DOI:10.19651/j.cnki.emt.2518088

基于特征增强的行人重识别算法*

姬晓飞 孙英超 宋京浩
(沈阳航空航天大学自动化学院 沈阳 110136)

摘 要:针对现有行人重识别算法过于依赖卷积神经网络作为主干网络,导致其过度关注具有显著特征的区域,而忽略了广义前景特征,进而全局特征信息不够丰富,对细微的判别特征关注度较差的问题,提出了一种基于特征增强的行人重识别算法。通过位置编码和多层多头注意力结构,更好地利用空间上下文信息,增强对空间相对位置的理解,有效捕捉空间结构信息,从而提升特征的表征能力,提升全局提取能力。局部分支利用空间向量关联的特征矩阵优化空间注意力,捕捉更加紧凑的广义外观特征,并通过建模不同通道间的关系加强通道维度特征表达,突出显著特征信息,从而增强判别性特征的关注度。最后,采用 softmax 损失、三元损失和中心损失在 Market-1501 和 DukeMTMC-ReID 数据集上进行了模型训练,实验结果充分证明了所提出算法的有效性和性能优势。

关键词:行人重识别;多头注意力;位置编码;向量关联的特征矩阵
中图分类号:TN391 **文献标识码**:A **国家标准学科分类代码**:510.04

Feature-enhanced based pedestrian re-identification algorithm

Ji Xiaofei Sun Yingchao Song Jinghao
(School of Automation, Shenyang Aerospace University, Shenyang 110136, China)

Abstract: Existing pedestrian re-identification algorithms heavily rely on convolutional neural networks as the backbone, which often leads to an overemphasis on regions with prominent features while neglecting broader foreground features. This results in insufficiently rich global feature representations and inadequate attention to subtle discriminative features. To address these issues, we propose a feature-enhanced pedestrian ReID algorithm. The global branch utilizes position encoding and a multi-layer, multi-head attention structure to better leverage spatial context information, enhance the understanding of relative spatial positions, and effectively capture spatial structural information, thereby improving feature representation and global feature extraction capability. The local branch optimizes spatial attention using feature matrices associated with spatial vectors, enabling the capture of more compact general appearance features. Furthermore, by modeling the relationships between different channels, it strengthens feature expression in the channel dimension, highlighting distinctive features and improving the attention to discriminative characteristics. Finally, the model is trained using softmax loss, triplet loss, and center loss on the Market-1501 and DukeMTMC-ReID datasets. Experimental results demonstrate the effectiveness and superior performance of the proposed algorithm.

Keywords: person re-identification; multi-head attention; position encoding; feature matrix with spatial vector association

0 引 言

行人重识别[1](person re-identification, ReID)也称行人再识别,旨在从不同摄像头捕捉的图像中识别并匹配同一行人[1]。由于其具备判断视频序列中是否存在特定行人的能力,行人重识别在智能监控、商场走失儿童查询和嫌疑人追逃等领域展现出了巨大的应用潜力。行人重识别技术主要涵盖特征提取与度量学习两大核心模块。特征提取模块采用卷积神经网络(convolutional neural networks, CNN)以提取行人的视觉特征,为度量学习模块提供基础的数据支撑。因此,特征提取模块在行人重识别中占据至关重要的地位,其特征捕获的能力直接决定了重识别系统的准确度与鲁棒性。

目前的行人重识别算法大多采用卷积神经网络作为主

收稿日期:2025-02-17
*基金项目:辽宁省教育厅重点攻关项目(LJKZZ20220033)资助

干网络,而卷积神经网络通过较大的卷积核来增加感受野。为减少模型参数量,通过大量堆叠小卷积核构成的重识别网络逐步涌现,但其少量的卷积堆核叠却限制了感受野的大小。因此,主流的重识别算法存在着过于关注显著特征而不能关注具有广义特征的前景信息的弊端。除此之外,大量注意力被引入到主干网络,由于注意力的降维操作使得重识别网络过于关注显著部位的特征信息,忽略了大量的其他部位的特征信息,导致网络无法聚焦于全局特征以及辨识度较高的细微特征,使得行人重识别的准确率提升不显著[2]。

为了获取带有更多的细微信息的全局特征,Zhao 等[3]提出了 Spindle NET。该网络在特征提取阶段,通过姿态估计算法将人体特征图划分为多个局部区域,并将局部肢体特征与全局特征进行融合,从而构建出更为精细的行人全局特征表示。该方法充分整合了全局特征与局部特征的互补优势。增强了特征的表达能力,为行人重识别任务提供了新的思路。但该算法无法关注前景信息,且融合后的特征存在较多的干扰信息,因此其在局部特征的补充和干扰信息的抑制方面仍有改进空间。

为了减少特征融合带来的干扰信息,多种基于局部特征的行人重识别方法相继涌现。Sun 等[4]提出了将卷积神经网络提取的特征图等比例切分成 6 份的 Part-Based Convolutional Baseline 网络,其中每一份特征图单独捕捉行人局部特征,通过互相独立的全连接层分别预测行人身份。此网络虽然得到了更为丰富的特征表达,减少了信息的干扰,但是其不考虑各个局部语义信息之间的关联性,导致其不能关注具有广义特征的前景信息,从而得到地匹配效果并不理想。

