

DOI: 10.13382/j.jemi.B2205837

基于局部细化多分支与全局特征共享的 无监督行人重识别方法*

钱亚萍^{1,2,3} 王凤随^{1,2,3} 熊磊^{1,2,3}

(1. 安徽工程大学电气工程学院 芜湖 241000; 2. 检测技术与节能装置安徽省重点实验室 芜湖 241000;
3. 高端装备先进感知与智能控制教育部重点实验室 芜湖 241000)

摘要:无监督行人重识别因其在真实监控场景中具备良好的可扩展性而备受关注。现有的无监督行人重识别方法主要通过基础骨干网络获取粗略的全局特征来训练网络,很少利用局部细化分支与全局特征共享来形成更具有鉴别性的特征描述符。本文提出一种基于局部细化多分支与全局特征共享的特征提取网络,该网络融合了粗略的全局特征和局部细化分支中的细腻特征来获取行人多样化的特征表达。另外,为了提升分支网络对潜在关键区域信息的捕获能力,在分支操作前放置通道细化信息融合的注意力块来增强网络对行人特征的关注力度,执行细化特征的专注学习。通过在 Market-1501、DukeMTMC-reID 和 MSMT17 数据集上的实验结果验证了所提方法的有效性,平均精度(mAP)分别提升了 4.4%、3.2%、6.4%,其中在 Market-1501 数据集上的平均精度达到了 83.3%。

关键词:行人重识别;无监督;多分支;局部细化;全局和局部特征;平均精度

中图分类号: TP391.4 **文献标识码:** A **国家标准学科分类代码:** 510.40

Unsupervised person re-identification method based on local refinement multi-branch and global feature sharing

Qian Yaping^{1,2,3} Wang Fengsui^{1,2,3} Xiong Lei^{1,2,3}

(1. School of Electrical Engineering, Anhui Polytechnic University, Wuhu 241000, China; 2. Anhui Key Laboratory of Detection Technology and Energy Saving Devices, Wuhu 241000, China; 3. Key Laboratory of Advanced Perception and Intelligent Control of High-end Equipment, Ministry of Education, Wuhu 241000, China)

Abstract: Unsupervised person re-identification was attracted much attention due to its good scalability in real surveillance scenarios. The existing unsupervised person re-identification methods mainly trained the network by obtaining rough global features through the basic backbone network, and rarely used local refinement branches and global feature sharing to form more discriminative feature descriptors. This paper proposed a feature extraction network based on local refinement multi-branch and global feature sharing. The network combined rough global features and fine features in local refinement branches to obtain diverse feature expressions of person. In addition, in order to improve the ability of the branch network to capture information of potential key areas, an attention block of channel refinement information fusion was placed before the branch operation to enhance the network's attention to person features and perform dedicated learning of refined features. The experimental results on Market-1501, DukeMTMC-reID and MSMT17 datasets verified the effectiveness of the proposed method. The average accuracy (mAP) was improved by 4.4%, 3.2% and 6.4% respectively, and the average accuracy on Market-1501 dataset reached 83.3%.

Keywords: person re-identification; unsupervised; multi-branch; local refinement; global and local characteristics; mAP

收稿日期: 2022-09-19 Received Date: 2022-09-19

* 基金项目:安徽省自然科学基金(2108085MF197、1708085MF154)、安徽高校省级自然科学研究重点项目(KJ2019A0162)、检测技术与节能装置安徽省重点实验室开放基金(DTESD2020B02)、安徽工程大学国家自然科学基金预研项目(Xjky2022040)、安徽高校研究生科学研究项目(YJS20210448、YJS20210449)资助

0 引言

行人重识别 (person re-identification, Re-ID) 试图在不重叠的监控系统中找出特定的行人。近年来, 由于其在视频监控和刑侦调查领域的显著贡献而成为研究热点。鉴于卷积神经网络强大的特征提取能力^[1], 有监督行人重识别方法在这项任务中获得了出色的性能^[2-3]。然而, 该方法依赖于足够的注释标签来训练网络, 这将面临昂贵的人工注释和繁重的标注操作, 限制了它在大规模现实 Re-ID 问题中的发展。于是, 从无标记数据中学习人物表征的纯无监督 (purely unsupervised) 行人重识别方法逐渐受到关注。

纯无监督方法通常利用聚类或图匹配的方法, 直接为未标记的目标数据集生成伪标签进行训练, 例如文献 [4-5]。由于训练过程缺乏标记数据的指导, 该类方法的整体性能大多比有监督方法和无监督域适应 (unsupervised domain adaptation, UDA) 方法要低。为了满足在实际场景中的可扩展性, 如何有效提升无监督 Re-ID 方法的识别精度成为研究的难点之一。

最近的纯无监督行人重识别方法大多利用整幅图像的全局特征进行图像检索。如 Wang 等^[6]利用 ResNet50 网络提取特征, 通过设计基于记忆的多标签分类损失函数 (the memory based multi-label classification loss, MMCL) 来辅助网络来预测行人的多类别标签。为了实现无监督学习, 该方法只在无标记源域中提取全局特征, 未考虑细化的特征差异对无监督分类网络性能的影响, 导致算法识别精度不高。Ge 等^[7]提出了一种自主对比学习 (self-paced contrastive learning, SPCL) 的框架, 该方法同样只采用残差网络用于学习全局特征表示, 再利用对比学习来更新网络的参数。其中, 全局特征提取方式虽然简单且耗时低, 但是特征单一, 缺乏局部细化的细粒度信息, 不利于在复杂场景下对无标记图像中的关键信息进行识别。Xuan 等^[8]提出了一种相机内和相机间相似度计算 (intra-inter camera similarity, IICS) 的方法, 分别用相机内和相机间的伪标签在两个阶段中训练 Re-ID 模型。Han 等^[9]提出将组抽样策略引入无监督行人重识别中, 它将来自同一类的样本组收集到一个小批次中。这样有助于维护每个类内的相似性结构并减少单个样本的影响, 获得了优越的性能。然而, 单一的全局特征可能会使图像中的次显著区域被忽略, 容易造成关键信息的弱化, 影响模型最终的表现。

