

DOI: 10.13382/j.jemi.B2407926

基于多细粒度双流网络的行人重识别^{*}

宋晓勇¹ 孙学宏^{1,2} 刘丽萍^{1,2} 覃国车¹ 余 彤¹ 李享国¹

(1. 宁夏大学电子与电气工程学院 银川 750021; 2. 宁夏沙漠信息智能感知重点实验室 银川 750021)

摘要:在实际监控场景中行人重识别任务面临诸多挑战,如部分图像遮挡(树木、人、汽车、小物体等)导致识别过程中关键信息丢失和识别精度下降。在遮挡行人重识别任务中,通常采用局部联合全局特征或姿态估计器的方法来解决识别精度低等问题,虽然在部分遮挡情况下利用单流网络有较好的识别性能,但在处理过程中未能充分挖掘剩余关键特征信息。为此,提出了一种基于多细粒度双流网络的遮挡行人重识别方法,通过设计多细粒度局部特征提取策略、双流特征处理网络和特征权重融合模块来增强关键特征信息提取能力。该方法采用视觉Transformer(ViT)提取全局特征,并将其划分为多组局部特征。随后,各组局部特征分别经过双流特征处理网络,将通过双流网络的特征进行特征权重融合,从而更有效地挖掘关键特征信息。在Occluded-Duke、Market-1501、DukeMTMC-reID和MSMT17数据集上实验结果证明所提方法的有效性与合理性,平均精度均值(mAP)/Rank-1指标分别达到了61.3%/68.3%、89.0%/95.2%、82.5%/91.1%和66.8%/84.5%。

关键词:遮挡行人重识别;多细粒度特征提取;双流网络;特征融合

中图分类号: TP391; TN911.7 文献标识码: A 国家标准学科分类代码: 510.40

Pedestrian re-identification based on a multi-granularity dual-stream network

Song Xiaoyong¹ Sun Xuehong^{1,2} Liu Liping^{1,2} Qin Guoche¹ Yu Tong¹ Li Xiangguo¹

(1. School of Electronic and Electrical Engineering, Ningxia University, Yinchuan 750021, China;

2. Key Laboratory of Intelligent Perception for Desert Information, Yinchuan 750021, China)

Abstract: In real surveillance scenarios, pedestrian re-identification tasks face numerous challenges, such as partial image occlusions (trees, people, cars, small objects, etc.) that lead to the loss of key information and a decline in recognition accuracy. To address issues like low recognition accuracy in occluded pedestrian re-identification tasks, methods that combine local and global features or use pose estimators are commonly employed. Although single-stream networks can achieve good recognition performance under partial occlusions, they fail to fully exploit the remaining critical feature information during processing. Therefore, we propose an occluded pedestrian re-identification method based on a multi-granularity dual-stream network. By designing a multi-granularity local feature extraction strategy, a dual-stream feature processing network, and a feature weight fusion module, the ability to extract key feature information is enhanced. This method employs a vision Transformer (ViT) to extract global features and divides them into multiple groups of local features. Subsequently, each group of local features is processed through a dual-stream feature processing network. The features obtained from the dual-stream network are then fused using a feature weight fusion mechanism, thereby more effectively mining key feature information. Experimental results on the Occluded-Duke, Market-1501, DukeMTMC-reID, and MSMT17 datasets demonstrate the effectiveness and validity of the proposed method, achieving mAP/Rank-1 indicators of 61.3%/68.3%, 89.0%/95.2%, 82.5%/91.1%, and 66.8%/84.5%, respectively.

Keywords: occluded person re-identification; multi-granularity feature extraction; dual-stream network; feature fusion

0 引言

行人重识别(person-re-identification, ReID)作为计算机视觉领域关键课题,在智能监控、公共安全等领域至关重要。该技术旨在解决跨设备、跨场景下,从大量复杂的行人图像中准确且高效地识别出同一行人的身份^[1]。在实际应用中,行人重识别技术面临光照变化、姿态多样性和复杂背景等挑战,这些因素显著影响其性能^[2]。同时,遮挡问题(如汽车、广告牌、树木等)进一步增加了该技术的复杂性。

