

# 基于连接注意力的行人重识别特征提取方法

魏紫薇<sup>1,2</sup>, 屈丹<sup>2</sup>, 柳聪<sup>2</sup>

(1. 郑州大学 软件学院, 郑州 450001; 2. 战略支援部队信息工程大学 信息工程学院, 郑州 450001)

**摘要:** 全民安全意识的逐步提高使得智能监控设备遍布各大公共场所, 行人重识别作为视频分析的关键技术之一, 被广泛应用于智能安防、自动驾驶等领域。为了提高真实环境下跨摄像头行人检索的识别精度, 提出一种基于注意力机制的行人重识别特征提取方法。在数据处理阶段, 考虑不同训练数据量下识别效果存在差异的问题, 对行人图片采用自动增强方法进行数据增强, 以提高数据集规模。在特征提取阶段, 将连接注意力模块与ResNet50残差网络相结合构成特征提取网络, 提取显著性更强的行人特征。在损失优化阶段, 采用三元组损失和圆损失对行人特征进行优化并完成距离度量, 最终根据距离的远近得到行人排序结果。实验结果表明, 在Market1501数据集上该方法的Rank-1值和mAP值分别达到95.90%和89.66%, 在DukeMTMC-reID数据集上Rank-1值和mAP值分别达到91.16%和81.24%, 在MSMT17数据集上Rank-1值和mAP值分别达到84.37%和62.73%, 与现有经典行人重识别方法PCB、MGN、Pyramid、OSNet等相比, 其网络识别性能评价指标均有明显提升。

**关键词:** 行人重识别; 深度学习; 卷积神经网络; 注意力机制; 自动增强

开放科学(资源服务)标志码(OSID):



中文引用格式: 魏紫薇, 屈丹, 柳聪. 基于连接注意力的行人重识别特征提取方法[J]. 计算机工程, 2022, 48(7): 220-226.

英文引用格式: WEI Z W, QU D, LIU C. Method for person re-identification feature extraction based on connected attention[J]. Computer Engineering, 2022, 48(7): 220-226.

## Method for Person Re-identification Feature Extraction Based on Connected Attention

WEI Ziwei<sup>1,2</sup>, QU Dan<sup>2</sup>, LIU Cong<sup>2</sup>

(1. Software College, Zhengzhou University, Zhengzhou 450001, China;

2. School of Information System Engineering, Strategic Support Force Information Engineering University, Zhengzhou 450001, China)

**[Abstract]** The gradual increase in national safety awareness has promoted the spread of intelligent monitoring equipment across major public places. Person re-identification, a key technique for video analysis, is widely used in intelligent security, automatic driving, and other fields. A method for person re-identification feature extraction based on the attention mechanism is proposed to improve the identification accuracy of cross camera person retrieval in real environments. During the data-processing stage, the automatic enhancement method was adopted for person pictures to enhance the data set size, considering the difference in the identification effect under different amounts of training data. During the feature extraction stage, the connection attention module and ResNet50 residual network were combined to form a feature extraction network to extract more significant person features. In the loss optimization stage, triple and circle losses were used to optimize the person characteristics and complete the distance measurement. Finally, the person ranking results were obtained according to the distance. The experimental results showed that the Rank-1 and mAP values of this method on the Market1501 data set were 95.90% and 89.66%, respectively. The Rank-1 and mAP values on the DukeMTMC-reID data set were 91.16% and 81.24%, respectively, and those on the MSMT17 data set were 84.37% and 62.73%, respectively. Compared to existing classical person re-identification methods (PCB, MGN, Pyramid, and OSNet), the network identification performance evaluation indexes improved significantly.

**[Key words]** person re-identification; deep learning; Convolutional Neural Network (CNN); attention mechanism; automatic enhancement

DOI: 10.19678/j.issn.1000-3428.0061884

## 0 概述

行人重识别也称行人再识别, 是指在无交叉视

域的多个摄像头监控系统中, 给定一个摄像机拍摄的行人图像, 利用深度学习、人工智能等技术从其他摄像机中重新识别出该行人。行人重识别作为智能图像处

基金项目: 国家自然科学基金(61673395, 62171470); 郑州市重大科技攻关项目(188PCXZX773)。

作者简介: 魏紫薇(1996—), 女, 硕士研究生, 主研方向为人工智能、智能信息处理、行人重识别; 屈丹, 教授、博士生导师; 柳聪, 硕士研究生。

收稿日期: 2021-06-09 修回日期: 2021-08-30 E-mail: qudanqudan@sina.com

理的关键技术之一,在智能安防、智能追踪、智能管理等应用场景中得到广泛关注与应用<sup>[1-2]</sup>。

行人重识别系统通常由图像预处理、图像特征提取、网络损失优化、行人排序4个部分组成:图像预处理对输入图片大小、分辨率等进行处理;图像特征提取旨在提取显著性更强的行人特征,是系统中的关键步骤;网络损失优化是在损失函数最小化原则下通过某种优化算法来寻找最优参数;行人排序是对行人图像之间的相似度依次进行排序并得到排序结果。

