

# 基于无参注意力和联合损失的行人重识别\*

程述立<sup>1,2</sup>, 汪烈军<sup>1†</sup>, 王有丹<sup>1</sup>

(1. 新疆大学 信息科学与工程学院, 新疆 乌鲁木齐 830017; 2. 新疆大学 数学与系统科学学院, 新疆 乌鲁木齐 830017)

**摘要:** 现阶段行人重识别一般只考虑二维特征, 将各个特征点统一处理, 存在特征提取不足的问题, 故提出基于无参注意力的行人重识别(PFNet)来解决上述问题. 该模型在 ResNet-50 网络上进行改进, 分别在第一个残差块和第三个残差块后引入无参注意力机制, 该注意力机制能根据图片本身特点赋予各特征点合适的权重, 可以保留更丰富的信息特征且不会引入额外参数. 接着使用自适应平均池化层保留主要特征且捕捉特定域的判别特征, 然后使用 ID 损失、三元组损失和自适应加权排序损失的联合损失函数来训练模型. 算法在 Market-1501、DukeMTMC-reID 和 CUHK03 三个主流的行人重识别数据集上的首位命中率分别达到 95.5%、90.9% 和 84.3%, 平均精度均值分别达到 89.6%、81.6% 和 82.0%. 实验结果表明, 使用注意力和联合损失函数的策略可以提高模型精度.

**关键词:** 行人重识别; 残差网络; 无参注意力; 联合损失函数

**DOI:** 10.13568/j.cnki.651094.651316.2022.07.09.0001

**中图分类号:** TP391 **文献标识码:** A **文章编号:** 2096-7675(2023)02-0202-08

**引文格式:** 程述立, 汪烈军, 王有丹. 基于无参注意力和联合损失的行人重识别[J]. 新疆大学学报(自然科学版)(中英文), 2023, 40(2): 202-209.

**英文引文格式:** CHENG Shuli, WANG Liejun, WANG Youdan. Person re-identification based on parameter-free attention and joint loss[J]. Journal of Xinjiang University(Natural Science Edition in Chinese and English), 2023, 40(2): 202-209.

## Person Re-Identification Based on Parameter-Free Attention and Joint Loss

CHENG Shuli<sup>1,2</sup>, WANG Liejun<sup>1</sup>, WANG Youdan<sup>1</sup>

(1. School of Information Science and Engineering, Xinjiang University, Urumqi Xinjiang 830017, China;

2. School of Mathematics and System Sciences, Xinjiang University, Urumqi Xinjiang 830017, China)

**Abstract:** At present, pedestrian re-recognition usually only considers two-dimensional features and deals with each feature point uniformly, which has the problem of insufficient feature extraction. In this paper, pedestrian re-recognition based on parameter-free attention (PFNet) is proposed to solve the above problems. The model is improved on ResNet-50 network, and the non-parametric attention mechanism is introduced after the first residual block and the third residual block respectively. The attention mechanism can assign appropriate weight to each feature point according to the characteristics of the image itself, and can retain richer information features without introducing additional parameters. Then, the adaptive average pooling layer is used to retain the main features and capture the discriminant features of a specific domain. Then, the model is trained with a combined loss function of ID loss, triplet loss and adaptive weighted ranking loss. The algorithm achieves 95.5%, 90.9% and 84.3%, respectively; On the three mainstream pedestrian re-recognition datasets of Market-1501, DukeMTMC-reID and CUHK03, and the average accuracy is 89.6%, 81.6% and 82.0%, respectively. Experimental results show that the strategy using attention and joint loss function can improve the model accuracy.

**Key words:** person re-identification; residual network; parameter-free attention; joint loss function

## 0 引言

随着基础设施的不断完善和公共需求的不断增长, 行人重识别作为视频监控系统的一项关键技术受到了广泛的关注, 并在众多领域都有很重要的应用价值. 行人重识别主要研究的是从不同的摄像机中检索具有相同身

\* 收稿日期: 2022-07-09

**基金项目:** 国家自然科学基金“音视频多模态协同的异常事件鲁棒性检测关键技术研究”(U1903213); 新疆维吾尔自治区自然科学基金“基于深度学习的遥感变化检测方法研究”(2022D01C82).

**作者简介:** 程述立(1990-), 男, 博士, 副教授, 从事多模态信息处理研究, E-mail: cslxju@xju.edu.cn.

† **通讯作者:** 汪烈军(1975-), 男, 博士, 教授, 从事计算机视觉研究, E-mail: wljxju@xju.edu.cn.

