

文章编号: 2095-2163(2022)11-0103-08

中图分类号: TP931.41

文献标志码: A

基于轻量多分支网络的行人重识别方法

罗丽洁, 韩华, 金婕, 黄丽

(上海工程技术大学 电子电气工程学院, 上海 201620)

摘要: 最新的行人重识别方法大都是基于卷积神经网络(CNN), 虽然这些网络在分类或目标检测等多项任务中有着不错的表现, 但这些方法大都侧重于图像最具辨别力的部分, 忽视了其他的一些相关特征, 而重识别任务需要更加丰富, 具有多样性的特征。本文提出了一种基于 OSNet(Omni-scale Network)的多分支网络结构, OSNet 是一个轻量级的 Re-ID 模型, 可将标准卷积分解为点卷积和深度卷积以便达到减少参数的目的。网络主干部分加入了注意力模块, 可以抑制无用信息; 而多分支的结构可以提取到更多样性的特征。该网络主要由全局分支、局部分支、顶部擦除分支和通道分支四个部分构成。全局分支用于提取图像的整体特征; 局部分支能学习到细粒度的特征; 顶部擦除分支通过擦除高激活性区域, 使得网络更加关注于激活性较差的区域; 通道分支使网络学习到更多与通道有关的信息。在 Market-1501、CUHK03、DukeMTMC-reID 三个公开数据集上的实验结果表明, 本文提出的方法针对行人重识别问题有着优秀的性能表现。

关键词: 行人重识别; 神经网络; 多分支网络; 注意力模块; 顶端擦除

Person re-identification method based on lightweight multi-branch network

LUO Lijie, HAN Hua, JIN Jie, HUANG Li

(School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

[Abstract] Most of the latest person re-identification (Re-ID) methods are based on convolutional neural networks (CNN). Although these networks have good performance in many tasks, such as classification or object detection, these methods often focus on the most discriminative part of an image and ignore some other relevant features. The Re-ID tasks need more abundant and diverse features. This paper proposes a multi-branch network structure based on OSNet (Omni-scale Network). OSNet is a lightweight Re-ID model, which decomposes the standard convolution into pointwise convolution and depthwise convolution in order to reduce parameters. The attention modules are added to the network backbone, which could inhibit useless information, and the multi-branch structure can extract more diverse features. It is mainly composed of four parts: global branch, local branch, top erased branch and channel branch. The global branch is used to extract the global features of pedestrian images; local branches can learn more fine-grained features; the top erased branch makes the network pay more attention to the low informative regions by erasing the most activated regions; channel branch can make the network learn more channel information. The experimental results on Market-1501, CUHK03 and DukeMTMC-reID show that the proposed method has excellent performance for person re-identification.

[Key words] person re-identification; neural network; multi-branch network; attention model; top DropBlock

0 引言

行人重识别技术是指通过利用一张查询图像, 从一个大规模的图像集中准确地找出属于同一个人的所有图像, 其中查询图像和图像集中的图像是来自不同相机拍摄到的没有重叠部分的图片。但是由于拍摄的摄像机参数设置存在着差异, 行人出现的地点也不相同, 还有不同的光影变化、行人姿态的改变、行人换衣、部分遮挡等一系列的问题, 造成拍摄

到的同一个行人图片之间会存在些许的差异, 为行人重识别的研究带来了诸多的挑战, 对行人重识别模型性能的提升带来影响。

为了能够解决上述的问题, 已陆续提出了基于度量学习^[1-2]的方法、针对样本不平衡^[3]方法等。近年来, 基于深度学习的行人重识别方法得到了越来越多的关注, 相较于传统的基于手工特征的方法, 用卷积神经网络^[4]可以提取到图像中更为显著、丰富的行人特征。通常, 卷积神经网络对提取到的全

基金项目: 国家自然科学基金(61305014); 上海市自然科学基金(22ZR1426200); 上海市教育委员会和上海市教育发展基金会“晨光计划”资助项目(13CG60)。

作者简介: 罗丽洁(1996-), 女, 硕士研究生, 主要研究方向: 计算机视觉、行人重识别; 韩华(1983-), 女, 博士, 教授, 主要研究方向: 目标识别与跟踪、行人重识别、智能计算等。

通讯作者: 韩华 Email: 2070967@mail.dhu.edu.cn

收稿日期: 2022-01-12

局特征具有鲁棒性,因为全局特征对外观变化和空间位置变化并不敏感,但上述的全局特征却容易忽略掉人体潜在的相关信息和细粒度特征。为了使得网络能够学习到更加显著的特征,部分研究加入了注意力机制^[5-6],产生的效果是十分明显的。通过注意力机制可以使网络更加专注于行人图像中有用的信息,同时,能够减少图像中背景噪声带来的干扰。此外,又相继研发出各种基于局部特征的方法,这些方法采用的策略大都是将身体划分为几个水平部分^[7-9],从而允许网络能够更多地聚焦到局部特征和细粒度特征上,以便减少因行人姿势的变化给网络特征提取带来的影响。文献[10]中提出了 DropBlock 方法,通过去除相关区域的特征,让网络加强学习其他区域的特征。与 DropBlock 的思想不同, Batch DropBlock^[11]采用的方法是通过在网络训练的过程中随机地丢弃掉一批特征图中的同一块区域,令网络加强对特征图其余部分的学习。

