

引用格式:朱宽堂,张建勋,谭暑秋. 基于全局特征和多种局部特征的行人重识别[J]. 微电子学与计算机, 2022, 39(2): 43-50.  
[ZHU K T, ZHANG J X, TAN S Q. Person re-identification based on global features and multiple local features[J].  
Microelectronics & Computer, 2022, 39(2): 43-50. ]DOI: 10. 19304/J. ISSN1000-7180. 2021. 0837

## 基于全局特征和多种局部特征的行人重识别

朱宽堂, 张建勋, 谭暑秋

(重庆理工大学 计算机科学与工程学院, 重庆 400054)

**摘要:** 能否提取具有区别度的行人特征是行人重识别问题的关键所在. 该文在 ResNet-50 的基础上, 提出一种使用多种方法提取特征的多分支网络结构(MMNet). 该网络将多种方法以一种巧妙的方式组合起来, 从而解决行人姿态变化、部分遮挡等引起的行人判别特征信息缺失问题. 在第一个分支中提取行人的全局特征, 在第二个分支中使用通道注意力模块提取想要关注的局部特征, 在第三个分支中将骨干网络提取的特征水平均匀的分割成不同的块, 从而提取出不同粒度的局部特征, 接着使用批量难样本三元组损失函数和 softmax 损失函数联合训练模型. 最终使用不同分支提取的特征串联在一起作为最终特征. 通过行人的全局特征和多种局部特征相互补充, 从而提取出更有区别度的行人特征. 算法在 Market-1501 和 DukeMTMC-reID 数据集上的平均精度均值和首位命中率分别达到 87.7%和 95.9%、79.9%和 89.2%. 试验结果表明, 使用多分支网络提取的特征具有互补性, 且平均精度均值和首位命中率比大多数行人重识别算法高.

**关键词:** 行人重识别; 多种方法; 多分支网络; 通道注意力; 局部特征

**中图分类号:** TP391

**文献标识码:** A

**文章编号:** 1000-7180(2022)02-0043-08

## Person re-identification based on global features and multiple local features

ZHU Kuantang, ZHANG jianxun, TAN Shuqiu

(College of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400054, China)

**Abstract:** The key of person re-identification is to extract distinctive pedestrian features. Based on ResNet-50, this paper proposes a multi-branch network structure (MMNet) which uses multiple methods to extract features. By combining various methods in a clever way, the problem of lack of pedestrian discrimination feature information caused by pedestrian posture change and partial occlusion can be solved. In the first branch, the global features of pedestrians are extracted. In the second branch, the local features that person re-identification want to pay attention to are extracted by the channel attention module. In the third branch, the features extracted from backbone are evenly divided into different blocks, so as to extract the local features with different granularity. Then the model is trained by using the bath hard triple loss function and softmax loss function. Finally, the features extracted from different branches are concatenated as the final features. By complementing the global features and various local features of pedestrians, distinctive pedestrian feature is extracted. The map and rank 1 of the algorithm on Market-1501 and DukeMTMC-reID datasets are 87.7% and 95.9%, 79.9% and 89.2% respectively. The experimental results show that the features extracted by the multi-branch network are complementary, and map and rank1 are higher than that of most person re-identification algorithms.

**Key words:** person re-identification; multiple methods; multi-branch network; channel attention; local feature

## 1 引言

随着社会的发展和人们安全意识的提高,大量的监控摄像头被安装在街道、商场、飞机场等公共场所,由此产生了庞大的监控视频数据.为了使用这些数据来保障人们的安全,提高人们的生活质量,研究者们提出了行人重识别技术来处理这些数据.行人重识别(Person Re-identification)也称行人再识别,是根据给定监控行人的图片,判断其他监控设备下是否存在这个行人的技术,是图像检索的一类子问题<sup>[1]</sup>.行人重识别可以与行人跟踪等技术相结合,其被广泛应用在刑侦、无人超市、安保等领域.近年来,随着深度学习的兴起,行人重识别技术成为计算机视觉领域的热点问题.

由于现实场景的复杂性,相比于其他计算机视觉技术,行人重识别存在着许多不同的挑战.包括:1)由于行人一般是处于运动状态的,导致目标行人的不同图片姿态大幅度变化,目标行人的背景差异大.2)由于摄像机本身质量的限制,得到的图片分辨率不高,比较模糊.3)由于光线的影响,图片的颜色和对比度发生变化.4)由于拍摄角度的问题,可能拍摄不到目标行人的正脸,行人也可能被车辆、雨伞等其他物体遮挡,从而得不到完整的行人.