份的人员或在同一摄像机的不同视频片段中识别出目标行人. 该任务的目的是通过图像检索实现目标匹配, 将目标人物与图片库中的人物信息进行比对, 筛选出图片库中含有目标人物的信息<sup>[1]</sup>. 然而该任务也存在许多难点, 如在复杂环境下出现行人部分部位遮挡的问题以及设备质量、距离导致的图像分辨率低等问题, 所以如何提取行人关键特征信息进行度量判断成为了研究重点.

近些年, 深度学习的方法在行人重识别任务上取得了一系列成果. 相较传统度量学习方法, 深度学习的方法代替了其特征提取配合特征度量学习的模式, 使得之后能够更好地判别特征信息. 基于卷积神经网络 (CNN) 在图像特征提取上彰显出的独特优势, 许多研究者将行人重识别的主干网络建立在 CNN 上, 如残差网络 (Residual Network, ResNet)<sup>[2]</sup>、稠密连接网络 (Dense Convolutional Network, DenseNet)<sup>[3]</sup>. 虽然 CNN 成为解决行人重识别任务的主流手段, 它也因为忽略了长距离的上下文信息之间的关系而存在着自身的局限性.

随着深度学习的不断发展, 注意力机制得到了广泛的研究与应用. 类似人类处理信息的方法, 该机制能够使网络快速聚焦一部分关键信息, 同时忽略其它信息. 注意力机制通常在输入的特征信息上计算权重, 根据权重对特征进行处理从而强化有效特征、抑制无效特征. 注意力机制从应用位置上可划分为三类: 1) 空间注意力. 它可以对全图进行处理从而定位出感兴趣的区域, 有效地弥补了卷积操作对于全局特征提取能力的不足之处, 代表性的网络有动态容量网络 (Dynamic Capacity Networks, DCN)<sup>[4]</sup>; 2) 通道注意力. 如压缩激发网络 (Squeeze-and-Excitation Networks, SENet)<sup>[5]</sup>, 它通过建模各个特征通道的重要程度, 针对不同的通道特征进行增强或者抑制; 3) 空间与通道混合注意力. 它可以同时从两个维度对特征进行选择, 代表性的网络有卷积块状注意力模块网络 (Convolutional Block Attention Module, CBAM)<sup>[6]</sup>、瓶颈注意力模块网络 (Bottleneck Attention Module, BAM)<sup>[7]</sup>等. 由于注意力机制在增强图片特征方面表现尤为突出, 行人重识别任务中出现了许多基于注意力机制的方法, 如自适应聚合网络 (Adaptive Aggregation Network, AANet)<sup>[8]</sup>、基于关系感知的全局注意力网络 (Relation-Aware Global Attention, RAG)<sup>[9]</sup>等. 一般来说, 对于行人重识别任务的优化主要是基于特征表示学习、深度度量学习和查询结果排序这三个方面. 研究人员基于这三个方面进行了大量工作, 有的改进了特征学习方法, 有的提出有效的度量损失函数, 也有的是在检索排序阶段进行优化.

一些研究人员通过引入局部特征学习或利用注意力机制聚焦于人体部位的关键信息来提取更饱满的行人特征. Yang 等<sup>[10]</sup>探索了高级特征提取方法, 目的是通过建模概念间关系来探索基于上下文的融合, 而这并不是简单的基于语义推理建模. 由于生成对抗网络 (Generative Adversarial Network, GAN)<sup>[11]</sup>在生成图像和学习特征方面具有良好性能, GAN 被广泛应用于人员再识别任务. 也有一些研究者利用人的三维形状信息, 更好地学习行人特征以提高行人重识别模型的鲁棒性. Ye 等<sup>[1]</sup>探索了封闭世界和现实世界的行人重识别系统, 在标准封闭世界中系统认为特征表示学习是该系统的三要素之一, 专注于开发特征构建策略, 结合行人的全局特征、局部特征、辅助特征和视频特征, 通过这四类特征的综合分析增强了最终的特征表示. Dai 等<sup>[12]</sup>则尝试设计新的网络模型, 该网络由 ResNet-50<sup>[2]</sup> 和一个特征丢弃网络构成, 同时学习两部分特征来增强模型的表达效果. 研究者们不仅在主干网络方面有所探究, 还在损失函数方面做了一些研究. 有研究者设计了不同的损失函数, 利用网络模型优化识别特征的学习<sup>[13]</sup>. 这些算法大多基于 ResNet-50 进行改进, 但存在参数繁多, 结构复杂的问题, 并且无法把握不同特征的重要性, 导致行人识别不够准确, 算法准确度有待加强.