虽然上述的方法在行人重识别问题上均已取得了不错的成绩与效果,但是却在一定程度上增加了模型整体的复杂性,而且还会耗费更多的计算成本和时间成本。因此,本文提出了一个基于 OSNet

(Omni-scale Network)^[12]的多分支网络结构,OSNet 是一个全尺度轻量级的 Re-ID 模型,可将标准卷积分解为点卷积和深度卷积,能够有效减少参数量,OSNet 参数与基于 ResNet50^[13]的模型相比,至少要降低一个数量级,但却能表现出更好的性能,所以在保证模型性能的前提下,基于 OSNet 的网络更加轻量,不仅减少了训练时间,还提高了训练效率。而多分支的网络结构相比单一的网络则可以学习到更加丰富、更细粒度的特征信息,有利于提升模型的性能。本文的网络包含了全局分支、局部分支、顶端擦除分支和通道分支四个部分,能够从多个维度进行特征提取,提取到的特征也更趋多样化。

1 网络模型

本文提出了一种基于 OSNet 的多分支网络结构,如图 1 所示,其中包含了全局分支(Global branch)、局部分支(Local branch)、顶端擦除分支(Top erased branch)和通道分支(Channel branch)四个部分。相较于单一的网络结构,多分支网络结构可以提取到更加细粒度、更具显著性的特征。

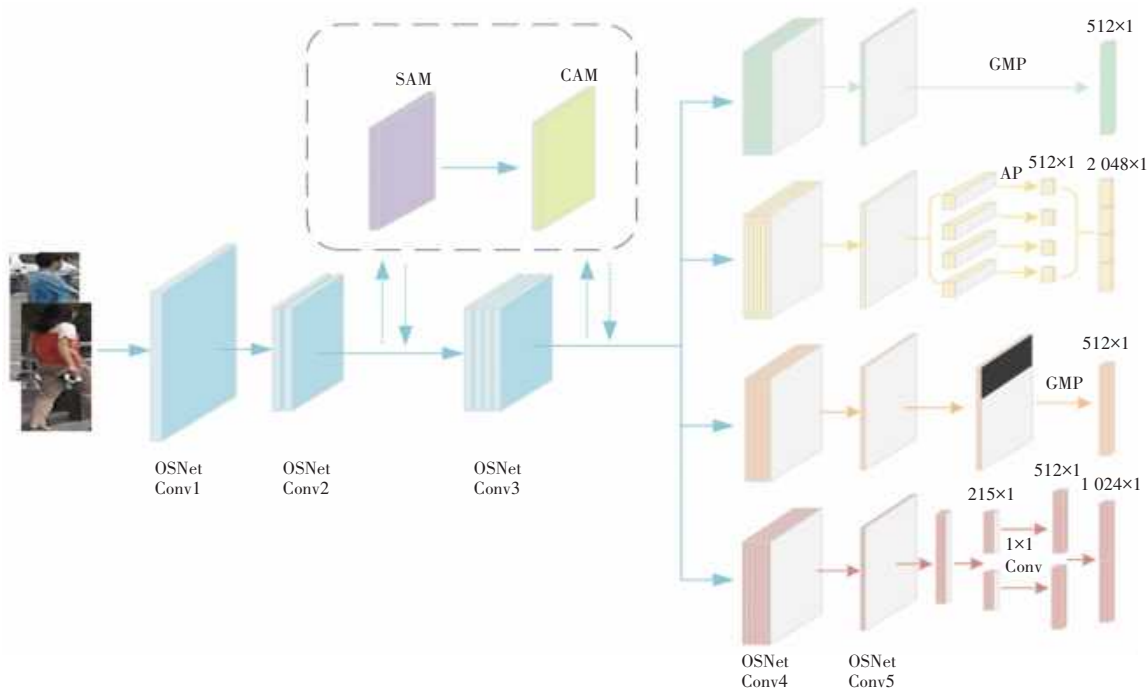


图 1 多分支网络结构图

Fig. 1 Multi-branch network structure diagram

1.1 注意力模块

注意力模型通常用于各种深度学习的任务中,其本质是通过权重参数加强网络对重要信息的学习,抑制无关的信息,在行人重识别领域中得到广泛

使用。为了更好地学习到有价值的图像特征,本文提出的网络结构与一般的 OSNet 网络相比,在 OSNet 网络的 conv2 和 conv3 层中加入了注意力模块,分别为空间注意力模块(Spatial Attention

Module, SAM)和通道注意力模块(Channel Attention Module, CAM)。从图 1 中可以看到,先由空间注意力模块处理特征,然后经由通道注意力模块,再到各个分支网络分别进行处理。