无论对于全局特征信息还是局部特征信息都存在提取不够完整的问题。为此,研究者们提出了多种融合全局与局部特征的多分支架构,以充分利用两者的互补性,提升特征表达的完整性和判别能力。Su 等[5]提出了 Pose-driven Deep Convolutional 网络,通过姿态估计划分六个身体部分区域,以仿射变换切分图像作为局部分支,全局与局部图像同时进行特征抽取,浅层共享双分支网络。Wang 等[6]提出了一种多粒度网络,该网络架构由 1 个全局分支和 2 个局部分支构成。通过将特征图水平分割为多个部分,并通过调整各局部分支的条纹数量,以实现多粒度局部特征的提取。Miao 等[7]提出的姿态引导特征对齐方法,专注于提升特征对齐的准确性,其结构包含 3 个分支。首先使用 ResNet-50 提取全局特征,然后将全局特征图水平分割为 6 个条形区域,以捕捉局部细节。第 3 个分支利用人体关键点信息,帮助模型聚焦未遮挡的区域,提升特征对齐效果。多分支结构的网络能够更好地提取行人的全局特征与局部细节,同时增加不同粒度特征之间的互补性,使得模型在行人识别任务中的准确性得到了提升。尽管多分支架构提高了特征的多样性和对细微特征的捕捉能力,使其生成更具辨识力的行人特征表示。然而,其主干网络的固定感受野在一定程度上制约了对长程依赖关系的捕捉,导致网络过于关注显著特征而忽视广义前景信息,影响了识别准确率和性能的提升。

综上所述,尽管当前行人重识别研究虽在多层次特征融合方面取得显著进展,但全局网络存在空间建模薄弱的问题,具体变现为上下文关联不够紧密,前景信息聚焦较差;而局部网络的动态关联性较差,无法聚焦于行人部位间空间依赖关系;尽管多分支网络在行人重识别任务中表现出优越的性能,其仍存在着过于关注显著特征而不能关注具有广义特征的前景信息,融合后的全局特征信息不够丰富以及网络模型对特征图中细微的判别特征关注度较差等问题。针对上述问题,本文提出了一种基于特征增强的行人重识别网络。本文将行人重识别的 OSNet 网络修改为双分支结构作为特征提取网络,动态结合局部全局信息,改善单个局部或者全局网络存在的动态关联性较差或空间建模薄弱问题;提出了全局特征增强模块,与 OSnet 网络结合,并有效结合多头注意力机制与三重注意力的掩码,提升了网络的长程依赖捕捉能力使其有效建模跨区域语义关联并通过位置编码的位置约束的注意力权重标定。同时,借鉴 Transformer 结构中的 MLP 模块,设计了增强广义特征的空间注意力与增强显著特征的通道注意力,将其作为局部分支。增强广义特征的空间注意力通过利用空向量构建关联的特征矩阵,并通过与空间注意力(spatial attention module,SAM)[8]结合从而捕捉更为紧凑的广义外观特征信息,而增强显著特征的通道注意力以通道作为节点重构通道矩阵,并结合通道注意力(channel attention module,CAM)[8]重构其特征掩码加强重要通道维度的特征表达,进一步提高了广义特征与显著特征之间的相关性,得到了表达能力更为丰富的全局特征。最终,通过交叉熵损失、三元损失、中心损失对模型进行了训练。Market-1501 和 DukeMTMC 数据集上的实验结果表明,所提出的算法在有效性和性能上均展现出显著优势性。

1 算法总体设计

基于特征增强的行人重识别网络由两个分支构成,分别是向量交互增强的全局分支、广义特征与显著特征联合提取的局部分支。为了搭建精简网络架构,本文采用了多尺度网络(omni-scale feature learning for person reidentification,OSNet)作为核心架构来获取行人的粗略特征表达,其网络主要由 5 个卷积块构成,Conv1、Conv2、Conv3、Conv4 以及 Conv5。选取 OSNet 的前 4 个卷积作为局部分支的主干网络。

向量交互增强的全局分支:首先将行人图片送入主干网络进行粗略的特征提取,将带有背景干扰信息的特征图送入 Conv5,通过 1×1 卷积、归一化与 Relu 去除背景信息,将特征图重塑为三维,分别进行位置编码(positional encoding)以及多层多头注意力块(multi-head attention)[9]

处理，加强特征之间的联系，融合后重构为四维特征图，最终以广义平均池化操作进行降维，通过归一化及全连接层得到全局特征向量。如图 1 的全局分支，全局分支能够有效地提取并强化行人的整体特征以及轮廓细节信息。