在无监督 Re-ID 中, 由于真实环境中的光照差异、拍摄角度和行人姿态的多变, 无法从无标记图像中获取充分的特征表达, 导致模型识别精度不高^[10]。另外, 他们大多忽略了行人重识别网络中所必需的细粒度局部特征

信息。这为重新识别一个人提供了更关键的线索。

在基于深度学习的有监督 Re-ID 方法中, 结合全局与局部特征共同改进识别网络的策略已经以先进的性能证明了它们的有效性。如 Sun 等^[11]提出的基于部分特征的卷积基线 (part-based convolutional baseline, PCB), 它将全局特征均匀的分割为 6 份以此获得更为丰富的行人表征, 获得了良好的识别性能。Wang 等^[12]提出了一个多粒度网络 (multiple granularity network, MGN), 通过联合全局和局部特征表示来获得具有多个粒度的局部特征描述符。然而, 却很少有人无监督学习中尝试。2022 年, Li 等^[13]提出一种多标签学习引导的自节奏聚类 (multi-label learning guided self-paced clustering, MLC) 框架用于无监督 Re-ID。该方法通过多尺度网络在全局和两个局部视图中生成了多粒度的特征表示。Pang 等^[14]提出一种基于选择性对比学习的无监督 Re-ID 方法。该方法联合全局和局部特征构造 3 个动态存储库来执行多样化的任务学习, 有效改善了原模型的性能。

为此, 本文不再遵循传统无监督学习中的单一特征描述, 提出了一种基于局部细化多分支与全局特征共享的无监督行人重识别方法。与上述只将行人分为上部和下部的局部分支不同, 本文以局部细化的方式, 将全局特征分割为更精细的局部来获取多样化的特征表达。同时, 设计通道细化信息融合的注意力块, 以增强网络对关键特征的关注力度。在后续研究中, 将继续从局部细化多分支与全局特征共享的角度出发, 尝试构造特征一致性约束, 进一步挖掘全局特征与各级分支中的关联信息, 以提升无监督行人重识别模型的整体性能。

1 本文算法

1.1 整体网络

本文提出了一种基于局部细化的无监督多分支行人重识别方法, 如图 1 所示。整体网络包含交替进行的聚类阶段和模型训练阶段。第 1 阶段中, 本文首先使用 ResNet50 来提取所有训练样本的全局特征, 然后使用聚类方法为全局所有训练实例分配伪标签。在文中, 全局特征与局部特征共享相同的伪标签。模型训练过程时, 遵循 HAN 等人提出的小组抽样策略对行人样本进行划分采样, 再构建局部细化的多分支网络获取多样化的特征表示。这些实例特征被送入相应的内存字典 (Memory) 中进行储存, 以便最终利用统一的对比损失函数监督网络的训练和更新。

1.2 局部细化的多分支网络

在无监督学习中, 图像关键标注信息的缺失使得行人重识别任务的鲁棒性大大降低, 于是图像中的其他信息

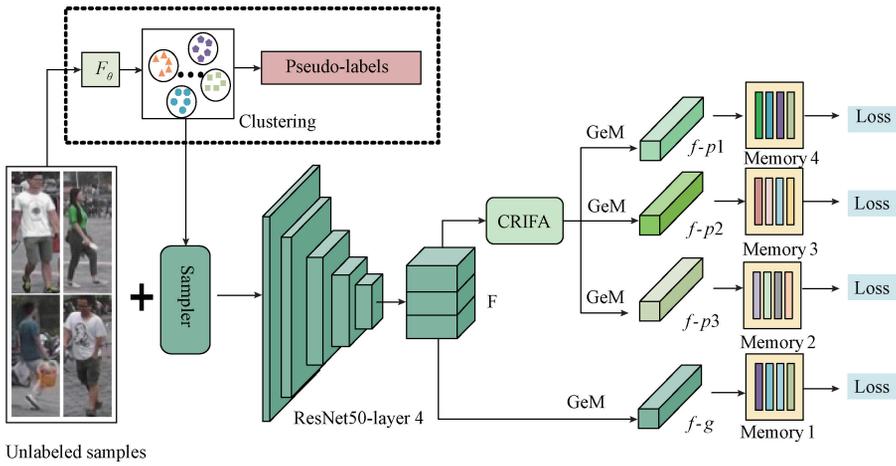


图 1 整体网络结构图

Fig. 1 Overall network structure diagram

显得异常关键。在深度学习中,辅助学习是为了帮助网络具备获取某项重要信息的能力,最终更好地完成学习任务。在 Re-ID 中,行人由定义明确的头部、躯干和腿部组成,由于姿态变化和局部差异引起的各种外观变化使得正确识别某类行人变得困难。而降低外观变化对识别模型产生影响的途径之一是使模型更加关注局部块的信息。如图 2 中,将行人图像由粗粒度特征至细粒度特征划分开,与单一的粗粒度全局特征相比,局部特征可以捕获细粒度的纹理信息。



