

引用格式: 莫太平, 覃汉岳, 孙鹏, 等. 融合注意力机制的多粒度行人再识别方法 [J]. 电子科技大学学报, 2025, 54(3): 393-400.
MO T P, QIN H Y, SUN P, et al. Multi-granularity person re-identification with attention mechanism fusion[J]. Journal of University of Electronic Science and Technology of China, 2025, 54(3): 393-400.

融合注意力机制的多粒度行人再识别方法



莫太平, 覃汉岳*, 孙鹏, 张向文, 孟春城

(桂林电子科技大学 电子工程与自动化学院, 桂林 541000)

摘要: 针对监控环境复杂, 行人在光照变化、视角变化和遮挡等不同条件下图像外观差异较大, 导致行人细节特征难以被提取的问题, 提出了一种融合注意力机制的多粒度行人再识别模型。该模型通过引入多分支结构, 提取包含不同尺度信息的特征图; 结合多粒度切分模块和注意力机制, 进一步提取特征图的局部判别性信息, 获取多样化的特征表示并实现特征的协调统一; 采用联合学习的方式对模型进行监督训练, 得到更全面的特征描述。在主流的行人再识别数据集 Market-1501、DukeMTMC-reID 和 CUHK03 上取得了优异的性能, mAP 分别达到了 88.42%、78.86% 和 76.70%, 结果表明了该模型的有效性。

关键词: 行人再识别; 多尺度多分支; 多粒度特征; 注意力机制; 特征融合

中图分类号: TP391

文献标志码: A

DOI: 10.12178/1001-0548.2024091

Multi-granularity person re-identification with attention mechanism fusion

MO Taiping, QIN Hanyue*, SUN Peng, ZHANG Xiangwen, and MENG Chuncheng

(School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin 541000, China)

Abstract: A multi-granularity person re-identification model integrating attention mechanism is proposed to address the complex monitoring environment and the significant differences in image appearance of pedestrians under different conditions such as lighting changes, perspective changes, and occlusion, which make it difficult to extract detailed pedestrian features. The model extracts feature maps containing information at different scales by introducing a multi-branch structure. Combining multi-granularity segmentation modules and the attention mechanism, the local discriminative information of the feature map is further extracted to obtain diverse feature representations and achieve the coordination and unity of features. The model is supervised and trained by using federated learning to obtain a more comprehensive feature description. Excellent performance has been achieved on the mainstream person re-identification datasets Market-1501, DukeMTMC-reID, and CUHK03, with mAP reaching 88.42%, 78.86%, and 76.70%, respectively, demonstrating the effectiveness of the proposed model.

Key words: person re-identification; multi scale and multi branch; multi granularity features; attention mechanism; feature fusion

行人再识别是采用计算机视觉技术, 从监控摄像头拍摄的图像或视频中检索特定行人的一种技术。因为行人图像的视角、光照变化大, 行人姿态不统一, 行人局部被遮挡等原因, 在监控视频中难以识别, 使得行人再识别技术成为当前计算机视觉和机器学习技术领域的研究热点, 并在智能视觉监控和无人驾驶汽车等领域有重要的应用价值^[1-2]。近年来, 随着卷积神经网络 (convolutional neural networks, CNN) 的发展, 基于深度学习的行人再

识别方法取得了显著进展, 相比于传统的手工特征提取方法具有更强大的特征提取能力以及更好的自动学习能力^[3-4]。早期的深度学习行人再识别模型主要采用全局特征表示学习。其中, 文献 [5] 提出的身份判别嵌入网络将每个身份视为一个类别, 把模型训练过程构建为一个分类问题, 为行人再识别任务展开了一个新的训练范式。在此基础上, 文献 [6] 提出了属性行人识别网络, 以身份标签和行人属性进行分类。全局特征模型在学习推理阶段计算快

收稿日期: 2024-04-15

基金项目: 国家自然科学基金 (62263006); 广西自动检测技术与仪器重点实验室基金 (YQ21107)

作者简介: 莫太平, 博士, 教授, 主要从事深度学习、图像处理等方面的研究。

*通信作者 E-mail: qhy@mails.guet.edu.cn

速,但是往往会忽略行人的细节特征。相反,采用局部特征表示学习的方法可以关注行人骨架、姿势、人体部件等细节,从而在复杂场景中更具鲁棒性^[7]。目前,在行人再识别领域中,主流的局部特征表示学习方法包括姿态估计、图像切块和基于注意力机制的方法。