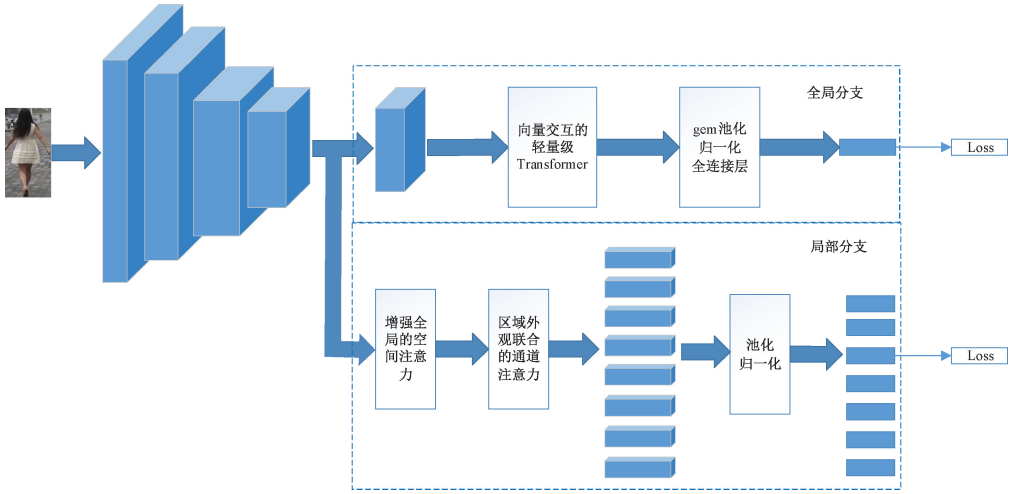


图 1　网络模型总体架构图

Fig. 1　Overall architecture diagram of the network model

广义特征与显著特征联合提取的局部分支：行人图片通过主干网络提取出带有背景干扰信息的粗略行人特征，将其输入到增强全局结构的空间注意力中，加强全局信息感知，去除背景干扰，得到更加关注于行人语义信息的特征。随后将其特征图送入到区域外观联合的通道注意力中，加强更为显著的局部语义信息表示，进而生成既包含行人全局外观细节[10]又强化了局部关键信息[11]的特征图。最终，特征图被分割成 8 个局部特征，每个特征都经过特征增强处理。这些局部特征通过全局平均池化(global avgpool)进行降维，并经过归一化处理，形成局部特征向量。整体网络模型架构如图 1 所示，局部分支能够捕捉到更为精细的行人细粒度特征信息。最终，分别对两个分支的特征向量计算损失。

2　算法网络结构及实现

2.1　OSNet 网络结构

本文采用全尺度网络(OSNet)[12]作为重识别提取行人特征的主干网络。该网络通过逐层堆叠其提出的轻量级 bottleneck 模块进行构建，避免了在网络的不同深度定制不同块的问题。OSNet 的设计使得网络能够在多个尺度上提取特征，从而克服了单一感受野下提取固定尺度特征的局限性。OSNet 由多个卷积模块组成，以支持多尺度信息的提取和更丰富的特征表示。其主干结构如表 1 所示。

表 1　OSNet 网络结构

Table 1　OSNet network architecture

stage	output	OSNet
Conv1	128×64,64	7×7 conv,stride=2
	64×32,64	3×3 max pool,stride=2
Conv2	64×32,256	Bottleneck×2
transition	64×32,256	1×1 conv
	32×16,256	2×2 average pool,stride=2
Conv3	32×16,384	Bottleneck×2
Transition	32×16,384	1×1 conv
	16×8,384	2×2 average pool,stride=2
Conv4	16×8,512	Bottleneck×2
Conv5	16×8,512	1×1 conv

2.2　向量交互增强的全局分支

针对绝大多数的主干卷积神经网络提取行人的特征感受野依然较为固定，只能提取出多范围的的局部特征，存在着过于关注显著特征而不能关注具有广义特征的前景信息的问题，提出了向量交互增强的全局分支，如图 1 中的全局分支。并设计了轻量型的向量交互增强 Transformer 结构。模块通过三重注意力(triplet attention)聚焦于特征图的宽度、高度及深度，获取更丰富的信息，掩码向多头注意力层进行投影，加强特征之间的向量交互。通过捕捉行人特征间的长程依赖关系，网络能够提取出更加紧密相关的全局特征信息，从而增强重识别网络的特征提取能力，使网络关注具有广义特征的前景信息。

此分支的具体结构设计如下：将 OSNet 网络输出的特征图并行送入位置编码与并行注意力模块，其中并行注意力模块利用多头注意力机制的优势在处理数据时捕捉全局上下文信息，以三重注意力获取长、宽及深度多维信息特征，并通过 6 层注意力层，使网络能够在每一层中对不同区域的特征进行综合，学习到全局性的信息。多头注意力本身往往缺乏空间位置感知能力。通过将位置编码，使得网络能够更好地理解每个特征在空间中的位置关系。将位置编码和注意力层并行融合，可以提高模型的灵活

性,使网络不仅增强了位置感知能力,还通过多种注意力提供了多样化的交互能力,得到更为丰富的信息表达特征。构建的轻量型的向量交互增强 Transformer 结构图如图 2 所示。