行人再识别旨在解决跨摄像头、跨场景下目标行人的关联与匹配, 作为智能视频监控系统的关键环节, 对维护社会公共秩序具有重大作用. 当前, 行人再识别研究视角多样, 如鸟瞰行人再识别<sup>[14]</sup>, 这类研究不仅趋向于水平视图捕获人员图像, 而且关注鸟瞰图拍摄的人员信息; 长时行人再识别研究<sup>[15]</sup>, 这类研究关注服装有关的生物特征, 通过服装状态意识网络, 嵌入服装状态意识来规范行人的描述; 跨模态行人再识别研究<sup>[16]</sup>, 旨在从数据库匹配出该行人的红外 (或可见光) 照片, 突破单模态行人再识别在黑暗环境中的局限性, 减轻模态差异的影响; 无监督行人重识别<sup>[17]</sup>, 旨在学习识别表征, 从无标签数据中检索行人; 遮挡行人重新识别<sup>[18]</sup>, 这是一种基于变换器的姿势引导特征分解方法, 利用姿势信息来明确分解语义部分; 域适应性行人重识别<sup>[19]</sup>, 研究表明时间上的连续性先验是有益的, 它为区分一些相貌相似的人提供了线索. 总之, 这些研究利用更多的补充信息来协助行人重识别任务.

总结上述行人重识别研究, 其核心在于网络模型的构建和度量学习策略的研究, 基于以上考虑, 本文从网络模型和损失函数两个角度进行算法设计, 提出一种基于无参注意力的行人重识别算法. 它是在 ResNet-

50 网络上进行改进, 分别在残差模块的末端引入无参注意力机制 (Simple, Parameter-Free Attention Module, SimAM)<sup>[20]</sup>, 这是一个简单而有效的注意力模块, 增强了网络特征提取及表达能力; 然后联合 ID 损失、Triplet 损失和 WRL 损失, 共同形成联合损失函数来优化网络参数; 最后在三个公开数据集 Market-1501、DukeMTMC-reID 和 CUHK03 上评估算法的识别性能. 所提模型参数量少、鲁棒性高, 具有较好的应用市场和研究价值.

## 1 PFNet 网络结构

本文将无参注意力模块应用在行人重识别模型中进行训练, 可以对任务中更为相关的内容自适应地赋予更高的权重, 使得该网络框架能有效提高行人识别的准确率. 整体框架结构如图 1 所示, 它由三个重要的部分组成: 主干网络 ResNet-50、SimAM 模块和联合损失函数.

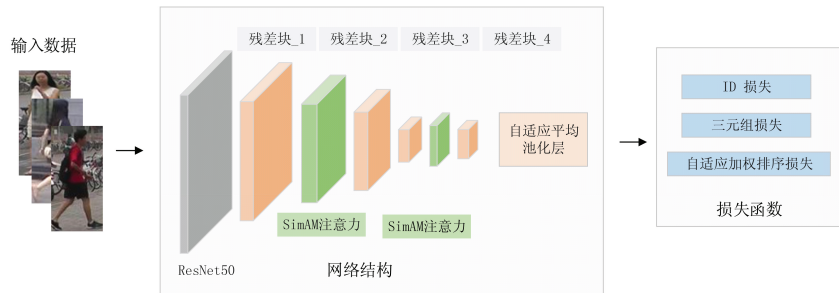


图 1 模型整体框架结构

由图 1 可知, 分别在第一个残差块和第三个残差块后加入 SimAM 模块, 这种方式可以在不同的网络深度提取到更为丰富的特征. SimAM 模块是一种通过能量函数赋予权值的无参注意力机制, 不会额外增加参数. 同时在损失函数前加入可训练池化层 (Generalized-Mean, GeM)<sup>[21]</sup> 压缩特征来提高检索能力. 最后采用 ID 损失、Triplet 损失和 WRL 损失组成的联合损失函数对网络进行训练. 将在下文对每一部分进行具体介绍.

### 1.1 主干网络 ResNet-50

残差网络可以有效抑制卷积网络中梯度爆炸的影响从而提高模型性能, 该优点使其在行人重识别领域有广泛应用. ResNet-50 结构简单, 主要由残差块构成, 在网络加深的同时有效减少梯度消失且不会增加实验误差, 使得训练结果有很好的提升. 残差块是规避梯度消失的关键, 它采用简单的跳跃式加法来保留更多信息, 只在原网络上增加很小的计算复杂度就能使模型在训练速度和网络性能方面均有所提升, 这对网络效果的提升有很大的价值. 常用的残差网络深度有 34 层、50 层、101 层等, 综合考虑模型的性能和复杂度, 本文使用 ResNet-50 网络模型作为算法的主干网络.