基于姿态估计的方法通过获取人体关节信息,预测人体姿态,旨在解决行人姿态不对齐的问题。文献[8]采用人体姿态估计算法来定位人体关节,并将关节分成了6个身体部位,随后对除了头部以外的部位采用仿射变换的方式进行对齐,并将对齐后的部位进行特征融合,该姿态估计方法会增加额外的计算开销。为此,文献[9]提出了离线学习的方法,仅在训练阶段引入额外的姿态提取监督学习,而在测试阶段不需要进行特征提取。基于姿态估计的方法能够较好地应对行人姿态不对齐的问题,但训练成本较高。相比之下,基于图像切块的方法通过对行人图像进行刚性切割,获得局部特征。文献[10]提出一种基于局部特征的卷积基线网络(part based convolutional baseline, PCB),将行人图像在水平方向上均分为6个部分,分别对6个部分进行分类和计算损失,分割方法简单灵活。文献[11]提出的多粒度网络(multiple granularity network, MGN)从多个层次的角度出发,采用不同的分块策略,以获取多粒度的全局特征和局部特征,进而获得相比PCB更丰富的特征表示。但是,图像切块的方式可能会导致各部分之间关联信息丢失。基于注意力机制的方法通过聚焦特征图中的相关区域,捕获复杂细节和上下文线索,得到判别特征。文献[12]提出交叉相关性注意力模块(cross-correlation Attention, CCA),通过两个特征图的注意力交叉映射学习关联性特征;文献[13]首次将Transformer模型应用于行人再识别领域,采用多头注意力机制解决行人不同部位特征信息不一致的问题。然而,这些注意力方法只能提取全局特征或固定尺度的局部特征,容易忽略小

尺度特征图的特征信息。

针对目前常见的行人再识别模型中相邻的局部特征信息容易因刚性切分导致特征丢失,以及小尺度特征容易被忽略的问题,本文提出了一种融合注意力机制的多粒度行人再识别模型。首先,采用了多分支结构来获取不同尺度的特征图,学习多尺度的行人特征。然后,设计了多粒度切分模块,关注相邻特征的特征信息,并与注意力机制相结合,以获取更丰富的局部细节信息。最后,将全局特征和局部特征进行融合,得到具有丰富行人信息的特征表示,通过结合图像切块和注意力机制的方法,有效地提高了行人再识别算法的性能。

1 行人再识别方法

1.1 网络框架

当前的行人再识别方法多使用ResNet-50作为骨干网络,其有效性已经在多个公开的大型数据集上进行了验证。ResNet-50引入了残差连接,能够有效缓解深层网络中的梯度消失和梯度爆炸问题,相比于传统的CNN网络(VGG-Net、GoogleNet等)更加稳定,参数量更少;相比于同样引入残差连接但是具有更深结构的ResNet-101, ResNet-50在模型复杂度和计算成本上取得更好的平衡。

本文以ResNet-50为骨干网络构建了一个多分支结构,网络框架如图1所示,包括多尺度特征模块、多粒度切分模块、注意力机制模块和损失计算。具体实现步骤如下:1)保持ResNet-50前3层的框架不变,在修改ResNet-50后两层的下采样环节时得到3个具有不同尺度信息的特征图;2)将3个分支的特征图输入多粒度切分网络,得到表示不同粒度信息的特征图;3)对这些特征图施加注意力机制模块,以探索具有判别性的信息;4)通过平均池化与最大池化的操作将这些特征图转化为特征向量,利用多任务学习多个损失共同监督网络优化。

