适用于无监督行人重识别的反向骨干网

于 鹏，朴 燕

（长春理工大学 电子信息工程学院，长春 130022）

摘 要：行人重识别（re-id）的目的是在不同的摄像机上识别同一个人的图像。虽然无监督模型比有监督模型有更好的泛用性，但无监督的聚类会更容易受到噪声干扰。针对这一问题，本文提出了一个可以减少噪声干扰的模型反向骨干网（RBNet），利用反向骨干网学习姿态检测模型输出的人体关键点，调整局部空间信息并生成掩码，用生成掩码增强指定位置注意力。实验结果表明：对比baseline在Market-1501到DukeMTMC-reID的跨域实验结果，mAP提升了7.0%，Rank-1提升了6.4%。强化对不同局部信息的注意力，可有效提升模型准确率。

关键词：人工智能；行人重识别；人体关键点；无监督域自适应

中图分类号：TP391.4 **文献标志码**：A **文章编号**：1671-5497（2024）11-3309-09

DOI：10.13229/j.cnki.jdxbgxb.20240916

Reverse backbone net for unsupervised person re-identification

YU Peng，PIAO Yan

（*School of Electronic and Information Engineering*，*Changchun University of Science and Technology*，*Changchun* 130022，*China*）

Abstract：The purpose of person re-identification（re-id）is to identify images of the same person on different cameras. Although unsupervised models have better generalization than supervised models，unsupervised clustering will be more susceptible to noise interference. To address this problem，this paper proposes a model reverse backbone net（RBNet）that can reduce noise interference，using RBNet to learn the keypoints of the human body output from the pose detection model，adjusting the local spatial information and generating masks，and augmenting the specified location attention with the generated masks. The experimental results show that comparing baseline′s cross-domain experimental results from Market-1501 to DukeMTMC-reID，mAP is enhanced by 7.0% and Rank-1 by 6.4%. Strengthening the attention to different local information can effectively improve the model accuracy.

Key words：artificial intelligence；person re-identification；human keypoints；unsupervised domain adaptation

收稿日期：2024-08-20.

基金项目：吉林省自然科学基金项目（20210101180JC）；吉林省科技厅科技发展计划项目（20180623039TC）.

作者简介：于鹏（1988-），男，博士研究生. 研究方向：机器视觉. E-mail：yup1212@mails.cust.edu.cn

通信作者：朴燕（1964-），女，教授，博士. 研究方向：三维成像及机器视觉. E-mail：piaoyan@cust.edu.cn

0 引 言

行人重识别(re-id)是给定一张人的图像后从不同时间、地点的摄像机所采集的图片中匹配特定人的检索任务,其在视频监控、智能安防等领域有着很重要的应用价值。随着深度学习领域的快速发展和大规模数据集的发布,基于深度神经网络的行人重识别方法取得了巨大的进展,但仍然存在挑战,比如如何减少背景杂乱、不同姿势等的干扰、跨域重识别。

目前,针对背景或姿势干扰等问题利用局部特征比较有效的研究如AANet[1]使用属性数据集[2]用GAP(Global average pooling)提取GFN(Global feature network)和PFN(Part feature network)获取局部特征。模型PCB[3]表明通过横向分割使得同一个特征图增加了关注点,从而提升了识别的准确率。Miao等[4]结合PCB和人体关键点检测的方法,进一步提升了PCB的准确率。使用图像分割的MMGA[5]在全局掩码和上下肢掩码去引导注意力也是有效的办法。多尺度和金字塔有助于局部特征的正确获取,如DetectoRS[6]使用递归特征金字塔获得局部特征。密集特征金字塔网络(DFPN)[7]针对人员重新识别任务容易受到尺度影响的问题,实现了无须预训练即可收敛到更好性能。RGA[8]使用注意力机制可以有效优化结果;OSNet[9]通过多尺度特征提升效果;Divyansh等[10]利用属性数据结合特征金字塔提取多尺度局部特征;Martinel等[11]利用特征金字塔整合局部特征和全局特征。

无监督学习有助于实现跨域重识别,无监督领域自适应UDA(Unsupervised domain adaptation)是将有标记的数据集视为源域(Source domain),未标记的数据集视为目标域(Target domain),通过从源域(Source domain)学习到的知识迁移到目标域(Target domain)以实现模型在真实环境下依然有良好的表现,但如Yang[12]所述,可以兼顾完全无监督的模型和无监督领域自适应的算法,完全无监督行人重识别学习实现跨域的结果并不比无监督领域自适应行人重识别好,因此,本文采用无监督域自适应的方法。基于无监督的平均教师方法(Mean teacher)[13]的MMT[14]利用更为健壮的软标签对伪标签进行在线优化,解决伪标签噪声从而增强跨域能力,该框架采用两个平行的网络起到一种相互监督的作用,避免单一网络的输出误差形成过拟合。