目前,使用全局特征和局部特征相结合的行人重识别算法大多数只使用某一种方法提取局部特征,不同方法提取的特征具有不同的特性,为了提取出更能区别不同行人的特征,将多种方法通过一种巧妙的方式组合起来,从而解决行人姿态变化、部分遮挡等引起的行人判别特征信息缺失问题.本文设计一种使用多种方法提取特征的多分支网络结构(MMNet),三个分支分别是全局分支、注意力分支和水平金字塔分支.本文使用全局分支提取行人的整体躯干信息,使用注意力分支提取感兴趣的行人局部信息,使用水平金字塔分支提取多粒度的行人局部信息,最后将三个分支得到的特征联合训练,提取出更有区别度的特征.

## 2 相关工作

在2014年以前,主要使用传统的方法进行行人重识别的研究,传统方法主要使用人工设计的特征,常见的人工特征包括颜色特征、形状特征和纹理特征等.由于人工特征的描述能力有限,因此传统的行人重识别算法在复杂环境中不能得到很好的效果.其次,虽然传统的算法在某个数据集中表现良好,但

是如果更换数据集,传统的算法不能保证依旧能提取出有区别度的特征.随着行人重识别的发展,其数据集的规模不断变大,这时传统的行人重识别算法的相似性度量求解变得很困难.

2014年,Li等人<sup>[2]</sup>提出了基于卷积神经网络的FPNN(Filter Pairing Neural Network)网络模型,并在行人重识别的数据集上取得良好的效果,行人重识别的重点开始转向深度学习.由于Resnet网络在图像分类领域取得优异的成绩,将行人图像输入微调后的Resnet网络提取行人图像的全局特征成为主要的研究方法.例如SVDNet<sup>[3]</sup>网络将奇异值分解引入全连接层.在后期的研究中,通常将Resnet网络作为基准框架.当研究行人图像的全局特征遇到性能瓶颈后,研究者开始设计能提取更精细特征的网络模型.

一种方法是通过姿态估计、关键点定位等先验知识定位局部区域,使得模型能够学习到具有语义信息的局部特征. Wei等人<sup>[4]</sup>提出了GLAD(Global-Local-Alignment Descriptor)网络模型,模型采用deep cut算法估计出行人的四个关键点,根据四个关键点把图像分成头部、上半身和下半身三大部分,接着将这三部分图像和全局图像同时输入网络,提取出四个特征,最后将它们串联起来作为最终特征,GLAD有效的解决了姿态差异和图像不对称问题. Su等人<sup>[5]</sup>提出了PDC(Pose-driven Deep Convolutional model)网络模型,模型由两个子网构成,第一个是特征嵌入子网(Feature Embedding sub-Net, FEN),FEN通过姿态估计的14个关键点将行人图像分成6个部分,然后对6个部分图像进行标准化,并调整图像的姿势.第二个是特征加权子网(Feature Weighting sub-Net, FWN),FWN自动学习不同部位的权重,从而可以更好的进行特征融合.

另一种方法是将网络提取的特征分割成预定义的条状.孙奕帆等人<sup>[6]</sup>提出PCB(Part-based Convolutional Baseline)网络将最终提取的特征水平均匀的分成6个部分,并通过多个softmax损失函数训练每个部分,同时提出RPP(Refined Part Pooling)策略解决语义对齐问题.旷视科技<sup>[7]</sup>提出的AlignedReID网络使用相同的方法将特征分成7个部分,但是它使用最短路径损失对齐行人的局部. Wang等人<sup>[8]</sup>提出一个多分支深度网络MGN,网络将ResNet-50的第四层及后面的部分分成三个支路,每个分支将特征纵向平均分割为不同的块,以此获得具有多个粒度的局部特征表示,并使用三元组损

失函数和交叉熵损失函数联合训练网络, MGN 在多个行人重识别数据集上取得了良好的效果.

还有一种方法是通过注意力机制提取感兴趣的局部区域. Liu 等人<sup>[9]</sup>提出 HPNet(HydraPlus network)网络,通过将注意力机制应用在不同的特征层,提取出更丰富的行人特征表示. Li 等人<sup>[10]</sup>提出参数量少的网络 HA-CNN,通过同时学习行人图像中的硬注意力和软注意力,得到行人不变性特征表示.

