

DOI:10.13364/j.issn.1672-6510.20230184

数字出版日期: 2024-06-21; 数字出版网址: <http://link.cnki.net/urlid/12.1355.N.20240621.1020.003>

基于区域特征补全和细粒度特征注意力的行人重识别方法

孙志伟, 吴广群, 马永军
(天津科技大学人工智能学院, 天津 300457)

摘要: 针对现有行人重识别方法对行人图像中遮挡区域特征利用率较低和行人图像中提取的特征细粒度不足的问题, 提出了一种基于区域特征补全和细粒度特征注意力的行人重识别方法。首先, 提出细粒度特征注意力模块, 通过对输入特征进行分割加权重组, 并添加注意力机制, 以获得细粒度特征; 其次, 引入区域特征补全模块, 通过将输入特征分块聚类, 使遮挡区域特征能够通过相同聚类的特征恢复; 最后, 使用身份损失、加权正则化三元组损失和中心损失对模型进行训练优化。在公开数据集 Market-1501、DukeMTMC-reID 上进行实验, 结果表明模型识别效果有所提升。

关键词: 行人重识别; 区域特征补全; 细粒度; 注意力机制

中图分类号: TP391 文献标志码: A 文章编号: 1672-6510(2024)05-0063-09

Person Re-Identification Method Based on Region Feature Completion and Fine-Grained Feature Attention

SUN Zhiwei, WU Guangqun, MA Yongjun
(College of Artificial Intelligence, Tianjin University of Science & Technology, Tianjin 300457, China)

Abstract: To address the issues of low utilization of occluded region features and insufficient granularity of extracted features in existing person re-identification methods, a person re-identification method based on region feature completion and fine-grained feature attention is proposed in this article. Firstly, a fine-grained feature attention module is proposed to segment, weight, and recombine the input features, with the addition of attention mechanisms to obtain fine-grained features. Secondly, a region feature completion module is introduced, which clusters the input features into blocks to enable the recovery of occluded region features through features in the same cluster. Finally, the model is optimized with the use of identity loss, weighted regularized triplet loss, and center loss. Experimental results on publicly available datasets Market-1501 and DukeMTMC-reID demonstrated that the proposed model achieved improved recognition performance.

Key words: person re-identification; region feature completion; fine-grained; attention mechanism

引文格式:

孙志伟, 吴广群, 马永军. 基于区域特征补全和细粒度特征注意力的行人重识别方法[J]. 天津科技大学学报, 2024, 39(5): 63-71.

SUN Z W, WU G Q, MA Y J. Person re-identification method based on region feature completion and fine-grained feature attention[J]. Journal of Tianjin university of science & technology, 2024, 39(5): 63-71.

行人重识别(person re-identification), 是一种基于计算机视觉的跨摄像头的图像检索技术, 其目标是匹配不同摄像机捕捉到的同一人物的图像或视频。近年来, 行人重识别技术由于在视频监控和刑事调查

等领域中的重要性而受到广泛关注。然而, 由于现有摄像机捕捉到的图像或视频存在分辨率低、背景杂乱、姿态变化、照明变化、物体遮挡和异构模态等问题, 行人重识别面临巨大挑战。因此, 从行人图像中

收稿日期: 2023-10-08; 修回日期: 2024-02-14

基金项目: 国家自然科学基金资助项目(61976156); 天津市自然科学基金资助项目(18JQNJC69500)

作者简介: 孙志伟(1973—), 男, 河北保定人, 副教授, zhwsun@tust.edu.cn

学习提取有效且鲁棒性强的特征,并设计适当的特征度量算法成为解决行人重识别问题的关键^[1]。

早期的行人重识别研究主要集中在手工特征构造和距离度量学习,然而这些传统的低级视觉特征提取方法难以获得有效和不变的特征,从而影响了基于这些特征的特征度量学习方法的准确性和有效性^[2]。随着深度学习技术的发展,越来越多的学者开始探索基于深度学习的行人重识别算法。相较于早期研究,利用深度学习解决行人重识别问题可以同时进行特征表示学习和度量学习,从而实现自动提取具有较高判别性的特征,并将这些特征映射到更合适的度量空间^[3]。

表征学习也称为特征表示学习,它通过将图像数据转化为高维特征向量,量化图像之间的差异,以提取图像特征。一方面,许多方法^[4-6]专注于提取未被遮挡的人物身体信息,以提取到更加具有判别性的特征。Zhu 等^[4]提出一种 ISP (identity-guided human semantic parsing) 模型,通过人体语义解析的方法定位和像素级别分类人体部位,仅使用人的身份标签作为输入。Zhang 等^[5]提出了一个关系感知全局注意力 (relation-aware global attention, RGA) 模块,可以使网络提取更具判别性的特征信息。Chen 等^[6]提出了一种注意力金字塔网络,通过叠加不同尺度的注意力特征,提取更加丰富的特征。这些方法^[4-6]可以提取到细粒度的特征,但也丢失了遮挡部分的信息。另一方面,也有一些方法^[7-9]专注于缓解遮挡部分的信息缺失,以提取到更加丰富的特征。Hou 等^[7]提出了一种使用生成对抗网络生成被遮挡区域的方法,先对图像进行补全,再与重识别网络结合进行行人重识别。Shi 等^[8]提出了一种属性解缠与配准网络,通过属性特征解缠挖掘非遮挡区域,充分利用了行人属性的高级语义概念缓解遮挡问题。Xu 等^[9]设计了一个特征恢复转换器,减少特征匹配中的噪声干扰和遮挡带来的信息丢失。Wang 等^[10]提出了一种特征完成转换器 (feature completion transformer), 隐式补充特征空间中被遮挡部分的语义信息。这些方法^[7-10]虽然可以提取到丰富的特征,但未能提取到具有细粒度或具有高细粒度的特征。