一个完整的行人重识别系统分为3个部分:行人检测、行人跟踪和行人检索。行人重识别系统使用目标检测模型获得人的图片,在系统中已经存在人体关键点,目标检测过程中的特征没有被充分利用。针对背景杂乱、不同姿势干扰等问题解决跨域重识别,本文提出了一个基于ResNet[15]的反向骨干网(RBNet),通过同时学习两种不同的任务实现为一个任务附加条件。通过融合多尺度特征掩码增强模型对特定区域的关注,在有监督训练学习过程中,解决了关键点特征不适配的问题。通过嵌入模型,实现有前提条件的无监督自适应学习模型。

1 本文方法

模型训练过程分3步(见图1):第一步获取特征;第二步基于重识别特征学习关键点特征;第三步利用反向模型构造的特征做掩码,减少非关键点特征的干扰,优化了原模型的特征提取。

1.1 模块结构

模型由普通的ResNet和一个反向尺寸的ResNet构成。本文通过扩大特征的覆盖范围和增强可解释性实现跨域的效果,采用较为常见的基础模型ResNet 50作为骨干网。首先,将图像重新调整为固定大小的图像,图像可以被打包成一批,然后输入网络中,用随机擦除等方法,使模型有效降低过度拟合和因遮挡而破裂的风险,传入第一步的图像做标准差归一化,第二步从主干获得的特征图经过全局平均池化(GAP)后,获得一个2 048维向量。

行人重识别模型应用于目标检测的下一步是在一帧视频里选出人的区域,利用已有的前置加工数据不会增加计算量,并且由于关键点检测模型和图像分割任务较为相似,所以前置的目标检测模型或姿态模型常常既可以做图像分割又可以检测人体关键点。例如,分割模型和姿态估计模型,姿态估计的重点和难点是定位人体关键点,可以使用姿态估计模型增强识别模型。本文选择用HRNet[16]获得的17个关键点特征作为条件引导学习。

在mmpose框架的模型库中,输入尺寸为256×192(与本文数据集的图片最为接近的尺

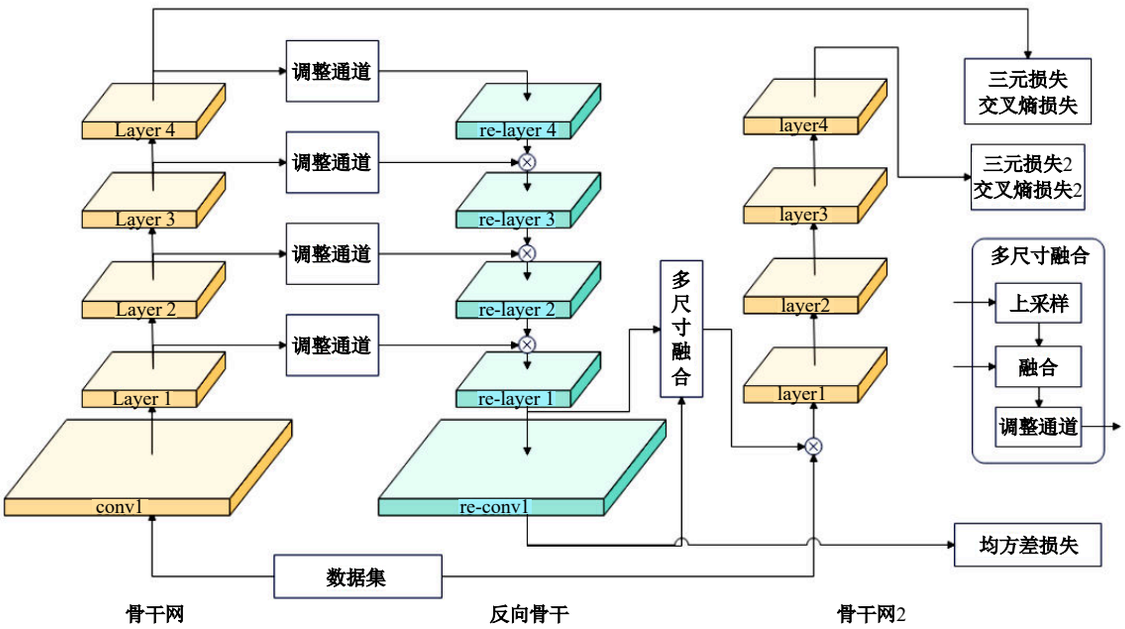

图 1　模型结构

Fig. 1　Model structure

寸)的关键点检测模型，AP(Average precision)排名前五的模型分别是 ViTPose-h 0.790、HRNet-w48+UDP 0.768、MSPN 4-stg 0.765、MSPN 4-stg 0.765 和 HRNet-w48 0.756 。虽然 VIT 排名第一,但和其他 4 个的差距不大,并且 HRNet 模型比 transformer 节省计算量,且本任务不需要获得精确到点的坐标,因此,本文选择 HRNet。