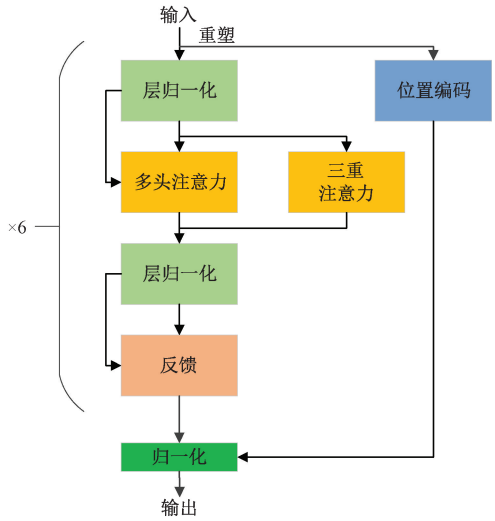


图 2 轻量型的向量交互增强 Transformer 结构

Fig. 2 Lightweight vector interaction Transformer

主干网络得到的特征图尺寸为 16×8×512,经过变换映射为 128×512 的向量送入至 6 层多头注意力块与位置编码。位置编码将其奇数维度与偶数维度分别采用余弦函数与正弦函数编码定义向量的位置信息。余弦函数定义为:

$$PE_{(pos,2i)} = \cos\left(\frac{pos}{10\ 000^{2/d}}\right) \tag{1}$$

正弦函数定义为:

$$PE_{(pos,2i+1)} = \sin\left(\frac{pos}{10\ 000^{(2i+1)/d}}\right) \tag{2}$$

其中,d 为特征向量的维度,pos 是位置索引,i 为维度索引。通过上述变换最终将位置信息映射到向量中,输出 128×512 的向量。

多头注意力机制首先将输入片段通过线性变换层转换为查询(query,Q)、键(key, K)和值(value,V)3 个向量。每个向量的维度设定为 $h \times dim$,其中 h 代表注意力头的数量,此处 $h=16$,而 $dim=16$ 表示每个注意力头的维度大小。

每一层通过注意力机制增强补丁间的关系,再通过前馈网络进行非线性变换。经过 6 层处理后,输出 128×512 的向量,并与位置编码相加融合,将特征图重新整合为 16×8×512 的维度。随后,采用广义均值池化动态维持特征的多样性,从而提取出 512 维的特征向量。此向量经过归一化处理及全连接层的转换后,用于损失函数的计算。

2.3 广义特征与显著特征联合提取的局部分支

针对局部分支切片存在的局部特征不对齐及细微特征信息提取较差的问题,提出了广义特征与显著特征联合提取的局部分支。此分支首先利用 OSNet 前四层提取的带有背景信息的粗略特征信息,然后将其特征图送入增强广义特征的空间注意力。首先对于特征进行降维,并沿不同的维度建立对称的向量节点,将其构建向量关联的特征矩阵来挖掘全局范围的相关性和语义信息。通过对长与宽的反向构建,去除复杂背景影响,将三分支叠加送入 SAM 中。使其重新聚焦于人体轮廓结构。其增强广义特征的空间注意力结构如图 3 所示。

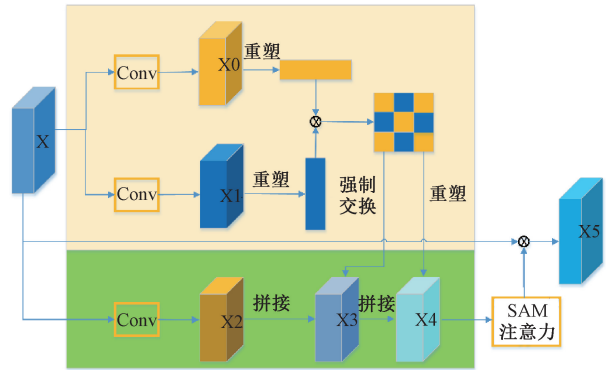


图 3 增强广义特征的空间注意力结构

Fig. 3 Enhanced generalized feature spatial attention

随后将加强人体外观轮廓提取的特征图送入增强显著特征的通道注意力,通过将长宽压缩构造通道节点,通道重构为对应节点的权重矩阵,沿不同方向反变化为四维特征图,与原特征拼接送 CAM 中,学习更为显著的通道掩码,在具有轮廓前景信息的基础上,突出更为显著的特征信息,构建的增强显著特征的通道注意力模块结构如图 4 所示。

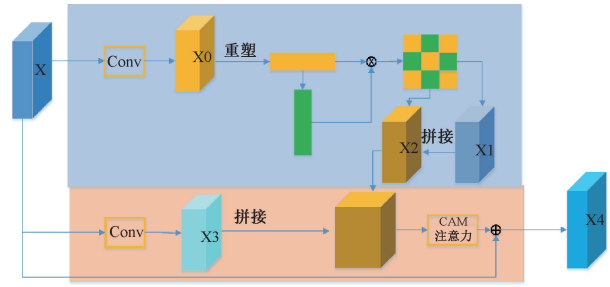


图 4 增强显著特征的通道注意力结构

Fig. 4 Enhanced salient feature channel attention

为了重构行人整体判别性更强的特征信息,将其特征图分割为 8 部分,此分支将多个局部分支的向量串联起来,构建公式如下:

$$\boldsymbol{g} = [\boldsymbol{g}_1^{\mathrm{T}}, \boldsymbol{g}_2^{\mathrm{T}}, \cdots, \boldsymbol{g}_n^{\mathrm{T}}]^{\mathrm{T}} \tag{3}$$