度量学习是深度学习的常用方法,通过定量方法,确定特征向量的相似程度。在网络的训练过程中,度量学习方法主要体现在损失函数上,更新深度网络的参数,从而提取更具判别力的图像特征。目前度量学习损失函数有交叉熵损失 (cross entropy loss)^[11]、三元组损失 (triplet loss)^[12]、四元组损失

(quadruplet loss)^[13]、中心损失 (center loss)^[14]、加权正则化三元组损失 (weighted regularization triplet loss)^[15]和自适应稀疏成对损失 (adaptive sparse pairwise loss)^[16]等。合理组合不同的损失函数可以发挥每个损失函数的优点,使它们能更好地度量样本间的相似性。

通过对行人重识别表征学习和度量学习方法的研究,发现上述方法^[4-10]没有利用遮挡区域特征提取到足够的细粒度特征。针对上述问题,本文提出了一种基于区域特征补全和细粒度特征注意力的方法 (region feature completion and fine-grained feature attention network, RFCFFA-Net), 该方法在补全特征的同时提取了细粒度特征,以保证提取到丰富和细粒度的特征。本文方法的主要工作如下:提出了细粒度特征注意力 (fine-grained feature attention, FFA) 模块,通过对特征进行分割和加权重组,并添加通道注意力引导网络学习细粒度的特征;引入区域特征补全 (region feature completion, RFC) 模块,在特征级补全遮挡区域特征;采用广义均值池化 (generalized mean pooling, GMP) 代替全局平均池化,并采用加权正则化三元组损失 (weighted regularization triplet loss, WRT) 代替三元组损失,进一步优化网络对特征的学习;在公开数据集 Market-1501、DukeMTMC-reID 上进行了实验,结果证明了本文方法的有效性。

1 本文方法

1.1 网络结构

本算法的基线是 Luo 等^[17]提出的强基线 (strong baseline) 算法,该算法结合多种训练技巧,在单个主干网络的全局特征上取得了良好的性能。该算法使用 ResNet50^[18]作为骨干网络,去除第 4 个卷积块后的全局平均池化层和全连接层。为了获得更大的特征图,将第 4 个卷积块的步长修改为 1,采用多个损失函数联合对网络进行训练,并提出了批归一化颈部 (BN neck) 结构,用于加快损失的收敛。强基线算法的特征提取部分只有 ResNet50 的 4 个卷积块,这些卷积块只对特征进行了简单的提取和处理。在卷积块之间添加特征处理模块可以提高提取特征表示的丰富程度,以提高特征的判别性,进而提升模型的识别效果。故本文模型的网络结构在基线结构的基础上增加了区域特征补全模块 RFC^[19]和细粒度注意力模块 FFA。同时,采用广义均值池化 GMP^[20]代替了

原始骨干网络后的全局平均池化, 以及采用加权正则化三元组损失^[15]代替原训练过程中的三元组损失。模型网络结构如图 1 所示。

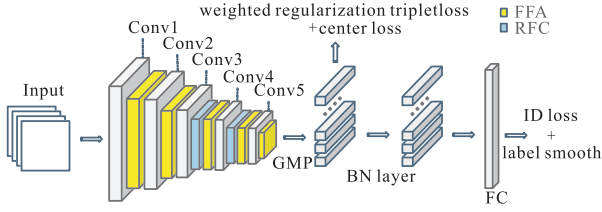


图 1 模型网络结构

Fig. 1 Network architecture of the model

将 ResNet50 的 5 层卷积层分别命名为 Conv1、Conv2、Conv3、Conv4、Conv5。在表征学习阶段, 输入 n 张图像, 经过图像预处理后输入骨干网络中, 依次经过 Conv1 层、FFA 模块、Conv2 层、FFA 模块、Conv3 层、RFC 模块、FFA 模块、Conv4 层、RFC 模块、FFA 模块、Conv5 层, 再通过广义均值池化得到图像特征。在度量学习阶段, 使用加权正则化三元组损失和中心损失对特征进行约束, 通过批归一化层 (BN layer) 将提取到的特征进行归一化, 再通过全连接层 (FC) 得到每张图像中行人身份的分类概率。对预测结果使用带有标签平滑 (label smooth) 的交叉熵损失 (也称 ID 损失) 进行处理。

1.2 区域特征补全模块

区域特征补全模块是 Hou 等^[19]提出的一种能够恢复特征空间中遮挡区域语义的模块, 该模块包括自适应分区单元 (adaptive partition unit, APU)、空间区域特征补全 (spatial RFC, SRFC) 模块和时态区域特征补全 (temporal RFC, TRFC) 模块。