在现有卷积神经网络(convolutional neural network, CNN)方法中,多数研究通过姿势估计^[3]、感兴趣区域^[4]和骨骼关键点^[5]等局部特征提取技术,构建具有空间感知能力的特征表征模型。为增强模型准确性,Chen 等^[6]、Tay 等^[7]引入注意力机制,增强模型对未被遮挡区域的关注,进而提高识别准确性;Miao 等^[8]通过结合姿态信息和特征聚合机制解决行人重识别中的遮挡问题,且提供了 Occluded-Duke 遮挡数据集;Sun 等^[9]提出了部分卷积基线网络(part-based convolutional baseline, PCB)和精细化部分池化方法(refined part pooling, RPP),通过精确定位和优化部分特征,处理遮挡和姿态变化的行人重识别任务;Zhou 等^[10]提出了一种姿态驱动可见性模型(pose-driven visibility model, PDVM),通过姿态估计和关键点检测来提取人体特征改善对齐问题,在遮挡的情况下提高行人重识别的准确性;贺晓东等^[11]提出了一种融合局部特征解析与视点感知的鲁棒性识别框架,通过语义分割将车辆解构为前、后、侧、顶 4 个语义部件构建细粒度表征,以视点感知网络动态预测视角概率分布生成自适应权重矩阵,实现特征平滑与增强。上述研究多聚焦于显性局部特征的独立优化或遮挡区域的规避策略,在特征处理过程中缺乏对关键判别性信息与全局上下文依存关系的深度挖掘,导致身份识别过程中易丢失部分关键特征。

随着视觉 Transformer(ViT)^[12]在计算机视觉、图像分类等任务中的显著成功,He 等^[13]突破传统卷积神经网络难以高效捕捉全局信息和长距离依赖关系的局限,首次提出了纯视觉 Transformer 物体重识别架构(Transformer-based object re-identification, TransReID),通过设计拼图补丁模块和侧信息嵌入模块提升物体重识别性能;Zhu 等^[14]提出了自动对齐行人重识别架构,通过引入特征图自适应对齐模块,确保在不同视角下提取具有判别性特征;Bian 等^[15]提出了遮挡特征恢复机制,有效增强了特征表示的完整性和鲁棒性;Li 等^[16]提出了一种端到端的部分感知 Transformer 模型,通过设计上下文感知 Transformer 编码器和部分原型 Transformer 解码器,

精准捕获不同身体部位;Wang 等^[17]提出了一种姿态引导的特征解耦技术,通过解耦行人图像中的关键特征与非关键特征,增强模型在遮挡情况下的鲁棒性和识别准确性;Zhou 等^[18]提出了一种运动感知 Transformer 架构,通过分析不同行人姿态照片所衍生的运动信息来识别主要身体部位;Zhu 等^[19]提出了一种双重跨注意力学习的方法,通过设计双重跨注意力模块在不同任务之间共享和强化关键信息,从而实现更具辨别力和鲁棒性的特征表示。以上研究虽在特征对齐、遮挡处理、部件感知等方面取得较好的性能,但其技术路线多聚焦于局部优化,忽略了在特征处理过程中对关键判别性信息的深入挖掘以及全局-局部特征联合建模不足等问题。

针对以上问题,受双流网络^[20]的启发和以往研究方法的不足,本文提出了一种纯 Transformer 架构的多细粒度双流网络(multi-granularity dual-stream network, MGDSN),该网络融合多细粒度局部特征提取策略、双流特征处理网络和特征权重融合模块。通过全局-局部协同机制构建层次化特征表示体系,将全局特征划分为 4 个粗粒度局部特征和 6 个细粒度局部特征并独立处理,既能捕捉局部细节信息(如鞋子、包包、帽子等),又能保留全局上下文信息(如整体姿态、衣着风格等)。纯 Transformer 架构的双流网络,充分利用浅层网络和深层网络对特征捕获能力的差异。利用浅层网络提取低级特征(如边缘、纹理和颜色等),深层网络提取高级特征(如语义信息或复杂模式等),分层特征提取方式可有效提升模型的性能、鲁棒性和泛化能力。本文最后设计了特征权重融合模块,实现特征的动态权重分配、强化关键信息、抑制噪声和冗余信息,从而促进特征间的互补性和提升模型性能。在多个公开数据集上验证了本文方法的有效性,在挖掘剩余关键特征信息和全局-局部特征联合建模等方面优于先进(state-of-the-art, SOTA)模型。

1 本文方法

本文提出的基于 Transformer 的多细粒度双流网络整体架构如图 1 所示,主要由以下几个关键模块构成:全局特征提取网络(以 ViT 作为全局特征提取网络)、多粒度局部特征提取模块、双流网络、特征权重融合模块以及行人身份预测模块。

1.1 全局特征提取

给定一副输入图像 $\mathbf{x} \in \mathbb{R}^{H \times W \times C}$,其中 H 、 W 、 C 分别表示其高度、宽度和通道数。首先将其划分为固定大小的一系列重叠图像块(patches)。设图像的高度为 H ,宽度为 W ,每个图像块的大小为 $P \times P$,滑动窗口的步长为 S 。则图像可以被划分为 N 个图像块,其中 N 的计算公式为:

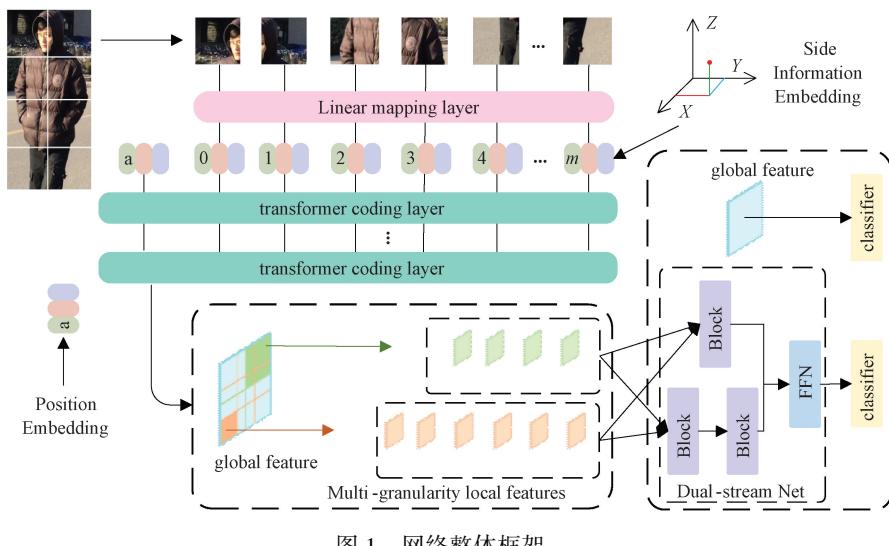


图 1 网络整体框架

Fig. 1 The overall network framework

$$N = \left\lfloor \frac{H + S - P}{S} \times \frac{W + S - P}{S} \right\rfloor \quad (1)$$

将每个图像块经过一个线性投影 F 映射到一个更高维的空间 D 中, 从而得到每个图像块的嵌入表示。引入一个可学习的类标记(class token) x_{cls} , 将其前置到输入序列中。通过添加可学习的位置嵌入(position embeddings), 空间信息被整合到模型中。输入序列表示为:

$$\mathbf{Z}_0 = [x_{cls}; x_p^1; x_p^2; \dots; x_p^N] + \mathbf{E}_{pos} \quad (2)$$

其中, $\mathbf{P} \in \mathbf{R}^{(N+1) \times D}$ 表示位置编码, 用于整合空间信息。

设输入序列 \mathbf{Z}_0 经过一系列 Transformer 层进行特征表示学习。每个 Transformer 层包含多头自注意力机制和前馈神经网络(multi-layer perceptron, MLP)。假设共有 L 个 Transformer 层, 则最后一层的输入特征表示为:

$$\mathbf{Z}^{L-1} = [z_0^{L-1}; z_1^{L-1}, z_2^{L-1}, \dots, z_N^{L-1}] \quad (3)$$

式中: z_0^{L-1} 为全局 token; z_i^{L-1} ($i=1, 2, 3, \dots, N$) 为第 i 个 patch 的特征; N 表示 patch tokens 的总数量(不包括全局 token)。

1.2 多粒度特征提取

对于处理遮挡行人重识别任务时, 由于部分细节特征易受遮挡干扰, 深度挖掘残余有效信息则尤为重要。本文为实现特征精细化表示, 将全局特征细分为多粒度局部特征(即将一个全局特征划分为一个 4 个局部特征和一个 6 个局部特征), 有助于模型捕捉更多细节信息。这些局部特征可以关注到不同行人的身体部位、纹理、颜色等, 从而增强模型的鲁棒性提升模型识别的准确率和适应性。经主干网络提取得到 $N+1$ 个特征向量, 引用 TransReID 中 JPM 思想, 将全局 token 外的序列进行移位

和打乱, 得到新的序列表示为:

$$Z_{JPM}^{L-1} = JPM([z_0^{L-1}; z_1^{L-1}, z_2^{L-1}, \dots, z_N^{L-1}]) \quad (4)$$

设经过多粒度特征提取后得到的局部特征划分索引表示为 $m_i = \left\lfloor \frac{N+1}{i} \right\rfloor$ ($i \in 4, 6$), 则各局部特征表示为:

$$L_a^i = \begin{cases} JPM([z_0^{L-1}; z_1^{L-1}, \dots, z_{k-1}^{L-1}]) \\ JPM([z_k^{L-1}; z_{k+1}^{L-1}, \dots, z_{2k-1}^{L-1}]) \\ \dots \\ JPM([z_{(i-1)k}^{L-1}; z_{(i-1)k+1}^{L-1}, \dots, z_k^{L-1}]) \end{cases} \quad (5)$$