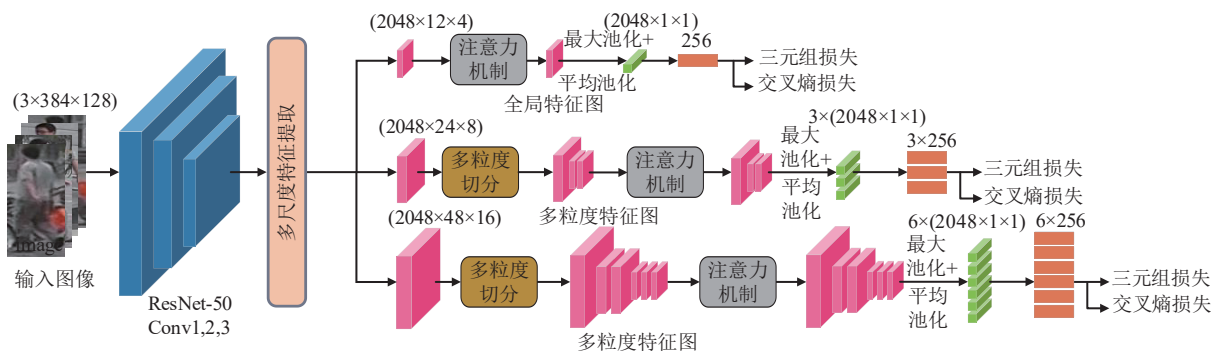


图1 融合注意力机制的多粒度行人再识别网络框架图

在测试阶段, 为了提升行人特征的表达能力, 将所有经过池化后的特征向量进行拼接融合, 得到最终的行人表示特征, 用于行人相似度度量计算。

1.2 多尺度特征提取模块

当前, 许多方法直接将整个人物图像输入骨干网络进行特征提取, 并使用骨干网络最高层输出的特征进行后续操作。尽管高层特征具有更强的语义信息, 但是经过多次卷积和池化操作后的高层特征会导致细节信息的丢失。为了在保留高层信息特征的同时关注低层的细节特征, 本文设计了多尺度特征提取模块, 如图2所示。该模块将骨干网络第3层之后的输出分为3个分支, 每个分支获取不同尺度的特征图。第1个分支沿用了 ResNet-50 的网络结构, 直接学习全局特征表示, 得到小尺寸特征图; 第2个分支删除了 ResNet-50 的第5层下采样, 保留了第4层的下采样环节, 得到中等尺寸的特征图; 与一些只在最后一层删除下采样环节的多尺度特征提取方法不同, 第3个分支删除了第4层和第5层的下采样环节, 得到大尺寸特征图, 相比于中等尺寸特征图保留了更多的空间信息。通过3个分支得到3个不同尺度的特征图, 确保模型可以表达整体行人信息的同时提升对行人局部特征的表征能力, 同时为后续的多粒度切分操作提供多尺度信息。

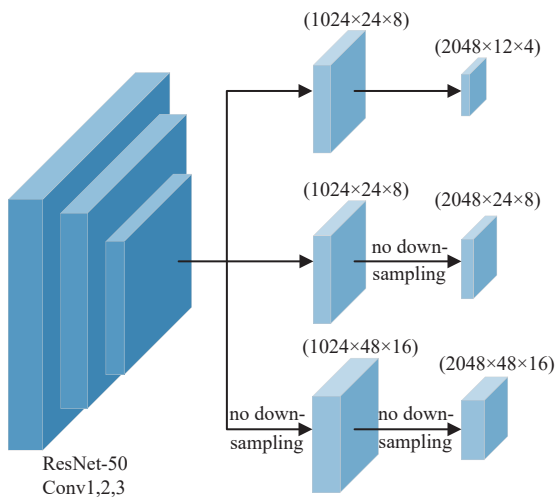


图2 多尺度特征提取框架图

1.3 多粒度切分模块

得到特征图之后, 如果直接使用全局特征进行学习, 会导致许多关键的局部信息被忽略, 而采用如 PCB、MGN 等刚性切块的方式提取局部特征可能会导致原本同属于一张特征图的特征被切分, 同时将不属于行人特征的背景噪声划分到行人特征当

中。为了关注更多关键的局部信息, 同时解决相邻特征信息容易被忽略的问题, 本文提出了多粒度切分策略, 对不同分支的特征图进行不同的切分方式, 如图3所示。具体操作如下。1) 给定当前输入特征图 $X_{h1} \in \mathbf{R}^{C \times H \times W}$, 其中 C 、 H 、 W 分别表示特征图的通道数、长度和宽度。2) 设切分数目为 n , 在空间层面对特征图进行 n 等分切块操作, 则最细粒度的特征图大小为 $X_{h2} \in \mathbf{R}^{C \times (H/n) \times W}$ 。3) 为了获得多粒度的特征表示, 将切分数目重新设置为 $n-1$, 并重新切分特征图, 大小为 $X_{h3} \in \mathbf{R}^{C \times (H/(n-1)) \times W}$ 。4) 依此类推, 直到切分数目 $n=1$, 此时不再进行切分操作, 这意味着已经得到了包含从粗粒度到细粒度特征的特征图。通过这种切分方式在一定程度上避免了水平分割对行人局部特征的影响, 并确保所获得的行人特征中包含的信息更加全面。在后续的特征融合操作中, 将相同粒度的相邻特征向量进行拼接, 有效解决硬性切分带来的特征丢失问题。

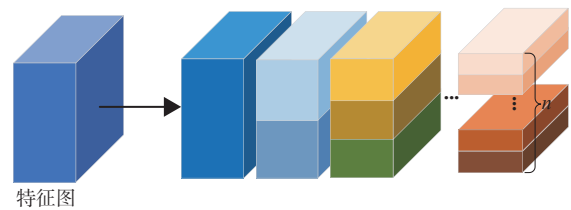


图3 多粒度切分模块框架图

在模型训练过程中, 对于直接使用 ResNet-50 结构的小尺寸特征图分支, 采用 $n=1$, 不进行切块操作, 保留骨干网络的全局特征信息。对于中等尺寸特征图和大尺寸特征图的分支, 分别采用 $n=2$ 和 $n=3$ 的多粒度切分方式, 既能提取到更细粒度的特征, 又能避免因切分数目过多引入大量背景噪声。