APU 模块可以自适应地将输入的特征图划分成不同的区域, 对应于特定的身体部位, 其原理如图 2

所示。只需要找出 4 个关键点 $(a_1^i, 0)$ 、 $(a_2^i, 0)$ 、 $(a_3^i, 0)$ 、 $(0, a_4^i)$, 即可对人物图像进行分割。该模块将寻找 4 个关键点的过程定义为 4 个分类问题, 故设计 4 个分类器寻找关键点。以寻找 $(a_1^i, 0)$ 为例, 特征 F_i 先通过卷积减少通道维度, 再沿水平方向进行平均池化, 然后通过分类器 Linear1 得到预测概率分布 $(p_1^i)_h \in R^H$, 其中 $(p_1^i)_h$ 为 $a_1^i = h$ 的概率。 a_1^i 可以通过 p_1^i 的最大值的索引获得。

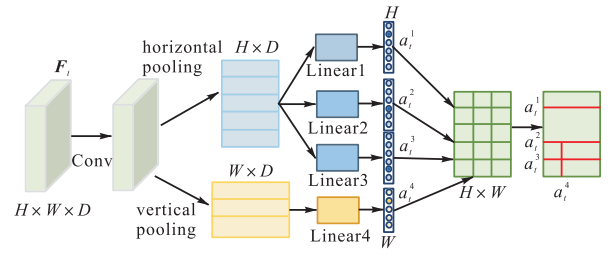


图 2 自适应分区模块

Fig. 2 Architecture of APU Module

RFC 模块中用于特征补全的模块为 SRFC 模块和 TRFC 模块, 两个模块采用互补线索, 分别侧重于空间和时间。由于本文实验数据集中的图像没有前后时间关系, 且 TRFC 模块输入的图像必须为连续的帧, 所以 TRFC 模块不适用于本文实验。因此, 只采用 RFC 模块中的 SRFC 模块进行特征补全。SRFC 模块恢复区域特征语义的原理如图 3 所示。RFC 模块中的自适应分区模块可以将输入的特征图自适应地划分为对应特定身体部位的不同区域; 在区域编码阶段, 通过分配矩阵 $S \in R^{N \times K}$ 将划分区域 $\{f^i\}_{i=1}^N$ 的特征聚集到几个聚类 $\{c^k\}_{k=1}^K$; S_{ik} 表示将区域 R^i 分配给 c^k 的概率。

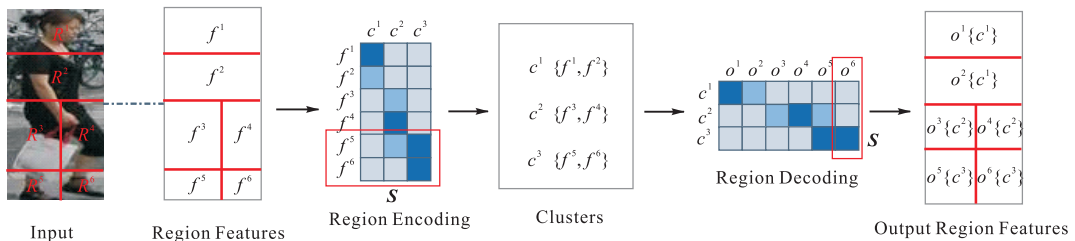


图 3 区域特征补全原理图

Fig. 3 Schematic diagram of RFC

在图 3 中, 较深的颜色表示较高的概率, SRFC 模块确保相似外观或接近位置的区域被分配到一个聚类中, 因此每个聚类收集了大部分相关区域的特征, 并且通常代表了身体的不同部位。例如, c^1 、 c^2

和 c^3 分别对应上衣、裤子和鞋子。在区域解码阶段, 通过 S 将 $\{c^i\}_{i=1}^K$ 分配给输出区域特征 $\{o^i\}_{i=1}^N$ 。因此, 遮挡区域可以使用相关的聚类恢复其特征。例如, R^5

可以使用 c^3 恢复其特征以表示裤子。

1.3 细粒度特征注意力模块

受到 SCNet^[21] 中拆分特征再重组方法的启发, 为了获得更细粒度的特征表示, 设计了一种名为细粒度特征注意力模块的混合注意力模块, 通过分割特征并加权重组, 再通过通道注意力得到更细粒度的特征, 其结构如图 4 所示。

FFA 模块包括 5 个步骤:

(1) 将经过卷积层后的输入特征 $X \in R^{C \times H \times W}$ 在通道维度分割为子张量 $\{X_i\}_{i=1}^N \in R^{\frac{C}{N} \times H \times W}$, 其中 C 为通道