式中: $i = 4, 6$; $a = 1, 2, \dots, i$ 。

1.3 双流网络

本文设计的双流网络由纯 Transformer 结构组成, 其中包含单分支浅层网络和单分支深层网络, 结构如图 2 所示。

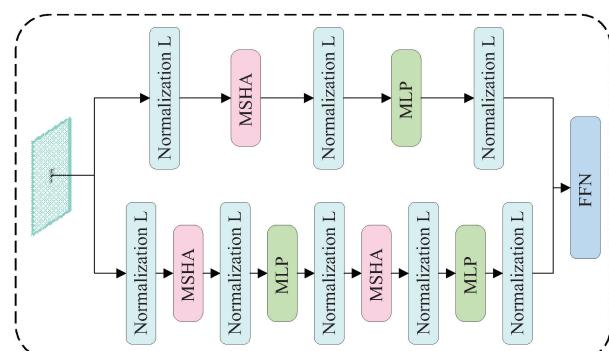


图 2 双流网络框架

Fig. 2 The dual-stream network framework

设双流网络处理后特征为 \mathbf{F}_a^i , 双流网络表示为 $T_b(\cdot)$ ($b \in (1, 2)$), 处理后的特征表示为:

$$\mathbf{F}_{b(a)}^i = T_b(\mathbf{L}_a^i) \quad (6)$$

式中: b 表示双流网络的分支; $i=4,6$; $a=1,2,\dots,i$ 。

将处理后的特征进行特征加权融合, 设融合后的特征为 \mathbf{F} 表示为:

$$\mathbf{F} = \lambda \mathbf{F}_{1(a)}^i + (1 - \lambda) \mathbf{F}_{2(a)}^i \quad (7)$$

式中: λ 是一个权重系数, 用于平衡两个分支的贡献。

1.4 损失函数

为了提高模型的分类能力和特征的判别性, 本文损失函数采用标准交叉熵损失和三元组损失^[21], 其中全局特征和两个局部特征分别用 L_{CE} 和 L_{tri} 监督训练。交叉熵损失可表示为:

$$L_{CE}(s, y) = -\log\left(\frac{e^{sy}}{\sum_{j=1}^K e^{sj}}\right) \quad (8)$$

式中: $s = [s_1, s_2, \dots, s_K]$ 为一个样本输出向量; y 为真实类别标签; sy 为正确预测得分; K 为类别总数; $\frac{e^{sy}}{\sum_{j=1}^K e^{sj}}$

为 y 的预测概率。

三元组损失函数可表示为:

$$L_{tri} = \max(d(a_i, p_i) - d(a_i, n_i) + \alpha, 0) \quad (9)$$

式中: a_i 表示为锚点样本; p_i 为正样本; n_i 为样本; $d(a_i, p_i)$ 为锚点与正样本之间的距离; $d(a_i, n_i)$ 为锚点与负样本之间的距离; α 是控制距离差参数; $\max(\cdot, 0)$ 表示输出值不小于 0。

最终损失函数可表示为:

$$L = L_{CE}(s, y) + L_{tri} + \frac{1}{k_1} \sum_1^{k_1} (L_{CE}(s, y) + L_{tri}) + \frac{1}{k_2} \sum_1^{k_2} (L_{CE}(s, y) + L_{tri}) \quad (10)$$

2 实验与分析

2.1 数据集与评价指标

本文在 4 个公开的行人重识别数据集上进行了实验, 包括 Market-1501^[22]、DukeMTMC-reID^[23]、Occluded-Duke 和 MSMT17^[24]。其中, Occluded-Duke 数据集是从 DukeMTMC-reID 数据集中处理得到的, 主要针对遮挡情况下的行人重识别问题。各数据集的使用摄像头数量(cam)、总行人 ID、训练集(train) 和测试集(test) 的行人 ID 及总图像数量(image) 信息如表 1 所示。

本文采用累积匹配特征(CMC) 和均值平均精度(mAP) 作为算法性能评估指标。CMC 关注查询图像的匹配结果, 衡量正确匹配在排名前 k 的数量, 直观反映模型的快速匹配能力; 而 mAP 则通过计算每个查询的平均精度, 综合评估模型的检索性能。

表 1 4 个公开数据集基本信息

Table 1 Basic information of four public datasets

数据集	Cam	ID	Train	Test	Images
Market1501	6	1 501	751	750	32 668
DukeMTMC	8	1 402	702	702	36 411
Occluded	8	1 402	702	519	35 489
MSMT17	15	4 101	1 041	3 060	126 441