### 1.2 无参注意力 SimAM

无参数的注意力模块 SimAM<sup>[20]</sup> 是目前较为先进的注意力机制, 能通过特征图本身推导出 3D 注意力权值, 可以更好地进行特征增强以提升模型性能, 同时不会引入额外的参数, 因此被称为“无参注意力”. SimAM 解决了通道空间注意力无法同步进行的问题. 由于人类视觉机制是“通道和空间注意力”同时进行的, SimAM 以能量函数的方式让通道和空间进行同步加权, 可以充分发掘每个神经元的重要性并赋予其合适的权值, 从而改善特征丢失的问题. SimAM 注意力结构如图 2 所示.  $c$  表示通道数,  $X$  表示每个通道输入  $H \times W$  矩阵的维度, 也是通道中神经元的数量.

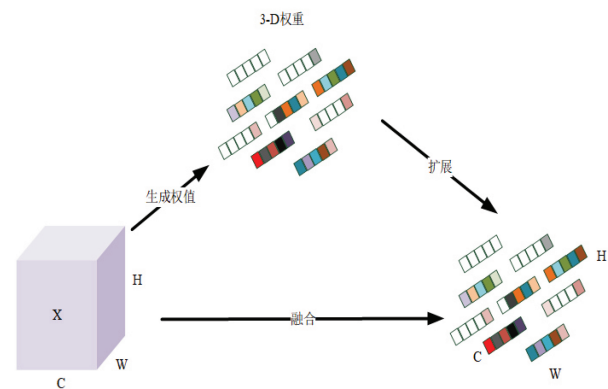


图 2 无参注意力模块<sup>[20]</sup>

一般来说, 信息丰富的神经元通常会抑制周围神经元, 具有抑制效应的神经元应当赋予更高的权重. 为了评估每个神经元的重要性, 通过能量函数的方式给每个神经元赋予合适的权值. 能量函数的表达式如公式 (1)

所示:

$$e_t(\omega_t, b_t, \gamma, x_i) = \lambda \omega_t^2 + \frac{1}{M-1} \sum_{i=1}^{M-1} (-1 - (\omega_t x_i + b_t))^2 + (1 - (\omega_t x_i + b_t))^2 \quad (1)$$

式中:  $t$  和  $x_i$  是输入特性  $X \in R^{C \times H \times W}$  中同一通道的目标神经元和其它神经元,  $(\omega_t t + b_t)$  和  $(\omega_t x_i + b_t)$  是关于目标神经元和其它神经元的线性变换,  $i$  是空间维度的索引,  $M$  是该通道上的神经元数量,  $\omega_t$  是变换的权值,  $b_t$  是变换的偏差. 通过计算得到  $\omega_t$  和  $b_t$  解析解和通道中所有神经元的均值和方差, 得到最小能量函数如公式 (2) 所示:

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (2)$$

式中:  $\hat{\mu} = \frac{1}{M} \sum_{i=1}^M x_i$ ,  $\hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^M (x_i - \hat{\mu})^2$ . 能量越低, 目标神经元与其它神经元的区别越大, 重要性越高, 通过对能量函数求倒数即可知道每个神经元的重要性. SimAM 模块充分考虑输入特征之间的差异性, 从神经元的角度考虑权值分配, 这种方式不会限制它们学习更多区别性线索的能力. 为了提取输入特征的关键信息并且正确区分该特征和其它特征之间的区别, 需要根据每个神经元的重要程度. 最后对特征进行加强即可得到所需的权值信息, SimAM 模块的表达式如公式 (3) 所示:

$$\tilde{X} = \text{sigmoid}\left(\frac{1}{E_n}\right) \odot X \quad (3)$$

式中:  $E$  表示为能量函数  $e_t^*$  跨通道和空间维度进行的分组, 加入  $\text{sigmoid}$  函数不但能使  $E$  值范围可控, 而且能正确映射每个神经元的相对重要性. 综上所述, SimAM 模块通过 3-D 权值对每个神经元进行重要性分析, 再对特征进行增强, 从而达到提取更丰富特征的目的. 该模块产生权值的过程不需要引入额外参数, 可以直接由特征图推导出, 相比其它注意力机制避免了过多的结构调整, 极大减少模型的复杂度.

### 1.3 损失函数

本文涉及的损失函数分别为 ID 损失 (Softmax Loss)、三元组损失 (Triplet Loss) 和自适应加权排序损失 (Adaptive Weighted Rank List Loss, WRLL) [21-22]. 使用的总损失函数如下:

$$L_{\text{total}} = L_{\text{ID}} + L_{\text{Triplet}} + L_{\text{WRLL}} \quad (4)$$

式中:  $L_{\text{ID}}$  代表交叉熵损失函数,  $L_{\text{Triplet}}$  代表三元组损失函数,  $L_{\text{WRLL}}$  代表自适应加权排序损失函数.