与 PCB[4] 网络不同,此分支根据拼接向量计算哪一部分用于产生 id 预测损失,其定义如下:

$$L_{id}^{p} = -\frac{1}{N_s}\sum_{N_s}^{i=1}\log\left(\frac{\exp((\boldsymbol{W}_p^{y_i})^{\mathrm{T}}\boldsymbol{g}_p^{i} + b_{y_i})}{\sum_j \exp((\boldsymbol{W}_p^{j})^{\mathrm{T}}\boldsymbol{g}_p^{j} + b_j)}\right) \tag{4}$$

其中，$\boldsymbol{W}^{j}$、$\boldsymbol{W}^{y_i}$、$\boldsymbol{W}_{p}^{y_i}$ 分别是权重矩阵 $\boldsymbol{W}$(g 的单一分类器)的第 y 列和 y_j 列。由于向量 $\boldsymbol{g}$ 包含了输入图像的全部信息，因此使用单个 id 预测损失可以驱动 $\boldsymbol{g}$ 学习到足够的判别信息。

广义特征与显著特征联合提取的局部分支的具体实施步骤如下：首先利用主干网络对行人图像进行初步的提取，从而生成尺寸为 16×8×512 的粗略特征。随后增强全局结构的空间注意力模块，利用卷积对空间维度进行压缩，获取 128×512 的特征图，对于 512 个通道，将其重构为 128 个节点，转换为 128×128 的相关特征图。通过空间注意力学习特征的空间相关性，将掩码反变换与特征图重构，得到与输入维度相同的加强全局关联性的特征图。再将其输入至区域外观联合的通道注意力，利用 1×1 卷积对通道维度进行压缩并重构其特征图，通过通道注意力加强学习关联性更强的通道，通过激活函数后将掩码反变换，并映射至加强全局关联性的特征图，得到丰富的特征信息表达的特征。随后，将特征图谱划分为 8 个尺寸为 2×8×512 的子特征图，通过全局平均池化操作以捕捉更丰富的细节信息，并将这些子特征图连接融合，最终形成一个 4 096 维的特征向量，计算其的损失函数。

2.4　损失函数设计

由于来自全局和局部分支的特征向量整体作为人员重识别的最终描述，为了更好的分类特征向量增强模型分类性能，本文采用 softmax 损失、软三边损失以及中心损失对模型进行约束。其中任意选择每批次的类别和图片，软边界三元组损失公式如下：

$$L_{triplet}=\sum_{i=1}^{P}\sum_{a=1}^{K}[\alpha+\|\boldsymbol{x}_a^{(i)}-\boldsymbol{x}_P^{(i)}\|_2-\|\boldsymbol{x}_a^{(i)}-\boldsymbol{x}_n^{(j)}\|_2]_+ \tag{5}$$

其中，$\boldsymbol{x}_a^{(i)}$ 代表自固定样本捕获的特征向量，是正样本捕获的特征向量；$\boldsymbol{x}_P^{(i)}$ 代表自负样本捕获的特征向量；负样本，$\boldsymbol{x}_n^{(j)}$ 代表锚点样本不属于同一身的某个样本的特征向量；α 为边缘超参数；而$[\cdot]_+$为 Hinge Loss 函数，即确保当括号内值为负损失为零。具体中心损失表达式如下：

$$L_{center}=\frac{1}{2N}\sum_{N}^{i=1}f_i-c_{y_i}^2 \tag{6}$$

其中，N 为训练样本的总量；f_i 表示第 i 个样本的特征向量；c_{y_i} 代表类别 y_i 的类别中心；网络的损失为三者之和。其具体表达式为：

$$L_{total}=L_{s-id}+\gamma_t L_{tiplet}+\gamma_c L_{center} \tag{7}$$

其中，γ_t、γ_c 为权重因子。

3　实验结果

3.1　数据集

为了验证所提算法效果，本文选取 ReID 公开数据集 Market-1501[13]、DukeMTMC[14]进行实验验证。

1) Market-1501 数据集。Market-1501 数据集是行人重识别领域的重要基准数据集。该数据集囊括了由 6 个独立监控摄像头捕捉的 1 501 名行人图像，总计 32 668 帧。该数据集被精心划分为 3 个主要部分：训练集、测试集及查询集。训练集汇聚了 751 名行人的 12 936 帧图像，旨在为模型提供学习材料；测试集则包含了 750 名行人的 19 732 帧图像，用以检验模型的泛化能力；查询集由 750 名行人的 3 368 帧图像构成，用于模拟实际查询场景，主要用于评估重识别模型在检索任务中的表现。Market-1501 部分图像如图 5 所示。

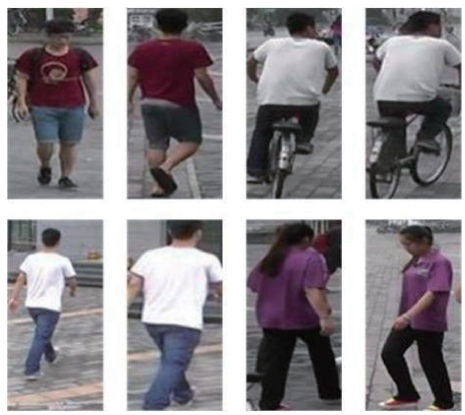


图 5　Market-1501 数据集图像

Fig. 5　Market-1501 dataset images

2) DukeMTMC 数据集。DukeMTMC 数据集是一个大规模、多目标、多摄像头的行人重识别数据集。该数据集由 8 个同步摄像头在不同的环境下采集，包含了来自不同视角的行人轨迹数据，涵盖了 1 812 个独立人物。DukeMTMC 数据集的规模十分庞大，共包括 36 411 张图像。此类图像是从原始视频中每 120 帧提取一张，确保了数据的多样性和覆盖面。其训练集涵盖 702 个人的 16 522 张图片，查询集包括 2 228 张图片。数据集提供了精确的边界框标注，同时还包含行人属性信息，如性别、是否背包等，能够为行人重识别任务提供了丰富的特征。部分图像如图 6 所示。