2.2 实验设置

本文实验在 64 位 Windows 操作系统下进行, 使用单张 NVIDIA RTX4090 显卡, CUDA11.8, Pytorch2.3.1, python3.8。本文采用 ViT-base 作为主干网络, 输入图像尺寸为 384×128。在训练过程中, 采用 SGD 优化器进行参数更新, 设置动量衰减因子为 0.9, 权重衰减为 1×10^{-4} , 初始学习率为 0.008, 训练过程中余弦衰减, 迭代次数为 120 次。

2.3 与最先进方法比较

为验证本文 MGDSN 模型的有效性, 针对遮挡行人重识别任务在 Occluded-Duke 数据集上进行了相关实验, 分析了本文算法与其他先进方法的对比, TransReID^[13] 结果为复现结果, 对比结果如表 2 所示。

表 2 在 Occluded-Duke 数据集上比较

Table 2 Comparison on the Occluded-Duke dataset

(%)

Backbone	方法	mAP	Rank-1
CNN	PCB ^[9]	33.7	42.6
	PGFA ^[8]	37.3	51.4
	HORID ^[5]	43.8	55.1
	PAT ^[16]	53.6	64.5
	DRL-Net ^[25]	53.9	65.8
	RTGAT ^[26]	50.1	61.0
	MVIP ^[27]	57.3	68.6
	MSOSNet ^[28]	57.6	68.6
	TransReID ^[13]	59.6	68.1
	PFD ^[17]	60.1	67.7
ViT	MAT ^[18]	58.8	66.2
	RFMT ^[29]	60.6	69
	DBAAT ^[30]	57.1	63.3
	AAformer ^[14]	58.2	67.1
	PAFormer ^[31]	60.4	66.4
	SSSC-TransReID ^[32]	61.0	69.2
	MGDSN(本文)	61.3	68.3

由表 2 所知, MGDSN 在 Occluded-Duke 数据集上的表现最佳。在 mAP 指标上, 相较于以 CNN 为主干网络的研究, MGDSN 比经典模型 PCB 高 27.6%, 比最新研究的模型 MSOSNet 高 3.7%; 相较于以 ViT 为主干网络的研究, MGDSN 比效果最好的模型 SSSC-TransReID 高 0.3%。在 Rank-1 指标上, 本文方法也接近 SOTA 水平, 较 PCB 提升了 25.7%。本文认为, 在遮挡行人重识别任

务中,部分遮挡会导致关键信息丢失,传统方法未能充分挖掘剩余信息,本文设计的模型能够深入挖掘剩余的关键信息,从而有效提升了模型的整体性能和鲁棒性。

表 3 在 Market-1501、DukeMTMC-reID 和 MSMT17 数据集上比较
Table 3 Comparison on the Market-1501, DukeMTMC-reID, and MSMT17 datasets (%)

Backbone	方法	Market1501		MSMT17		DukeMTMC-reID	
		Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
CNN	PCB ^[9]	92.3	77.4	68.2	40.4	88.7	78.4
	HOReID ^[5]	94.2	84.9	—	—	86.9	75.6
	PAT ^[16]	95.4	88.0	—	—	88.8	78.2
	DRL-Net ^[25]	94.7	86.9	78.4	55.3	88.1	76.6
	RTGAT ^[26]	95.3	88.2	—	—	85.6	74.3
	MVIP ^[27]	95.3	87.9	83.9	61.4	91.9	80.9
	MSOSNet ^[28]	95.5	88.4	81.2	59.6	91.6	79.2
ViT	TransReID ^[13]	95.0	88.8	84.6	66.6	90.4	81.8
	PFD ^[17]	95.5	89.6	82.7	65.1	90.6	82.2
	MAT ^[18]	95.3	89.4	—	—	89.3	81.8
	DCAL ^[19]	94.7	87.5	83.1	64.0	89.0	80.1
	RFMT ^[29]	95.4	89.2	—	—	90.5	82.4
	DBAAT ^[30]	95.1	88.8	—	—	90.6	82.0
	AAformer ^[14]	95.4	88.0	84.4	65.6	90.1	80.9
MGDSN(本文)		95.2	89.0	84.5	66.8	91.1	82.5

由表 3 所知,本文设计的 MGDSN 在 MSMT17 数据集上效果最优,相较于以 CNN 为主干网络的最新且性能最好的模型 MVIP, mAP 和 Rank-1 指标分别提升了 5.4% 和 0.6%;相较于以 ViT 为主干网络的模型 AAformer,本文方法的 mAP 提升了 1.2%,Rank-1 相差无几;与经典模型 PCB 方法相比,mAP 和 Rank-1 分别提升了 26.4% 和 16.3%。MGDSN 在 DukeMTMC-reID 数据集上也表现最佳,相比 MVIP 模型,mAP 指标提升了 1.6%、Rank-1 指标下降 0.8%;相较于以 ViT 为主干网络的所有模型,本文方法的 mAP 和 Rank-1 都达到最佳性能,比次优模型 RFMT 的 mAP、Rank-1 分别提高了 0.1%、0.6%。MGDSN 在 Market1501 数据集上也接近目前 SOTA 结果。对比大部分经典算法,MGDSN 整体性能表现最优,证明了其充分挖掘剩余关键信息的重要性和本方法的适用性。