1.1.1 空间注意力模块

空间注意力模块更加关注于空间维度内的相关特征,能够更好地聚合空间域中与语义相关像素,如

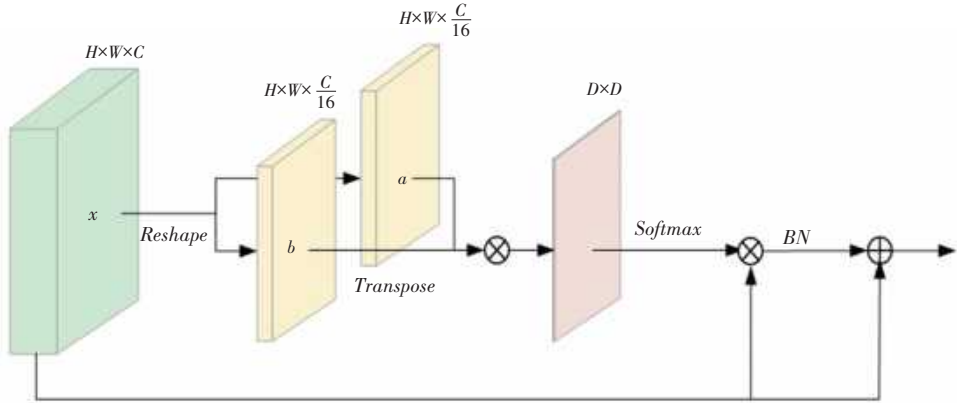


图 2 空间注意力模块

Fig. 2 Spatial attention module

1.1.2 通道注意力模块

通道注意力模块用于探寻不同通道间的特征映射关系,使得网络能够更好地关注具有价值的特征,提取到更具显著性的特征。通道注意力模块(CAM)如图 3 所示。本文的通道注意力模块是基于压缩-激发块(Squeeze-and-Excitation block)^[14],而与文献[14]不同的是,删除了初始 block 中的全局平均池化(Global Average Pooling, GAP),有助于将空间信息保留到注意力 block 中。

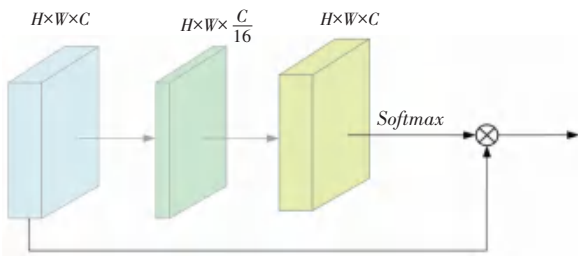


图 3 通道注意力模块

Fig. 3 Channel attention module

图 3 中,输入维度为 $H \times W \times C$ 的特征图,经过第一个卷积后维度变为 $H \times W \times \frac{C}{16}$,再经过第二个卷积后维度变为 $H \times W \times C$,接着经过 *Softmax* 函数,得到通道注意力张量。

1.2 全局分支

全局分支的优势主要体现在学习行人图像的整体

图 2 所示。图 2 中,输入维度为 $H \times W \times C$ 的特征图 x , H 、 W 、 C 分别表示高度、宽度和通道数,经过卷积后得到 2 个维度为 $H \times W \times \frac{C}{8}$ 的特征图 a 和 b 。特征图 a 被重塑为 $D \times C$ 的张量,特征图 b 被重塑为 $C \times D$ 的张量,其中 $D = H \times W$ 。经过 *Softmax* 函数后,再进行批量归一化处理,得到空间注意力张量。

信息方面,但对一些细粒度特征的区分度并不大。本文的全局分支是由 OSNet 网络的 *conv4* 层和 *conv5* 层组成的,输入的图像特征经过一个全局最大池化(Global Max Pooling, GMP)层后生成了一个 512 维的向量。在局部分支中采用的是平均池化(Average Pooling, AP),而在全局分支中采用全局最大池化,主要是为全局分支和局部分支机构提供了彼此间的功能多样性。

1.3 局部分支

相较于全局分支,局部分支更擅长提取细粒度特征,减少因行人姿态带来的变化,能够弥补全局分支的不足。局部分支采用与全局分支较为相似的层结构,但使用平均池化代替了全局最大池化。为了实现特征的多样化,将特征图水平分成 4 个条带,并将 4 个 512 维的特征向量连接起来,生成一个 2 048 维的特征向量,与 PCB (Part-based Convolutional Baseline) 网络采用多个 ID 预测损失、并且每个部件都具有独立的 ID 预测损失不同的是,本文的局部分支只使用一个 ID 预测损失。

1.4 顶端擦除分支

卷积神经网络往往只关注图像中最具辨别力的部分,如图 4 所示。高激活区域大都集中在图像的上半部分,这会导致网络不能够很好地学习到图像其他部分的特征,所以本文采用顶端(高激活区域)擦除的方式,而不是随机擦除,让网络可以关注到其他区域。根据 BDB (Batch DropBlock) 网络中提出的