#### 1.3.1 ID 损失

行人重识别模型中, ID 损失就是 Softmax 损失, 这是由 Softmax 函数和交叉熵损失函数组成的损失函数. Softmax 损失用来衡量预测样本和真实样本的误差, 所得到的损失函数值越小表明误差越小.

#### 1.3.2 三元组损失

三元组损失是数据集中随机样本的损失和集, 而随机样本的损失由目标样本与正负样本之间的欧式距离判别. 三元组损失主要使具有相同标签的图片在空间位置上更加接近, 同时令具有不同标签的图片在空间位置上更加疏远[23]. 该功能在个体级别上有很显著的效果, 使三元组损失函数在行人重识别领域得到广泛应用.

#### 1.3.3 自适应加权排序损失

WRLL 是一个可以同时优化正样本和负样本的自适应损失函数, 可以根据样本与输入图片的欧氏距离自适应分配权重, 有效拉近正样本的距离并拉开负样本的距离. 在训练集  $X = \{(x_i, y_i)\}_{i=1}^N$  中, 有  $C$  类样本,  $(x_i, y_i)$  表示第  $i$  个样本和样本标签,  $y_i$  的样本标签取值为  $[1, 2, \dots, C]$ . 训练集中所有样本可表示为  $\{(x_i^c)\}_{i=1}^{N_c}$ , 其中正样本集合为  $P_{c,i}^*$ , 负样本集合为  $N_{c,i}^*$ . 若给定一张目标图片, 损失函数  $L_m(x_i, x_j; f)$  的目的是使目标图片的正负样本最少拉开  $m$  的距离, 正样本聚集到  $\alpha - m$  的超球面范围内, 负样本在  $\alpha$  的距离外.  $L_m(x_i, x_j; f)$  表达式如公式 (5) 所示, 其中  $d_{ij} = \|f(x_i) - f(x_j)\|_2$  是目标图片与样本之间的欧氏距离. 当  $y_i = y_j$  时,  $y_{ij}$  值为 1, 反之则为 0.

$$L_m(x_i, x_j; f) = (1 - y_{ij})|\alpha - d_{ij}| + y_{ij}|d_{ij} - (\alpha - m)| \quad (5)$$

$L_P(x_i^c; f)$  表示正样本损失函数, 主要是优化输入与正样本集之间的距离, 如公式 (6) 所示. 由于负样本数量远远大于正样本数量, 采用  $\text{softmin}$  函数可以有效缓解网络泛化能力差的问题, 如公式 (7) 所示:

$$L_P(x_i^c; f) = \frac{1}{|P_{c,i}^*|} \sum_{x_j \in P_{c,i}^*} L_m(x_i^c; x_j^c; f) \quad (6)$$

$$w_{ij} = \frac{\exp(-d_{ij}^n)}{\sum_{x_j^c \in N_{c,j}^*} \exp(-d_{ij}^n)} \quad (7)$$

式中:  $d_{ij}^n$  是输入图片和负样本的欧氏距离.  $w_{ij}$  会根据输入图片和  $d_{ij}^n$  的距离为负样本损失函数动态地赋予权重.  $d_{ij}^n$  越小, 输入图片和负样本越接近, 越容易误判, 导致性能下降. 采用 softmin 函数为负样本损失函数赋予一个较大的权值来提取负样本信息. 负样本损失函数的表达式如公式 (8) 所示:

$$L_N(x_i^c; f) = \sum_{x_j^k \in |N_{c,j}^*|} \frac{w_{ij}}{\sum_{x_j^k \in |N_{c,j}^*|} w_{ij}} L_m(x_i^c; x_j^k; f) \quad (8)$$

本文损失函数中, 既优化了正样本又优化了负样本, 表达式如公式 (9) 所示:

$$L_{\text{WRLL}}(x_i^c; f) = L_P(x_i^c; f) + L_N(x_i^c; f) \quad (9)$$

## 2 实验结果与分析

本节首先介绍 PFNet 模型的数据集和评价指标, 然后详细说明实验环境以及参数配置, 再将 GeM、三元组损失、WRLL 和 SimAM 注意力机制依次叠加到基线上进行消融实验. 最后将 PFNet 与其它先进行人重识别算法进行对比, 实验表明 PFNet 模型有较为优秀的表现.

### 2.1 数据集

为了验证 PFNet 模型的有效性, 本文在 Market-1501、DukeMTMC-reID 和 CUHK03 三个权威行人重识别数据集上进行了大量实验. 每个数据集均被分为三个集合: 训练集、测试集和查询集. 模型通过训练集进行训练, 然后通过查询集进行评估和模型调优, 最后通过测试集进行测试和验证. Market-1501、DukeMTMC-reID 和 CUHK03 数据集的具体信息如表 1 所示.

