文章编号：2095-2163(2024)01-0095-07 中图分类号：TP391.41 文献标志码：A

基于视频预训练和注意力特征融合的行人重识别方法

南 灏，吴丽君
（福州大学 物理与信息工程学院，福州 350116）

摘 要：行人重识别是跨摄像头追踪的关键环节之一，主流方法多采用 ImageNet 进行预训练，忽视了数据集的域间差异，且以结构庞大的多分支模型居多，模型复杂度较高。本文设计一种行人重识别方法，采用基于原始视频带噪声标签参与监督的方式进行预训练，减少域间差异以提升特征表达能力；以基于注意力的特征融合方式取代残差网络的跳接映射，增强网络的特征提取能力；在网络中嵌入坐标注意力机制，在低复杂度的情况下强化关键特征，抑制低贡献特征；采用随机擦除对输入数据做数据增强以提高泛化能力，联合分类损失、三元组损失和中心损失函数对网络进行监督训练。在公开数据集 Market-1501 和 Duke-MTMC 上完成了消融实验，与主流方法对比实验表明本方法在不需要复杂多分支逻辑结构的前提下，仍可达到较高的精度。

关键词：行人重识别；预训练；残差网络；特征融合；注意力机制

Person re-identification based on video pre-training and attentional feature fusion

NAN Hao, WU Lijun
(College of Physics and Information Engineering, Fuzhou University, Fuzhou 350116, China)

Abstract: Person re-identification is one of the key steps in cross camera tracking. Most mainstream methods use ImageNet for pre training, ignoring the difference between domains of data sets, and most of them are multi branch models with large structures, which have high complexity. In this paper, a pedestrian re recognition method is designed, which adopts the method of pre training based on the original video with noisy tags to participate in the supervision, and reduces the difference between domains to improve the feature expression ability; The attention based feature fusion method is used to replace the jump mapping of the residual network, which enhances the feature extraction ability of the network; Embed coordinate attention mechanism in the network to strengthen key features and suppress low contribution features in the case of low complexity; At the same time, random erasure is used to enhance the input data to improve the generalization ability. The network is supervised by combining classification loss, triple loss and central loss functions. Ablation experiments have been completed on Market-1501 and Duke MTMC public datasets. The comparison experiments with mainstream methods show that this method can still achieve high accuracy without complex multi branch logic structure.

Key words: person re-identification; pre-training; residual network; feature fusion; attention mechanism

0 引 言

社会公共安全的重要性凸显，智能监控与追踪在智慧城市建设中扮演着越来越重要的角色。在跨摄像头监控追踪领域，受监控视频像素低、实际环境复杂多变等因素影响，人脸识别技术难以提供可靠的检测和识别，因此基于整个行人图像特征的行人重识别方法受到了广泛关注。因光照、遮挡、不同摄像机参数、行人姿态和衣着变化等一系列因素的影响，行人重识别工作仍然面临诸多挑战。

传统行人重识别方法通过人工操作进行特征提取，如直方图或局部最大特征等算法，除此之外还提出了许多基于度量学习的方法，如 Koestinger[1] 等人提出利用大规模度量提升模型效果。随着数据量的日益剧增和深度学习的兴起，基于深度神经网络可提取更具判别性的深度特征，2014 年 Li 等[2] 首次利用深度学习解决行人重识别问题，将行人重识别看作目标分类和检索任务，以卷积神经网络为基础

基金项目：福建省自然科学基金(2022H0008，2021J01580)；福州市科技计划项目(2021-P-030，2021-P-059)。

作者简介：南 灏(1998-)，男，硕士研究生，主要研究方向：计算机视觉、行人重识别。

通讯作者：吴丽君(1984-)，女，博士，副教授，主要研究方向：计算机视觉。Email：lijun.wu@fzu.edu.cn

收稿日期：2023-01-03

架构的深度学习方法在行人重识别研究中不断取得突破。研究表明此类方法甚至在大量情形下超过了人类水平[3]。2018 年 Sun 等[4]提出了经典的 PCB+RPP(Part-based Convolutional Baseline)算法,将行人图像分块,融合各块特征提高判别性。然而该方法分别提取不同分块特征且需要结合空间注意力对分割界限进行优化,网络并非端到端,增加了模型复杂度。Wang 等[5]在融合局部特征的思想上,设计了 MGN(Multiple Granularity Network)算法,同时考虑全局特征和多粒度局部特征,但需要维护全局和局部分支,网络结构仍然复杂;Luo 等[6]注意到仅考虑全局特征也可以达到优秀的效果,提出了 BoT(Bag of Tricks)算法,并为重识别领域总结了许多训练策略;She 等[7]提出在残差网络上引入 CBAM(Convolutional Block Attention Mechannism)注意力机制模块进行重识别,将通道和空间注意力同时嵌入以提升特征表示能力,但 CBAM 采用 7×7 卷积提取局部信息,始终无法实现信息更大感受野的长范围依赖。

