2872-2893.

[2] 吴彦丞,陈鸿昶,李邵梅,等. 基于行人属性异质性的行人再识别神经网络模型[J]. 计算机工程,2018,44(10):196-203.
WU Y C,CHEN H C,LI S M,et al. Pedestrian re-identification neural network model based on pedestrian attribute heterogeneity[J]. Computer Engineering,2018,44(10):196-203. (in Chinese)

[3] LI R,ZHANG B P,TENG Z,et al. A divide-and-unite deep network for person re-identification[J]. Applied Intelligence,2021,51(3):1479-1491.

[4] 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. (in Chinese)

[5] ZHENG L,YANG Y,HAUPTMANN A G. Personre-identification:past,present and future[EB/OL]. [2021-10-09]. https://arxiv.org/abs/1610.02984v1.

[6] HERMANS A,BEYER L,LEIBE B. In defense of the triplet loss for person re-identification[EB/OL]. [2021-10-09]. https://arxiv.org/abs/1703.07737.

[7] ZHENG Z D,ZHENG L,YANG Y. A discriminatively learned CNN embedding for person reidentification[J]. ACM Transactions on Multimedia Computing,Communications,and Applications,2018,14(1):13-21.

[8] SUN Y F,ZHENG L,YANG Y,et al. Beyond part models:person retrieval with refined part pooling (and a strong convolutional baseline). [EB/OL]. [202110-09]. https://arxiv.org/abs/1711.09349.

[9] FU Y,WEI Y C,ZHOU Y Q,et al. Horizontal pyramid matching for person re-identification[C]//Proceedings of AAAI Conference on Artificial Intelligence. Hawaii,USA:AAAI Press,2019:8295-8302.

[10] ZHAO H Y,TIAN M Q,SUN S Y,et al. Spindle net:person re-identification with human body region guided feature decomposition and fusion [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C.,USA:IEEE Press,2017:907-915.

[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] ZHENG M,KARANAM S,WU Z Y,et al. Re-identification with consistent attentive siamese networks[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C.,USA:IEEE Press,2019:5728-5737.

[13] LI W,ZHU X T,GONG S G. Harmonious attention network for person re-identification[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C.,USA:IEEE Press,2018:2285-2294.

[14] DAI Z Z,CHEN M Q,GU X D,et al. Batch dropblock network for person re-identification and beyond[C]//Proceedings of IEEE/CVF International Conference on Computer Vision. Washington D. C.,USA:IEEE Press,2019:3690-3700.

[15] YANG W J,HUANG H J,ZHANG Z,et al. Towards rich feature discovery with class activation maps augmentation for person re-identification[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C.,USA:IEEE Press,2019:1389-1398.

[16] LUO H,GU Y Z,LIAO X Y,et al. Bag of tricks and a strong baseline for deep person re-identification [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Washington D. C.,USA:IEEE Press,2019:1487-1495.

[17] ULYANOV D,VEDALDI A,LEMPITSKY V. Instance normalization:the missing ingredient for fast stylization [EB/OL]. [2021-10-09]. https://arxiv.org/abs/1607.08022.

[18] PAN X G,LUO P,SHI J P,et al. Two at once:enhancing learning and generalization capacities via IBN-Net[C]//Proceedings of the European Conference on Computer Vision. Berlin,Germany:Springer,2018:464-479.

[19] HU J,SHEN L,ALBANIE S,et al. Squeeze-and-excitation networks[C]//Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence. Washington D. C.,USA:IEEE Press,2018:2011-2023.

[20] ZHENG L,SHEN L,LU T,et al. Scalable person re-identification:a benchmark[C]//Proceedings of 2015 IEEE International Conference on Computer Vision. Washington D. C.,USA:IEEE Press,2015:1116-1124.

(下转第295页)

（上接第287页）

[21] RISTANI E, SOLERA F, ZOU R, et al. Performance measures and a data set for multi-target, multi-camera tracking [C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 17-35.

[22] LI W, ZHAO R, XIAO T, et al. DeepReID: deep filter pairing neural network for person re-identification [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2014: 152-159.

[23] SUN Y F, ZHENG L, DENG W J, et al. SVDnet for pedestrian retrieval [C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 3800-3808.

[24] WANG G S, YUAN Y F, CHEN X, et al. Learning discriminative features with multiple granularities for person re-identification [C]//Proceedings of the 26th ACM International Conference on Multimedia. New York, USA: ACM Press, 2018: 274-282.

[25] ZHONG Z, ZHENG L, CAO D L, et al. Re-ranking person re-identification with k-reciprocal encoding [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 3652-3661.

编辑　赖玉玲