3) MSMT17 数据集。MSMT17 是面向行人重识别研究的大规模基准数据集，由浙江大学团队于 2018 年构建发布。该数据集通过跨场景多模态采集系统，在北京的交通枢纽区域部署 15 个监控节点(含 12 个户外节点与 3 个室内节点)，在 4 个典型时间段(清晨、正午、午后、黄昏)及多种天气条件下采集得到连续 15 天的监控资源。相较数据集 Market-1501 和 DukeMTMC，该数据集在规模与多样性维度实现显著提升。其训练集包含 1 041 个行人身份的 32 621 张检测图像，测试集则包含 3 060 个身份的 93 820 张图像，总计达 126 441 张标注样本。每个身份平均出现在 2～3 个独立摄像头视角下，有效增强了跨视角匹配的复杂性。MSMT17 数据集部分图像如图 7 所示。

图 6 DukeMTMC 数据集图像

Fig. 6 DukeMTMC dataset images

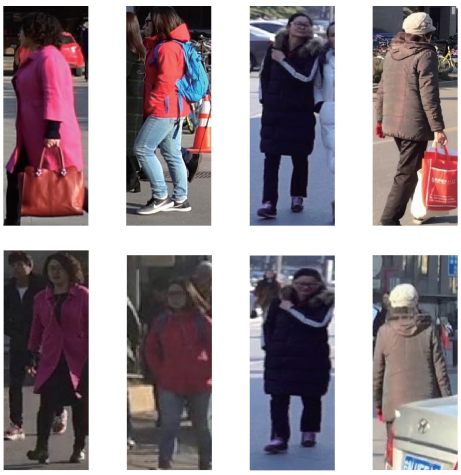


图 7 MSMT17 数据集图像

Fig. 7 MSMT17 dataset images

3.2 参数配置与评价指标

1) 参数配置。在本文实验中,所有输入图像均被调整为 256 pixel×128 pixel,批次处理 64 张图像。训练过程中采用了随机翻转和随机擦除技术,训练周期共计 150 轮。采用 Adam 优化算法[15],初始学习率设为 3.5×10^{-5}。在训练初期,即前 20 轮,采用线性热身策略逐步将学习率提升至 3.5×10^{-4}。随后,在完成 60 轮训练后,学习率降至 3.5×10^{-5},并在 90 轮后进一步调整至 3.5×10^{-6}。测试阶段,图像同样被归一化并调整至 256 pixel×128 pixel。

2) 评价指标。本研究选用平均精度均值 mAP 和首位命中率 Rank-1 作为性能评价的标准。其中,Rank-1 反映了在行人重识别任务中,mAP 查询目标的正确身份在检索结果中位列第一的概率。则是一个综合性的评价指标,它考虑了所有正确匹配图像在排序结果中的平均位置。相关计算公式如下:

$$mAP=\frac{\sum_{c}^{k=0}AP_k}{C} \tag{8}$$

其分子表示各类别的平均精度 AP 的总值,而 C 表示类别数总和。

3.3 消融实验

为了研究所提双分支网络中的单独分支对于网络模型整体性能的影响,本文在 Market-1501 数据集上进行了消融测试,实验在同相同平台下运行 5 次,取其结果最好一次中的数据作为实验结果。其测试结果如表 2 所示,在 Market-1501 数据集上,无论是全局分支还是局部分支相较于主干网络均有改善。局部分支改善更为明显,Rank-1 提升了 0.9%,mAP 提高了 4%。在此基础上,通过联合构建全局分支,使得识别准确率得到了进一步提升。实验结果表明,本文提出的双分支联合框架是有效的,其不仅加强了行人整体轮廓信息与语义信息的关联,还得到了更具判别性的细粒度局部特征,在保证较为精简的参数下,有效提升了网络的识别性能。

表 2 各分支性能的对比

Table 2 Comparison of the performance of each branch

方法	DukeMTMC	
	Rank-1/%	mAP/%
主干分支	94.8	84.9
向量交互增强的全局分支	95.0	86.7
广义特征与显著特征联合提取的局部分支	95.7	88.9
本文	96.3	89.5

3.4 与其他先进方法对比

如表 3 所示,本文提出方法与目前其他先进方法在两个数据集上作比较。其中 BoT 网络[14]采用了许多有效的方法进行训练。OSNet[12]是一个堆叠轻量级瓶颈模块的小架构网络。Cheng 等[16]所提出的 SCR 网络时采用多分支结构共同训练的网络模型。HAT[11]是结合了卷积神经网络(CNN)与 Transformer 模型的高度网络。