因为 HRNet 输出的人体关键点的范围比较小,不能直接适用于行人重识别,所以设计一个反向骨干(块)(Re-Backbone)用学习人体关键点解决这个问题。将骨干网 的 layer 1、layer 2、layer 3、layer4、输出的特征分别经过 change_chanel 模块传入 re_layer 1、re_layer 2、re_layer 3、re_layer 4,通过改变通道将原模型特征转化为分组特征，out 1 用于学习姿态特征和生成掩码。为了让有效的特征调整到一个通道,change_chane 1 用 1×1 卷积和 relu 实现。

使用 pose 模型定位关键点重识别模型特征对齐,但是关键点没有覆盖到全部特征,所以本文模型在骨干网后加入一个反向骨干块,人体关键点更适配重识别模型。反向骨干块的结构和 ResNet 类似,表 1 展示了原 ResNet 50 每层发生改变的地方,B、C、H、W 分别是批尺寸、通道数、高、宽,反向骨干块的特征变化情况与原骨干网一致,在每个 layer 的第 0 个块做池化,反向骨干块在与骨干网的池化层相同的位置做上采样恢复特征尺寸。即每个 re-layer 的最后一个输出块的位置上采样,最后输出通道设置为人体关键点个数 17,再将不同尺寸的特征融合用以减小图片中无关信息的影响。

表 1 模型参数

Table 1　Model parameters

block	orderedDict	I/O	B	C	H	W	B	C	H	W	I/O	orderedDict	block
conv 1	conv1	input	64	3	256	128	64	17	256	128	output	conv1	re-conv 1
		output	64	64	128	64	64	64	128	64	input		
	maxpool	input	64	64	128	64	64	64	128	64	output	upsample	
		output	64	64	64	32	64	64	64	32	input		
layer 1	0	input	64	64	64	32	64	64	64	32	output	2	re-layer 1
	0	output	64	256	64	32	64	256	64	32	input	2	
layer 2	0	input	64	256	64	32	64	256	64	32	output	3	re-layer 2
	0	output	64	512	32	16	64	512	32	16	input	3	
layer 3	0	input	64	512	32	16	64	512	32	16	output	5	re-layer 3
	0	output	64	1 024	16	8	64	1 024	16	8	input	5	
layer 4	0	input	64	1 024	16	8	64	1 024	16	8	output	2	re-layer 4
	0	output	64	2 048	16	8	64	2 048	16	8	input	2	

骨干网向反向骨干块传递的特征图使用1×1卷积对齐通道数。使用深度可分离卷积可在满足区分不同关键点的前提下减少计算量,深度可分离卷积需要一个全部特征过程,本模型直接用骨干网的同尺寸特征图引导。第三步将从ResNet传到反向骨干块的特征用分组卷积将其分为17组,用于空间注意力的权重。

第二步 反向骨干块上采样后融合ResNet传来的特征,由于反向骨干块的目的是让模型注意到人体关键点及临近像素,所以使用双线性插值上采样(Bilinear interpolation)得到损失值,最后特征融合时使用分组1×1卷积齐通道数并将上采样调整到和re-conv 1一致。经过训练测试,掩码用conv 1和layer 1融合比用所有层融合都好,为了减少分类关键点损失函数对分类ID损失函数,只在底层conv 1和layer 1层融合。

第三步 re-layer 1输出的特征图当作掩码再通过改变通道传入骨干网的conv 1输入,第三步利用生成的掩码让骨干网的重新学习特征记为骨干网2,经过GAP后获得的2 048维向量用三元损失函数计算损失值。第三步将骨干网和反向骨干块的同层特征融合到骨干网2中,通过对反向骨干块控制通道实现调节原模型特征的效果,第三步获得的三元损失函数经过加权。由于关键点特征定位的范围不适配重识别需要的特征,而本文设计的模型只是利用关键点特征获得一组局部的特征并需要将获得的关键点转化为一个区域,即需要构造一个具有如下语义的模型:获取一组掩码分N个部分,每个部分代表一个关键点信息,且所有部分加和应该完全覆盖到可以用于区分是不是同一人的全部特征。因此,本文设计的模型的输出为一组掩码。

RBNet模型使用的三元损失函数如式(1)所示,其中:$\mathcal{L}_{\text{tri}}$为三元损失函数,f_i为预测值,f_i^+和f_i^-分别为正样本和负样本,正样本和负样本之间的距离m设置为0.3,$\|*\|$为欧氏距离,M为骨干网2输出掩码的函数。

$$\begin{cases}\mathcal{L}_{\text{tri}}=\dfrac{1}{N}\sum_{i=1}^{N}\left[\left\|f_i-f_i^+\right\|-\left\|f_i-f_i^-\right\|+m\right]\\ \mathcal{L}_{\text{tri2}}=\dfrac{1}{N}\sum_{i=1}^{N}\left[\left\|f_i(x\cdot M(x))-f_i^+(x\cdot M(x))\right\|\right.\\ \qquad\left.-\left\|f_i(x\cdot M(x))-f_i^-(x\cdot M(x))\right\|+m\right]\end{cases}\tag{1}$$