### 3 多分支网络

#### 3.1 三个分支介绍

本文以 ResNet50 作为骨干网络. 与最初的 ResNet50 的不同之处在于: 删除最后的池化层和全

连接层. 在保持 res\_conv4\_2 之前的所有模块不变的情况下, 将后面的模块分成三个独立的分支. 如图 1 所示, 从上到下分别是全局分支(Global branch), 注意力分支(attention branch), 水平金字塔分支(HP branch). 其中, 图中的长方体表示特征图; 含有 CA 标志的矩形框表示通道注意力模块; 含有 bottleneck 标志的矩形框表示 bottleneck 层; 含有 Triplet 标志的矩形框表示难样本三元组损失函数; 含有 softmax 标志的矩形框表示 softmax 损失函数; 含有 256 标志的矩形框表示 256 维的特征; 含有 2048 标志的矩形框表示 2048 维特征; 含有 test 标志的方框表示测试阶段. 为了防止噪音的干扰, 整个网络只共享骨干网络的部分参数, 其余参数都不共享.

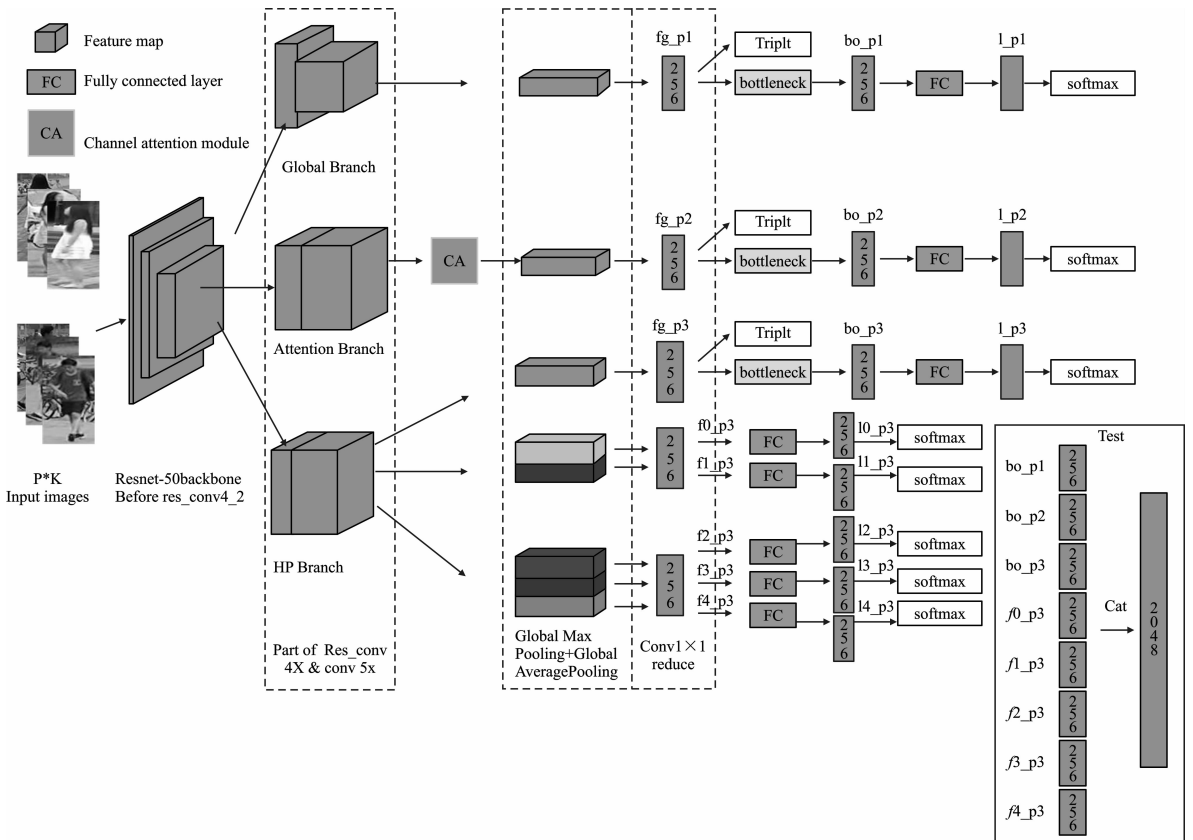


图 1 多分支网络结构(MMNet)

Fig. 1 Multi-branch network(MMNet)

局分支(Global branch)并不改变骨干网络的网络结构. 最大池化可以从特征图中提取出最具有区别度的信息, 平均池化可以从特征图中提取出每个空间单位的全局信息<sup>[11]</sup>. 由于只使用某一种池化可能会丢失图像的部分信息, 所以本文同时使用全局平均池化和全局最大池化作为骨干网络之后的池

化层, 并使用它们的和作为最后的池化结果. 在该分支, 数据经过骨干网络和卷积核为  $12 \times 4$  的池化层之后, 接着通过  $1 \times 1$  的卷积层将特征图维数从 2048 维降低到 256 维得到特征  $fg\_p1$ , 特征  $fg\_p1$  通过 bottleneck 层得到特征  $bo\_p1$ , 最后通过全连接层得到特征  $l\_p1$ .