擦除大小为特征图像高度的三分之一、且宽度相同的区域效果为最好,故本文也采用了这种方法。

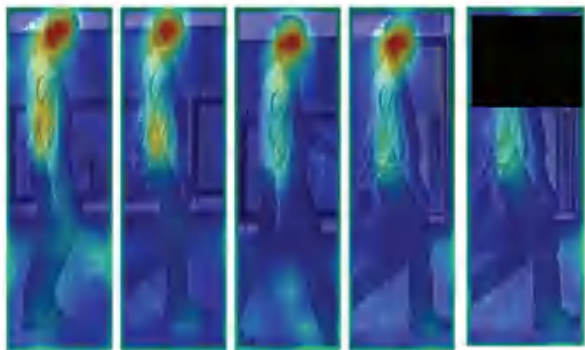


图4 激活图与顶端擦除

Fig. 4 Activation map and top erased map

每个训练批次输入行人图像,经过神经网络,输出特征图 F , 根据文献[15]提出的定义,可以将特征图像 F 转换为激活图 A , 即:

$$A = \sum_{i=1}^c |F_i| \quad (1)$$

其中, F_i 表示每个大小为 $H \times W$ 的张量, c 表示通道数。

在 A 的基础之上,将每个条带 r_j 的相关性 R 定义为第 j 行上值的平均值,对此可表示为:

$$r_j = \frac{\sum_{k=1}^w A_{j,k}}{w} \quad (2)$$

再将 r_j 值最大的行归零。

1.5 通道分支

加入通道分支可以使得网络提取到关于通道的更多信息,丰富整个网络提取到特征的多样性。首先,通过卷积的方式生成一个 512 维的向量,然后将这个向量划分为 2 个长度为 256 维的向量,再使用 1×1 的卷积对特征进行缩放,得到 2 个 512 维向量分别为 c_1 和 c_2 。这里, 1×1 Conv 的参数在 2 个通道部分之间共享,可以减少参数的数量,节约计算时间。

1.6 损失函数

为了使模型能够得到更好的训练,总的损失函数为每个分支的损失函数之和,包括了三元组损失 $L_{triplet}$, ID 损失 L_{id} 和中心损失 L_{center} , 数学表达式为:

$$L_{sum} = \lambda_1 L_{id} + \lambda_2 L_{triplet} + \lambda_3 L_{center} \quad (3)$$

其中, $\lambda_1, \lambda_2, \lambda_3$ 为权重参数,实验中的取值分别为 $\lambda_1 = 1, \lambda_2 = 1, \lambda_3 = 0.005$ 。

进一步地,给出 ID 损失公式具体如下:

$$L_{id} = -\frac{1}{N} \sum_{i=1}^N q_i \log(p_i) \quad (4)$$

其中, N 表示样本的数量; p_i 表示预测行人 ID 为 i 的概率; q_i 表示真实的标签。三元组损失分别从每一个 $batch$ 中抽取 P 个行人的 K 张图片,推导得到的公式为:

$$L_{triplet} = \sum_{i=1}^P \sum_{a=1}^K [\alpha + \max_{p=1 \dots K} \|x_a^{(i)} - x_p^{(i)}\|_2 - \min_{\substack{n=1 \dots K \\ j=1 \dots P, j \neq i}} \|x_a^{(i)} - x_n^{(i)}\|_2]_+ \quad (5)$$

其中, $x_a^{(i)}, x_p^{(i)}, x_n^{(i)}$ 分别表示从锚点样本、正样本和负样本中提取到的特征, α 表示边缘超参数。

为了能够提高网络对特征的鉴别能力,对类内距离进行约束,采用中心损失作为特征提取的依据,中心损失的数学定义公式可写为:

$$L_{center} = \frac{1}{2} \sum_{i=1}^m \|x_i - c_{y_i}\|_2^2 \quad (6)$$

其中, m 表示 mini-batch 大小; x_i 表示样本特征; c_{y_i} 表示第 y_i 个类别的特征中心。

2 实验结果与分析

2.1 数据集和评价指标

本文的实验是在 Market-1501^[16]、CUHK03^[17]、DukeMTMC-reID^[18] 三个公开数据集上进行的。表 1 中的数据为这 3 个数据集的基本信息。表 1 中, Market-1501 数据集是通过 6 个摄像头采集到的 1 501 个行人的图片。DukeMTMC-reID 数据集用 8 个摄像头采集到 1 812 个行人的图片,下文由 DukeMTMC 表示。CUHK03 是通过 5 对摄像头采集到的 1 467 个行人的图片,其中 CUHK03-L 的行人框由人工进行标注,而 CUHK03-D 的行人框则由机器进行标注。

表 1 数据集信息

Tab. 1 Dataset information

数据集	训练集/ 张	训练集 /(人·个 ⁻¹)	测试集/ 张	测试集 /(人·个 ⁻¹)
DukeMTMC	16 522	702	17 661	702
Market-1501	12 936	751	19 732	750
CUHK03-L	7 368	767	5 328	700
CUHK03-D	7 365	767	5 332	700

上述的数据集分别采集于不同地点、不同季节、不同时间段,采集图片的设备包含了高清摄像头和低清摄像头。数据集中的行人图片有着复杂的背

景、明显的光照变化、不同的清晰程度, 贴近现实场景。

本文采用 2 个常用的评价指标: 首位命中率 ($rank1$) 和平均精度值 (Mean Average Precision, mAP) 来对行人重识别模型的性能进行评价。其中, $rank1$ 表示的是与每张查询图片所对应的返回的图片集中排在第一张是准确的概率, 而 mAP 表示的是返回查询图片结果平均精度 (Average Precision, AP) 的平均值。