如表 3 所示,本文方法在性能上显著优于 OSNet[12],在 Market-1501、DukeMTMC 以及 MSMT17 数据集上均实现了显著提升。具体而言,在 Market-1501 数据集上,Rank-1 指标提升了 1.5%,mAP 提升了 4.6%;在 DukeMTMC 数据集上,Rank-1 提升了 2.6%,mAP 提升了 8.5%。

与采用多分支结构联合训练的 SCR 模型[16]相比,本文方法在 Market-1501 数据集上表现更为突出,Rank-1 提升了 0.6%,mAP 提升了 0.5%。在 DukeMTMC 数据集上,本文方法同样表现优异,Rank-1 提升了 0.1%,mAP 提升了 0.6%,实现了性能的全面超越。

表 3　与其他先进方法在图像数据集结果对比

Table 3　Comparison of results on image datasets with other state-of-the-art methods

方法	Market-1501		DukeMTMC		MSMT17	
	Rank-1/%	mAP/%	Rank-1/%	mAP/%	Rank-1/%	mAP/%
BoT[14]	94.5	85.9	86.4	76.4	—	—
OSNet[12]	94.8	84.9	88.6	73.5	—	—
SCR[16]	95.7	89.0	91.1	81.4	—	—
HAT[11]	95.6	89.5	91.4	80.4	86.3	61.2
本文	96.3	89.5	91.2	82	86.3	61.7

对比于近年新出现的卷积神经网络与 Transformer 联合网络 HAT[11]，本文以 Rank-1 提高 0.7% 的水平在 Market-1501 数据集实现了超越。在更具挑战性的 MSMT17 数据集上 mAP 提高了 0.5%，但 Rank-1 提高了 0.2%。在 DukeMTMC 数据集 mAP 提高了 1.6%，但 Rank-1 降低了 0.2%。但与 HAT[11] 网络此类 Transformer 与卷积联合架构相比，如表 4 所示，本文网络的参数与计算量远大幅度降低，虽然参数量相较于骨干网络的 OSNet 提高较多，但其精准度提升较大，其推理速度仍存在较强的竞争力，因此本文方法在需平衡精度与计算资源的场景具有较强可行性。

表 4　与其他方法参数对比

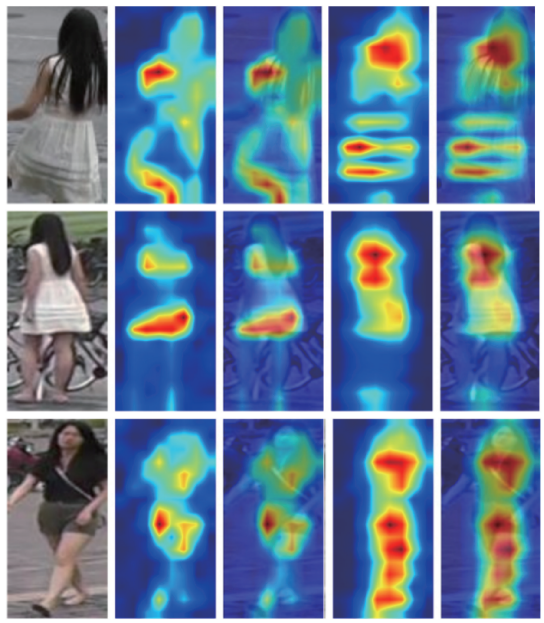
Table 4　Comparison of parameters with other methods

方法	参数量/M	测试时间/(s/batch)
OSNet[12]	2.17	19.92
HAT[11]	49.45	28.56
本文	13.32	21.32

3.5　特征可视化分析

为了验证所提算法的性能，本文在行人重识别图像数据集 Market-1501 上进行了可视化操作，相关结果如图 8 所示。

图 8 第 1 列展示了原始图像，第 2 和 3 列分别展示了经过全局分支处理后的特征图可视化效果及其热图加权到原始图像的结果。全局分支的特征图主要集中在行人的躯干和四肢区域，表明该分支能够有效捕捉行人的全局信息，帮助模型理解行人的整体姿态和结构。这种全局信息的捕捉不仅增强了特征之间的关联性，还为后续的局部特征提取提供了重要的上下文支持。第 4 和 5 列展示了局部分支的特征图可视化效果及其热图加权到原始图像的结果。局部分支利用空间注意力和通道注意力机制，能够有效提取行人的局部关键特征信息，如头部、肩部、手部等细节特征。从热图中可以看出，局部分支的特征图更加聚焦于行人的关键部位，尤其是在头部和手部区域表现出较高的关注度。这种局部特征的捕捉能力使得模型能够更好地识别行人的细微判别特征[17]，增强了模型对行人关键部位的识别能力，能够帮助模型在全局信息的基础上进一步聚焦于行人的局部细节，提升特征的判别性。