注意力分支(attentionbranch)与全局分支具有相似的结构,但是存在两处不同的地方.第一处是在 ResNet-50 的 res\_conv5\_1 区域没有使用下采样,这使得骨干网络输出的特征图增大了一倍,此方法只增加了很小的计算成本,却保留了更多的行人细粒度特征.第二处是在 ResNet-50 的第五层之后增加了通道注意力模块.如图 2 所示.通道注意力模块通过显式地建模通道之间的相互依赖关系,自适应地重新校准通道的特征响应<sup>[12]</sup>.即给输入特征图的每个通道添加权重,而权重值的大小表示该特征通道行人信息的重要程度,根据重要程度抑制无关的特征,从而提取出想要关注的行人特征.若输入特征图的尺寸大小为  $W \times H \times C$  ( $W$  表示特征图的宽度,  $H$  表示特征图的高度,  $C$  表示特征图的通道数).模块的具体实现过程为:输入特征图首先通过一个全局平均池化层将特征图压缩为  $1 \times 1 \times C$ ,再通过一个两级的全连接神经网络对压缩后的特征图进行线性变换,接着通过 sigmoid 得到特征图每个通道的权重,最后使用 scale 操作将每个权重乘到输入特征图上,得到尺寸大小不变的输出特征图.在该分支,通道注意力模块的输出特征经过核为  $24 \times 8$  的池化层和  $1 \times 1$  的卷积层得到 256 维的特征 fg\_p2,特征 fg\_p2 通过 bottleneck 层得到特征 bo\_p2,最后通过全连接层得到特征 l\_p2.

为了确定使用的注意力分支,本文分别在骨干网络之后分别设置通道注意力机制、空间注意力机制和 CBAM(Convolutional Block Attention Module)<sup>[13]</sup>进行实验,最终得到的 map 分别为 87.7、87.0 和 86.9,根据实验结果,最终在注意力分支应用通道注意力机制.

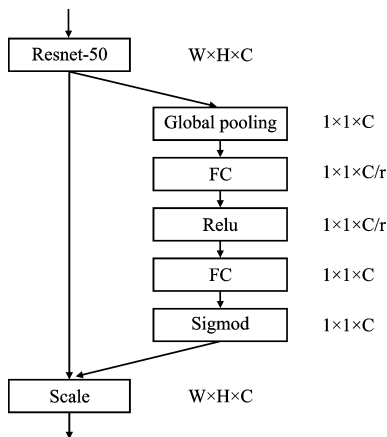


图 2 通道注意力模块

Fig. 2 Channel attention module

水平金字塔分支(HP branch)中骨干网络的 res\_conv5\_1 区域同注意力分支一样没有使用下采样.这个分支根据池化层的卷积核的大小不同,以各种粒度水平均匀的切割特征图,使得网络可以充分利用人体各部分的信息.在缺少某些关键部分的情况下,网络也仍然可以识别出正确的行人<sup>[14]</sup>.将该分支的骨干网络之后的部分分成三个独立的小分支,第一个小分支使用卷积核为  $24 \times 8$  的池化层把特征图分成一块,之后经过  $1 \times 1$  的卷积层得到降维后的特征 fg\_p3,特征 fg\_p3 通过 bottleneck 层得到特征 bo\_p3,最后通过全连接层得到特征 l\_p3.第二个小分支和第三个小分支分别使用卷积核为  $12 \times 8$ 、 $8 \times 8$  的池化层把特征图水平均匀的分成两块和三块,然后将其送往  $1 \times 1$  的卷积层,其中第二个小分支得到特征 f0\_p3 和 f1\_p3,最后经过全连接层得到特征 l0\_p3 和 l1\_p3.第三个小分支得到特征 f2\_p3、f3\_p3 和 f4\_p3,最后经过全连接层得到特征 l2\_p3、l3\_p3 和 l4\_p3.每个小分支的在学习过程中并不共享参数,从而增强特定粒度下人体各部分的信息的区别度.