交叉熵损失通常用于多分类,如式(2)所示,因此,本模型使用交叉熵损失函数分类不同的id。骨干网2的输入乘以掩码后再正向传导,C是分类的总数,y是类别标签,p是预测的类别,i是样本编号,c表示类别编号即行人id号。

$$\begin{cases}\mathcal{L}_{\text{id}}=\dfrac{1}{N}\sum_{i}\sum_{c=1}^{C}y_{ic}\log(p_{ic})\\ \mathcal{L}_{\text{id2}}=\dfrac{1}{N}\sum_{i}\sum_{c=1}^{C}f_{ic}(x\cdot M(x))\log(p_{ic})\end{cases}\tag{2}$$

re-conv 1输出的17个通道的特征与从HRNet获得的17个人体关键点位置的特征使用均方差函数计算出Loss值,如式(3)所示,张量尺寸为(-1,17,128,64),y_i是特征上的所有点,利用类似特征金字塔的结构让反向骨干块把骨干网的特征当作数据输入,且反向骨干块使用深度可分离卷积将输入拆分为17组的特征。

$$\mathcal{L}_{\text{MSE}}=\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i-y_i)^2\tag{3}$$

将模型最后输出的Loss加权相加,如式(4)所示,为了防止由损失函数过小导致反向传导时对参数的修正失效,训练时将调整λ_3使$\lambda_3\mathcal{L}_{\text{MSE}}$和$\lambda_2\mathcal{L}_{\text{triplet}}$大小接近。

$$L=\lambda_1\mathcal{L}_{\text{id}}+\lambda_2\mathcal{L}_{\text{stri}}+\lambda_3\mathcal{L}_{\text{MSE}}+\lambda_4\mathcal{L}_{\text{id2}}+\lambda_5\mathcal{L}_{\text{tri2}}\tag{4}$$

1.2 无监督自适应(UDA)

只用有监督的模型泛用性普遍不如无监督好,因此,很多模型会在ResNet后接聚类算法。无监督领域自适应的任务是指在源域训练好的模型直接适应新的未训练的目标域。基于图像或特征相似度的伪标签法,首先用聚类算法对无标签的目标域图像特征进行聚类,从而生成伪标签,然后模型在目标域上用伪标签监督学习,以上两步循环直至收敛。本文跨域实现部分基于平均模型MMT,MMT利用离线优化的硬伪标签与在线优化的软伪标签进行联合训练,硬伪标签由聚类生成。

在每个epoch前进行单独更新,使用另一个网络的过去平均模型来监督可以避免误差放大,用时间平均模型来生成可靠的软伪标签监督其他网络。因为平均教师模型可以看作对网络过去的参数进行平均,RBNet的骨干网也是过去的参数,所以在基于MMT的模型实现跨域时骨干网2改为参数做平均(骨干网2的参数另存在一个新的实例中),如图2所示,在Net 1和Net 2输出后添加反向骨干块,使Mean Net增加对特定部位的

关注。伪标签生成方法使用DBSCAN[17](Density-based spatial clustering of applications with noise),θ_1为Net 1当前网络参数,θ_2是Net 2当前网络参数,使用平均参数法实现监督,如式(5)(6)所示,其中:$E^{(T-1)}[\theta_1]$和$E^{(T-1)}[\theta_2]$平均模型表示对过去的参数进行平均,初始时间平均参数为$E^{(0)}[\theta_1]=\theta_1$,$E^{(0)}[\theta_2]=\theta_2$。

$$^{(T)}[\theta_1]=aE^{(T-1)}[\theta_1]+(1-a)[\theta_1] \tag{5}$$

$$E^{(T)}[\theta_2]=aE^{(T-1)}[\theta_2]+(1-a)[\theta_2] \tag{6}$$

式中,x为输入的图像;$F(\cdot)$为特征提取器,用$M(\cdot)$调整后的Mean Net 2输入数据,得到强化局部特征的f_2;$C(\cdot)$为分类器,$C_1^t(f_1)$和$C_2^t(f_2)$分别为Net 1和Net 2的伪标签分类器,用于监督的软伪标签$\mathcal{L}_{\text{sid}}^t(\theta_1|\theta_2)$生成,用类似的方式获得$f_1$和$\mathcal{L}_{\text{sid}}^t(\theta_2|\theta_1)$,如式(8)所示:

$$\begin{cases} f_2=F\left(M\left(x_i^{t}|\theta_2\right)\cdot x_i^{t}|E^{(T)}[\theta_2]\right) \\ \mathcal{L}_{\text{sid}}^t(\theta_1|\theta_2)=-\dfrac{1}{N_t}\displaystyle\sum_{i=1}^{N_t}\left(C_2^t(f_2)\log C_1^t\left(F\left(x_i^t|\theta_1\right)\right)\right) \end{cases} \tag{7}$$

$$\begin{cases} f_1=F\left(M\left(x_i^{t}|\theta_1\right)\cdot x_i^{t}|E^{(T)}[\theta_1]\right) \\ \mathcal{L}_{\text{sid}}^t(\theta_2|\theta_1)=-\dfrac{1}{N_t}\displaystyle\sum_{i=1}^{N_t}\left(C_1^t(f_1)\log C_2^t\left(F\left(x_i^t|\theta_2\right)\right)\right) \end{cases} \tag{8}$$