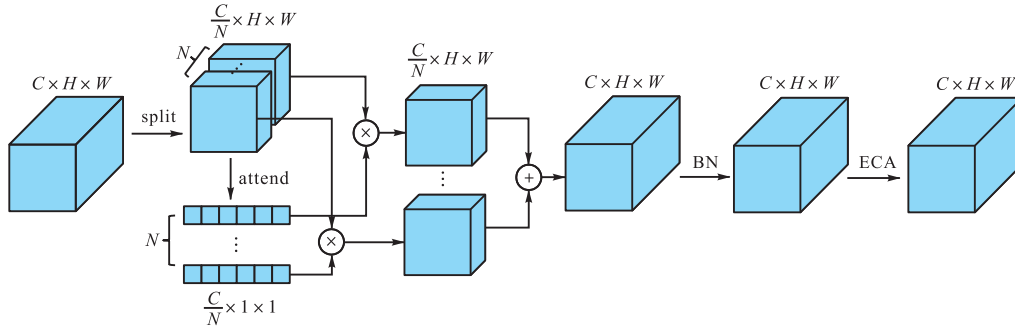


图 4 细粒度特征注意力模块结构

Fig. 4 Architecture of FFA Module

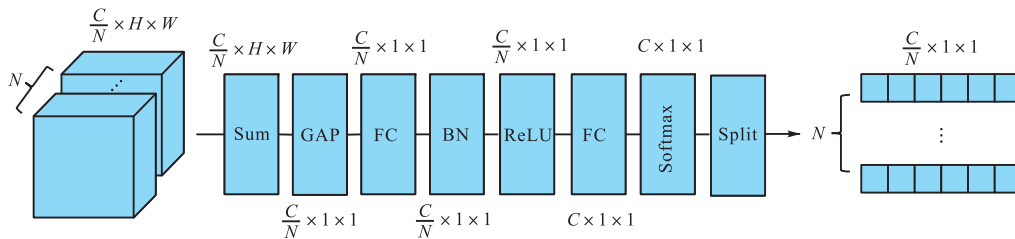


图 5 加权操作流程图

Fig. 5 Flowchart of weighted operation

(3) 使用注意力分布 $\{A_i\}_{i=1}^N$ 对分割后的子张量 $\{X_i\}_{i=1}^N$ 进行加权, 在通道维度进行拼接, 得到合并特征 $X = [A_i(X_i)]_{i=1}^N$ 。

(4) 经过 BN (batch normalization) 进行归一化操作。

(5) 通过 ECA (efficient channel attention) 模块^[18] 进一步提取通道信息后得到输出特征, 并将输出特征输入下一个卷积块。

ECA 模块是 Wang 等^[22] 提出的一种有效的通道关注模块, 该模块通过非线性自适应确定的一维卷积组成, 只使用了少量的参数, 获得了较好的性能。该模块的结构如图 6 所示, 输入特征为经过归一化后的合并特征 $X \in R^{C \times H \times W}$, 可得输出特征 X 的公式为

数, H 和 W 为特征图的高度和宽度, N 为分割数量。

(2) 对特征 $\{X_i\}_{i=1}^N$ 进行加权操作, 得到每个分割的子张量对应的注意力分布 $\{A_i\}_{i=1}^N \in R^{\frac{C}{N} \times H \times W}$ 。加权操作过程如图 5 所示, 可得 $\{A_i\}_{i=1}^N$ 的公式为

$$\{A_i\}_{i=1}^N = s \left(\sigma \left(\text{FC}_2 \text{ReLU} \text{BN} \left(\text{FC}_1 \text{Avgpool} \left(\sum_{i=1}^N X_i \right) \right) \right) \right) \quad (1)$$

其中 s 表示 Split 拆分操作, 沿通道维度将权重向量拆分为 N 份。

$$X = X \times \text{Sigmoid}(\text{Conv}(\text{Avgpool}(X))) \quad (2)$$

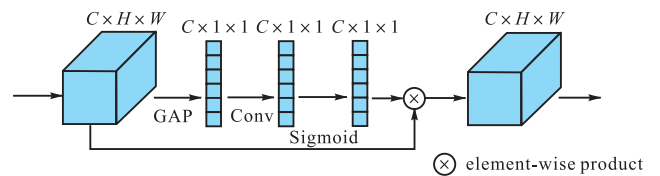


图 6 ECA 模块结构图

Fig. 6 Architecture of ECA module

1.4 广义均值池化模块

广义均值池化模块^[20]用于提取图像或特征图的全局特征。广义均值池化采用了一种参数化的非线性函数, 将特征图中的每个像素值进行变换, 然后计算所有像素的平均幂值, 并对其取幂值的倒数。这种池化方法可以控制幂的参数, 用于调整池化操作的非

线性程度。

广义均值池化相较于最大平均池化的优点在于它具有灵活性和可调节性。通过调整参数,可以在池化操作中引入更多的非线性变换,从而捕捉到更细粒度的特征信息。此外,广义均值池化还具有较好的数学性质,能够保持特征的空间结构,减轻特征的维度损失。因此,引入广义均值池化代替基线算法中的最大平均池化。公式为

$$f = [f_1, \dots, f_k, \dots, f_K]^T$$

$$f_k = \left(\frac{1}{|X_k|} \sum_{x_i \in X_k} x_i^{P_k} \right)^{\frac{1}{P_k}} \quad (3)$$

其中: f_k 表示特征图, K 是最后一层中特征图的数量; X_k 是特征映射 $k \in \{1, 2, 3, \dots, k\}$ 的 $W \times H$ 激活集; P_k 是池化参数,当 $P_k = 1$ 时,上述操作近似于平均池化,当 $P_k \rightarrow \infty$ 时,上述操作近似于最大池化。由于无法确定 P_k 的大小,故需要经过训练,即在反向传播过程中学习求取最佳 P_k 。