1.4 注意力机制模块

注意力机制能够让模型更好地关注特征图中的重要信息, 提高模型的识别准确性。当前许多融合注意力机制的行人再识别模型只采用了空间注意力机制, 忽略了不同特征图通道间的语义信息, 导致识别精度不高。本文的注意力机制模块由空间注意力机制和通道注意力机制并联构成。空间注意力用于捕捉特征在空间维度上的重要信息并进行处理, 而通道注意力则关注具有更强判别性特征的通道, 并增强这些通道的特征表达。采用并联结构能使模型同时关注通道和空间层面, 同时避免注意力机制的串联导致计算成本增加。

空间注意力机制如图4所示, 首先将输入特征

$X_h \in \mathbf{R}^{C \times H \times W}$ 经过线性变换得到 $Q \in \mathbf{R}^{(H \times W) \times 1}$, $K \in \mathbf{R}^{(H \times W) \times 1}$, $V \in \mathbf{R}^{(H \times W) \times C}$ 。进一步地, 对特征 Q 进行尺寸变换得到 $Q_r \in \mathbf{R}^{1 \times (H \times W)}$, 将 Q_r 与 K 做矩阵相乘操作, 并通过一个 Softmax 层进行归一化, 得到代表空间信息的注意力得分 $F_s \in \mathbf{R}^{(H \times W) \times (H \times W)}$, 表达式为:

$$F_s = \text{Softmax}(Q_r \times K) \quad (1)$$

收集到注意力得分后, 用来和特征 V 进行加权求和并得到 $D \in \mathbf{R}^{(H \times W) \times C}$, 以提升判别性特征的权重, 学习空间注意力信息。最终, 为了使这些特征能够嵌入到后续网络中, 对 D 进行线性变换, 调整回输入时的形状, 从而得到具有空间注意力的特征 $O_s \in \mathbf{R}^{C \times H \times W}$ 。

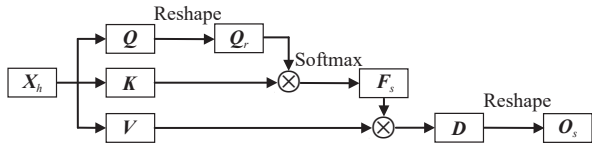


图 4 空间注意力机制框架图

通道注意力机制如图 5 所示, 首先, 将输入特征 $X_h \in \mathbf{R}^{C \times H \times W}$ 分别进行平均池化和最大池化操作, 将特征的高度和宽度压缩为 1, 使每个通道内的特征图展平为一维向量, 即经过平均池化和最大池化的特征分别为 $X_{\text{avg}} \in \mathbf{R}^{C \times 1 \times 1}$ 和 $X_{\text{max}} \in \mathbf{R}^{C \times 1 \times 1}$ 。然后将得到的特征分别经过全连接层、ReLU 激活层、全连接层、Sigmoid 激活层, 得到带有通道注意力机制的权重参数 $F_c \in \mathbf{R}^{C \times 1 \times 1}$, 表达式为:

$$F_c = S[L_2(\text{Re}(L_1(X_{\text{avg}}))) + L_2(\text{Re}(L_1(X_{\text{max}})))] \quad (2)$$

式中, S 为 Sigmoid 激活函数; L_1 、 L_2 分别为不同大小的全连接层; Re 为 ReLU 激活函数。

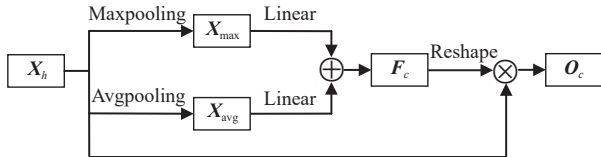


图 5 通道注意力机制框架图

最后在高度和宽度两个维度上扩展权重参数 F_c , 与输入特征 X_h 进行逐元素相乘, 得到具有通道注意力的特征 $O_c \in \mathbf{R}^{C \times H \times W}$ 。

1.5 损失计算

如图 1 所示, 使用了平均池化和最大池化操作来处理 3 个分支输出的特征图。平均池化能够对整体特征进行平滑处理, 保留完整的特征信息, 而最

大池化则强调局部最显著的特征。通过综合使用这两种池化操作, 能够融合全局和局部的信息, 提供更丰富的特征表示能力。之后采用 Add 操作对两种池化后的特征进行特征融合, 经过 1×1 卷积层、归一化和 ReLU 激活函数来降低特征图的维度, 以得到用于训练的特征向量。为了提高网络的学习能力并避免权重失衡等问题, 本研究采用了分类损失和对比损失联合学习的方法来监督网络的优化。将对比损失应用于网络直接输出的特征向量, 以进行行人特征的度量学习。同时, 将分类损失应用于所有经过全连接层的特征向量, 以进行行人特征的特征学习。