在诸多行人特征提取方法中,全局特征提取<sup>[3]</sup>较早被应用到行人重识别网络中,其缺点是不易捕捉图片之间的细微区别。在这之后,很多研究人员提出全局特征与局部特征相结合的方法<sup>[4-6]</sup>,使得行人重识别性能得到大幅提升。尽管行人重识别技术取得了很大的进展,但是受遮挡、姿态变化、光线变化、图像分辨率等因素影响,其识别精度仍难以满足现实需求<sup>[7-8]</sup>。为解决该问题,很多研究人员提出基于注意力机制的行人重识别网络<sup>[9-11]</sup>,该类网络利用注意力机制、深度模型来关注行人图像的显著特征,从而降低各种环境变化和背景干扰所带来的负面影响<sup>[12-14]</sup>。

文献[15]提出包含非局部注意力模块(Non-Local Attention Module, A)、广义平均池化(Generalized-Mean Pooling, G)、加权三元组损失(Weighted Regularization Triplet, W)的AGW网络,该网络在提取某处特征时利用其周围点的信息,获得了更大的感受野以及更广泛的信息分布。然而,AGW网络是对当前卷积模块进行注意学习从而训练模型,在训练过程中,当前注意块没有前面的监督信息来进行指导。文献[16]提出连接注意力模块,其通过当前注意块来从前面注意块中收集信息,并将其传递给下一个注意块,使得注意块之间的信息可以相互配合,从而提高注意力模块的学习能力,最终大幅提升网络的特征学习效果。

针对真实环境下行人重识别网络识别精度低的问题,本文提出一种基于连接注意力机制的行人重识别特征提取方法,旨在提取显著性更强的行人图像特征以实现行人识别。使用自动增强技术<sup>[17]</sup>对图像进行预处理,获取关于平移、旋转、剪切等多种图像处理功能的融合策略,以实现有效的数据增强,进而提升模型的鲁棒性。采用基于连接注意力机制的行人重识别网络模型对行人图像进行特征提取,获得显著性更强的行人特征进行相似度度量。此外,利用三元组损失和圆损失进行损失优化,以获得更高的行人重识别精度。在Market1501<sup>[18]</sup>、DukeMTMC-reID<sup>[19]</sup>、MSMT17<sup>[20]</sup>这3个主流的行人重识别数据集上进行实验,以验证本文基于连接注意力的行人重识别网络的识别性能。

## 1 行人重识别特征提取网络

如图1所示,本文行人重识别特征提取网络分为4个步骤:

1)对输入的行人图片进行预处理。

2)将预处理后的图片输入以ResNet50作为主干网络、卷积块2~卷积块5分别结合连接注意力模块的整体网络结构中进行特征提取。

3)对提取的行人特征进行损失优化并完成相似度度量。

4)进行行人排序,得到行人重识别系统的识别结果。

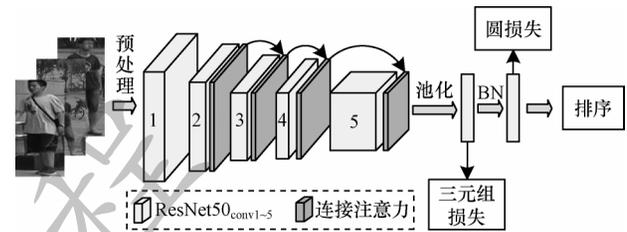


图1 行人重识别特征提取网络结构

Fig.1 Person re-identification feature extraction network structure

### 1.1 基于自动增强的图像预处理

在基于深度学习的网络模型训练过程中,网络模型由最初的浅层网络发展到目前更深、更宽的网络,因此,对数据量的需求也大幅提高,从而避免模型的过拟合问题。在真实环境中,由于数据标注成本等因素影响,导致训练数据欠缺,需要对数据集进行数据扩增。传统的数据增强方法包括尺寸变换、颜色增强等,随着深度学习中更深网络层的出现,简单的数据增强方法已经无法满足各类场景的应用需求。在行人重识别场景中,光照、角度、分辨率等因素的影响使得深度网络的学习更为困难。

自动增强是一种自动搜索合适数据进行标注数据拓展的算法,其包含2个部分:

1)搜索空间:确定图像变换增强的选择范围。

2)搜索算法:确定图像增强选择范围内的优选结果。

搜索空间中的一个图像增强策略包含5个子策略,在训练时为每个小批量中的每张图像选择1个子策略,因此,数据扩展为原始数据的2倍。在搜索空间中,共有16种图像运算方法,包括剪切(沿X轴或Y轴)、平移(沿X轴或Y轴)、旋转、最大化图像对比度、反色、均衡、像素点反转、像素点反转叠加、色调分离、对比度、饱和度、亮度、锐度以及随机擦除。每个子策略包含2种图像运算方法,图像运算包含该方法的应用概率和应用幅度,即以一定的概率选择使用该方法并在图像上采用一定的幅度使用该方法。图像运算方法选择具有随机性,导致图像增强后的图片存在多种结果,因此,确定多种结果中更为合适的一个结果尤为重要。为此,本文采用基于控制器RNN的搜索算法来采样数据增强策略,该策略具有关于图像变换方式、变换概率和变换幅度的信

息。在实验过程中,利用文献[17]中基于 ImageNet 的增强策略来训练网络模型。综上,本文采用基于自动增强的方法对训练数据进行增强,增加数据量以避免模型过拟合,最终利用更深的网络来提取更细粒度的图片特征。

## 1.2 基于注意力机制的行人特征提取

### 1.2.1 连接注意力模块

连接注意力模块是在通道注意力(Channel Attention, CA)<sup>[21]</sup>的基础上通过相邻注意块的信息互连而形成的。通道注意力模块结构如图2所示,其通过学习通道注意力系数对卷积块进行加权,以获得最终的卷积输出。

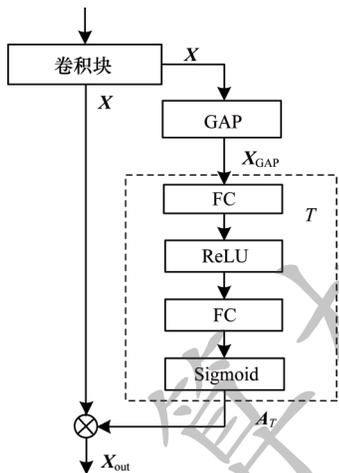


图2 通道注意力模块结构

Fig.2 Channel attention module structure

对由卷积块产生的特征图  $X \in \mathbb{R}^{C \times W \times H}$  进行全局平均池化(Global Average Pooling, GAP)和特征转换,如式(1)所示:

$$A_T = T(X_{GAP}, \omega_t) \quad (1)$$

其中:  $C \times W \times H$  表示特征图大小,即通道数、图像宽度和图像高度;  $\omega_t$  是转换参数;  $X_{GAP} \in \mathbb{R}^C$  为全局池化后的特征;  $T$  为特征转换,该转换通过双层前向全连接网络来实现,2个全连接层的非线性激活函数分别为 ReLU 和 Sigmoid;  $A_T \in \mathbb{R}^C$  是转换后的通道注意力系数输出。注意力引导输出  $X_{out} \in \mathbb{R}^{C \times W \times H}$ ,如式(2)所示:

$$X_{out} = A_T \odot X \quad (2)$$

其中:  $\odot$  表示通道点乘运算,即每个通道注意力系数对每个通道内的像素点进行点乘。最终通道注意力模块输出可表示为:

$$X_{out} = T(X_{GAP}, \omega_t) \odot X \quad (3)$$

连接注意力模块结构如图3所示,其将前序通道注意力块的通道注意力系数  $A_T$  与当前注意力块的全局池化结果进行融合,这种融合能够确保当前卷积块的通道注意力系数,是由其全局池化结果和前序卷积块通道注意力信息的联合学习而得到。