2.2 实验细节

本文实验是在 PyTorch 1.7.1 上展开, 编程语言选用了 Python 3.8, 硬件设备为配有一块 NVIDIA RTX 3090 GPU 的服务器, 系统是 Ubuntu 20.04 版本。

将输入图像调整为 256×128 像素, 训练期间使用随机水平翻转和擦除^[19]作为数据增强方式, 采用 Adam^[20]作为优化器。训练批量大小为 48, 一个批量由 8 个样本组成, 其中每个样本包含 6 个行人的身份。训练时使用 WarmUp 策略, 一共进行 110 个 $epoch$ 的训练, 初始学习率设置为 $6e-04$, 到第 50 个 $epoch$ 的时候学习率降为 $6e-05$; 到第 80 个 $epoch$ 的时候, 学习率降为 $6e-06$; 最后, 到第 110 个 $epoch$ 的时候, 学习率下降到 $6e-07$ 。

2.3 消融实验

加入多分支会在一定程度上增加模型的参数、增加计算成本, 因此, 必须有充分理由, 才可增加网络分支。为探究多分支网络各个分支对模型性能的影响, 在 Market-1501 和 CUHK03-D 两个数据集上进行消融实验。表 2 显示的是网络在不同分支组合下的性能。表 2 中, G 表示全局分支, P 表示局部分支, T 表示顶端擦除分支, C 表示通道分支。结果表明, 使用多分支的结构相比于单分支有所提升, 在 CUHK03-D 上表现比较明显, $rank1$ 有 3.1 个百分点的提升, mAP 有 4.1 个百分点的提升。从实验结果可以看出, 每增加一个网络分支在一定程度上对网络的性能都有所提升, 多个分支在网络结构中起到了互补的作用。在二分支网络中, 全局加上局部分支的效果最好。这样一来, 全局分支用以提取图像的一般特征, 加上局部分支后, 改进了网络对细粒度特征的学习。在三分支的网络结构中, 全局加局部、再加顶端擦除分支的组合效果是最好的, 顶端擦除分支的加入可以令网络关注到容易被忽略的低信息区域, 学习到更加丰富的特征。因此, 实验结果表明, 多分支的网络结构可以相互弥补不足, 加强彼此之间对复杂特征的学习, 增强网络对特征的提取能

力。

表 2 各个分支对网络性能的影响

Tab. 2 The influence of each branch on the network performance

Branch	Market-1501		CUHK03-D	
	$rank1$	mAP	$rank1$	mAP
G	94.9	88.7	80.5	76.7
G+P	95.3	90.2	82.9	79.3
G+T	95.2	90.3	82.4	79.0
G+C	95.1	90.2	81.5	78.8
G+P+T	95.5	90.3	83.3	80.6
G+P+C	95.4	90.4	83.2	80.4
G+C+T	95.6	90.4	83.0	80.1
G+P+T+C	95.9	90.6	83.6	80.8

使用注意力机制可以令网络提高行人重识别模型的性能, 加强网络对重要信息的学习, 抑制无关信息。因此, 在这部分的消融实验中, 在 Market-1501 上评估了注意力模块对网络的影响。注意力模块对网络性能的影响见表 3。

表 3 注意力模块对网络性能的影响

Tab. 3 The influence of attention modules on the network performance

Attention Module	Market-1501	
	$rank1$	mAP
No	95.2	89.5
SAM	95.4	90.1
CAM	95.4	90.0
SAM+CAM	95.8	90.9

从表 3 的实验结果可以看出, 注意力模块有助于网络抑制无用信息, 与未加入注意力模块相比, 加入后模型的 $rank1$ 提升了 0.6 个点, mAP 提升了 1.4 个点。

为了验证顶端擦除 (Top DropBlock) 的方式要比随机擦除 (Random DropBlock) 的方法更加优越, 在 Market-1501 和 DukeMTMC 上进行了消融实验。顶端擦除对网络性能的影响见表 4。

表 4 顶端擦除对网络性能的影响

Tab. 4 The influence of top DropBlock on the network performance

Method	Market-1501		DukeMTMC	
	$rank1$	mAP	$rank1$	mAP
Random DropBlock	95.8	90.6	90.9	81.7
Top DropBlock	96.2	91.1	91.6	82.9

从表 4 的实验结果可以看出, 顶端擦除要比随机擦除在 2 个数据集上的表现更好。顶端擦除可以

提高网络对低信息区域的关注,提取到更为丰富的信息,相比随机擦除更加具有针对性。

2.4 与其他方法对比

将本文的网络与先进方法在3个数据集上进行比较,实验结果均未使用 Re-ranking 方法,取3次实验的平均值。其中,BoT (Bag of Tricks)^[21]是一个加入了很多训练技巧的基线网络;OSNet (Omni-scale Network)^[12]是一个轻量级的行人重识别网络;MHN (Mixed High Order Attention Network)^[22]让网络学习混合高阶注意力;BDB (Batch DropBlock Network)^[11]对特征图进行了随机擦除;ABD (Attentive But Diverse Network)^[23]结合了通道和位置注意力;Pyramid^[24]是一个学习不同粗细粒度特征的金字塔模型;SCSN (Saliency-Guided Cascaded Suppression Network)^[25]挖掘图像中潜在的显著性特征;SCR (Spatial and Channel Partition Representation Network)^[26]是一个对全局和局部特征进行联合训练的网络。