使用分类损失中的交叉熵损失进行监督学习, 交叉熵损失函数的表达式为:

$$L_{\text{cls}} = -\frac{1}{M} \sum_{i=1}^M \sum_{k=1}^N \log(q(k_i))p'(k_i) \quad (3)$$

式中, M 为特征向量个数, 文中设置为 10; N 为训练样本中的行人总数; k_i 为第 i 个行人特征向量; $q(k_i)$ 为行人身份的预测概率; $p'(k_i)$ 为改进的行人图像标签分布, 采用了标签平滑的方式进行处理, 防止过拟合现象。 $p'(k_i)$ 的表达式为:

$$p'(k_i) = \begin{cases} (1-\varepsilon)p(k_i) + \varepsilon/N & i = y \\ \varepsilon/N & i \neq y \end{cases} \quad (4)$$

式中, $p(k_i)$ 为原始的行人图像标签分布; ε 是超参数, 文中设置为 0.1; y 为行人图像的真实标签。

使用对比损失中的难样本三元组损失进行监督学习。难样本三元组损失是三元组损失的一种改进, 能够选择批次中更具挑战性的正负样本进行训练, 提高泛化能力^[14]。难样本三元组损失函数的表达式为:

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

式中, α 是人为设置的一个参数; M 为特征向量个数, 文中设置为 10; P 为行人图像的标签数量; K 为每个标签对应的行人图像数量; F^a 、 F^p 、 F^n 分别是原样本、正样本和负样本的特征向量表示, 如 $F_{i,l}^a$ 是第 i 个特征向量的第 l 幅图像对应的特征。

通过交叉熵损失与难样本三元组损失的共同优

化, 能够提高模型的训练效果。总的损失函数表达式为:

$$L = L_{cls} + \mu L_{tri} \quad (6)$$

式中, μ 为难样本三元组损失的权重参数。

2 实验结果与分析

2.1 数据集与评价指标

本研究在行人再识别领域的 3 个主流数据集 Market-1501^[15]、DukeMTMC-ReID^[16] 和 CUHK03^[17] 上进行实验, 并将本文方法与其他行人再识别方法进行了对比。

Market-1501 数据集包含了由 6 个摄像头拍摄的 1 501 个行人类别。该数据集分为训练集、测试集和查询集 3 个部分, 测试集和训练集不包含重复的行人类别。训练集包含了 751 个不同的行人, 共有 12 936 幅图像; 测试集包含了 750 个不同的行人, 共有 19 732 幅图像; 查询集中的行人图像来自测试集, 每个类别的行人至多有 6 幅图像, 共有 3 368 幅图像。

DukeMTMC-ReID 数据集是 DukeMTMC 数据集的一个子集, 由 8 个摄像头拍摄, 共有 36 411 幅图像和 1 402 个行人类别。其中, 训练集包含 16 522 幅图像, 涵盖了 702 个不同行人; 测试集包含 17 661 幅图像, 同样涵盖 702 个行人; 查询集有 2 228 幅图像, 与测试集中的行人类别相同。

CUHK03 数据集包括了由 10 个摄像头拍摄的 1 467 个行人的 14 096 幅图像。该数据集有两种划分方式, 根据手动标记的行人边界框和 DPM 检测的行人边界框进行划分。训练集包含 7 368 幅图像, 涵盖了 767 个不同行人; 测试集包含 5 328 幅图像, 涵盖了 700 个不同行人; 查询集有 1 400 幅图像, 与测试集中的行人类别相同。

为了验证本文提出的行人再识别模型的有效性, 使用了业内常用的累计匹配特性曲线 (cumulative match characteristic, CMC) 和查询平均精度 (mean average precision, mAP) 作为评价标准。CMC 能够反映模型的检索精度, 常以 Rank- n 的形式出现, 表示按照相似性降序排列, 识别结果中前 n 个结果包含目标的准确率。CMC 的表达式为:

$$CMC(n) = \begin{cases} 1 & k_i \leq n \\ 0 & k_i > n \end{cases} \quad (7)$$

式中, k_i 为第 i 张查询图像的第 k 个匹配结果。

mAP 综合考虑了模型的准确率和召回率, 其

表达式为:

$$mAP = \int_0^1 P(R) dR \quad (8)$$

式中, P 是预测的准确率, 表示和查询图像同一个 ID 的图片在查询结果中的比例; R 表示召回率, 表示和查询图像同一个 ID 的图片出现在查询结果中的数量占总数的比例。