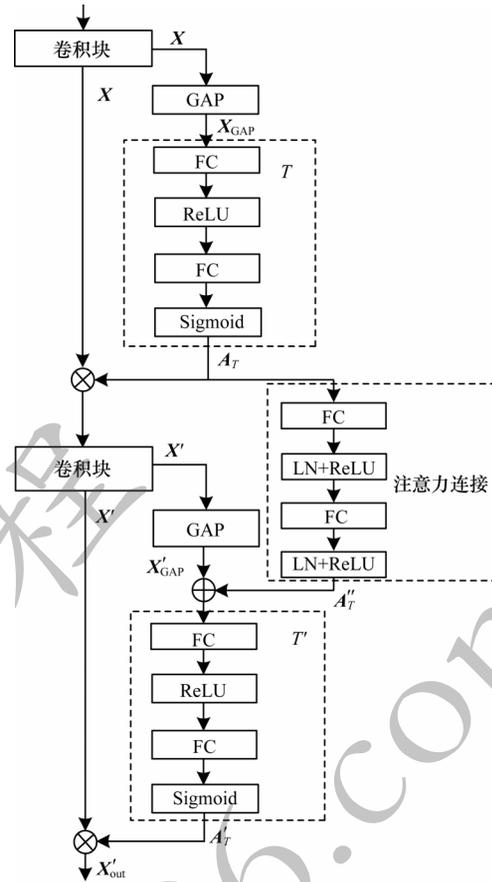


图3 连接注意力模块结构

Fig.3 Connected attention module structure

在不同阶段由卷积块所产生的特征图通道数不同,与之相关联的注意力图通道数也不同,为了解决通道匹配问题,本文将先前注意块的输出  $A_T$  采用注意力连接模块进行通道数匹配,该注意力连接模块包括全连接层和归一化层(Layer Normalization, LN),归一化后都采用 ReLU 非线性函数。为减少参数量,2个全连接层在进行通道匹配时,通道数由  $(C, \frac{C'}{r})$  转换成  $(\frac{C'}{r}, C)$ ,其中,  $C$  和  $C'$  分别代表先前和当前通道数量,  $r$  为通道衰减参数,其对模型性能的影响将在后续实验中详细分析(可参见表3)。最终,注意力模块的输出  $X'_{out}$  如式(4)所示:

$$X'_{out} = T'(f(\alpha X'_{GAP}, \beta A_T), \omega_t) \odot X' = T'(\alpha X'_{GAP} + \beta A_T, \omega_t) \odot X' \quad (4)$$

其中:  $X'_{GAP}$  为卷积块特征图  $X'$  全局池化后的输出;  $A_T$  为先前的注意力系数  $A_T$  经通道匹配之后的转换输出;  $\alpha$  和  $\beta$  是可学习的参数,实验过程中  $\alpha$  和  $\beta$  均设置为 1;  $f$  为连接函数,即将前序注意力信息与当前注意力信息进行融合,本文实验中  $f$  采用普通加法函数;  $T'$  为当前卷积块  $X'$  的特征转换。

从上文分析可以看出,通道注意力通过自主学习的方式得到每个特征通道的重要性程度,赋予重要的特征通道更大的权重,赋予作用较小的特征通道更小的权重,但其只考虑当前注意力通道的重要

性程度, 忽略了前序注意块的信息, 存在特征显著性不强的问题。连接注意力模块在其基础上将之前注意块信息融入到当前注意块中, 使得注意块之间的信息相互配合, 从而提高了注意力模块的学习能力和网络的特征学习效果。

### 1.2.2 特征提取网络

本文以 ResNet50 作为主干网络, 分别在卷积块 2~卷积块 5 中, 在残差块引入连接注意力机制, 让其在卷积块之间进行信息融合, 最终构成特征提取网络, 结构如图 4 所示。

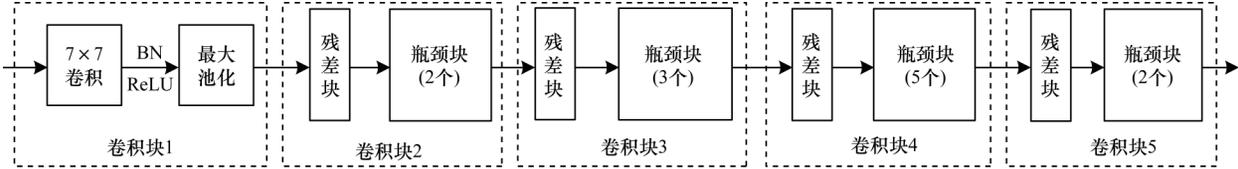


图4 特征提取网络结构

Fig.4 Feature extraction network structure

本文特征提取网络由 5 个卷积块构成, 其中, 卷积块 1 包括一个 7×7 卷积和最大池化层, 其余卷积块均由残差块和瓶颈块组成, 不同之处在于瓶颈块叠加的个数不同。以卷积块 2 为例, 其结构如图 5 所示, 残差块包含 1×1 卷积、3×3 卷积以及连接注意力模块, 连接注意力模块在 1×1 卷积层之后, 瓶颈块包

含 1×1 卷积、3×3 卷积以及通道注意力模块。为了避免网络层数过多带来的梯度消失问题, 残差块采用 1×1 卷积进行残差连接, 连接注意力模块和通道注意力模块对特征通道赋予不同的权重, 以获得显著性更强的行人特征, 从而得到识别精度更高的特征提取网络。