添加RBNet后软三元损失数与有监督的三元损失函不同,求得与负样本的距离d^-见公式(9)和与正样本的距离d^+[公式(10)]后进行归一化得到损失函数见式(11)。

$$d^-=f_i(x\cdot M(x))-f_i^-(x\cdot M(x)) \tag{9}$$

$$d^+=f_i(x\cdot M(x))-f_i^+(x\cdot M(x)) \tag{10}$$

$$\mathcal{L}_{\text{stri}}=-\sum_{i=1}^{N}\log\left(\frac{\exp(d^-)}{\exp(d^+)-\exp(d^-)}\right) \tag{11}$$

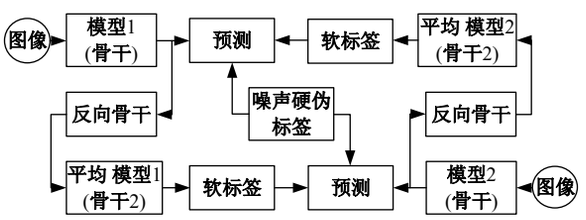

图2　基于MMT设计的跨域模型

Fig. 2　A cross-domain model based on MMT

2　实　验

在本节测试RBNet的效果。为了分析本文方法的有效性,对常用的数据集进行了广泛的实验,并使用RBNet与其他最先进的方法进行比较,通过跨数据集实验测试模型的泛化能力,用消融实验探讨有效性。

2.1　实验设置

实验在3个广泛使用的人re-id数据集上评估了本文提出的RBNet,即Market-1501[18]、DukeMTMC-reID[19]和MSMT17[20]。

Market-1501:由6台相机的12 936张图像组成训练集,其中包含了751个身份标签的图像750个身份的19 732张图像用于测试。所有图像都是由一所大学里的5台高分辨率相机和1台低分辨率相机拍摄。

DukeMTMC-reID:该数据集是DukeMTMC数据集的一个子集,由杜克大学的8台高分辨率摄像机拍摄,其中包含1 404个身份。选择具有16 522张图像的702个身份作为训练集,剩下的具有19 889张图像的702个身份作为测试集。

MSMT17:是最具挑战性和大规模的数据集,包括15台摄像机拍摄的126 441个边界盒和4 101个图像,其中32 621张1 041个身份的图像被用于训练。

本文实验使用了基于PyTorch行人重识别常用的框架及提供的工具FastReID和torchreid(框架是OSNet模型的文章的一部分),操作系统为64位Ubunt 20,显卡为NVIDIA GeForce GTX 3090。开始时学习速率为0.001 2,优化器使用Adam训练60个epoch,每20个epoch衰减10。输入的图像大小调整为256×128,使用的数据增强包括随机翻转、随机裁剪,DBSCAN[17]算法生成伪标签,不需要指定聚类簇的数目,仅需要设置半径ϵ参数和密度阈值MinPoints。虽然完全无监督不会遇到有标签数据集的限制,DBSCAN中的超参数,两个样本之间的最大距离计算后初始值大约为0.42,聚类簇中的最小样本数设置为4。使用杰卡德距离(Jaccard distance)与邻近的k=30个样本特征进行聚类。对模型性能的评估,采用在大多数人的重新识别文献中常用的标准Rank-N指标和平均精度(mAP)。

2.2　实验结果

2.2.1　非跨域实验结果

与近几年多个效果较好模型的mAP和R-1对比见表2,本文模型优于大多数模型,其性能也明显优于大多数已发布的方法,在Market-1501

表 2 在Market-1501、DukeMTMC-reID 和 MSMT17数据集上的性能对比

Table 2 Performance comparison on Market-1501, DukeMTMC-reID and MSMT17 datasets.

Method	Market-1501		DukeMTMC-reID		MSMT17	
	mAP	R-1	mAP	R-1	mAP	R-1
GCP [21]	88.9	95.2	78.6	89.7	—	—
MMGA [5]	87.2	95.0	78.1	89.5	—	—
RGA-SC [8]	88.4	96.1	—	—	57.5	80.3
AANet-50 [1]	82.45	93.89	72.56	86.42	—	—
PCB [3]	81.6	93.8	69.2	83.3	—	—
OSNet [9]	84.9	94.8	73.5	88.6	52.9	78.7
FastReID [22]	88.2	95.4	79.8	89.6	59.9	83.3
CBDB-Net [23]	85.0	94.4	74.3	87.7	—	—
PAT [24]	88.0	95.4	78.2	88.8	—	—
CDNet [25]	86.0	95.1	76.8	88.6	54.7	78.9
RBNet(conv 1)	88.5	95.6	80.0	89.7	60.8	84.9
RBNet(relayer 1)	**89.1**	**96.1**	**81.1**	**90.0**	**62.7**	**85.1**
RBNet(relayer 2)	84.3	94.0	73.9	83.3	49.4	77.1