图 2 行人特征的细致划分

Fig. 2 The refinement of person characteristics

为此,本文构建了一个简单而有效的局部细化多分支网络对全局和局部特征进行分支细化,用来提取四组多样化的特征表示,如图 1 所示。将未标记的行人数据集记为 $D = \{x_i\}_{i=1}^N$, 其中 x_i 为第 i 张图像, N 为图像数。行人样本经 ResNet50 的 layer4 后得到粗略的共享全局特征,表示为 $F_\theta(x_i) \in R^{C \times H \times W}$, 其中 C, H, W 分别表示通道的大小、特征图的高度和宽度。然后网络将分为两路,由全局特征分支 $f-g$ 和包含 CRIFA 模块的 3 个局部特征分支 $f-p_n$ ($n=1, 2, 3, n$ 为分支总数) 联合构建,用于学习丰富且多样化的行人表示。在全局分支中,通过对输出特征 F 应用可学习的广义均值 (generalized mean, GEM) 池化作为一种细粒度的实例检索,得到全局特征 f_g^N 。在

局部分支中,将输出特征 F 经 CRIFA 模块后划分为 3 个水平条带,同样应用 GEM 得到局部细化的特征 f_{p-n}^N , 此时的局部特征中融合了更多的全局高语义共享信息,表达能力更强。最终得到全局和局部实例特征集,利用式 (1) 和 (2) 表示:

$$F_g = \{f_g^1, \dots, f_g^N\} \quad (1)$$

$$F_{p-n} = \{f_{p-n}^1, \dots, f_{p-n}^N\} \quad (2)$$

其中, t 为训练周期。

为了保存得到的全局和局部实例特征,采用 4 个结构相同的内存字典 Memory k ($k=1, 2, 3, 4$) 来存储所有样本的实例特征,并在训练过程中通过相应分支的输出特征进行更新,如式 (3) 和 (4) 所示。

$$M_i^g = \alpha \cdot M_i^g + (1 - \alpha) \cdot f_g^N \quad (3)$$

$$M_i^{p-n} = \alpha \cdot M_i^{p-n} + (1 - \alpha) \cdot f_{p-n}^N \quad (4)$$

其中, $M \in R^{N \times d}$, d 是特征维数; f_g^N 和 f_{p-n}^N 分别表示不同分支的输出特征; M_i^g 和 M_i^{p-n} 分别表示存储在各自 Memory 中、累积不同输出特征后的聚类特征; $\alpha \in [0, 1]$ 为更新速率。

1.3 CRIFA 模块

考虑到切分全局特征可能会使局部特征中的重要信息丢失,设计了 CRIFA 模块在 ResNet50 的 layer4 和全局平均池化层之间关注高层特征并加权。这些来自高层丰富的语义信息能够帮助多分支网络细化行人之间的细微差异,更好地辨别行人身份。

随着网络的加深,捕获的特征级别也越高,信息共享能力就越强。本文在利用残差加深网络学习层的基础上,针对原残差结构中缺乏对局部跨通道交互信息的关注,在 CRIFA 模块中将基础全局特征和加权后的高层重要特征融合,以此来去除冗余的多余信息,加强关键特

征,同时有利于后续局部多分支细化学习的开展,进一步 提高网络的性能。模块结构如图 3 所示。

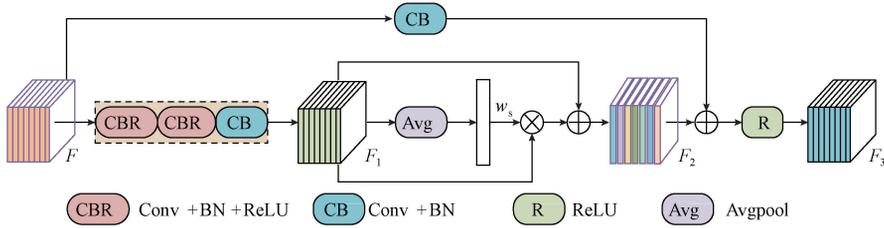


图 3 CRIFA 模块结构

Fig. 3 CRIFA module structure diagram

首先,CRIFA 模块的输入为中间特征图 $F \in R^{H \times W \times C}$, 输出特征记为 $F_3 \in R^{H \times W \times C}$, 其中 C, H, W 分别表示通道的大小、特征图的高度、宽度。将特征图 F 通过 3 组二维卷积块获得深层特征图 F_1 ; 其次,对 F_1 增加局部跨通道交互信息的关注,即利用全局平均池化(global average pooling, GAP)获取通道信息,然后在 GAP 之后的特征上引入权重共享的一维卷积进行细化学习,通过使用 Sigmoid 激活函数得到各个通道之间的权重 w_s 。然后,将所获权重与输入特征图 F_1 的对应元素相乘,最终与 F_1 的对应元素相加,得到 F_2 。具体的实现过程如式(5)所示。

$$F_2 = [\text{Sigmoid}(C1D_k(\text{GAP}(F_1))) \otimes F_1] \oplus F_1 \quad (5)$$

其中, $C1D_k$ 表示通过卷积核大小 $k=5$ 的一维卷积过程,GAP 表示全局平均池化, \otimes 表示逐元素乘法, \oplus 表示逐元素加法。