1.5 损失函数

强基线算法的损失函数为多损失联合,其中包括交叉熵损失 (L_{ID})、三元组损失 (L_{Tri}) 和中心损失 (L_C)。强基线算法模型的总损失函数公式为

$$L = L_{ID} + L_{Tri} + \beta L_C \quad (4)$$

其中 β 是中心损失的平衡权重,且 $\beta = 0.0005$ 。

在训练预测阶段使用标签平滑的交叉熵损失,计算公式为

$$L_{ID} = \sum_{i=1}^N -q_i \log(p_i) \quad (5)$$

$$q_i = \begin{cases} 1 - \frac{N-1}{N} \varepsilon & y = i \\ \frac{\varepsilon}{N} & y \neq i \end{cases} \quad (6)$$

其中: N 为数据集中的行人数量; p_i 为输出的行人身份的预测概率; y 为行人身份标签; ε 为超参数,设为 0.1。

强基线算法使用三元组损失和中心损失约束广义均值池化后的图像特征。三元组损失函数公式为

$$L_{Tri} = [d_p - d_n + \alpha]_+ \quad (7)$$

其中: d_p 为正样本间距离, d_n 为负样本间距离, α 为一个大于 0 的常数。

三元组损失只考虑 d_p 和 d_n 的差值,忽略了它们的绝对值。为了弥补三元组损失的不足,基线算法在训练中引入了中心损失,同时学习每个类的深度特征中心,并对深度特征与其对应类中心之间的距离进行

惩罚。中心损失函数为

$$L_C = \frac{1}{2} \sum_{j=1}^B f_{i_j} - c_{y_j}^2 \quad (8)$$

其中: y_j 为小批量中第 j 个图像的标签, c_{y_j} 为深度特征的第 i 类中心, B 为批量大小。中心损失最小化可以提高类内紧凑性。

采用加权正则化三元组损失 (L_{WRT}) 代替原三元组损失,加权正则化继承了正负对之间相对距离优化的优点,同时避免了引入任何额外的超参数。 L_{WRT} 的公式为

$$L_{WRT}(i) = \log \left(1 + \exp \left(\sum_j w_{ij}^p d_{ij}^p - \sum_k w_{ik}^n d_{ik}^n \right) \right) \quad (9)$$

$$w_{ij}^p = \frac{\exp(d_{ij}^p)}{\sum_{d_{ij}^p \in P_i} \exp(d_{ij}^p)} \quad (10)$$

$$w_{ik}^n = \frac{\exp(-d_{ik}^n)}{\sum_{d_{ik}^n \in N_i} \exp(-d_{ik}^n)} \quad (11)$$

其中: (i, j, k) 表示每个训练批中的硬三元组;对于锚点 i , P_i 为相应的正集合, N_i 为负集合; d_{ij}^p / d_{ik}^n 表示正/负样本对的成对距离。

本算法模型的总损失函数包括交叉熵损失、加权正则化三元组损失和中心损失,公式为

$$L = L_{ID} + L_{WRT} + \beta L_C \quad (12)$$

其中: β 采用强基线中的参数值, $\beta = 0.0005$ 。

2 实验结果与分析

2.1 数据集及评价指标

为了验证本文模型的有效性,分别在两个公开数据集 Market-1501 数据集和 DukeMTMC-reID 数据集上进行了实验。

Market-1501 数据集包含了清华大学校园中 6 个摄像头拍摄的行人图像,共有 1501 个被标注的行人和 32668 张图像。其中,训练集包含 751 个行人和 12936 张图像,测试集包含 750 个行人和 19732 张图像,查询集包含 750 个行人和 3368 张图像。

DukeMTMC-reID 数据集中的行人图像来自 Duke 大学校园的 8 个静态摄像头,共有 1812 个行人的 36411 张图像。训练集包含 702 个行人和 16522 张图像,测试集包含 703 个行人和 408 个干扰行人,共 17661 张图像,查询集包含测试集中的 702 个行人,并在每个摄像头中随机选择一张图像,共有 2228 张图像。

为了评估本文模型的效果,使用行人重识别评价

指标中的累计匹配特征曲线(cumulative matching characteristics, CMC)和平均准确均值(mean average precision, mAP, 用符号 P_{mAP} 表示)。CMC 曲线表示在前 N 张搜索结果中出现正确结果的概率, 其中 Rank-1 表示搜索结果中置信度最高的图像的正确率, Rank-5 表示在置信度排名前 5 的图像中的正确率。mAP 则表示排名靠前的结果的准确性, 反映了整体多张图像的检索能力, 其公式为

$$P_{mAP} = \frac{\sum_{i=1}^{N_q} P_{av,i}}{N_q} \quad (13)$$

其中 N_q 为查询图像的数量, $P_{av,i}$ 为第 i 张图像的平均查准率。

2.2 实验环境及参数设置

实验使用的深度学习框架为 PyTorch 1.13.1, Python 版本为 3.8, CUDA 版本为 11.6, 训练所用显卡为 12 GB 显存的 NVIDIA-RTX3060。

实验参数设置: 骨干网络为 ResNet50, 并使用在 ImageNet-1k 上预训练的参数初始化骨干网的参数; 训练批次为 64, 测试批次为 128; 输入图像尺寸设置为 256×128 , 并采用随机水平翻转和随机裁剪等数据增强方法对图像进行预处理, 以增强模型的泛化能力; 实验总共训练 120 个周期(epoch); 在训练过程中, 前 10 个 epoch 内学习率从 3.5×10^{-5} 线性增加到 3.5×10^{-4} , 接着学习率在第 40 个和第 70 个 epoch 时分别衰减到 3.5×10^{-5} 和 3.5×10^{-6} ; 本文模型优化使用了自适应梯度下降法(Adam)。