表 1 数据集信息描述

数据集	训练集 (行人/图片)	测试集 (行人/图片)	查询集 (图片)
Market-1501	751/13 936	750/19 732	3 368
DukeMTMC-reID	702/16 522	702/17 661	2 228
CUHK03	767/7 356	700/5 333	≤700

### 2.2 评价指标

在行人重识别领域, 首位命中率 (Rank-1) 和平均检测精度 (mAP) 是两个广泛使用的评价指标. 行人重识别模型以待检测图片作为输入, 最终输出一个排序列表. Rank-1 就是输出列表中第一张匹配结果为正确的精度, 是行人重识别领域评价模型性能优劣的一项重要指标. mAP 是多个带检测图片的精准率均值, 适用于多标签数据集, 是行人重识别中常用的度量标准.

### 2.3 实验环境与参数配置

本实验基于 PyTorch 深度学习框架进行训练, 实现代码为 Python, 操作系统为 Ubuntu16.04, 在 GPU 配置为 4 片 Tesla v100 的终端进行训练与测试. 在训练阶段, 将输入图片大小调整为  $256 \times 128$ , 使用水平翻转和随机擦除进行数据增强. 用 Adam 优化器训练所有模型, 批量大小为 48, 训练周期为 50, 初始学习率设置为 0.000 5. 为了验证 PFNet 的有效性, 在三个主流数据集 Market-1501、DukeMTMC-reID 和 CUHK03 上进行实验, 并采用 Rank-1 和 mAP 作为评价指标.

### 2.4 消融实验

本文分别基于池化层 GeM、损失函数 Triplet、损失函数 WRLL 以及注意力机制 SimAM 四个部分在 Market-1501、DukeMTMC-reID 和 CUHK03 这三个数据集上进行实验, 实验结果如表 2 所示. 由表 2 可知, 本文使用 ResNet-50 和交叉熵损失函数作为基线. 在三个数据集上, 本文算法较基线均有很明显的提升, Rank-1 分别提升了 1.8%、5.3% 和 25.0%, mAP 分别提升了 5.7%、6.8% 和 27.0%. 在 CUHK03 数据集上效果尤为明显, 主要是 WRLL 能充分挖掘样本信息, 为每个类学习超球面保护样本结构从而提升模型性能. 在基线上叠加 GeM、Triplet、WRLL 和 SimAM 这四个模块的算法精度高于其它实验, 达到最优效果. 表明 SimAM 也会使行人重识别精度有所提升, 这是因为该模块从神经元的角度赋予每个特征点合适的权重, 一定程度上丰富输入特征, 提升模型性能.

表 2 在三个数据集上的消融实验

ResNet-50 + $L_{ID}$ (基线)	GeM	Triplet	WRLL	SimAM	Market-1501		DukeMTMC-reID		CUHK03	
					Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
					/%	/%	/%	/%	/%	/%
✓					93.7	83.9	85.6	74.8	59.3	55.0
✓	✓				94.1	84.5	86.8	75.4	62.1	56.4
✓	✓	✓			95.1	87.9	89.9	80.0	63.8	62.7
✓	✓	✓	✓		95.3	89.6	90.7	80.8	83.3	80.9
✓	✓	✓	✓	✓	95.5	89.6	90.9	81.6	84.3	82.0

实验过程中, 验证了注意力模块在四个残差块中不同位置插入的情况, 实验结果表明, 在残差块 1、3 之后插入注意力模块, 所提出的算法性能最优. 因此, 将 SimAM 注意力模块放在残差块 1 和残差块 3 之后, 此时算法获得最优的识别性能. 相关消融实验的实验结果如表 3 所示.

由表 3 可知, 注意力模块嵌入在四个残差块的任何一个后面, 或者同时嵌入在四个残差块之后, 算法均未获得最优性能. 而当 SimAM 注意力模块放在残差块 1 和残差块 3 之后, 算法获得最优的识别性能, 因此在表 2 以及后续的实验, 注意力模块默认的嵌入位置为残差块 1 和残差块 3 之后.