本实验采用 Rank-1 作为评测指标, 并采用单查询模式进行实验与对比。

2.2 参数设置

本实验在一台搭载了深度学习框架的服务器上进行。服务器配置了 24 GB 的 4090 显卡和型号为 AMD EPYC 9654 的 CPU。服务器的内存容量为 16 GB, 操作系统为 Ubuntu20.04。所有实验均使用 Pytorch 深度学习框架。

在数据预处理阶段, 统一将行人图像尺寸裁剪为 384×128 , 并采用了随机翻转、随机擦除和标准化等图像增强操作。在训练阶段, 使用了在 ImageNet 上预训练的 ResNet-50 模型作为初始框架。在损失函数方面, 难样本三元组损失函数的参数 α 表示正负样本之间与原样本的间隔, 在训练中希望原样本与正样本的间隔尽可能小, 与负样本的间隔尽可能大, 因此根据实际训练情况将 α 设置为 1.2。为了平衡交叉熵损失与三元组损失对模型的优化程度, 根据实际训练情况将 μ 设置为 1。在参数更新方面, 采用了 Adam 优化器, 并将权重衰减因子设置为 0.000 5, β_1 设置为 0.9, β_2 设置为 0.999。此外, 将训练批次大小设置为 32, 根据模型复杂度和实际训练模型收敛速度, 设定初始学习率为 0.000 2, 并在第 320 和第 370 轮时将学习率衰减为当前学习率的 0.1 倍, 总共进行 400 轮训练。

2.3 消融实验

为了验证本文方法的有效性, 在 Market-1501 和 DukeMTMC-reID 上进行了消融实验, 单独使用了只包含全局特征的小尺寸特征图分支, 去除了注意力机制的网络作为 Baseline。所有对比实验都使用了相同的参数设置。

为验证多分支结构的有效性, 进行了 4 组对比实验。第 1 组实验为 Baseline, 仅采用了小尺寸特征图分支提取特征; 第 2 组实验采用了小尺寸特征图分支和大尺寸特征图分支共同提取特征; 第 3 组实验选用中尺寸特征图分支和大尺寸特征分支共同提取特征; 作为对照, 第 4 组实验同时使用了 3 个

分支提取特征。消融实验中,用 L_2 表示中尺寸特征图分支,用 L_3 表示大尺寸特征图分支。表 1 展示了实验结果,实验在 Market-1501 数据集和 DukeMTMC-reID 数据集上呈现相同规律。与 Baseline 相比,采用多分支结构的实验准确率显著提高,其中同时使用 3 个分支的效果最好,因为多分支结构能够通过不同的多粒度切分模块,提取不同粒度的行人特征,获得更丰富的特征信息。从第 2 组和第 4 组实验与第 3 组和第 4 组实验的比较中可以发现,将局部特征更细粒度地划分可以取得更好的效果。

表 1 多分支结构的影响 %

方法	Market-1501		DukeMTMC-reID	
	mAP	Rank-1	mAP	Rank-1
Baseline	76.37	90.08	67.58	83.30
Baseline+ L_2	85.60	94.18	69.70	83.80
L_2+L_3	86.24	94.27	71.09	84.07
Baseline+ L_2+L_3	87.28	94.71	78.49	88.29

为验证多分支结构的有效性,进行了 4 组对比实验, L_{att} 表示采用了注意力机制的模型。从表 2 中可以看出,将注意力机制模块嵌入网络中,模型的表现有所提升。在 Market-1501 数据集上,本文模型与去除注意力机制的模型相比,Rank-1 和 mAP 分别提高了 1.11% 和 1.14%。在 DukeMTMC-reID 数据集上,也分别提高了 0.54% 和 0.37%,证明了将注意力机制嵌入模型中能够进一步提高性能。

表 2 注意力机制的影响 %

方法	Market-1501		DukeMTMC-reID	
	mAP	Rank-1	mAP	Rank-1
Baseline	76.37	90.08	67.58	83.30
Baseline+ L_{att}	78.60	91.12	68.70	83.42
本文模型(无 L_{att})	87.28	94.71	78.49	88.29
本文模型	88.42	95.82	78.86	88.83

为验证各模块对模型计算速度的影响,本文在 Market-1501 数据集上进行了实验。在相同硬件条件下,测试各方法训练运行一个批次(32 张图像)的时间。从表 3 中可以看出,本文模型添加的注意力机制模块和多分支结构都会对计算速度造成影响,其中多分支结构对计算速度的影响较大,因为不同分支不同粒度特征图的特征提取会增加模型计算量。虽然本文模型的计算速度相比于