2.3 FFA 模块实验

2.3.1 分割数量对比实验

为了探究图 4 中的分割数量对结果的影响, 在 Market-1501 数据集上进行了实验。实验中将特征分割为 1 份、2 份、4 份、8 份和 16 份, 并观察实验结果。如图 7 所示, 可以观察到在分割数量为 2 时, mAP 的值超过其他分割数量的结果, Rank-1 的值只略低于分割数为 8 的值。综合不同分割数量的 mAP 和 Rank-1, 将 FFA 模块中的分割数量设置为 2。

2.3.2 注意力模块对比实验

为了探究在图 4 中特征加权重组后是否需要再添加注意力以及应该使用何种类型的注意力, 在 Market-1501 数据集上进行了实验, 分别比较使用 ECA 模块、不加注意力模块(NO)以及常见的注意力模块 CBAM^[23]、SE^[24]和 CA^[25]的效果, 实验结果如图 8 所示。从结果中可以观察到, 添加 ECA 模块获得了最好的综合效果。因此, 本算法模型在特征重组后添

加 ECA 模块可以提升模型性能。

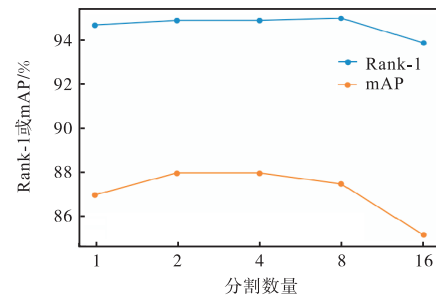


图 7 分割数量对比结果

Fig. 7 Comparison results of split number

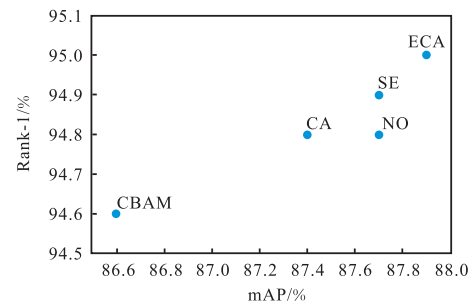


图 8 注意力模块对比结果

Fig. 8 Comparison results of attention module

2.4 消融实验

为了验证每个引入模块的有效性, 在 Market-1501 和 DukeMTMC-reID 两个公开数据集上进行了消融实验, 实验结果见表 1。

在 DukeMTMC-reID 数据集上, 将每个模块单独引入基线(Baseline)模型进行实验。添加 RFC 模块时, mAP 和 Rank-1 提升了 2.1% 和 1.6%; 添加 FFA 模块时, mAP 和 Rank-1 提升了 2.5% 和 1.8%; 采用 GMP 模块时, mAP 和 Rank-1 提升了 0.8% 和 0.9%; 采用 WRT 模块时, mAP 和 Rank-1 提升了 1.0% 和 1.0%。结果表明: 在卷积块中插入 RFC 或 FFA 的识别效果比不插入任何模块的识别效果好, 采用 GMP 的识别效果比采用全局平均池化的识别效果好, 采用 WRT 的识别效果比采用三元组损失的识别效果好。

当同时引入 RFC 和 FFA 模块时, mAP 和 Rank-1 较单独引入 RFC 时提升了 0.5% 和 0.5%, 较单独引入 FFA 时提升了 0.1% 和 0.3%, 表明 RFC 和 FFA 模块组合是有效的。在此基础上再引入 GMP 模块替换原平均池化, mAP 和 Rank-1 提升了 0.8% 和 0.7%, 表明 GMP 模块可以进一步提高特征的判别能力。最后, 引入加权正则化三元组损失替换三元组损失, mAP 和 Rank-1 提升了 0.4% 和 0.4%, 说明加权正则

化三元组损失相对于三元组损失可以更好地约束特征, 减少相似特征间的距离。

表 1 消融实验结果

Tab. 1 Results of ablation experiment

模型	Market-1501		DukeMTMC-reID	
	mAP	Rank-1	mAP	Rank-1
Baseline	85.9	94.5	76.4	86.4
+RFC	87.6	94.8	78.5	88.0
+FFA	87.9	95.0	78.9	88.2
+GMP	86.6	94.4	77.2	87.3
+WRT	86.7	94.2	77.4	87.4
+RFC+FFA	88.1	95.1	79.0	88.5
+RFC+FFA+GMP	88.2	95.3	79.8	89.2
本文模型	89.4	95.3	80.2	89.6

2.5 相关方法对比

在 Market-1501 与 DukeMTMC-reID 两个公开数据集上, 将 RFCFFA-Net 模型与近几年提出的行人重识别方法进行对比, 结果见表 2。由表 2 可知, 本算法模型在 Market-1501 数据集上的 mAP 高于其他算法, Rank-1 的值略低于 ABD-Net 和 PAT, 在 DukeMTMC-reID 数据集上 mAP 和 Rank-1 都高于其