(a) 原图 (a) Original image　(b) 全局分支 (b) Global branch　(c) 局部分支 (c) Local branch

图 8　可视化热图

Fig. 8　Visualization heatmap

全局分支和局部分支的特征图可视化效果具有明显的互补性。全局分支侧重于捕捉行人的整体结构信息，能够有效提取行人的全局姿态[18]和轮廓特征，而局部分支则聚焦于行人的局部细节，捕捉了行人的关键部位信息[19]。与文献[20]强制叠加各类不同特征不同的，这种全局与局部信息的结合使模型能够兼顾行人的整体外观和细节特征，从而生成更加全面且多样化的特征表达。因此在行人重识别任务中，全局分支可以帮助模型识别行人的整体姿态和轮廓，而局部分支则可以帮助模型识别行人的头部、肩部等关键部位。这种全局与局部信息的协同作用，显著提升了模型对行人特征的表达能力，增强了模型的鲁棒性。

4　结　　论

本文提出了一种基于特征增强的行人重识别方法，以

解决现有方法忽略广义特征前景信息、全局特征信息不足以及对细微判别特征关注度较差等问题。该方法以 OSNet 为主干网络,设计了基于向量交互增强的全局分支和广义特征与显著特征联合提取的局部分支两个分支。全局分支通过增强特征交互和远距离特征依赖,提升了全局特征表达能力;局部分支通过特征映射向量重构关联性特征图,使网络更聚焦于行人轮廓,同时通过通道压缩强化判别性特征,实现了对行人轮廓和局部特征的丰富表达。实验结果显示,该方法在 Market-1501 数据集和 DukeMTMC 数据集上表现优异,与现有先进方法相比具有竞争力。未来研究将致力于开发更精准聚焦人体结构的新算法,以进一步提升特征提取能力。

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. Research progress on pedestrian re-identification based on deep learning[J]. Acta Automatica Sinica, 2019, 45(11): 2032-2049.

[2] GRAD D, TAO H. Viewpoint invariant pedestrian recognition with an ensemble of localized features[C]. Computer Vision-ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part I 10. Springer Berlin Heidelberg, 2008: 262-275.

[3] ZHAO H Y, TIAN M Q, SUN SH Y, et al. Spindle net: Person re-identification with human body region guided feature decomposition and fusion[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1077-1085.

[4] SUN Y, ZHHENG L, YANG Y, et al. Beyond part models: Person retrieval with refined part pooling(and a strong convolutional baseline)[C]. European Conference on Computer Vision (ECCV), 2018: 480-496.

[5] SU C, LI J, ZHANG S, et al. Pose-driven deep convolutional model for person re-identification[C]. IEEE International Conference on Computer Vision, 2017: 3960-3969.

[6] WANG G AN, YANG SH, LIU H Y, et al. High-order information matters: Learning relation and topology for occluded person re-identification[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 6449-6458.

[7] MIAO J Q, WU Y, LIU P, et al. Pose-guided feature alignment for occluded person re-identification[C]. IEEE/CVF International Conference on Computer Vision, 2019: 542-551.

[8] DOSOVITSKIY A, BEEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale[J]. ArXiv preprint arXiv:2010.11929, 2020.

[9] HE S, LUO H, WANG P, et al. Transformer-based object re-identification[C]. IEEE/CVF International Conference on Computer Vision, 2021: 15013-15022.

[10] LU Y, JIANG M, LIU Z, et al. Dual-branch adaptive attention transformer for occluded person re-identification[J]. Image and Vision Computing, 2023, 131: 104633.

[11] ZHANG G, ZHHANG P, QI J, et al. Hierarchical aggregation transformers for person re-identification[C]. 29th ACM International Conference on Multimedia, 2021: 516-525.

[12] ZHAO K, YANGY, CAVALLAR A, et al. Omni-scale feature learning for person re-identification[C]. IEEE/CVF International Conference on Computer Vision, 2019: 3702-3712.

[13] ZHENG L, SHEN L, TIAN L, et al. Scalable person re-identification: A benchmark [C]. IEEE International Conference on Computer Vision, 2015: 1116-1124.

[14] LUO H, GU Y, LIAO X, et al. Bag of tricks and a strong baseline for deep person re-identification[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019.

[15] ZHENG F, DENG C, SUN X, et al. Pyramidal person re-identification via multi-loss dynamic training[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 8514-8522.

[16] CHENG R, WANG L K, WEI M R. Learning discriminative and generalizable features with multi-branch for person re-identification[J]. Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology, 2022, 42(6): 5987-6001.

[17] RISTAN 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, 2016: 17-35.

[18] MA Z, ZHAO Y, LI J. Pose-guided inter-and intra-part relational transformer for occluded person re-identification[C]. 29th ACM International Conference on Multimedia, 2021: 1487-1496.

[19] MAXWELL J C. A treatise on electricity and magnetism[J]. Nature, 1872, 7(182): 478.

[20] 姬晓飞,赵帅,宋京浩,等. 基于姿势估计和特征融合的行人重识别算法[J]. 电子测量与仪器学报,2024,38(4):187-194.
JI X F, ZHAO SH, SONG J H, et al. Person reidentification algorithm based on pose estimation and feature fusion[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(4): 187-194.

作者简介

姬晓飞(通信作者),博士,副教授,主要研究方向为视频分析与处理、模式识别理论等。
E-mail:jixiaofei7804@126.com

孙英超,硕士研究生,主要研究方向为图像处理和模式识别。
E-mail:1879261353@qq.com

宋京浩,硕士研究生,主要研究方向为视频分析。
E-mail:1879261353@qq.com