最终,将原始中间特征图 F 通过一组二维卷积块后与 F_2 进行融合,得到增强后的特征图 F_3 。CRIFA 模块的公式表示为:

$$F_3 = \text{ReLU}[CB(F) \oplus F_2] \quad (6)$$

1.4 损失函数

本文算法将全局特征水平分割并分别计算全局分支与局部分支的对比损失。为了平衡各分支中的损失权重,利用式(7)作为本文的模型损失。其中, $L_c^{f_i-g}$ 表示全

局分支部分的损失, $L_c^{f_i-pm}$ 表示局部分支部分的损失 ($n=3$), $1/3$ 为权重因子; N 表示实验批次数量。

$$L = L_c^{f_i-g} + \frac{1}{3} \sum_{i=1}^N L_c^{f_i-pm} \quad (7)$$

在分配一个伪标签后,利用实例特征来构造统一的对比损失来监督网络的学习,其中统一对比损失函数定义如式(8)所示。

$$L_c = -\log \frac{\exp(\langle f, c^+ \rangle / \tau)}{\sum_{k=0}^{K-1} \exp(\langle f, c_k \rangle / \tau)} \quad (8)$$

其中, f 表示实例特征,在本文代表全局特征 $f_g^{N^t}$ 和局部分支特征 $f_{p-n}^{N^t}$; c^+ 表示 f 所属聚类簇的质心(特征中心); c_k 表示 C_k 所属聚类簇的质心, k 是聚类簇的数量, τ 为超参数, \langle, \rangle 用来计算两个特征向量之间的内积来度量它们的相似性。

2 实验结果与分析

2.1 数据集描述和评价指标

本文在 3 个大规模的行人重识别数据集 Market-1501、DukeMTMC-reID 和 MSMT17 上进行了一系列实验来评估所提方法的有效性,并将实验结果与其他先进的无监督行人重识别方法进行比较。其中,每个数据集中行人图像库的划分如表 1 所示。

表 1 不同数据集在无监督 Re-ID 任务中的训练和测试数据统计

Table 1 Training and test data statistics of different datasets in unsupervised Re-ID tasks

数据集	训练身份	训练图像	测试身份	查询图像	总图像	摄像机个数
Market-1501	751	12 936	750	3 368	32 668	6
DukeMTMC-reID	702	16 522	702	2 228	36 441	8
MSMT17	1 041	32 621	3 060	11 659	126 441	15

Market-1501 数据集是由安置在清华大学的 6 个不同相机采集的 1 501 个行人的图像。该数据集划分为训练集和测试集两个部分,训练集行人数量为 751,共有 12 936 张图像;测试集行人数量为 750,共有 19 732 张

图像。

DukeMTMC-reID 是从杜克大学内的 8 个不重叠的相机视图中收集,包含 1 404 个行人,训练集行人数量为 702,共有 16 522 张图像;测试集行人数量为 702,共有

17 661 张图像。

MSMT17 数据集是目前 Re-ID 中最赋有挑战性的综合跨场景大数据集,它包含由 12 个室外相机和 3 个室内相机所采集的 126 411 张行人图像,4 101 个行人身份,并且涵盖了多场景和多时段的行人图像。这些图像在不

同环境的相机之间和不同光照、场景、姿态和背景之间存在很大的差异,如图 4 所示。与其他两个数据集相比,该数据集中行人和摄像头数目更多,场景和照明变化更大,时间跨度更广,因此,在该数据集下模型的表现更具挑战性。



图 4 MSMT17 数据集行人图像示例图

Fig. 4 MSMT17 dataset person image example diagram

在评估指标方案中,选择通用的匹配精度 Rank-*n* 和平均精度均值 mAP 作为性能评估的标准。其中,Rank-*n* 表示待查询图像与候选图像按照相似度从大到小排序的前 *n* 张候选图片中存在与查询图像属于同一身份的准确率。mAP 反映了检索的行人在数据库所有正确的图片中处在顺序列表前面的程度,它是将多分类任务中的平均精度 AP 求和再取平均值后的结果,计算过程如式(9)所示。其中,AP_{*i*} 表示每个类别的平均精度,C 表示类别数,每次实验采用随机测试集重复 10 次以获取平均性能。

$$mAP = \frac{\sum_{i=1}^n AP_i}{C} \quad (9)$$

2.2 实验设置

该文实验环境配置为显卡 GeForce RTX 2080Ti、CPU 为 4 核 Intel(R) Xeon(R) Silver 4110 CPU @ 2.10 GHz、操作系统为 Ubuntu 18.04、CUDA10.0 和深度学习框架 Pytorch1.1.0。在训练中,每个小批量样本中包含 256

张、代表 16 个行人身份的图像,行人图像预处理为 256×128 大小;数据增强策略采用对图片进行随机翻转、裁剪和擦除操作;在每次迭代中,使用 DBSCAN 和 Jaccard 距离对 *k* 个最近相邻的样本进行聚类,设置 *k* = 30;聚类过程中,实例对之间的距离阈值设置为 0.6,形成聚类的最小实例数设为 4;将对损失函数中的超参数 τ 设为 0.05,动量系数设为 0.2;其他实验参数设置如表 2 所示。本文所有的实验均在单查询设置下进行,不采用任何后处理技术,如重排序。

本文在每个 epoch 计算结束后计算当前的 mAP 值,并在迭代次数达到 50 时计算各 Rank-*n* 的指标值。在 Market-1501、DukeMTMC-reID 和 MSMT17 数据集集中的训练时间大约分别需要 3.63、4.43、11.35 h,增加多分支前后的训练时间对比结果如表 3 所示,在 3 个数据集中加入多分支网络和 CRIFA 模块都增加了网络的训练时间,表明网络花费了额外的时间去学习多样化的局部细化特征。