2.4 消融实验

为进一步验证各模块在 MGDSN 方法中的贡献,首先在 Occluded-Duke 数据集上对多细粒度局部特征提取策略进行实验验证。在相同实验设置下以复现 TransReID 结果为基础进行验证,Baseline 表示复现 TransReID 结果,其中全局特征分为 4 个局部特征,经过双流网络的浅层单分支网络处理;m1 表示将全局特征分为 6 个局部特征,经过双流网络的浅层单分支网络处理;m2 表示将全局特征同时划分为 4 个和 6 个局部特征,经过双流网络的浅层单分支网络处理;m3 表示将 4 个局部特征经过双流网络处理;m4 表示将 6 个局部特征经过双

为进一步验证本文设计模型的有效性,本文在 Market-1501、DukeMTMC-reID 和 MSMT17 数据集上展开相关实验,实验结果如表 4 所示。

表 4 不同细粒度实验结果

Results of experiments with varying granularities (%)				
方法	mAP	Rank-1	Rank-5	Rank-10
Baseline	59.6	68.1	82.2	86.3
m1	60.6	68.5	82.6	87.3
m2	60.6	68.1	82.4	86.7
m3	60.4	68.1	83.1	87.1
m4	60.6	68.4	81.9	86.3
m5	60.5	67.8	82.2	87.2
m6	61.3	68.3	82.8	86.2

由表 4 可知,在模型 m1 中,mAP、Rank-1、Rank-5 和 Rank-10 分别提升 1%、0.4%、0.4% 和 1%。本文认为,将全局特征划分为 6 个局部特征时,单个局部区域的尺寸减小,每个特征更加专注于局部细节,使网络能够捕捉到更为微小且重要的特征(如衣物、鞋子和背包等边缘细节)。在模型 m2 中,相较于基线模型,mAP、Rank-5 和 Rank-10 分别提升 1%、0.2% 和 0.4%,Rank1 保持不变;与 m1 模型相比,mAP 保持不变,Rank-1、Rank-5 和 Rank-10 则分别下降 0.4%、0.2% 和 0.6%。本文认为,在基线

模型的基础上,将全局特征同时划分为 4 和 6 个局部特征时,模型的信息捕捉能力有所增强,基于模型 m1,Rank 指标下降主要归因于以下两点:首先,将 4 和 6 个局部特征分别通过相同的浅层单一网络处理,这限制了特征提取能力仅停留在低层次,导致一定程度的低级特征冗余;其次,不同大小的局部特征也可能引发部分特征信息的冗余。在模型 m3、m4 中,较基线模型 m3 的 Rank-1 保持不变,mAP、Rank-5 和 Rank-10 分别提升 0.8%、0.9% 和 0.8%。m4 的 Rank-10 保持不变,mAP 和 Rank-1 分别提升 1% 和 0.3%。以上结果表明,本文设计的双流网络在充分挖掘局部特征信息方面具有较明显优势。在模型 m6 中,相较基线模型, mAP、Rank-1 和 Rank-5 分别提升 1.7%、0.2% 和 0.6%;相较 m2 模型, mAP、Rank-1 和 Rank-5 分别提升 0.7%、0.2% 和 0.4%。以上结果均证明多细粒度双流网络在高级和低级特征的结合提取方面展现出较显著的优势;相较于 m5 模型, mAP、Rank-1 和 Rank-5 分别提升 0.8%、0.5% 和 0.6%,证明加入特征权重融合模块的有效性。

其次在 Occluded-Duke 数据集上对双流网络的单分支浅层网络和单分支深层网络进行实验验证。与多细粒度局部特征提取策略实验相同,Baseline、m1 和 m2 模型表示双流网络的浅层单分支网络进行处理;m6 表示最终模型;p1 表示将全局特征分为 4 个局部特征,经过双流网络的深层单分支网络处理;p2 表示将全局特征分为 6 个局部特征,经过双流网络的深层单分支网络处理;p3 表示将全局特征同时划分为 4 个和 6 个局部特征,经过双流网络的深层单分支网络进行处理,且将两个细粒度特征使用特征权重进行融合,表 5 是双流网络各模块实验结果。

表 5 双流网络各模块实验结果

Table 5 Experimental results of each module of the dual-stream network (%)