为了确定水平金字塔分支的第二和第三小分支,本文在 Market1501 数据集上进行实验,分别将第二、第三小分支的分割粒度设置为 2 块和 3 块、2 块和 4 块、3 块和 4 块.得到的 map 分别为 87.7、86.8、87.1.根据实验结果,最终设置第二和第三小分支的分割粒度为 2 块和 3 块.

在算法的测试部分,本文使用了全局分支、注意力分支和水平金字塔分支的第一个小分支中经过 bottleneck 层之后得到的特征  $\{bo\_p1, bo\_p2, bo\_p3\}$ ,接着使用水平金字塔分支的第二和第三小分支中经过降维之后的特征  $\{f0\_p3, f1\_p3, f2\_p3, f3\_p3, f4\_p3\}$ ,把这些特征直接串联在一起得到 2048 维的最终特征  $f$ .如公式(1)所示:

$$f = bo\_p1 + bo\_p2 + bo\_p3 + f0\_p3 + f1\_p3 + f2\_p3 + f3\_p3 + f4\_p3 \quad (1)$$

### 3.2 分支和损失函数

在网络的训练阶段,本文使用  $L_{softmax}$  (softmax 损失函数)用于分类,其如公式(2)所示.其中,  $W_k^T$  代表数据集中行人类别  $k$  的权重向量,  $f_i$  代表行人的特征向量,  $C$  代表数据集,  $N$  代表训练过程时小批量梯度下降中批量的大小.本损失函数应用于每个分支的全连接层之后的特征,分别是  $\{l\_p1, l\_p2, l\_p3, l0\_p3, l1\_p3, l2\_p3, l3\_p3, l4\_p3\}$

$$L_{\text{softmax}} = - \sum_{i=1}^N \ln \frac{e^{w_{y_i}^T f_i}}{\sum_{k=1}^C e^{w_k^T f_i}} \quad (2)$$

本文使用  $L_{\text{triplet}}$  (batch hard triplet loss)<sup>[7]</sup> 用于度量学习,其如公式(3)所示. 其中,在一个批量大小为  $P \times k$  的小批量梯度下降中( $p$  表示从数据集中选取的行人数量,  $k$  表示每个行人选择的照片数量),  $f_a$  是随机选择的行人(anchor)的特征,  $f_p$  是与 anchor 具有相同行人身份的特征,  $f_n$  是与 anchor 具有不同行人身份的特征,  $\| * \|$  是欧式距离,  $\alpha$  是控制两个欧式距离之间最小间隔的超参数. 本损失函数设置在全局分支、注意力分支和水平金字塔分支的第一个小分支上. 应用于  $1 \times 1$  卷积层之后的特征, 分别是  $\{fg\_p1, fg\_p2, fg\_p3\}$ .

$$L_{\text{triplet}} = - \sum_{i=1}^P \sum_{a=1}^K [\alpha + \max_{p=1 \dots K} \|f_a^{(i)} - f_p^{(i)}\|_2 - \min_{\substack{n=1 \dots K \\ j=1 \dots P \\ j \neq i}} \|f_a^{(i)} - f_n^{(j)}\|_2]_+ \quad (3)$$

对于嵌入空间中的图像对,三元组损失函数主要优化欧几里得距离,而 softmax 损失函数主要优化余弦距离<sup>[11]</sup>. 若使用 softmax 损失函数和难样本三元组损失函数同时优化某一特征向量,可能出现一种损失减少,而另一种损失正在振荡. 为了解决这种问题,本文在全连接层前面增加了批处理归一化层(bottleneck 层),批处理归一化层可以平衡特征向量的每个维度并保持同身份人的特征的紧凑分布. 由于行人未对齐等问题,在经过水平分割后的行人图片的局部特征上使用三元组损失函数,可能使得模型在训练过程中学习到奇怪的特征,因此本文没有在水平金字塔的第二和第三小分支设置三元组损失函数.

本文算法的总损失为每种损失函数的算术平均数之和,为了减少训练时分支之间的相互干扰,本文在三元组损失的算术平均数前面添加权重参数. 总损失如公式(4)所示,其中  $\lambda$  表示三元组损失函数的权重. 本文在 Market1501 数据集上实验时,为了降低复杂度,将  $\lambda$  的值分别设置为 0.1、0.5、0.8、0.9、1、2、10. 最终得到的 map 分别为 86.2、87.2、87.3、87.7、87.4、86.8、81.2. 根据实验结果,最终设置  $\lambda$  的值为 0.9.

$$L_{\text{total}} = \frac{\sum_{i=1}^8 L_{\text{softmax}}}{8} + \lambda \frac{\sum_{i=1}^3 L_{\text{triplet}}}{3} \quad (4)$$