表5是本文所提出的方法在 Market-1501 和 DukeMTMC 两个公开的数据集上与最近几年在行人重识别领域中先进方法的比较结果。从表5中可以看出,同样使用 OSNet 作为主干网络,本文的方法在 Market-1501 数据集上比 OSNet 的 *rank1* 高出 1.3 个百分点,*mAP* 高出 6.3 个百分点;在 DukeMTMC 数据集上比 OSNet 的 *rank1* 高出 2.9 个百分点, *mAP* 高出 9.6 个百分点。与采用随机擦除策略的 BDB 网络相比,在 Market-1501 数据集上本文的方法比 BDB 的 *rank1* 高出 0.8 个百分点,*mAP* 则高出 4.5 个百分点;在 DukeMTMC 数据集上比 BDB 的 *rank1* 高出 2.5 个百分点, *mAP* 高出 4.5 个百分点。和采用特征金字塔的方法 (Pyramid)、挖掘潜在图像显著特征的方法 (SCSN)、对全局特征和局部特征进行联合训练的方法 (SCR) 等先进方法相比,本文提出的方法也有着更佳的表现。

表6是本文的方法在 CUHK03-D 和 CUHK03-L 上与先进方法的比较。由表6可知,同样与 OSNet 网络相比,本文的方法在 CUHK03-D 上有着更好的表现, *rank1* 高出 11.2 个百分点, *mAP* 高出 12.9 个百分点,在 CUHK03-L 上 OSNet 没有实验数据。与 BDB 网络相比,本文的方法在 CUHK03-D 上, *rank1* 高出了 7.1 个百分点, *mAP* 高出 7.2 个百分点;并且在 CUHK03-L 上, *rank1* 高出 6.1 个百分点, *mAP* 高出 6.8 个百分点。与其他的先进方法相比,本文的方法在 CUHK03-D 上比 SCSN 在 *rank1* 上低了 1.2 个

百分点, *mAP* 上低了 0.3 个百分点,而在 CUHK03-L 上,比 SCSN 在 *rank1* 上低了 1.3 个百分点,在 *mAP* 上低 0.5 个百分点。结果表明,本文提出的方法在 CUHK03-D 和 CUHK03-L 数据集上,虽然没有能够在 *rank1* 和 *mAP* 两项性能评价指标上都达到最佳,但在行人重识别问题上仍然有着不错的性能表现。

表5 Market-1501 和 DukeMTMC 上与最先进方法比较

Tab. 5 Comparison of state-of-the-arts on Market-1501 and DukeMTMC

Method	Market-1501		DukeMTMC	
	<i>rank1</i>	<i>mAP</i>	<i>rank1</i>	<i>mAP</i>
BoT	94.5	85.9	86.4	76.4
OSNet	94.8	84.9	88.6	73.5
MHN	95.1	85.0	89.1	77.2
BDB	95.3	86.7	89.0	78.6
ABD	95.6	88.3	89.0	78.5
Pyramid	95.7	88.2	89.0	79.0
SCSN	95.7	88.5	91.0	79.0
SCR	95.7	89.0	91.1	81.4
The proposed	96.1	91.2	91.5	83.1

表6 CUHK03-D 和 CUHK03-L 上与最先进方法比较

Tab. 6 Comparison of state-of-the-arts on CUHK03-D and CUHK03-L

Method	CUHK03-D		CUHK03-L	
	<i>rank1</i>	<i>mAP</i>	<i>rank1</i>	<i>mAP</i>
BoT	-	-	-	-
OSNet	72.3	67.8	-	-
MHN	71.7	65.4	77.2	72.4
BDB	76.4	73.5	79.4	76.7
ABD	-	-	-	-
Pyramid	79.9	74.8	78.9	76.9
SCSN	84.7	81.0	86.8	84.0
SCR	82.2	77.6	83.8	80.4
The proposed	83.5	80.7	85.5	83.5

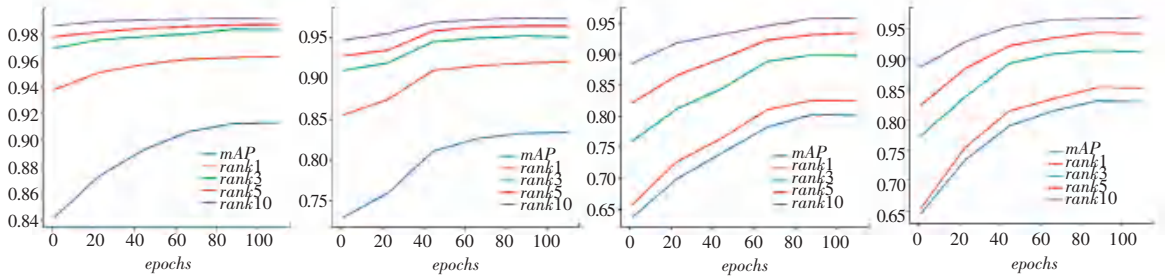
通过在行人重识别领域3个公开数据集上与其他先进方法进行比较发现,本文提出的多分支网络在性能上有着一定的优越性,尤其是在 Market-1501 和 DukeMTMC 两个数据集上, *rank1* 和 *mAP* 都有着最佳的性能,在 CUHK03 数据集上虽然没有都达到最佳性能,但也有着不错的表现。相比于单一的网络结构,多分支网络可以挖掘出行人图像中更多的具有相关性的特征,也丰富了特征的多样性,更加适合行人重识别任务。加入注意力模块后,可以令网络更多地关注到有用的信息,抑制一些不具有