和DukeMTMC-reID上的测试排名已经超过了大多数其他模型的水平,研究模型的可解释性和通用性更为重要。

有监督模型下的3个常用数据集的mAP,分别从88.2%、79.8%和59.9%增加到89.1%(+0.9%)、81.1%(+1.3%)和62.7%(+2.8%),这说明姿态信息的引导原模型提高了特征的鉴别能力。

2.2.2 可视化

本模型实验使用了OSNet提供的torchreid框架及其提供的工具,修改输入后获得热图(见图3)。图3(a)表示原始关键点特征的热图,图3(d)是反向骨干块生成的掩码,图3(b)(c)分别是骨干网在conv 1和layer4的输出,图3(e)(f)分别是添加掩膜后的conv 1和layer 4的输出。图3(b)(e)相比明显关注到了更多人体部位,对比图3(e)(b)可以看出,在掩膜的诱导下图3(e)额外关注了更多的关键点。使用注意力逐渐覆盖全部人体的特征当作掩码进行训练,可以有效减少模型受到的干扰,屏蔽掉无关信息可有效提升效果。从图3(c)(f)可以看出,通过掩码调整了模型的关注范围后,比原FastReID输出的特征覆盖范围更大、更贴合人身体。

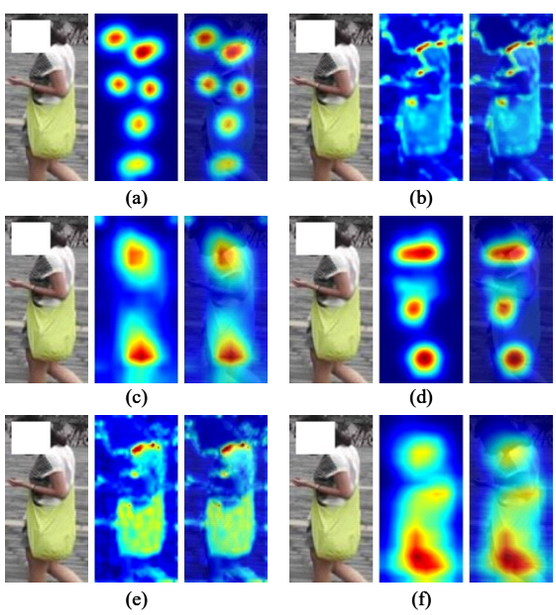

图 3 显示模型注意力

Fig.3 Displays model attention

将中间过程的特征图可视化,可以发现一些关注点过于集中的情况。虽然这样也可以实现区分已知数据集的不同人,但当跨域时遇到相同身体位置有相同局部特征的不同人时就会降低准确率,利用掩码实现了对全身所有局部特征的高度注意,相比与baseline准确率有明显提升,这证明了本文所提方法的有效性。

跨域热图对比如图4所示。模型的源域是Mar-ket-1501,图片显示了模型在目标域Duke-MTMC的注意力,图4(a)为未使用本文模块的注意力示意图,图4(b)为使用了本文模块的注意力示意图。相比baseline的热图,添加了人体关键点为条件的图其关注的位置更准确了。

2.2.3 跨域实验结果

聚类跨域实验结果如表3所示。反向骨干块的姿态引导模型生成的掩码可以自由地选择位

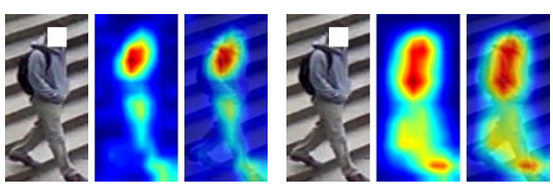

图 4 不同域时模型的注意力

Fig. 4 Attention of the model when different domains

表 3 RBNet在DukeMTMC-reID to Market-1501和Market-1501 to Duke任务中的使用聚类跨域实验结果

Table 3 Results of RBNet′s cross-domain experiments using clustering in the DukeMTMC-reID to Market-1501 and Market-1501 to DukeMTMC-reID tasks (M: Market-1501, D: DukeMTMC-reID)

M->D					
基于模型	model	mAP	Rank-1	Rank-5	Rank-10
ResNet-50	MMCL [27]	51.4	72.4	—	—
E_{app} (ResNet 50)	DG-Net++ [28]	63.8	78.9	—	—
ResNet-50	GDS [29]	55.1	73.1	—	—
ResNet 50	GPP [30]	54.4	74.0	—	—
ResNet 50	MMT-500 [14]	63.1	76.8	88.0	92.2
MMT (ResNet 50)	MetaCam [12]	65.0	79.5		
fast-reid	GLT [26]	69.2	82.0	90.2	92.8
MMT (ResNet 50)	MMT+RBNet	**70.1**	**83.2**	**90.8**	**92.9**
D->M					
基于模型	model	mAP	Rank-1	Rank-5	Rank-10
ResNet-50	MMCL	60.4	84.4	—	—
E_{app}(ResNet-50)	DG-Net++	61.7	82.1	—	—
ResNet-50	GDS	61.2	81.1	—	—
ResNet 50	GPP+	63.8	84.1	—	—
ResNet 50	MMT-500	71.2	87.7	94.9	96.9
MMT (ResNet 50)	MetaCam	76.5	90.1	—	—
fast-reid	GLT	79.5	92.2	96.5	97.8
MMT (ResNet 50)	MMT+RBNet	**79.8**	**92.4**	**96.7**	**97.9**