## 4 实验结果与分析

本文算法由基于 python 的深度学习框架 Pytorch 实现. 在 GPU 配置是英伟达 3070 的电脑上进行模型训练和测试评估. 在训练阶段,将输入图像的大小调整为  $384 \times 128$ ,并以 0.5 的概率使用随机擦除法进行数据增强. 模型总共训练 300 个 epoch,采用小批量梯度法进行梯度下降,每个 minibatch 中随机选择 8 个人,并为每个人随机选择 4 张图像作为小批量样本. 初始学习率设置为 0.0002,每 60 个 epoch 学习率降低为原来的 0.1 倍. 模型使用 Adam 作为优化器,三元组损失函数的超参数设置为 0.3. 在评估过程中,批量大小设置为 16. 采用余弦距离计算样本之间的距离. 为了验证本文模型的效果,分别在行人重识别常用的两个数据集 Market1501 和 DukeMTMC-reID 上进行实验,并采用 map 和 rank n 作为评价标准.

### 4.1 数据集

Market1501 数据集<sup>[15]</sup>是在清华大学校园中使用六台摄像机(包括 5 个高清摄像头和 1 个低清摄像头)采集的. 整个数据集含有 1501 个行人,总共有 32668 个检测到的行人矩形框,其中训练集有 751 个行人,含有 12936 张图像,测试集有 750 个行人,含有 19732 张图像. market1501 数据集使用 DPM 自动检测并切割出行人,其结果包含一些检测误差(可以更接近实际使用). 每个行人至少被 2 个摄像头拍到、最少被 6 个摄像头拍到,并且每个行人在一个相机下可能具有多个图像.

DukeMTMC 数据集是在 Duke 大学校园拍摄的. 包含了 8 个摄像头拍摄的 85 分钟的高分辨率视频. 而 DukeMTMC-Reid<sup>[16]</sup>是使用 DukeMTMC 专门制作的行人重识别数据集. 整个数据集是从原始数据集中每 120 帧采样一张图片,总共得到了 36411 张图像. 其中有 408 个行人只出现在一个摄像头下,1404 个行人出现在不只一个摄像头下. 测试集包含 702 个行人,一共有 17661 张图像,训练集有 702 个行人,一共有 16522 张图像.

### 4.2 评价标准

Rank n 表示当查询图像与底库中所有图片计算相似度并按从大到小排序之后,排序结果中前 n 张图有正确结果的概率. 例如:rank 1 表示首位命中率,即排序结果的第一位是不是他本人. rank 5

表示排序结果的前五位有没有他本人. 一般来说, 行人重识别从 rank  $n$  中选择几个关键点作为评价标准. 本实验选择 rank1 和 rank5 作为评价标准. Map(mean average precision) 平均精度均值平衡了查准率和查全率. 不同于 rank  $n$ , map 并不是只看排序结果的前  $n$  个的正确概率, 而是反映在底库中所有与查询图像同类的图片在排序结果队列前面的程度. 几张正确的图片排序靠前并不能使 map 的值变高, 所以 map 能更加全面的衡量行人重识别算法效果.

### 4.3 实验仿真与结果分析

#### 4.3.1 模型有效性验证

根据分支的数量和类型可把网络结构分成七种. 首先是仅有单分支的网络结构: 全局分支, 注意力分支, 水平金字塔分支. 其次是两个分支互相组合的网络结构: 全局分支和注意力分支, 全局分支和水平金字塔分支, 注意力分支和水平金字塔分支. 最后是三个分支一起组合的网络结构: 全局分支、注意力分支和水平金字塔分支. 在 Market1501 数据集和 DukeMTMC-reID 数据集上的相关实验如表 1 和表 2 所示.

从表 1 和表 2 可知: 在 Market1501 数据集和 DukeMTMC-reID 数据集上, 单分支的网络结构的各项精度相对于其他两种网络结构都是最低的, 所有的两分支组合在一起的网络结构的各项精度都比原有的单分支的高. 这说明每个分支都能提取到其他分支提取不到的特征. 而三个分支提取的特征相互补充, 使得三分支组合的网络结构的各项精度达到最高.