表 2 不同数据集在无监督 Re-ID 任务中的训练和测试数据统计

Table 2 Training and test data statistics of different datasets in unsupervised Re-ID tasks

数据集	GPU	迭代次数	批次	学习率	权重衰减率	组抽样大小
Market-1501	1	50	64	0.00035	0.1	256
DukeMTMC-reID	1	50	64	0.00035	0.1	128
MSMT17	4	50	256	0.00005	0.1	1024

表 3 增加分支模块后各网络的训练时间对比

Table 3 Comparison of training time of each network after adding branch module

训练时间(时:分:秒)	Market-1501	DukeMTMC-ReID	MSMT17
全局分支	3:24:08	4:10:43	10:36:49
全局分支+局部多分支	3:33:25	4:19:46	11:03:20
全局分支+局部多分支+CRIFA	3:38:16	4:26:02	11:21:04

2.3 精度对比实验

表 4 中展示了所提方法与基线方法在 3 个数据集中的性能对比数据,能够看出本文方法相较于基线方法,所提方法相较于基线方法 mAP 值分别提升了 4.4%、

3.2%、6.4%, Rank-1 值分别提升了 1.9%、2.0%、1.2%,表明在无监督特征学习中,利用多分支之间的互补性有利于形成更完整的行人表示,帮助基线网络提升整体性能。

表 4 在不同数据集中的精度对比实验结果

Table 4 Accuracy comparison experimental results in different datasets

(%)

Methods	Market-1501				DukeMTMC-reID				MSMT17			
	mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10
Baseline	78.9	91.4	96.1	97.3	68.9	82.6	90.6	92.7	21.0	47.2	58.6	63.8
Ours	83.3	93.3	97.4	98.5	72.1	84.6	91.9	94.0	27.4	58.4	69.0	73.3

2.4 消融实验

为验证该文方法各模块的作用,本小节在 Market-1501 数据集上开展消融实验。将文献[9]中的框架作为基线(baseline)网络,每个对比实验的参数设置和训练策略与其保持相同。实验结果如表 5 所示。

对比基线中仅学习单一特征的全局分支,结合全局与局部多分支学习的 Ours1 模型的 mAP 和 R-1 分别达到 81.4%、92.3%,相较于基线网络分别提升了 2.5%、0.9%,表明多分支辅助网络学习到行人图像各个部位的

细粒度信息,形成了更完整的特征表示。嵌入 CRIFA 模块的多分支网络模型(Ours2)相较于 Ours1 提升了 1.9% (mAP)和 1.0%(Rank-1),证明 CRIFA 模块能够有效地增强多分支网络捕捉行人关键特征的能力,因此能学习到更有利于判别行人的细节信息。与基线网络进行对比,在网络中执行局部细化的多分支学习后,mAP 和 R-1 最终达到了 83.3%、93.3%,相较于基线网络分别提升了 4.4%、1.9%,验证了所提方法的有效性。

表 5 消融实验结果

Table 5 Results of ablation experiments

(%)

Methods	Global	Local multi-branches	CRIFA	mAP	R-1	R-5	R-10
Baseline	✓			78.9	91.4	96.1	97.3
Ours1	✓	✓		81.4	92.3	97.3	98.0
Ours2	✓	✓	✓	83.3	93.3	97.4	98.5

另外,为探究 CRIFA 模块和不同局部多分支之间作用对网络性能的影响,本文在 Market-1501 数据集中展开了对比实验。通过依次在基础全局分支 f-g 中增添不同的局部多分支(f-p1、f-p2、f-p3、f-p4)来逐步寻找使得模型性能最优的精准分支搭配。对比实验结果如表 6 所示。

分析表 6 可得:

1)对比包含全局分支 f-g 的基线网络和包含 f-g 并将特征图切割为 n (n 取 1、2、3、4)部分的局部特征分支后,改进前后性能均有不同程度的提升。证明本文联合全局-局部分支的结构有效激发了粗粒度特征与细粒度特征之间的互补性,因此学习到更利于辨别行人身份的判别信息,所以性能有所提升。

2)对比在 1)设置下,在不同分支网络中增加 CRIFA 模块设计的网络性能。发现 CRIFA 模块能够进一步地提升分支网络的性能,有效地辅助细化后的局部分支捕获行人的次显著特征,丰富了特征描述符。

3)当模型复杂度提高(即加入更多的分支)后,发现对局部特征切分的越精细,算法性能反而有所下降。说

明过于精细的切分使得多个局部块的信息被重复利用,导致特征的大量冗余,网络很难捕获到有用的鉴别特征。

4)对比表中所有网络设计的性能,f-g 与增加 CRIFA 模块的多分支(f-p1+f-p2+f-p3)联合学习的设计最优。表明精细、准确的切分和施加有效的关注能够学习到不同水平区域间最具判别的互补视觉特征,帮助无监督网络获得最佳的识别性能。

表 6 对比不同网络设计对模型性能影响的结果

Table 6 Comparing the effects of different network designs on model performance

(%)