方法	mAP	Rank-1	Rank-5	Rank-10
Baseline	59.6	68.1	82.2	86.3
m1	60.6	68.5	82.6	87.3
m2	60.6	68.1	82.4	86.7
p1	59.9	66.3	81.5	86.7
p2	60.6	68.4	82.5	87.5
p3	60.7	68.2	82.6	86.8
m6	61.3	68.3	82.8	86.2

由表 5 可知,在模型 p1 中较 Baseline 模型,mAP 和 Rank-10 分别提升 0.3% 和 0.4%;在模型 p2 中较 m1 模型,mAP 保持不变,Rank-10 提升 0.2%。本文认为,更深的模型通常能够学习更复杂的特征表示,因为每一层都可以在前一层的基础上提取更高层次的特征,因此适当的深层次网络效果会更好些;在模型 p3 中较 m2 模型,

mAP、Rank-1,Rank-5 和 Rank-10 分别提升 0.1%、0.1%、0.2% 和 0.1%。本文认为由 p1,p2 和 p3 实验整体证明,适当的深层次网络提取特征性能优于浅层次网络。m6 模型较 m2 和 p3 模型,证明同时使用深、浅层网络可增强对关键特征的提取能力。

为进一步验证深、浅层网络对特征捕获能力的差异,对深、浅层网络和整体双流网络提取的特征图进行了可视化,特征图可视化结果如图 3 所示。

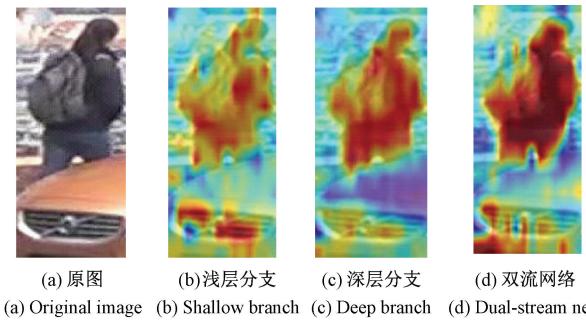


图 3 特征图可视化结果

Fig. 3 The dual-stream network framework

对比图 3(b)、(c)可知,深层次网络能较好的提取关键特征。由图 3(d)可知,双流网络提取关键特征的效果最好。

3 结 论

针对遮挡行人重识别任务中不能充分挖掘剩余关键特征信息和全局-局部特征联合建模不足等问题,本文提出了基于多细粒度双流网络的行人重识别方法。结合全局-局部协同机制构建层次化特征表示体系,通过捕获不同细节和整体信息,增强特征的多样性和表达能力;设计的双流网络和特征权重融合模块,通过浅层网络提取低级特征,深层网络提取高级特征,形成互补性特征表达,充分挖掘剩余关键信息,最大程度提升模型的性能、鲁棒性和泛化能力。在 4 个经典行人重识别数据集上,实验证明本文所提 MGDSN 模型在充分挖掘剩余关键特征信息等方面优于 SOTA 模型。在未来的工作中,将考虑建立遮挡预测网络,通过生成遮挡掩码定位图像中被遮挡的区域(如背包、遮挡物等),从而提升模型对有效特征的辨别能力。

参考文献

- [1] 钱亚萍,王凤随,熊磊. 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[J]. 电子测量与仪器学报, 2023, 37(1): 106-115.
QIAN Y P, WANG F S, XIONG L. Unsupervised person reidentification method based on local refinement multi