Baseline 有一定差距,但是从表 3 中的 mAP 值中体现出,添加的模块和网络结构可以提供更丰富的特征表示,整体上能够提升模型性能,对行人再识别任务中识别精度的提升具有更大价值。

表 3 在 Market-1501 数据集上的计算速度对比

方法	计算速度/batch s ⁻¹	mAP/%
Baseline	0.131	76.37
Baseline+ L_{att}	0.139	78.60
Baseline+ L_2	0.178	85.60
L_2+L_3	0.292	86.24
本文模型(无 L_{att})	0.311	87.28
本文模型	0.406	88.42

2.4 本算法与主流算法对比

为验证本文算法的先进性,将其与目前多种主流的行人再识别算法在数据集 Market-1501、DukeMTMC-ReID 和 CUHK03 上进行了比较,具体结果见表 4。这些算法包括基于行人姿态估计的方法,如 PGFA^[18](姿态引导特征对齐)、PGR^[9](姿态引导表示)等;基于局部特征切块的方法,如 PCB^[10](基于局部特征切块的卷积基线)、MGN^[11](多颗粒度网络)、DSA^[19](密集语义对齐)等;以及加入注意力机制的方法,如 AMFE^[20](融合注意力与多层次特征提取的行人再识别方法)、OAMN^[21](遮挡感知注意力网络)和 AaP-ReID^[22](行人注意力感知网络)等。

从实验结果上看,在 Market-1501 数据集上,本文模型的 Rank-1 和 mAP 分别为 95.82% 和 88.42%,表现出了优越的性能。与 MGN 相比,本文的多粒度切块方式兼顾了不同尺度特征图的信息,同时充分融合了多粒度特征与注意力机制的优点,取得了更加优越的表现。对于采用了两种方式(Label 和 Detected)划分的 CUHK03 数据集,本文模型的 Rank-1 和 mAP 分别为 76.70%、78.84% 和 76.30%、78.39%。在 CUHK03-Detected 数据集上,相比于同样提取多尺度特征的 AMFE,本文的 Rank-1 和 mAP 分别提升 10.29% 和 10.80%。在具有严重遮挡现象的 DukeMTMC 数据集上 TransReID 取得了更好的性能,但是与其他融合注意力机制的模型相比(AMFE、OAMN),本文模型仍取得了较为明显的性能提升,验证了本文模型能够更好地解决遮挡带来的影响。结果表明,本文模型能够有效提高行人再识别的准确率,超过当前的大多数方法。

表4 本文方法与当前主流 Re-ID 方法的性能对比

%

方法	Market-1501		DukeMTMC-reID		CUHK03-Detected		CUHK03-Labeled	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
PGR ^[9]	77.21	93.87	65.98	83.63	—	—	—	—
PGFA ^[18]	76.80	91.80	65.50	82.60	—	—	—	—
PCB ^[10]	81.60	93.80	69.20	83.30	57.50	63.70	—	—
MGN ^[11]	86.90	95.70	78.40	88.70	66.00	66.80	67.40	68.00
DSA ^[19]	87.60	95.70	74.30	86.20	73.10	78.20	75.20	78.90
AMFE ^[20]	82.10	93.00	71.80	83.90	65.50	68.10	—	—
OAMN ^[21]	79.80	93.20	72.60	86.30	—	—	—	—
AaP-reID ^[22]	86.30	94.60	76.20	87.60	72.40	74.90	—	—
TransReID ^[13]	88.10	94.90	81.30	90.20	—	—	—	—
本文模型	88.42	95.82	78.86	88.83	76.30	78.39	76.70	78.94

2.5 可视化分析

为了更加直观地展现本文模型的性能, 将本文模型与 MGN 模型进行可视化结果对比, 从 Market-1501 数据集中输入两幅查询图像进行比较, 检索结果如图 6 所示。检索得到的图片根据相似度降序进行排序, 图 6a 与图 6b 为一组对照实验, 图 6c 与图 6d 为一组对照实验。



图6 本文与 MGN 模型可视化结果对比

从第 1 组对照实验中可以看出, 图 6a 中的第 10 张图像不是查询图像本人, 而图 6b 的检索全部正确; 在第 2 组对照实验中, 选用了下半身有遮挡、行人姿态变化大的查询图像。可以看出, 图 6c 中的第 5、8、9、10 幅图像都不是待查询图像本人, 而图 6d 中只有第 8 幅图像识别错误。由此可知, 本文算法在识别明显无遮挡、姿态差异小的行人时, 能够实现更加准确地检索; 当行人间存在遮