效信息的区域、如图片的背景,减少对网络进行特征提取的干扰。

2.5 结果可视化

图 5 是本文的模型在一次训练的过程中 $rank1$ 、 $rank3$ 、 $rank5$ 、 $rank10$ 和 mAP 随着 $epoch$ 迭代的变化趋势。从图 5 中可以看出,各项指标都在随着 $epoch$

次数的增加逐渐地上升,前 80 个 $epoch$ 的上升趋势比较明显,后 30 个 $epoch$ 上升趋势比较缓慢。根据本文实验的设置,到第 80 个 $epoch$ 的时候降低网络的学习率,模型逐渐开始收敛,到第 100 个 $epoch$ 的时候,各项评价指标逐渐趋于平稳,所以本文的实验一共取 110 个 $epoch$,对实验结果进行分析。



(a) 在 Market-1501 上的运行结果 (b) 在 DukeMTMC 上的运行结果 (c) 在 CUHK03-D 上的运行结果 (d) 在 CUHK03-L 上的运行结果

图 5 $rank1$ 、 $rank3$ 、 $rank5$ 、 $rank10$ 和 mAP 随着 $epoch$ 的变化

Fig. 5 $rank1$ 、 $rank3$ 、 $rank5$ 、 $rank10$ and mAP change with epochs

激活图与查询图像检索结果如图 6 所示。从图 6 中可以看到,正确的结果以绿色框突出显示,而不正确的结果以红色框突出显示,行人图像中红色越深的部分代表激活的程度越高,越蓝的部分代表激活程度越低。本文使用 Top DropBlock,对高激活区

域进行擦除,而不是随机的 DropBlock,可以使激活的部分更加分散到行人身上,这有助于提取到更好的特征表示。而加入注意力模块可以令网络更多地关注到行人身上,减少背景对提取特征时的干扰,可以使得网络更加关注到图像的有效区域。

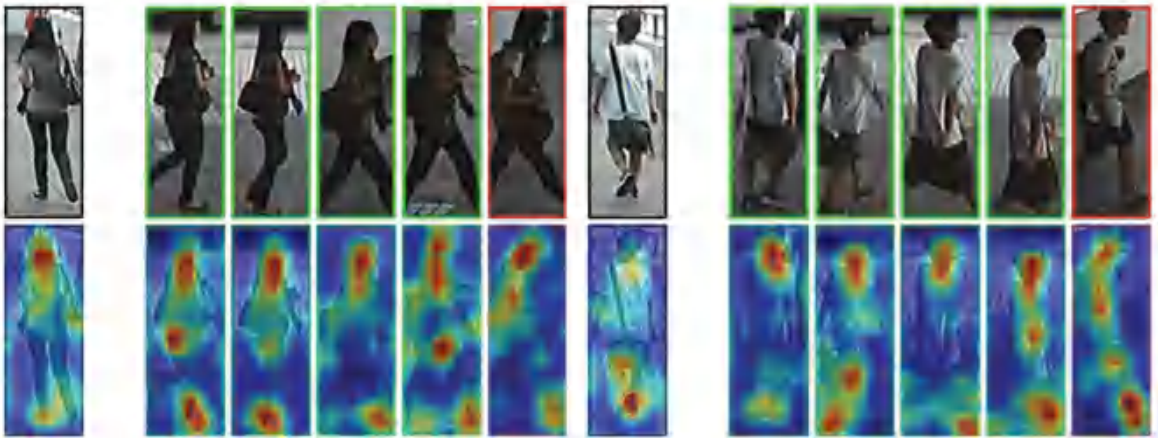


图 6 激活图与查询图像检索结果

Fig. 6 Activation map and query image retrieval results

从图 6 中看到,模型根据查询图像检索到的匹配图片中,对于一些与查询图片十分相似的负样本,模型在判断上仍会出现错误,负样本困难也一直是行人重识别研究的难题,但排在前面的大都是正确结果,说明模型还是具有不错的性能。

3 结束语

本文提出了一种基于轻量多分支网络结构的方法,用于解决行人重识别特征提取的问题。使用轻量级网络 OSNet 作为主干网络,主要由全局分支、局

部分支、顶部擦除分支和通道分支四个部分构成。其中,全局分支提取行人的一般特征;局部分支将特征图进行分割,学习更细粒度的特征;顶部擦除分支通过丢弃特征图的高激活区域使得网络更加关注于激活性较差的区域,能够学习到更加丰富的信息;通道分支用于获取更多的通道信息。使用多分支的网络结构能够使网络学习到更具多样性、且区分度更高的特征。通过在 3 个公开数据集上的仿真实验证明,本文提出的轻量多分支网络模型在行人重识别领域有着优秀表现,并且与更加复杂的 SOTA 方法