置,因此,本文采取无监督领域自适应方法实现跨域检索时,输出到了不同的ResNet 50。在数据集Market-1501和Duke上对比两种UDA方法,分别是添加了聚类算法的ResNet 50方法(如fastnet)和MMT方法。基于fastnet的UDA方法很多,有GLT [26](Group-aware label transfer for domain adaptive person re-identification)等。

对比测试有无的UDA(MMT)如表3所示。从源域为Market-1501、目标域为DukeMTMC-reID和源域为Duke、目标域为Market-1501的两个UDA的实验结果可以看出,与无反向骨干块的base-line对比,Market 1501到Duke的mAP,分别从63.1%和71.2%增加到70.1%(+7.0%)和79.8%(+8.6%);Duke到Market-1501的Rank 1,分别从76.8%和87.7%增加到83.2%(+6.4 %)和92.4%(+4.7%),和近几年优秀的无监督领域自适应任务相比依然有提升,这说明部分信息的集成提高了特征的鉴别能力。

2.3 消融实验

由于无监督学习到的特征具有一定的域泛化效果,所以为了只让RBNet关键模块(见图1)影响域泛化效果,在这一部分的消融实验中,将关键模块部分使用有监督学习训练并测试。非聚类跨域实验结果(见表4)分别在Market-1501和DukeMTMC-reID上测试并对比其他不使用聚类实现跨域实验结果,本文模型RBNet的模块相比base-line,mAP分别从17.3%和17.4%增加到27.4%和27.5%。

表4 RBNet在DukeMTMC-reID to Market-1501和Market-1501 to DukeMTMC-reID任务中的不使用聚类跨域实验结果

Table 4 Results of RBNet′s non-clustering cross-domain experiments in the DukeMTMC-reID to Market-1501 and Market-1501 to DukeMTMC-reID tasks

M->D				
Method	mAP	Rank-1	Rank-5	Rank-10
APNet-c3 [31]	22.8	37.7	52.4	59.0
OSNet [9]	26.7	48.5	62.3	67.4
resnet 50	17.3	32.7	47.1	53.4
resnet 50+RBNet	**27.4**	**49.4**	**62.9**	**67.7**
D->M				
Method	mAP	Rank-1	Rank-5	Rank-10
APNet-c3	23.7	50.9	66.6	72.6
OSNet	26.1	57.7	73.7	80.0
resnet 50	17.7	43.9	50.9	67.5
resnet 50+RBNet	**27.5**	**60.6**	**75.7**	**80.8**

对比rebcakbone使用不同层的特征作掩膜时对mAP的影响见表4。re-conv 1计算热图Loss,conv 1的特征传给conv 1时,Market-1501数据集上提升0.3,DukeMTMC-reID数据集上提升0.2;re-conv 1计算热图Loss,re-layer 1的特征传给conv 1时,Market-1501数据集上提升0.9,DukeMTMC-reID数据集上提升1.3。对比有无反向骨干块的情况,发现RBNet可以将跨域准确率提升到接近同样不使用聚类算法的其他模型。

原FastReID上实验直接跨域的结果为15%,利用关键点更大的感受野区域选择,可以实现排除无效的干扰并提升模型准确率。

3 结束语

本文提出了一种简单而有效的模型反向骨干网(RBNet),通过学习另一个模型的输出对不同尺度的全局范围结构信息进行建模、获得掩码,以提高模型推理的准确性,并且对有监督的跨域和无监督领域自适应行人重识别的效果都有提升,该方法主要包括全局特征融合模块和基于聚类结果的有监督学习模块。具体来说,基于MMT的反向骨干块对id和行人身份id进行无监督分析,解决了同一行人在不同环境、不同视角下的统一成像风格问题。特别是对于每个特征位置,通过指定模型的通道来学习指定的身体部位,增加了模型的可解释性,并且只比较存在相同部位的通道的距离以提高精度。大量实验证明了本文方法的有效性,该方法可以在3个数据集上实现最先进的结果,可用于完全无监督的re-id和领域自适应re-id。对比baseline在Market-1501到Duke MTMC-reID的跨域实验结果,mAP提升了7.0%,Rank-1提升了6.4%。在DukeMTMC-reID到Market-1501的跨域实验结果,mAP提升了8.6%,rank-1提升了4.7%。

参考文献:

[1] Chiat P T, Sharmili R, Kim H Y. AANet: attribute attention network for person re-identifications[C]// 32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 7127-7136.

[2] Lin Y T, Zheng L, Zheng Z D, et al. Improving person re-identification by attribute and identity learning [J]. Pattern Recognition, 2019, 95: 151-161.