网络设计	mAP	R-1	R-5	R-10
f-g	78.9	91.4	96.1	97.3
f-g+f-p1+f-p2	81.9	93.0	96.9	98.0
f-g+f-p1+f-p2+CRIFA	83.0	93.3	97.3	98.3
f-g+f-p1+f-p2+f-p3	81.4	92.3	97.3	98.0
f-g+f-p1+f-p2+f-p3+CRIFA	83.3	93.3	97.4	98.5
f-g+f-p1+f-p2+f-p3+f-p4	80.7	92.3	96.8	97.8
f-g+f-p1+f-p2+f-p3+f-p4+CRIFA	81.7	92.4	97.1	97.8

2.5 可视化对比

本文通过在 Market-1501 数据集上选择 3 位行人进

行实验,可视化展示了本文方法与基线方法对同一行人检索前 10 张图像的对比结果,如图 5 所示。其中,每张图的第 1 列 Query 代表随机挑选的待查询行人图像,其

他 1~10 列图像表示行人重识别的搜索结果。另外,每张图像上方的数字颜色代表该检索结果是否正确,数字中带框的表示错误,不带框的表示正确。

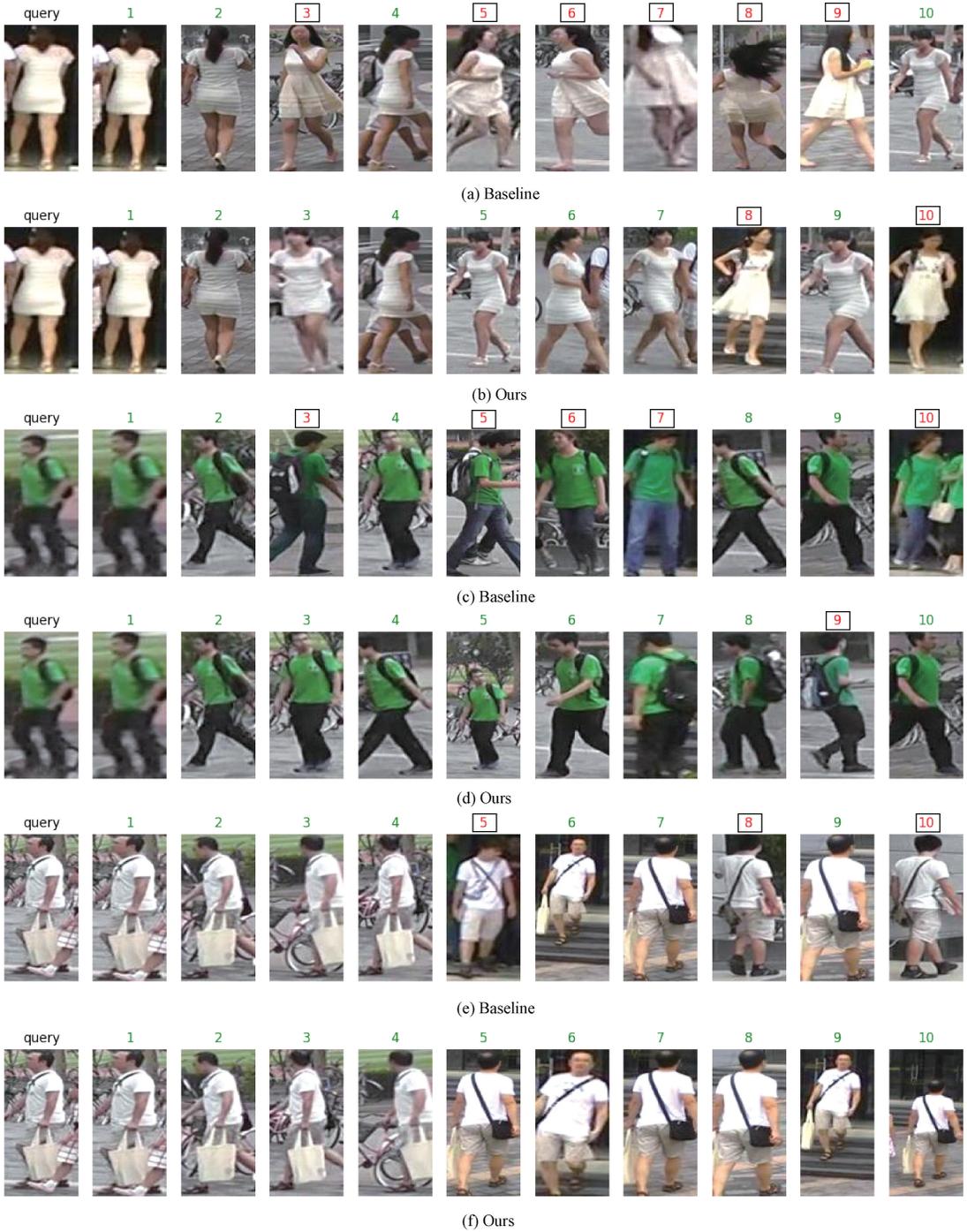


图 5 在 Market-1501 数据集上的可视化结果

Fig. 5 Visualization results on the Market-1501 dataset

当行人特征较为单一,如图 5(a)中的行人。由于在全局特征中不能充分关注到图像中的局部细节,如裙摆形状、头发状态和衣袖款式,最终出现了 6 个错误的识别

结果。与其相比,本文方法图 5(b)中只有两个错误的识别结果,充分验证了本文方法能够得到更具鉴别性的丰富特征描述符,使得网络整体识别能力更强。同时相比

于 baseline, 本文方法检索到正确行人的位置更加靠前, 表明了本文方法的先进性。另外, 基线方法在检索图 5 (c) 中的行人时, 获得了 5 个错误的查询结果, 错误结果图中的行人与 Query 中的行人特征十分相似, 表明仅采取全局特征会使网络很难关注到某些关键信息, 如识别错误的第 6、7、10 位行人衣服上的 logo 细节。而增加将特征图切割为 3 部分的局部分支后, 网络获得了更加细致的特征关注点, 最终得到了仅 1 个错误的查询结果。在图 5(e) 中, 当行人的状态复杂, 如身处多光照环境下, 背包和拎包的状态。基线方法无法准确判别错误识别结