他算法。分析表 2 中的数据可知本文模型的识别效果有所提升。

表 2 对比实验结果

Tab. 2 Comparison experiment results

模型	Market-1501		DukeMTMC-reID	
	mAP	Rank-1	mAP	Rank-1
BAT-Net ^[26]	85.5	94.1	77.3	87.7
ABD-Net ^[27]	88.3	95.6	78.5	89.0
HoReID ^[28]	84.9	94.2	75.6	86.9
ISP ^[4]	88.6	95.3	80.0	89.6
CDNet ^[29]	86.0	95.1	76.8	88.6
PAT ^[30]	88.0	95.4	78.2	88.5
Mos ^[31]	86.8	94.7	77.0	88.7
PGFL-KD ^[32]	87.2	95.3	79.5	89.6
DRL-Net ^[33]	86.9	94.7	76.6	88.1
FED ^[34]	86.3	95.0	78.0	89.4
本文模型	89.4	95.3	80.2	89.6

2.6 可视化实验结果

为了可视化地展示模型的训练效果, 在 Market-1501 数据集上选择部分行人进行识别, 识别结果如图 9 所示。

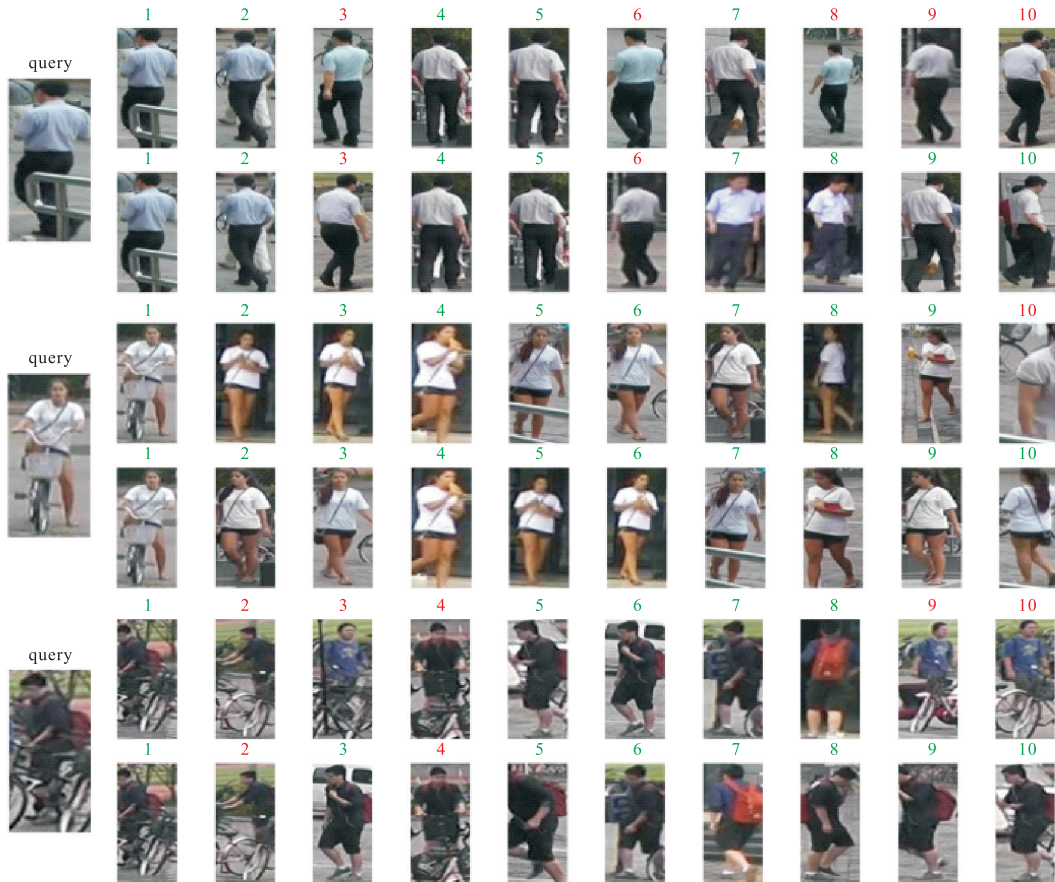


图 9 在 Market-1501 数据集上部分查询结果可视化

Fig. 9 Visualization of partial query results on Market-1501 datasets

在图 9 中, query 为待查询人物的图像, 第 1 行为基线模型识别结果, 第 2 行为本文模型的识别结果, 绿色序号代表识别结果中的人物 ID 与待检测人物的 ID 相同, 红色为不同。以第一幅图为例, 原识别结果中有 5 张识别错误, 本文模型的识别结果仅有 2 张识别错误。从排序结果可以看出, 面对不同事物不同程度的遮挡, 本文模型可以提高人物识别的准确率。

3 结 语

本文提出了一种有效的基于区域特征补全和细粒度特征注意力的行人重识别方法。为了弥补遮挡部分特征的缺失, 引入了 RFC 模块, 通过将特征分块聚类, 使遮挡区域的特征可以使用相关的聚类恢复其特征。为了捕获更细粒度的特征, 引入了细粒度特征注意力模块, 通过对输入特征进行分割加权重组, 再进一步学习通道特征, 可以获得更细粒度的特征。未来行人重识别还有很多工作要做, 包括结合其他图像任务的网络以增强特征, 如行人属性识别、语义分割等, 以及融合多个特征, 如多路径特征融合、多尺度特征融合等, 以提高模型的准确性和泛化性。

参考文献:

- [1] 王素玉, 肖塞. 行人重识别研究综述[J]. 北京工业大学学报, 2022, 48(10): 1100-1112.
- [2] 李擎, 胡伟阳, 李江昀, 等. 基于深度学习的行人重识别方法综述[J]. 工程科学学报, 2022, 44(5): 920-932.
- [3] 冯霞, 杜佳浩, 段仪浓, 等. 基于深度学习的行人重识别研究综述[J]. 计算机应用研究, 2020, 37(11): 3320-3326.
- [4] ZHU K, GUO H Y, LIU Z W, et al. Identity-guided human semantic parsing for person re-identification [EB/OL]. [2023-06-09]. <http://arxiv.org/pdf/2007.13467.pdf>.
- [5] ZHANG Z, LAN C, ZENG W, et al. Relationaware global attention for person re-identification[C]//IEEE. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2020: 3186-3195.
- [6] CHEN G, GU T, LU J, et al. Person re-identification via attention pyramid[J]. IEEE Transactions on image processing, 2021, 30: 7663-7676.
- [7] HOU R, MA B, CHANG H, et al. Vrste: occlusion-free video person re-identification[C]//IEEE. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2019: 7183-7192.
- [8] SHI Y, LING H, WU L, et al. Attribute disentanglement and registration for occluded person re-identification[J]. Neurocomputing, 2022, 470: 226-235.
- [9] XU B, HE L, LIANG J, et al. Learning feature recovery transformer for occluded person re-identification[J]. IEEE Transactions on image processing, 2022, 31: 4651-4662.
- [10] WANG T, LIU H, LI W, et al. Feature completion transformer for occluded person re-identification[EB/OL]. [2023-06-09]. <https://doi.org/10.48550/arXiv.2112.02466>.
- [11] VARIOR R R, HALOI M, WANG G. Gated siamese convolutional neural network architecture for human re-identification[C]//Computer Vision-ECCV 2016: 14th European Conference. Berlin: Springer International Publishing, 2016: 791-808.
- [12] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: a unified embedding for face recognition and clustering[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2015: 815-823.
- [13] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep learning[M]. Cambridge: MIT Press, 2016.
- [14] WEN Y, ZHANG K, LI Z, et al. A discriminative feature learning approach for deep face recognition[C]//Computer Vision-ECCV 2016: 14th European Conference. Berlin: Springer International Publishing, 2016: 499-515.
- [15] 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, 2021, 44(6): 2872-2893.
- [16] ZHOU X, ZHONG Y, CHENG Z, et al. Adaptive sparse pairwise loss for object re-identification[C]//IEEE. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2023: 19691-19701.
- [17] LUO H, GU Y, LIAO X, et al. Bag of tricks and a strong baseline for deep person re-identification[C]//IEEE. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New York: IEEE, 2019: 9025455.

- [18] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 770–778.
- [19] HOU R, MA B, CHANG H, et al. Feature completion for occluded person re-identification[J]. IEEE Transactions on pattern analysis and machine intelligence, 2021, 44(9): 4894–4912.
- [20] BERMAN M, JÉGOU H, VEDALDI A, et al. Multigrain: a unified image embedding for classes and instances[EB/OL]. [2023-06-09]. <https://doi.org/10.48550/arXiv.1902.05509>.
- [21] LIU J J, HOU Q, CHENG M M, et al. Improving convolutional networks with self-calibrated Convolution [C]//IEEE. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2020: 10096–10105.
- [22] WANG Q, WU B, ZHU P, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]//IEEE. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2020: 11534–11542.
- [23] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[EB/OL]. [2023-06-09]. https://openaccess.thecvf.com/content_ECCV_2018/papers/Sanghyun_Woo_Convolutional_Block_Attention_ECCV_2018_paper.pdf.
- [24] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 7132–7141.
- [25] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//IEEE. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2021: 13713–13722.
- [26] FANG P, ZHOU J, ROY S K, et al. Bilinear attention networks for person retrieval[C]//IEEE. Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2019: 8030–8039.
- [27] CHEN T, DING S, XIE J, et al. ABD-Net: attentive but diverse person re-identification[C]//IEEE. Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2019: 8351–8361.
- [28] WANG G, YANG S, LIU H, et al. High-order information matters: learning relation and topology for occluded person re-identification[C]//IEEE. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2020: 6449–6458.
- [29] LI H, WU G, ZHENG W S. Combined depth space based architecture search for person re-identification[C]//IEEE. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2021: 6729–6738.
- [30] LI Y, HE J, ZHANG T, et al. Diverse part discovery: occluded person re-identification with part-aware transformer[C]//IEEE. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2021: 2898–2907.
- [31] JIA M, CHENG X, ZHAI Y, et al. Matching on sets: Conquer occluded person re-identification without alignment[J]. Proceedings of the AAAI conference on artificial intelligence, 2021, 35(2): 1673–1681.
- [32] ZHENG K, LAN C, ZENG W, et al. Pose-guided feature learning with knowledge distillation for occluded person re-identification[C]//ACM. Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM, 2021: 4537–4545.
- [33] JIA M, CHENG X, LU S, et al. Learning disentangled representation implicitly via transformer for occluded person re-identification[J]. IEEE Transactions on multimedia, 2022, 25: 1294–1305.
- [34] WANG Z, ZHU F, TANG S, et al. Feature erasing and diffusion network for occluded person re-identification [C]//IEEE. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2022: 4754–4763.