相比性能更加优越。

参考文献

- [1] HAN Hua, ZHOU Mengchu, SHANG Xiwu, et al. KISS+ for rapid and accurate pedestrian re-identification [J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 22(1): 394-403.
- [2] HAN Hua, ZHOU Mengchu, ZHANG Yujin. Can virtual samples solve small sample size problem of KISSME in pedestrian re-identification of smart transportation? [J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(9): 3766-3776.
- [3] HAN Hua, MA Wenjin, Zhou Mengchu, et al. A novel semi-supervised learning approach to pedestrian reidentification [J]. IEEE Internet of Things Journal, 2020, 8(4): 3042-3052.
- [4] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [J]. Advances in Neural Information Processing Systems, 2012, 25: 1097-1105.
- [5] CHEN Tianlong, DING Shaojin XIE Jingyi, et al. Abd-net: Attentive but diverse person re-identification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South); IEEE, 2019: 8350-8360.
- [6] ZHANG Zhizheng, LAN Cuiling, ZENG Wenjun, et al. Relation-aware global attention for person re-identification [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Seattle, WA, USA; IEEE, 2020: 3183-3192.
- [7] SUN Yifan, ZHENG Liang, YANG Yi, 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. ECCV 2018. Lecture Notes in Computer Science (). Cham: Springer, 2018, 11208: 501-518.
- [8] WANG Guanshuo, YUAN Yufeng, CHEN Xiong, et al. Learning discriminative features with multiple granularities for person re-identification [C]//Proceedings of the 26th ACM international conference on Multimedia. Seoul, Republic of Korea; ACM, 2018: 274-282.
- [9] PARK H, HAM B. Relation network for person re-identification [C]//Proceedings of the AAAI Conference on Artificial Intelligence. AAAI, 2020, 34(07): 11839-11847.
- [10] GHIASI G, LIN T Y, LE Q V. Dropblock: A regularization method for convolutional networks[J]. arXiv preprint arXiv:1810.12890, 2018.
- [11] DAI Zuoqiao, CHEN Mingqiang, GU Xiaodong, et al. Batch dropblock network for person re-identification and beyond[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South); IEEE, 2019: 3690-3700.
- [12] ZHOU Kaiyang, YANG Yongxin, CAVALLARO A, et al. Omni-scale feature learning for person re-identification [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South); IEEE, 2019: 3701-3711.
- [13] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA; IEEE, 2016: 770-778.
- [14] HU Jie, SHEN Li, SUN Gang. Squeeze-and-excitation networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA; IEEE, 2018: 7132-7141.
- [15] KOMODAKIS N, ZAGORUYKO S. Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer [C]//5th International Conference on Learning Representations (ICLR). Toulon, France; dblp, 2017: 1-13.
- [16] ZHENG Liang, SHEN Liyue, TIAN Lu, et al. Scalable person re-identification: A benchmark [C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile; IEEE, 2015: 1116-1124.
- [17] LI Wei, ZHAO Rui, XIAO Tong, et al. DeepReID: Deep filter pairing neural network for person re-identification [C]//Proceedings of the 2015 IEEE International Conference on Computer Vision. Washington, DC: IEEE Computer Society, 2014: 152-159.
- [18] RISTANI E, SOLERA F, ZOU R S, et al. Performance measures and a data set for multi-target [C]//Proceedings of the 2016 European Conference on Computer Vision. Cham: Springer, 2016: 17-35.
- [19] ZHONG Zhun, ZHENG Liang, KANG Guoliang, et al. Random erasing data augmentation [C]//Proceedings of the AAAI Conference on Artificial Intelligence. AAAI, 2020, 34(07): 13001-13008.
- [20] KINGMA D P, BA J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.
- [21] LUO Hao, GU Youzhi, LIAO Xingyu, et al. Bag of tricks and a strong baseline for deep person re-identification [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Long Beach, USA; IEEE, 2019: 1487-1495.
- [22] CHEN Binghui, DENG Weihong, HU Jiani. Mixed high-order attention network for person re-identification [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South); IEEE, 2019: 371-381.
- [23] CHEN Tianlong, DING Shaojin, XIE Jingyi, et al. Abd-net: Attentive but diverse person re-identification [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea; IEEE, 2019: 8351-8361.
- [24] ZHENG Feng, DENG Cheng, SUN Xing, et al. Pyramidal person re-identification via multi-loss dynamic training [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seoul, Korea (South); IEEE, 2019: 8514-8522.
- [25] CHEN Xuesong, FU Canmian, ZHAO Yong, et al. Saliency-guided cascaded suppression network for person re-identification [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA; IEEE, 2020: 3297-3307.
- [26] CHEN Hao, LAGADEC B, BREMOND F. Learning discriminative and generalizable representations by spatial-channel partition for person re-identification [C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Snowmass, CO, USA; IEEE, 2020: 2483-2492.