图 6 在 Market-1501 数据集上的可视化热力图

Fig. 6 Visual heat map on Market-1501 dataset

2.6 与其他方法对比

为了验证本文方法的先进性, 将本文方法在公开数据集中与其他先进的纯无监督行人重识别方法进行比

果图中第 5、8、10 位行人手部拿书的状态, 而本文方法能够在复杂的特征表达中获取关键的鉴别性信息, 检索出前 10 位正确的行人, 验证了所提方法能够展示出色的行人重识别效果。

为验证本文 CRIFA 模块的有效性, 本文对模块输出的区域注意力特征进行可视化分析, 可视化结果如图 6 所示。由图可看出, 该分支能够更加注重行人肢体、手中物体、裙摆、背包、衣服纹理等多样化的细节特征, 说明了 CRIFA 模块能够有效增强网络对潜在关键区域信息的捕获能力。

表 7 与其他先进的无监督行人重识别方法的对比结果

Table 7 Comparison with other advanced unsupervised person re-identification methods

(%)

Methods	Market-1501				DukeMTMC-reID				MSMT17			
	mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10
MMCL ^[6]	45.5	80.3	89.4	92.3	40.2	65.2	75.9	80.0	11.2	35.4	44.8	49.8
Ref. [14]	54.4	82.2	89.9	92.6	47.2	69.9	79.7	82.2	13.3	41.4	53.6	58.7
JVTC+ ^[15]	47.5	79.5	89.2	91.9	50.7	74.6	82.9	85.3	17.3	43.1	53.8	59.4
ACAN ^[16]	50.6	73.3	87.6	91.8	46.6	65.1	80.6	85.1	11.2	27.1	40.9	47.3
HCT ^[17]	56.4	80.0	91.6	95.2	50.7	69.6	83.4	87.4	-	-	-	-
DSCE ^[18]	61.7	83.9	92.3	-	53.8	73.8	84.2	-	15.5	35.2	48.3	-
MLC ^[13]	66.2	86.7	93.5	95.6	52.3	73.6	82.3	85.5	12.7	39.2	49.4	53.9
JGCL ^[19]	66.8	87.3	93.5	95.5	62.8	82.9	87.1	88.5	21.3	45.7	58.6	64.5
IICS ^[8]	72.9	89.5	95.2	97.0	64.4	80.0	89.0	91.6	18.6	45.7	57.7	62.8
SpCL ^[7]	73.1	88.1	95.1	97.0	65.3	81.2	90.3	92.2	19.1	42.3	55.6	61.2
C ³ AB ^[20]	77.0	90.0	96.3	97.6	67.0	81.2	90.3	92.6	-	-	-	-
HCD ^[21]	78.1	91.1	96.4	97.7	65.6	79.8	88.6	91.6	26.9	53.7	65.3	70.2
Baseline ^[9]	79.2	92.3	96.6	97.8	69.1	82.7	91.1	93.5	21.0	47.2	58.6	63.8
CACL ^[22]	80.9	92.7	97.4	98.5	69.6	82.6	91.2	93.8	23.0	48.9	61.2	66.4
Ref. [23]	81.7	92.4	97.4	98.1	69.0	82.9	90.9	93.0	24.6	50.2	61.4	65.7
Ours	83.3	93.3	97.4	98.5	72.1	84.6	91.9	94.0	27.4	58.4	69.0	73.3

上述方法大多都是基于 ResNet50 网络获取行人的全局特征表示, 采用粗粒度的全局特征往往将关注点放置行人主干部位, 容易忽略四肢、脚部等局部细节, 导致模型性能有限。与大多数现有的只利用全局特征的

无监督方法相反, 同时使用全局特征和局部特征分支来表示图像, 通过与各类方法的对比结果可见, 本文方法在 3 个数据集上均取得了较高的精度, 超过了目前大多数方法。

3 结 论

本文针对无监督行人重识别提出了一种局部细化的多分支特征提取网络,该网络重点关注单一特征表达能力的欠缺,采取 3 个局部视图的特征分支来削弱传统全局特征中固有的粗粒度偏差,并利用多分支之间的互补性形成了多样化的细化表征,从而提升模型的重识别精度。此外,本文还设计了通道细化信息融合的注意力模块,用于强化网络对细粒度区域的激活响应,提升了网络定位关键信息的能力。最终在 3 个公共数据集中验证了所提方法提取特征的多样性和有效性,具有更好的鲁棒性和识别效果,为无监督行人重识别提供了可行的研究思路。在下一步的工作中,将考虑在局部多分支学习中构造特征一致性约束,进一步挖掘各级特征之间的空间关系,从而获得更好的识别效果。