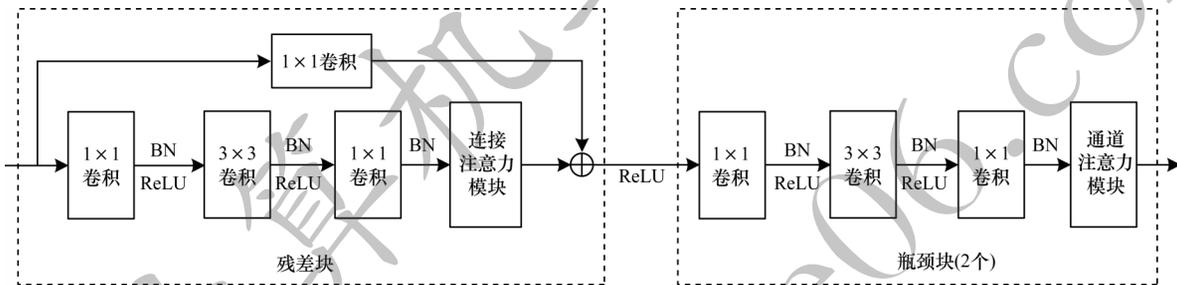


图5 卷积块 2 的结构

Fig.5 Structure of convolution block 2

### 1.3 损失函数

本文实验使用的损失函数为三元组损失函数<sup>[22]</sup>和圆损失函数<sup>[23]</sup>, 两者能够学习高判别力的行人特

$$L_{Tri} = \sum_{i=1}^P \sum_{a=1}^K \left[ \max_{p=1,2,\dots,K} (\|F(x_i^a) - F(x_i^p)\|_2^2) - \min_{\substack{j=1,2,\dots,P \\ n=1,2,\dots,K \\ j \neq i}} (\|F(x_i^a) - F(x_i^n)\|_2^2) + \alpha \right]_+ \quad (5)$$

其中:  $x^a$  为固定样本;  $x^p$  为正样本, 即  $x^p$  与  $x^a$  具有相同的行人 ID;  $x^n$  为负样本, 即  $x^n$  与  $x^a$  具有不同的行人 ID;  $F$  为特征提取函数, 即利用连接注意力优化的残差网络进行特征提取后再进行池化 (如图 1 所示);  $\|\cdot\|_2^2$  表示正负样本对特征之间的欧氏距离;  $\alpha$  为手动设置的距离阈值, 用于限制正负样本对之间的距离, 实验过程中其取值 0.3;  $[\cdot]_+$  表示  $\max(\cdot, 0)$  函数。对于每一个训练批次, 挑选出  $P$  个行人, 每个行人随机挑选  $K$  张图片, 即一个批次包含  $P \times K$  张图片, 对于每一张图片, 挑选出一个与其最不相似的正样本 (即与固定样本特征距离最大的样本) 和一个最相似的负样本 (即与固定样本特征距离最小的样本), 最终计算三元组损失。

圆损失旨在学习类别区分性信息, 即最大化类内相似度同时最小化类间相似度。假设与  $x$  相关的类内相似度分数有  $K$  个, 类间相似度分数有  $L$  个, 则

征。三元组损失旨在一定距离上将正负样本分开, 即最小化正样本对之间的距离, 同时最大化负样本对之间的距离, 其目标函数如式 (5) 所示:

圆损失的目标函数如式 (6) 所示:

$$L_{circle} = \log_a \left[ 1 + \sum_{j=1}^L \exp(\gamma \alpha_n^j (s_n^j - \Delta_n)) \sum_{p=1}^K \exp(-\gamma \alpha_p^j (s_p^j - \Delta_p)) \right] \quad (6)$$

其中:  $s_n$  和  $s_p$  分别为关于样本空间  $x$  的类间相似度和类内相似度, 采用余弦相似度来计算类内和类间相似度分数;  $\Delta_n$  和  $\Delta_p$  分别为类间和类内间隔, 实验过程中分别取值 0.25 和 0.75;  $\alpha_n$  和  $\alpha_p$  为非负整数权重因子;  $\gamma$  为尺度因子, 实验过程中取值 1。

为了最大化地发挥损失函数的作用, 本文实验过程中采用总损失  $L_{total}$ , 如式 (7) 所示:

$$L_{total} = m L_{Tri} + n L_{circle} \quad (7)$$

其中:  $m$  和  $n$  分别为损失函数的权重因子, 其对实验结果的影响见表 5, 实验过程中  $m$  和  $n$  均取值 1。

## 2 实验结果与分析

### 2.1 数据集与评价指标

为了验证本文所提方法的有效性,分别在Market1501<sup>[18]</sup>、DukeMTMC-reID<sup>[19]</sup>、MSMT17<sup>[20]</sup>这3个公开数据集上进行测试。Market1501于2015年在清华大学内采集,图像来自6个不同的摄像头,其中1个摄像头为低分辨率,训练集包括751个人的12 936张图片,测试集包括750个人的19 732张图片。DukeMTMC-reID于2017年在杜克大学内采集,图像来自8个不同的摄像头,训练集包括702个人的16 522张图片,测试集包括702个人的19 889张图片。MSMT17在北京大学内采集,图像来自15个不同的摄像头,训练集包括1 041个人的30 248张图片,测试集包括3 060个人的93 820张图片,其中,11 659张图片被随机选出作为查询图片,剩下的82 161张图片作为图库。