表 1 在数据集 Market1501 上, 七种网络结构的对比

Table 1 The comparison of seven network structures on market1501

对比网络结构	Map	Rank 1	Rank 5
Global branch	73.0	88.7	96.2
Attention branch	76.2	90.0	96.9
HP branch	83.8	91.7	96.3
Global branch + Attention branch	81.3	92.5	97.3
Global branch + HP branch	85.1	93.3	97.1
Attention branch + HP branch	85.5	93.9	97.1
Global branch + Attention branch + HP branch	87.7	95.9	98.3

表 2 在数据集 DukeMTMC-reID 上, 七种网络结构的对比

Table 2 The comparison of seven network structures on DukeMTMC-reID

对比网络结构	Map	Rank 1	Rank 5
Global branch	64.3	80.3	90.2
Attention branch	67.3	82.0	91.5
HP branch	76.1	85.9	92.9
Global branch + Attention branch	72.2	85.0	93.1
Global branch + HP Branch	77.6	87.9	93.6
Attention branch + HP branch	77.4	87.5	93.9
Global branch + Attention branch + HP branch	79.9	89.2	95.1

部分实验数据的可视化结果如图三所示, 其中第一列 Query 表示行人查询图像, 后面十列表示对应图片的查询结果, 红框表示查询结果出错.



图 3 部分行人查询结果图

Fig. 3 Query results of some pedestrian

#### 4.3.2 本文算法与其他先进算法的对比

为了进一步验证本文算法的有效性, 本文选取了 6 个先进的行人重识别算法与本文算法进行实验对比. 分别是: 使用全局特征的方法 (SVDNet, OursNIR<sup>[17]</sup>, DBF<sup>[18]</sup>), 使用先验知识提取局部特征的方法 (GLAD)、使用注意力机制的方法 (HACNN, TSNet<sup>[19]</sup>) 和将最终特征分割为预定义条状的局部特征的方法 (AlignedReID、PCB + RPP、MGN). 在 Market1501 数据集和 DukeMTMC-reID 数据集的对比结果如表 3 所示.

从表 3 可知: 在 Market1501 数据集和 DukeMTMC-reID 数据集上, 本文算法的 map 分别达到 87.7% 和 79.9%, Rank 1 分别达到 95.9% 和 89.2%. MGN 算法和本文的算法性能表现相似, MGN 算法也是使用全局特征结合局部特征的方法, 但是它只使用一种方法提取局部特征, 而本文使用两种

不同的方法提取局部特征. 在 Market1501 数据集上,本文算法的 map 比 MGN 算法提高 0.8%, Rank 1 提高 0.2%. 在 DukeMTMC-reID 数据集上,本文算法的 map 比 MGN 算法提高 1.5%, Rank 1 提高 0.4%. 根据实验结果,相比于其他先进的行人重识别算法,本文的算法的各项精度最高,证明了本文算法的有效性.

表3 与先进方法的对比

Table 3 Comparison with advanced methodss

方法	Market1501		DukeMTMC-reID	
	Map	Rank 1	Map	Rank 1
SVDNet <sup>[2]</sup>	62.1	82.3	56.8	76.7
GLAD <sup>[3]</sup>	73.9	89.9	—	—
HA-CNN <sup>[9]</sup>	75.7	91.2	63.8	80.5
AlignedReID <sup>[6]</sup>	77.7	90.6	67.4	81.2
PCB+RPP <sup>[5]</sup>	81.6	93.6	69.2	83.3
MGN <sup>[7]</sup>	86.9	95.7	78.4	88.7
OursNIR <sup>[17]</sup>	83.1	—	—	—
DBF <sup>[18]</sup>	85.6	95.0	77.1	87.1
TSTNet <sup>[19]</sup>	86.39	94.74	75.01	86.01
本文方法	87.7	95.9	79.9	89.2

## 5 结束语

本文提出一种融合全局分支、注意力分支和水平金字塔分支的算法,把行人的全局特征与不同方法提取的局部特征相结合,通过各种特征的互补作用,在无需姿态估计等先验知识的情况下,使用 softmax 损失函数和 batch hard 三元组损失函数联合训练模型,提取出具有区别度的行人特征. 最后在 Market1501 数据集和 DukeMTMC-reid 数据集上进行了相关实验,其实验结果表明,本文的算法具有有效性,提高了行人重识别的精度.

## 参考文献:

[1] 宋婉茹, 赵晴晴, 陈昌红, 等. 行人重识别研究综述 [J]. 智能系统学报, 2017, 12(6): 770-780. DOI: 10.11992/tis.201706084.  
SONG W R, ZHAO Q Q, CHEN C H, et al. Survey on pedestrian re-identification research [J]. CAAI Transactions on Intelligent Systems, 2017, 12(6): 770-780. DOI: 10.11992/tis.201706084.