表 3 注意力模块的嵌入位置消融实验

注意力模块的嵌入位置	Market-1501		DukeMTMC-reID		CUHK03	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
	/%	/%	/%	/%	/%	/%
放在残差块 1、2 之后	95.1	89.0	89.8	79.7	83.3	81.5
放在残差块 1、3 之后	<b>95.5</b>	<b>89.6</b>	<b>90.9</b>	<b>81.6</b>	<b>84.3</b>	<b>82.0</b>
放在残差块 1、4 之后	94.7	88.9	89.6	80.9	82.8	81.3
放在残差块 2、3 之后	94.6	88.7	89.1	80.8	83.0	81.3
放在残差块 2、4 之后	95.3	89.2	90.5	81.3	83.9	81.5
放在残差块 3、4 之后	94.2	88.4	88.4	80.5	82.7	81.0
放在残差块 1 之后	93.4	87.9	86.6	79.2	81.2	80.1
放在残差块 2 之后	93.7	88.0	87.7	80.3	81.1	79.5
放在残差块 3 之后	94.1	87.3	87.4	79.9	80.7	78.8
放在残差块 4 之后	93.8	87.2	87.9	79.7	81.3	79.9
放在残差块 1、2、3、4 之后	94.1	88.5	89.3	80.7	82.5	81.1

## 2.5 与先进实验进行对比

为了验证所提模型的优越性, 本文选择 5 个较为先进的行人重识别算法进行对比分析. 采用联合损失函数训练模型的方法 (BagTricks<sup>[24]</sup>、AGW<sup>[1]</sup>)、注意力机制的方法 (AANet<sup>[24]</sup>、GLWR<sup>[22]</sup>) 以及多尺度、多分支的方法 (Auto-ReID+<sup>[25]</sup>), 主干网络均是 ResNet-50. 表 2 和表 4 展示了算法在 Market-1501、DukeMTMC-reID 和 CUHK03 这三个数据集上的性能对比结果.

由表 4 可知, 在三个数据集上, 本文算法的 Rank-1 分别达到 95.5%、90.9% 和 84.3%, mAP 分别达到 89.6%、81.6% 和 82.0%. GLWR 算法和本文算法模型相似, 也是采取将注意力机制插入到残差网络不同层级中的方法, 以此获取更为丰富的特征, 但是它的注意力只是从通道和空间的维度获取特征, 而本文的注意力还能从图片本身提取特征. 在 Market-1501 数据集上, 本文算法的 mAP 比 GLWR 算法提升了 0.1%. 在 DukeMTMC-reID 和 CUHK03 数据集上, 本文算法的 Rank-1 比 GLWR 算法分别提升了 0.2% 和 2.0%, mAP 分别提升了 0.2% 和 3.1%. 根据实验结果分析, 本算法在三个数据集上表现都优于其它先进算法, 是一个拥有较好的鲁棒性和先进性的行人重识别模型.

由表 4 可知, 本文提出的网络架构相对 BagTricks、GLWR 参数略有提升, 但本文算法相对 BagTricks 在 Rank-1 上提升了 25.5%, 在 mAP 上提升了 25.4%. 相对 GLWR, 本文算法在 Rank-1 和 mAP 上分别提升了

2.0% 和 3.1%。虽然 AGW 整体参数量较少,但 AGW 的 Rank-1 和 mAP 整体得分相对本文算法较低,这说明 AGW 难以获得到特征图中大部分重要的前景信息。整体来看,相对其它网络,本文在微增参数的前提下 Rank-1 和 mAP 取得了较大的改善。

表 4 在三个数据集上与前沿算法的对比

方法	Market-1501		DukeMTMC-reID		CUHK03		网络复杂度
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	
	/%	/%	/%	/%	/%	/%	
BagTricks <sup>[24]</sup>	94.5	85.9	86.4	76.4	58.8	56.6	63.15 M
AGW <sup>[1]</sup>	95.1	87.8	89.0	79.6	63.6	62.0	25.11 M
AANet <sup>[24]</sup>	93.9	82.5	86.4	72.6	\	\	\
GLWR <sup>[22]</sup>	95.5	89.5	90.7	81.4	82.3	78.9	64.17 M
Auto-ReID+ <sup>[25]</sup>	95.8	88.2	90.1	80.1	78.1	74.2	\
本文	95.5	89.6	90.9	81.6	84.3	82.0	64.28 M

## 2.6 实验可视化

本文将算法的排序结果进行了可视化展示,如图 3 所示,第 1 列红色框为输入的待检索图片,后 7 列图片是输入样本(红色框图片)经过模型可视化得到的最终结果。可以观察到,本文的算法没有匹配错误,且在一些特殊情况下也能准确识别。不仅对正常行走且周围无遮挡的行人能准确识别其身份,而且对面部遮挡的戴帽子行人或只有背部特征的行人也能精准匹配其行人身份。另外通过查询结果的第 3 组第 3 张图片也可以观察到,在含有并行行人信息的情况下,本算法也能准确进行识别。



图 3 行人检索排序展示

通过客观性能对比和主观可视化分析,实验结果显示,本文的算法能够在行人所处环境相对复杂的情况下进行高准确率的识别工作,验证了本算法具有较好的识别性能、较强的鲁棒性优势。