本文采用目前行人重识别领域最常用的2种评价指标,即首位准确率(Rank-1)和平均精确均值(mean Average Precision, mAP)。Rank-1反映了待选行人图像中匹配值排在首位的图像是待查询行人的概率,mAP综合考虑精确率和召回率,能够反映图像库中所有正确的图片排在检索列表最前面的程度。

### 2.2 实验设置

实验采用ResNet50作为主干网络,并利用ImageNet预训练参数进行初始化。为了保留图像的分辨率,ResNet50最后一个瓶颈层的步长设置为1,每个瓶颈层后均叠加一个注意力块。在训练期间,采用自动增强的数据增强策略,输入图片大小为384×128像素,批次大小是64,其包含16个行人类别,每个类别包含4个行人图像。采用随机梯度下降优化器进行网络优化,动量因子为0.9,权重衰减为0.000 5。初始学习率设置为0.01,采用预热学习策略,在训练40个回合后下降10倍。模型共训练80个回合,为了更好地初始化分类器的参数,在前10个回合只训练分类器参数,即用来计算损失的特征参数,同时冻结网络参数,经过10个回合迭代后释放网络参数,后70个回合训练整个网络模型。

### 2.3 结果分析

本文采用Rank-1和mAP这2种性能指标来衡量所提方法的性能,Rank-1和mAP的值越高,说明模型性能越好,行人重识别精度越高。

#### 2.3.1 本文方法与现有方法的性能比较

在Market1501、DukeMTMC-reID、MSMT17这3个公开行人重识别数据集上进行测试,将本文方法与PCB<sup>[3]</sup>、MGN<sup>[4]</sup>、Pyramid<sup>[5]</sup>、OSNet<sup>[7]</sup>、MHN<sup>[9]</sup>、SONA<sup>[10]</sup>、AGW<sup>[15]</sup>、FastReID<sup>[24]</sup>等基于深度学习的经典行人重识别方法进行比较,结果如表1所示。

表1 不同行人重识别方法的性能比较

Table 1 Performance comparison of different person re-identification methods %

方法	Market1501		DukeMTMC-reID		MSMT17	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
PCB <sup>[3]</sup>	92.30	77.40	81.80	66.10	—	—
MGN <sup>[4]</sup>	95.70	86.90	88.70	78.40	—	—
Pyramid <sup>[5]</sup>	95.70	88.20	89.00	79.00	—	—
OSNet <sup>[7]</sup>	94.80	84.90	88.60	73.50	78.70	52.90
MHN <sup>[9]</sup>	95.10	85.00	89.10	77.20	—	—
SONA <sup>[10]</sup>	95.70	88.70	89.30	78.10	—	—
AGW <sup>[15]</sup>	95.10	88.30	89.00	78.60	68.30	49.30
FastReID <sup>[24]</sup>	95.40	88.20	89.60	79.80	83.30	59.90
本文方法	95.90	89.66	91.16	81.24	84.37	62.73
本文方法+QE	95.87	93.90	92.73	88.73	—	—
本文方法+Rerank	95.99	94.96	93.40	91.18	—	—

从表1可以看出:在Market1501数据集上,本文方法的Rank-1和mAP值较其他对比方法均最优;在DukeMTMC-reID、MSMT17数据集上,本文方法的Rank-1和mAP值同样取得了最优结果,超出其他方法1.5个百分点左右。此外,本文在测试阶段对重排序(Rerank)<sup>[25]</sup>和查询扩展(Query Expansion, QE)这2个测试技巧进行实验,从结果可以看出,在Market1501和DukeMTMC-reID这2个数据集上,本文方法的Rank-1和mAP值较对比方法均有明显提升。由此可以看出,在3个公开行人重识别数据集上,本文方法在行人重识别精度方面极具竞争力,这主要得益于连接注意力模块将上下文信息进行融合,提取显著性更强的行人特征进行距离度量,最终获得了较好的识别效果。