[2] LI W, ZHAO R, XIAO T, et al. DeepReID: deep filter pairing neural network for person re-identification [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 152-159. DOI: 10.1109/CVPR.2014.27.

[3] SUN Y F, ZHENG L, DENG W J, et al. SVDNet for pedestrian retrieval [C]//IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 3820-3828. DOI: 10.1109/ICCV.2017.410.

[4] WEI L H, ZHANG S L, YAO H T, et al. GLAD: global-local-alignment descriptor for Scalable person re-identification[J]. IEEE Transactions on Multimedia, 2019, 21(4): 986-999. DOI: 10.1109/TMM.2018.2870522.

[5] SU C, LI J N, ZHANG S L, et al. Pose-driven deep convolutional model for person re-identification [C]//Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 3980-3989. DOI: 10.1109/ICCV.2017.427.

[6] SUN Y F, ZHENG L, YI Y, et al. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline) [C]//Proceedings of the 2018 15th European Conference on Computer Vision. Munich: Springer, 2018: 501-518. DOI: 10.1007/978-3-030-01225-0\_30.

[7] ZHANG X, LUO H, FAN X, et al. AlignedReID: surpassing human-level performance in person re-identification[J]. arXiv: 1711.08184, 2017.

[8] 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. Seoul: Association for Computing Machinery, 2018: 274-282. DOI: 10.1145/3240508.3240552.

[9] LIU X H, ZHAO H Y, TIAN M Q, et al. HydraPlus-Net: attentive deep features for pedestrian analysis [C]//Proceeding of IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 350-359. DOI: 10.1109/ICCV.2017.46.

[10] LI W, ZHU X T, GONG S G. Harmonious attention network for person re-identification [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 2285-2294. DOI: 10.1109/CVPR.2018.00243.

[11] 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 the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Long Beach: IEEE, 2019: 1487-1495. DOI: 10.1109/CVPRW.2019.00190.

[12] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).



- Salt Lake City: IEEE, 2018: 7132-7141. DOI: 10.1109/CVPR.2018.00745.
- [13] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich: Springer, 2018: 3-19. DOI: 10.1007/978-3-030-01234-2\_1.
- [14] 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 (1): 8295-8302. DOI: 10.1609/aaai.v33i01.33018295.
- [15] ZHENG L, SHEN L Y, TIAN L, et al. Scalable person re-identification: a benchmark[C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE, 2015: 1116-1124. DOI: 10.1109/ICCV.2015.133.
- [16] RISTANI E, SOLERA F, ZOU R, et al. Performance measures and a data set for multi-target, multi-camera tracking[C]//European Conference on Computer Vision. Amsterdam: Springer, 2016: 17-35. DOI: 10.1007/978-3-319-48881-3\_2.
- [17] 王林, 杨乐. 嵌套池化三元组卷积神经网络的行人再识别[J]. 微电子学与计算机, 2019, 36(9): 73-78. DOI: 10.19304/j.cnki.issn1000-7180.2019.09.014.
- WANG L, YANG L. Pedestrian re-identification of nested pooling triple convolutional neural networks [J]. Microelectronics & Computer, 2019, 36(9): 73-78. DOI: 10.19304/j.cnki.issn1000-7180.2019.09.014.
- [18] 熊伟, 熊子婕, 杨获椿, 等. 基于深层特征融合的行人重识别方法[J]. 计算机工程与科学, 2020, 42(2): 358-364. DOI: 10.3969/j.issn.1007-130X.2020.02.022.
- XIONG W, XIONG Z J, YANG D C, et al. Pedestrian re-identification based on deep feature fusion [J]. Computer Engineering & Science, 2020, 42(2): 358-364. DOI: 10.3969/j.issn.1007-130X.2020.02.022.
- [19] 张晓艳, 张宝华, 吕晓琪, 等. 深度双重注意力的生成与判别联合学习的行人重识别[J]. 光电工程, 2021, 48 (5): 200388. DOI: 10.12086/oe.2021.200388.
- ZHANG X Y, ZHANG B H, LV X Q, et al. The joint discriminative and generative learning for person re-identification of deep dual attention [J]. Opto-Electronic Engineering, 2021, 48(5): 200388. DOI: 10.12086/oe.2021.200388.

#### 作者简介:

朱宽堂 男, (1997-), 硕士研究生. 研究方向为数字图像处理与分析.

张建勋(通讯作者) 男, (1971-), 博士, 教授. 研究方向为数字图像处理与分析、实时计算机图形学.

E-mail: 466908695@qq.com

谭暑秋 女, (1982), 博士, 讲师. 研究方向为图像处理、人脸识别.