- branch and global feature sharing [J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(1) : 106-115.
- [2] ZHANG X, YAN Y, XUE J H, et al. Semantic-aware occlusion-robust network for occluded person re-identification [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 31(7) : 2764-2778.
- [3] 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. Pedestrian re-identification algorithm based on pose estimation and feature fusion [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(4) : 187-194.
- [4] HE B, LI J, ZHAO Y, et al. Part-regularized near-duplicate vehicle re-identification [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 3992-4000.
- [5] WANG G, YANG S, LIU H, et al. High-order information matters: Learning relation and topology for occluded person re-identification [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 6449-6458.
- [6] CHEN B H, DENG W H, HU J N. Mixed high-order attention network for person re-identification [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 371-381.
- [7] TAY C P, ROY S, YAP K H. Aanet: Attribute attention network for person re-identifications [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 7134-7143.
- [8] MIAO J, 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, 2019: 542-551.
- [9] 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 (ECCV), 2018: 480-496.
- [10] ZHOU S, WU J, ZHANG F, et al. Depth occlusion perception feature analysis for person re-identification [J]. Pattern Recognition Letters, 2020, 138: 617-623.
- [11] 贺晓东, 王春艳, 孙昊, 等. 基于局部特征与视点感知的车辆重识别算法 [J]. 仪器仪表学报, 2022, 43(10) : 177-184.
- HE X D, WANG CH Y, SUN H, et al. Local-features and viewpoint-aware for vehicle re-identification [J]. Chinese Journal of Scientific Instrument, 2022, 43(10) : 177-184.
- [12] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale [C]. The 9th International Conference on Learning Representations, 2021.
- [13] HE S, LUO H, WANG P, et al. TransReID: Transformer-based object re-identification [C]. The IEEE/CVF International Conference on Computer Vision, 2021: 14993 - 15002.
- [14] ZHU K, GUO H, ZHANG S, et al. AAformer: Auto-aligned transformer for person re-identification [J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(12) : 17307-17317.
- [15] BIAN Y, LIU M, WANG X, et al. Occlusion-aware feature recover model for occluded person re-identification [J]. IEEE Transactions on Multimedia, 2024, 26: 5284-5295.
- [16] LI Y L, HE J F, ZHANG T Z, et al. Diverse part discovery: Occluded person re-identification with part-aware transformer [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 2898-2907.
- [17] WANG T, LIU H, SONG P, et al. Pose-guided feature disentangling for occluded person re-identification based on transformer [C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(3) : 2540-2549.
- [18] ZHOU M, LIU H Y, LV Z K, et al. Motion-aware transformer for occluded person re-identification [J]. ArXiv preprint arXiv:2202.04243, 2022.
- [19] ZHU H W, KE W J, LI D, et al. Dual cross-attention learning for fine-grained visual categorization and object re-identification [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 4692-4702.
- [20] YAN Q S, GONG D, ZHANG Y N. Two-stream convolutional networks for blind image quality assessment [J]. IEEE Transactions on Image Processing, 2018, 28 (5) : 2200-2211.
- [21] HERMANS A, BEYER L, LEIBE B. In defense of the triplet loss for person re-identification [J]. ArXiv preprint arXiv:1703.07737, 2017.
- [22] 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, 2015: 1116-1124.
- [23] RISTANI 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. Cham: Springer International Publishing, 2016: 17-35.
- [24] WEI L H, ZHANG S L, GAO W, et al. Person transfer GAN to bridge domain gap for person re-identification[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 79-88.
- [25] JIA M X, CHENG X H, LU S J, et al. Learning disentangled representation implicitly via transformer for occluded person re-identification[J]. IEEE Transactions on Multimedia, 2022, 25: 1294-1305.
- [26] HUANG M Y, HOU C P, YANG Q Y, et al. Reasoning and tuning: Graph attention network for occluded person re-identification [J]. IEEE Transactions on Image Processing, 2023, 32: 1568-1582.
- [27] DONG N, YAN S L, TANG H, et al. Multi-view information integration and propagation for occluded person re-identification[J]. Information Fusion, 2024, 104: 102201.
- [28] ZHANG Y Z, YANG Y H, KANG W L, et al. Multi-scale occlusion suppression network for occluded person re-identification[J]. Pattern Recognition Letters, 2024, 185: 66-72.
- [29] YANG Z Z, CHEN Y N, YANG Y P, et al. Robust feature mining transformer for occluded person reidentification[J]. Digital Signal Processing, 2023, 141: 104166.
- [30] LU Y H, JIANG M Z, LIU Z, et al. Dual-branch adaptive attention transformer for occluded person reidentification[J]. Image and Vision Computing, 2023, 131: 104633.
- [31] JUNG H, LEE J, YOO J, et al. Paformer: Part aware transformer for person reidentification[J]. ArXiv preprint arXiv:2408.05918, 2024.
- [32] JI Z J, CHENG D L, FENG K. Exploring stronger transformer representation learning for occluded person re-identification [J]. ArXiv preprint arXiv: 2410.15613, 2024.

作者简介



宋晓勇,2023年于北方民族大学获得学士学位,现为宁夏大学硕士研究生,主要研究方向为模式识别与图像处理、行人重识别。

E-mail: sxy55552022@163.com

Song Xiaoyong received his B. Sc. degree from North Minzu University in 2023. He is now a M. Sc. candidate at Ningxia University. His main research interests include pattern recognition and image processing and person re-identification.



孙学宏(通信作者),2016年于北京邮电大学获得博士学位,现为宁夏大学教授、博士生导师,主要研究方向为新一代无线通信技术与网络、Pol-InSAR 图像信号处理、大数据分析与处理。

E-mail: sunxh@nxu.edu.cn

Sun Xuehong (Corresponding author) received his Ph. D. degree from Beijing University of Posts and Telecommunications in 2016. He is now a professor and Ph. D. supervisor in Ningxia University. His main research interests include new-generation wireless communication technologies and networks, Pol-InSAR image signal processing, and big data analysis and processing.