#### 2.3.2 消融实验

本文消融实验均以Market1501数据集为例进行测试。为了探索不同分支对实验结果的影响,将自动增强方法和连接注意力模块嵌入ResNet50网络中进行识别,其中,M1表示数据预处理中的自动增强,M2表示连接注意力机制。从表2可以看出,2个分支对实验结果均有一定程度的提升,合并在一起对实验结果的提升更为明显,Rank-1值提升1个百分点左右,mAP值提升3个百分点左右,因此,增加数据集规模并提取显著性更强的行人特征有助于提升识别精度。

表2 消融实验结果

Table 2 Ablation experimental results %

网络结构	Rank-1	mAP
ResNet50	94.71	86.48
ResNet50+M1	94.86	87.00
ResNet50+M2	95.78	89.20
ResNet50+M1+M2	95.90	89.66

在连接注意力模块中,为了匹配通道大小,引入了通道衰减 $r$ ,为了探索 $r$ 值对模型性能的影响,本文进行实验测试。从表3可以看出:当 $r$ 值为2时,实验结果最好; $r$ 值为4和直接采用一个全连接层(即对应表中第一行FC层)的实验结果相比 $r$ 值为2时稍差。

表3 不同通道衰减 $r$ 值下的实验结果  
Table 3 Experimental results under different channel attenuation  $r$  values

通道衰减 $r$ 值	Rank-1	mAP
1 FC	95.55	89.22
2 FC( $r=16$ )	95.67	89.03
2 FC( $r=8$ )	95.43	89.02
2 FC( $r=4$ )	95.55	89.26
2 FC( $r=2$ )	95.90	89.66

本文还测试了不同主干网络对实验结果的影响,ResNet18、ResNet34、ResNet50分别代表主干网络层数分别为18、34、50的典型网络,IBN(Instance Batch Normalization)网络是将ResNet50等典型网络中的BN层替换为IBN层之后形成的网络。从表4可以看出,当主干网络为ResNet50时,更深的网络结构提取了更高层次的行人特征,识别效果更好。为了验证本文所采用损失函数的有效性,对 $m$ 和 $n$ 的取值进行测试。从表5可以看出,三元组损失和圆损失联合使用时实验效果更好,这主要得益于三元组损失使得正样本对距离更近同时负样本对距离更远,而圆损失在优化距离的同时赋予类间样本与类内样本不同的权重因子,使得其优化方向更为准确。

表4 不同主干网络下的实验结果

Table 4 Experimental results under different backbone networks

主干网络	Rank-1	mAP
ResNet18	93.02	82.68
ResNet34	94.71	86.46
ResNet50	95.90	89.66
ResNet18-IBN	92.22	80.67
ResNet34-IBN	94.15	85.11
ResNet50-IBN	95.69	89.07

表5 不同损失函数下的实验结果

Table 5 Experimental results under different loss functions

$m$ 和 $n$ 的取值	Rank-1	mAP
$m=1, n=0$	95.18	88.76
$m=0, n=1$	95.51	89.02
$m=1, n=1$	95.90	89.66

### 2.3.3 行人重识别结果示例

为了更加清晰地呈现行人重识别系统的识别效果,在Market1501数据集上对识别结果进行可视化,如图6所示,query为待查询图片,在图像库中寻找与其为同一个人的图片,显示结果为前10张相似度最高的行人图片,其中,虚线框内为错误的识别结果,其余为正确的识别结果。从图6可以看出,错误的检索图片与待查询图片在外观方面非常相似,在提取行人外观特征时,网络模型捕获不到细微的差别,导致了错误的识别结果。因此,如何提取更加细粒度的行人特征,使得网络模型关注到行人图像的细小差别显得尤为重要,这也是本文下一步的研究重点。



图6 部分行人重识别结果

Fig.6 Some person re-identification results

### 3 结束语

在现实环境中,行人重识别容易受光照、姿态等外在变化和不同背景的影响,导致数据量欠缺以及行人重识别系统识别率低的问题。为此,本文提出一种基于连接注意力的行人重识别特征提取方法。将自动增强策略引入图像预处理过程中以提高模型的鲁棒性,将连接注意力模块嵌入 ResNet50 网络进行特征提取,从而增强行人特征的显著性,在此基础上,进行相似度度量以得到识别精度较高的行人重识别模型。在3个主流数据集上的实验结果表明,该方法的行人重识别性能评价指标优于 PCB、MGN 等方法。在本文方法的基础上,利用预训练模型和 transformer 网络来提取更加细粒度的特征以及提升跨域行人重识别模型的精度,将是下一步的研究方向。