挡、姿态变化时, 本文模型能充分发挥多粒度与注意力机制的优势, 从图像中获取更多有效信息, 达到更高的识别准确率。

3 结束语

本文提出了一种融合注意力机制的多分支多粒度行人再识别方法, 该方法通过对特征图进行多粒度切分, 获取具有丰富性和鲁棒性的行人特征信息。嵌入注意力机制模块自适应地关注重要特征, 并进一步提高分类准确性。通过联合损失优化模型, 进一步提高模型的泛化能力, 学习到更具有区分性的特征表示。与其他方法相比, 本文方法实现了更合理的分割方式, 既保留了全局特征信息又充分利用了局部特征, 同时减少了背景噪声的干扰。在公共数据集上, 进行了消融实验, 与现有方法的比对实验和可视化实验, 验证了本文方法的鲁棒性和优越性。在未来的工作中, 将进一步优化特征图的分割方式, 并减少注意力机制的冗余计算, 以进一步提高模型的性能。

参考文献

- [1] LI X Z, ZHOU W G, ZHOU Y, et al. Relation-guided spatial attention and temporal refinement for video-based person re-identification[C]//Proceedings of the AAAI Conference on Artificial Intelligence. New York: AAAI, 2020: 11434-11441.
- [2] LAI S Q, CHAI Z H, WEI X. Transformer meets part model: Adaptive part division for person re-identification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 4150-4157.
- [3] DAI Z Z, WANG G, YUAN W, et al. Cluster contrast for unsupervised person re-identification[C]//Proceedings of the Asian Conference on Computer Vision. Macao, China: ACCV, 2022: 1142-1160.
- [4] 王松, 纪鹏, 张云洲, 等. 自适应感受野网络的行人重识

- 别[J]. 控制与决策, 2022, 37(1): 119-126.
- WANG S, JI P, ZHANG Y Z, et al. Adaptive receptive network for person re-identification[J]. Chinese Control and Decision, 2022, 37(1): 119-126.
- [5] ZHENG Z D, ZHENG L, YANG Y. A discriminatively learned CNN embedding for person re-identification[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2017, 14(1): 1-20.
- [6] LIN Y T, ZHENG L, ZHENG Z D, et al. Improving person re-identification by attribute and identity learning[J]. Pattern Recognition, 2019, 95: 151-161.
- [7] 李聪, 蒋敏, 孔军. 基于多尺度注意力机制的多分支行人重识别算法[J]. 激光与光电子学进展, 2020, 57(20): 37-45.
- LI C, JIANG M, KONG J. Multi-branch person re-identification based on multi-scale attention[J]. Chinese Laser & Optoelectronics Progress, 2020, 57(20): 37-45.
- [8] 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: 3960-3969.
- [9] LI J N, ZHANG S L, TIAN Q, et al. Pose-guided representation learning for person re-identification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 44(2): 622-635.
- [10] SUN Y, ZHENG L, YANG Y, et al. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)[C]// Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 480-496.
- [11] 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. Sanya: ACM, 2018: 274-282.
- [12] ZHOU J M, ROY S K, FANG P F, et al. Cross-correlated attention networks for person re-identification[J]. Image and Vision Computing, 2020, 100: 103931.
- [13] HE S T, LUO H, WANG P C, et al. Transreid: Transformer-based object re-identification[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 15013-15022.
- [14] SOMERS V, DE VLEESCHOUWER C, ALAHI A. Body part-based representation learning for occluded person re-identification[C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Paris: IEEE, 2023: 1613-1623.
- [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. Santiago: IEEE, 2015: 1116-1124.
- [16] ZHENG Z D, ZHENG L, YANG Y. Unlabeled samples generated by gan improve the person re-identification baseline in vitro[C]// Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 3754-3762.
- [17] 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.
- [18] MIAO J X, WU Y, LIU P, et al. Pose-guided feature alignment for occluded person re-identification[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 542-551.
- [19] ZHANG Z Z, LAN C L, ZENG W J, et al. Densely semantically aligned person re-identification[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 667-676.
- [20] 张荣, 王进, 张天奇, 等. 融合注意力与多层次特征提取的行人再识别方法[J]. 传感器与微系统, 2023, 42(12): 68-71.
- ZHANG R, WANG J, ZHANG T Q, et al. Pedestrian re-identification method fusing attention and multi-level feature extraction[J]. Chinese Transducer and Microsystem Technologies, 2023, 42(12): 68-71.
- [21] CHEN P X, LIU W F, DAI P Y, et al. Occlude them all: Occlusion-aware attention network for occluded person re-id[C]// Proceedings of the IEEE/CVF international conference on computer vision. Montreal: IEEE, 2021: 11833-11842.
- [22] GAUTAM V, PRASAD S, SINHA S. AaP-ReID: improved attention-aware person re-identification[EB/OL]. [2023-09-27]. <https://arxiv.org/pdf/2309.15780.pdf>.

编辑 刘飞阳