## 3 结论

针对行人重识别中存在的遮挡和相似外观等问题,本文设计了一个新的行人重识别算法,经过实验验证,发现所提算法在一定程度上缓解了以上问题。该算法的主干网络为 ResNet-50,通过在网络的第一层残差块和第三层残差块后嵌入 SimAM 注意力模块为每个神经元赋予合适的权重,并采用 GeM 模块获取具有区分性的特征,最后基于提出的联合损失函数来训练该网络模型。这里,SimAM 注意力模块属于无参注意力模块,整体而言,与基线对比算法相比,本文在增加少量参数的前提下,所提出的算法在 Rank-1 和 mAP 指标上取得了较大的性能改善。所提模型在三个公共数据集上进行测试,并将其和目前较先进的算法进行对比,实验结果表明本文设计的基于无参注意力机制的行人重识别模型具有很高的精准度,其性能领先目前大部分行人重识别算法。后期将进一步轻量化行人重识别网络架构,以促进本文网络在较低网络复杂度的前提下取得更好的性能。

## 参考文献:

- [1] YE M, SHEN J, LIN G, 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] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016.
- [3] HUANG G, LIU Z, LAURENS V, et al. Densely connected convolutional networks[C]//2017 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017.

- [4] ALMAHAIRI A, BALLAS N, COOIJMANS T, et al. Dynamic capacity networks[C]//The 33rd International Conference on Machine Learning (ICML). New York: ACM, 2016.
- [5] JIE H, LI S, GANG S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
- [6] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//The European Conference on Computer Vision (ECCV). München: Springer, 2018.
- [7] PARK J, WOO S, LEE J Y, et al. BAM: bottleneck attention module[C]//British Machine Vision Conference (BMVC). Newcastle: Springer, 2018.
- [8] TAY C P, YAP K H, ROY S. AANet: attribute attention network for person re-identifications[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019.
- [9] ZHANG Z, LAN C, ZENG W, et al. Relation-aware global attention for person re-identification[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020.
- [10] YANG B, SHAN Y, PENG R, et al. A feature extraction method for person re-identification based on a two-branch CNN[J]. Multimedia Tools and Applications, 2022, 81(1): 39169-39184.
- [11] WEI L, ZHANG S, WEN G, et al. Person transfer GAN to bridge domain gap for person re-identification[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City: IEEE, 2018.
- [12] DAI Z, CHEN M, GU X, et al. Batch dropblock network for person re-identification and beyond[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019.
- [13] MENG Z, KARANAM S, WU Z, et al. Re-identification with consistent attentive siamese networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019.
- [14] YAN C, PANG G, WANG L, et al. BV-person: a large-scale dataset for bird-view person re-identification[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021.
- [15] GU X, CHANG H, MA B, et al. Clothes-changing person re-identification with RGB modality only[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE, 2022.
- [16] YANG M, HUANG Z, HU P, et al. Learning with twin noisy labels for visible-infrared person re-identification[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE, 2022.
- [17] CHO Y, KIM W J, HONG S, et al. Part-based pseudo label refinement for unsupervised person re-identification[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE, 2022.
- [18] WANG H, SHEN J, LIU Y, et al. NFormer: robust person re-identification with neighbor transformer[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE, 2022.
- [19] WANG T, LIU H, SONG P, et al. Pose-guided feature disentangling for occluded person re-identification based on transformer[C]//Association for the Advancement of Artificial Intelligence (AAAI). Vancouver: AAAI, 2022.
- [20] YANG L, ZHANG R Y, LI L, et al. SimAM: a simple, parameter-free attention module for convolutional neural networks [C]//International Conference on Machine Learning (ICML). Lugano: ACM, 2021.
- [21] RADENOVIĆ F, TOLIAS G, CHUM O. Fine-tuning CNN image retrieval with no human annotation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(7): 1655-1668.
- [22] GONG Y, WANG L, LI Y, et al. A discriminative person re-identification model with global-local attention and adaptive weighted rank list loss[J]. IEEE Access, 2020, 8: 203700-203711.
- [23] HU B, XU J, WANG X. Learning generalizable deep feature using triplet-batch-center loss for person re-identification[J]. Science China Information Sciences, 2021, 64: 120111.
- [24] LUO H, JIANG W, GU Y, et al. A strong baseline and batch normalization neck for deep person re-identification[J]. IEEE Transactions on Multimedia, 2020, 22(10): 2597-2609.
- [25] GU H Y, FU G Y, LI J M, et al. Auto-ReID+: searching for a multi-branch ConvNet for person re-identification[J]. Neurocomputing, 2021, 435: 53-66.