#### 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. (in Chinese)
- [ 2 ] 李幼蛟,卓力,张菁,等. 行人再识别技术综述[J]. 自动化学报,2018,44(9):1554-1568.  
LI Y J,ZHUO L,ZHANG J,et al. A survey of person re-identification[J]. Acta Automatica Sinica,2018,44(9):1554-1568. (in Chinese)
- [ 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]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2018:480-496.
- [ 4 ] 董亚超,刘宏哲,徐成. 基于显著性多尺度特征协作融合的行人重识别方法[J]. 计算机工程,2021,47(6):234-244,252.  
DONG Y C,LIU H Z,XU C. Person re-identification method based on joint fusion of saliency multi-scale features[J]. Computer Engineering, 2021, 47 ( 6 ) : 234-244, 252. (in Chinese)
- [ 5 ] ZHENG F,DENG C,SUN X,et al. Pyramidal person re-identification via multi-loss dynamic training [C]// Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2019:8506-8514.
- [ 6 ] 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, 2019,33:8295-8302.
- [ 7 ] ZHOU K Y,YANG Y X,CAVALLARO A,et al. Omni-scale feature learning for person re-identification [C]// Proceedings of IEEE/CVF International Conference on Computer Vision. Washington D. C. , USA: IEEE Press, 2019:3701-3711.
- [ 8 ] ZHUO J X,LAI J H,CHEN P J. A novel teacher-student learning framework for occluded person re-identification [EB/OL]. [2021-05-05]. <https://arxiv.org/pdf/1907.03253.pdf>.
- [ 9 ] CHEN B H,DENG W H,HU J N. Mixed high-order attention network for person re-identification[C]//Proceedings of IEEE/CVF International Conference on Computer Vision. Washington D. C. , USA: IEEE Press, 2019:371-381.
- [ 10 ] BRYAN B,GONG Y,ZHANG Y Z,et al. Second-order non-local attention networks for person re-identification [C]// Proceedings of IEEE/CVF International Conference on Computer Vision. Washington D. C. , USA: IEEE Press, 2019:3759-3768.
- [ 11 ] HOU R B,MA B P,CHANG H,et al. Interaction-and-aggregation network for person re-identification [C]// Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2019:9309-9318.
- [ 12 ] 李佳宾,李学伟,刘宏哲,等. 基于局部特征关联与全局注意力机制的行人重识别[J]. 计算机工程,2022,48(1):245-252.  
LI J B,LI X W,LIU H Z,et al. Person re-identification based on local feature relation and global attention mechanism[J]. Computer Engineering, 2022,48(1):245-252. (in Chinese)
- [ 13 ] CHEN T L,DING S J,XIE J Y,et al. ABD-net: attentive but diverse person re-identification [C]//Proceedings of IEEE/CVF International Conference on Computer Vision. Washington D. C. , USA: IEEE Press, 2019:8350-8360.
- [ 14 ] QUAN R J,DONG X Y,WU Y,et al. Auto-ReID: searching for a part-aware ConvNet for person re-identification [C]// Proceedings of IEEE/CVF International Conference on Computer Vision. Washington D. C. , USA: IEEE Press, 2019:3749-3758.
- [ 15 ] YE M,SHEN J,LIN G,et al. Deep learning for person re-identification: a survey and outlook [EB/OL]. [2021-05-05]. <https://arxiv.org/abs/2001.04193>.
- [ 16 ] MA X,GUO J D,TANG S H,et al. Learning connected attentions for convolutional neural networks [C]//Proceedings of IEEE International Conference on Multimedia and Expo. Washington D. C. , USA: IEEE Press, 2021:1-6.
- [ 17 ] CUBUK E D,ZOPH B,MANE D,et al. AutoAugment: learning augmentation policies from data [EB/OL]. [2021-05-05]. <https://arxiv.org/abs/1805.09501>.
- [ 18 ] ZHENG L,SHEN L Y,TIAN L,et al. Scalable person re-identification: a benchmark [C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. , USA: IEEE Press, 2015:1116-1124.
- [ 19 ] 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.
- [ 20 ] WEI L H,ZHANG S L,GAO W,et al. Person transfer GAN to bridge domain gap for person re-identification [C]// Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2018:79-88.
- [ 21 ] HU J,SHEN L,SUN G. Squeeze-and-excitation networks [C]// Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2018:7132-7141.
- [ 22 ] CHEN W H,CHEN X T,ZHANG J G,et al. Beyond triplet loss: a deep quadruplet network for person re-identification [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2017:1320-1329.
- [ 23 ] SUN Y F,CHENG C M,ZHANG Y H,et al. Circle loss: a unified perspective of pair similarity optimization [C]// Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2020:6397-6406.
- [ 24 ] HE L X,LIAO X Y,LIU W,et al. FastReID: a pytorch toolbox for general instance re-identification [EB/OL]. [2021-05-05]. <https://arxiv.org/abs/2006.02631>.
- [ 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.