参考文献

- [1] 朱阳光,刘瑞敏,黄琼桃. 基于深度神经网络的弱监督信息细粒度图像识别[J]. 电子测量与仪器学报, 2020,34(2):115-122.
ZHU Y G, LIU R M, HUANG Q T. Fine-grained image recognition with weakly supervised information based on deep neural network [J]. Journal of Electronic Measurement and Instrumentation, 2020, 34 (2): 115-122.
- [2] 周航,黄春光,程海. 基于全局多粒度池化的可见光红外行人重识别[J]. 电子测量技术, 2022, 45 (1): 122-128.
ZHOU H, HUANG CH G, CHEN H. Visible infrared person re-identification based on global multi-granularity pooling[J]. Electronic Measurement Technology, 2022, 45(1): 122-128.
- [3] 刘仁春,孟朝晖. 基于孪生对抗 SGAN 的行人重识别研究[J]. 电子测量技术, 2019, 42(15): 155-166.
LIU R CH, MENG CH H. Research of person re-identification based on SGAN [J]. Electronic Measurement Technology, 2019, 42(15): 155-166.
- [4] SRIDHAR R S, MUNAGA V N K P, RAMADOSS B. Spatio-temporal association rule based deep annotation-free clustering (STAR-DAC) for unsupervised person re-identification [J]. Pattern Recognition, 2022, 122: 108287.
- [5] VALEM L P, PEDRONETTE D C G. Person re-id through unsupervised hypergraph rank selection and fusion [J]. Image and Vision Computing, 2022, 123: 104473.
- [6] WANG D K, ZHANG SH L. Unsupervised person re-

- identification via multi-label classification [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 10978-10987.
- [7] GE Y X, ZHU F, CHEN D P, et al. Self-paced contrastive learning with hybrid memory for domain adaptive object re-id [C]. Proceedings of the 34th Conference on Neural Information Processing Systems (NIPS), 2020: 11309-11321.
- [8] XUAN SH Y, ZHANG SH L. Intra-inter camera similarity for unsupervised person re-identification [C]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 11921-11930.
- [9] HAN X M, YU X H, NAN J, et al. Rethinking sampling strategies for unsupervised person re-identification [EB/OL] [2021-8-28] [2022-9-19] <http://arxiv.org/abs/2107.03024>.
- [10] ZHANG B H, ZHU S Y, ZHOU Y F, et al. A novel unsupervised person re-identification algorithm based on soft multi-label and compound attention model [J]. Multimedia Tools and Applications, 2022, 81 (17): 24081-24098.
- [11] 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 the 2018 European Conference on Computer Vision (ECCV), 2018: 501-518.
- [12] WANG G SH, YUAN Y F, CHEN X, et al. Learning discriminative features with multiple granularities for person re-identification [C]. Proceedings of the 2018 ACM Multimedia Conference, 2018: 274-282.
- [13] LI Q, PENG X J, QIAO Y, et al. Unsupervised person re-identification with multi-label learning guided self-paced clustering [J]. Pattern Recognition, 2022, 125: 108521.
- [14] PANG B, ZHAI D M, JIANG J J, et al. Fully unsupervised person re-identification via selective contrastive learning [J]. ACM Transactions on Multimedia Computing, Communications and Applications, 2022, 18(2): 1-15.
- [15] LI J N, ZHANG S L. Joint visual and temporal consistency for unsupervised domain adaptive person re-identification [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2020: 483-499.
- [16] QI L, WANG L, HUO J, et al. Adversarial camera alignment network for unsupervised cross-camera person re-identification[J]. IEEE Transactions on Circuits and

- Systems for Video Technology, 2022, 32 (5): 2921-2936.
- [17] ZENG K W, NING M N, WANG Y H, et al. Hierarchical clustering with hard-batch triplet loss for person re-identification [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 16-18, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 13654-13662.
- [18] YANG F X, ZHONG ZH, LUO ZH M, et al. Joint noise-tolerant learning and meta camera shift adaptation for unsupervised person re-identification [C]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 4853-4862.
- [19] CHEN H, WANG Y H, BENOIT L, et al. Joint generative and contrastive learning for unsupervised person re-identification [C]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 2004-2013.
- [20] LI M K, SUN H, LIN CH Q, et al. The devil in the tail: Cluster consolidation plus cluster adaptive balancing loss for unsupervised person re-identification [J]. Pattern Recognition, 2022, 129: 108754.
- [21] ZHENG Y, TANG SH X, TENG G L, et al. Online pseudo label generation by hierarchical cluster dynamics for adaptive person re-identification [C]. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2021: 8351-8361.
- [22] LI M K, LI CH G, GUO J. Cluster-guided asymmetric contrastive learning for unsupervised person re-identification [J]. IEEE Transactions on Image Processing, 2022(31): 3606-3617.
- [23] CHEN D, ZHOU J Y, WANG N N, et al. Hybrid dynamic contrast and probability distillation for unsupervised person re-id [J]. IEEE Transactions on Image Processing, 2022(31): 3334-3346.

作者简介



钱亚萍, 硕士研究生, 主要研究方向为计算机视觉、行人重识别。

E-mail: 2554913824@qq.com

Qian Yaping, M. Sc. candidate. Her main research interests include computer vision and person re-identification.



王凤随(通信作者), 博士, 副教授, 主要研究方向为图像与视频信息处理、视觉计算与智能分析。

E-mail: fswang@ahpu.edu.cn

Wang Fengsui (Corresponding author), Ph. D., associate professor. His main research interests include image and video information processing, visual computing and intelligent analysis.



熊磊, 硕士研究生, 主要研究方向为计算机视觉、目标检测。

E-mail: 319250240@qq.com

Xiong Lei, M. Sc. candidate. His main research interests include computer vision and target detection.