[3] Sun Y F, Zheng L, Yang Y, et al. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline) [C] //15th European Conference on Computer Vision, Munich, Germany, 2018: 501-518.

[4] Miao J X, Wu Y, Yang Y. Identifying visible parts via pose estimation for occluded person re-identification[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(9): 4624-4634.

[5] Cai H L, Wang Z G, Cheng J X. Multi-scale body-part mask guided attention for person re-identification [C]//32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 1555-1564.

[6] Qiao S Y, Chen L C, Alan Y. Detectors: detecting objects with recursive feature pyramid and switchable atrous convolution[C] //IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, USA, 2021: 10208-10219.

[7] Hou S Q, Yin K N, Liang J, et al. Gradient-supervised person re-identification based on dense feature pyramid network[J]. Complex & Intelligent Systems, 2022, 8(6I): 5329-5342.

[8] Zhang Z Z, Lan G L, Zeng W J, et al. Relation-aware global attention for person re-identification[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 3183-3192.

[9] Zhou K Y, Yang Y X, Cavallaro A, et al. Omni-scale feature learning for person re-identification[C]// IEEE/CVF International Conference on Computer Vision, Seoul, South Korea, 2019: 3701-3711.

[10] Divyansh G, Netra P, Thomas T, et al. Towards explainable person re-identification[C]//IEEE Symposium Series on Computational Intelligence, Orlando, USA, 2021: 9660071.

[11] Niki M, Gian L F, Christian M. Deep pyramidal pooling with attention for person re-identification[J]. IEEE Transactions on Image Processing, 2020, 29: 7306-7316.

[12] Yang F X, Zhong Z, Luo Z M, et al. Joint noise-tolerant learning and meta camera shift adaptation for unsupervised person re-identification[C] //IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, USA, 2021: 4853-4862.

[13] Antti T, Harri V. Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results[C]//31st Annual Conference on Neural Information Processing Systems, Long Beach, USA, 2017: 1195-1204.

[14] Ge Y X, Chen D P, Li H S. Mutual mean-teaching: pseudo label refinery for unsupervised domain adaptation on person re-identification[C]//8th International Conference on Learning Representations, Addis Ababa, Ethiopia, 2020: 200101526.

[15] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2016: 770-778.

[16] Sun K, Xiao B, Liu D, et al. Deep high-resolution representation learning for human pose estimation [C]//32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA,

2019: 5686-5696.

[17] Martin E, Hans P K, Jorg S, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[C]//2nd International Conference on Knowledge Discovery and Data Mining, Portland, USA, 1996: 226-231.

[18] Zheng L, Shen L Y, Tian L, et al. Scalable person re-identification: a benchmark[C]//IEEE International Conference on Computer Vision, Santiago, USA, 2015: 1116-1124.

[19] Ergys R, Francesco S, Roger Z, et al. Performance measures and a data set for multi-target, multi-camera tracking[C]//14th European Conference on Computer Vision, Amsterdam, Netherlands, 2016: 17-35.

[20] Wei L H, Zhang S L, Gao W, et al. Person transfer GAN to bridge domain gap for person re-identification [C]//31st IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 79-88.

[21] Park H J, Ham B. Relation network for person re-identification[C]//34th AAAI Conference on Artificial Intelligence, New York, USA, 2020: 11839-11847.

[22] He L X, Liao X Y, Liu W, et al. Fastreid: a pytorch toolbox for general instance re-identification[J]. Proceedings of the 31st ACM International Conference on Multimedia, 2020, 10: 9662-9667.

[23] Tan H C, Liu X P, Bian Y H, et al. Incomplete descriptor mining with elastic loss for person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(1): 160-171.

[24] Li Y L, He J F, Zhang T Z, et al. Diverse part discovery: occluded person re-identification with part-aware transformer[C] //IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, USA, 2021: 2897-2906.

[25] Li H J, Wu G J, Zheng W S. Combined depth space based architecture search for person re-identification [C] //IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, USA, 2021: 6725-6734.

[26] Zheng K C, Liu W, He L X, et al. Group-aware label transfer for domain adaptive person re-identification[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, USA, 2021: 5306-5315.

[27] Wang D K, Zhang S L. Unsupervised person re-identification via multi-label classification[J]. International Journal of Computer Vision, 2022, 130(12): 2924-2939.

[28] Zou Y, Yang X D, Yu Z D, et al. Joint disentangling and adaptation for cross-domain person re-identification[C]//16th European Conference on Computer Vision, Glasgow, UK, 2020: 87-104.

[29] Jin X, Lan C L, Zeng W J, et al. Global distance-distributions separation for unsupervised person re-identification[C]//16th European Conference on Computer Vision, Glasgow, UK, 2020: 735-751.

[30] Zhong Z, Zheng L, Luo Z M, et al. Learning to adapt invariance in memory for person re-identification [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(8): 2723-2738.

[31] Chen G Y, Gu T P, Lu J W, et al. Person re-identification via attention pyramid[J]. IEEE Transactions on Image Processing, 2021, 30: 7663-7676.