除了针对网络结构的优化外,如何使用预训练模型对网络权重进行有效的初始化,以优化网络收敛情况,也成为了一个研究方向。已有工作习惯使用 ImageNet 进行预训练,由于 ImageNet 的图片和行人重识别任务中的图片有着很大的域间差异,此种方法并不会给网络带来提升,因此 Fu 等[8]基于 7 万个街景视频,利用 YOLO-V5 作为检测器搜集行人图片,制作了大型数据集 LUperson,结合数据增强,以动量对比学习的方式进行无监督预训练,提升了网络精度;Yang[9]等又提出了内部身份正则化,并引入全局一致性约束来继续优化无监督框架,但这些方法皆没有利用数据集原始视频的时间和空间相关性;Fu 等[10]在上述成果基础上提出了利用原始视频来训练的 PNL(Pre-training with Noisy Labels)方法,对此类信息加以利用。

基于上述分析,本文设计基于 ResNet50 的行人重识别方法,利用 LUperson 数据集原始视频的带噪声标签训练初始化主干网络,促进网络收敛;以基于注意力的特征融合模块替换残差网络中的恒等映射残差跳接,增强网络特征提取能力;最后,嵌入坐标注意力 CA(Coordinate Attention)[11]模块,对特征重新赋予权重,进一步提升特征判别效果。本网络不含多分支的复杂结构,仅利用网络输出的全局特征。在公开数据集 Market-1501 和 Duke-MTMC 上进行实验,验证此方法的合理性和可行性。

1 网络模型

本文整体网络模型如图 1 所示。首先,采用基于原始视频搜集的数据集 LUperson-NL,以基于 PNL 预训练框架的训练方法对主干网络进行初始化;其次,引入改良后的注意力特征融合 AFF(Attentional Feature Fusion)[12]模块并进行优化,取代残差块中的跳接映射,不再将输出和残差跳接直接相加;最后,在网络中融入坐标注意力 CA 模块,对特征输出重新赋予权重,进一步优化网络的特征提取能力,提升判别性。

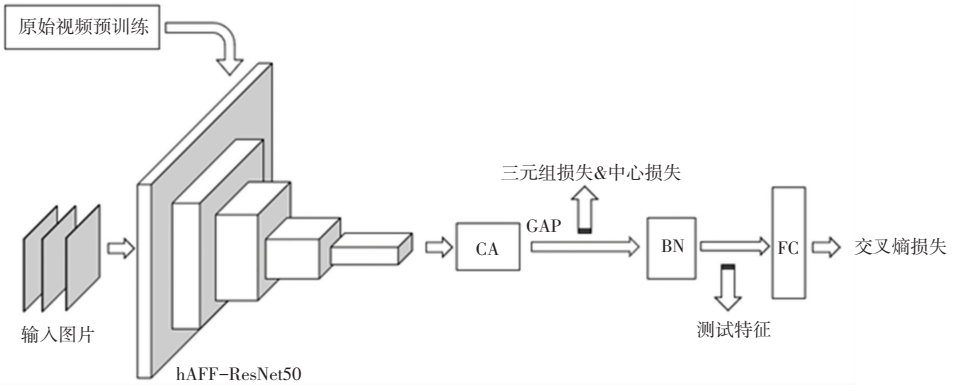

图 1 本文网络结构

Fig. 1 Structure diagram of the proposed method

1.1 原始视频预训练方法

本文采用基于 PNL 框架的预训练方法来预训练主干网络。为了利用原始视频的时间和空间相关性,数据集 LUperson-NL 是一系列行人轨迹序列,以先进目标跟踪算法从数据集 LUPerson 的原始视频中获取,并加以姿态筛选和采样过滤,对每个轨迹赋予相同身份,从 21 000 个场景中收集的 1 000 万张大量不同身份的图像组成。针对此种大规模含噪

声标签的数据集，采用双分支对比网络结构，采取不同数据增强策略，通过不共享权重的两个相同编码器到输出特征，基于动量对比学习，将无监督常用的对比损失方法，优化为融合分类监督损失、原型对比损失和标签引导对比损失为一体的弱监督训练方法。

1.1.1　分类监督损失

主分支的输出特征 q_i 通过全连接层及 Softmax 分类器，得到特征所属于对应标签类别的概率 p_i，并利用标签 y_i 进行分类损失的计算，损失函数如式(1)所示：

$$L_c^i = -\log(p_i[y_i]) \tag{1}$$

其中，i 表示对应的类别。

标签 y_i 随网络进程进行更新，具体体现在原型对比损失模块。本文在分类损失基础上加入标签平滑，进一步降低标签噪声的影响。

1.1.2　原型对比损失

在每次训练过程中，对 K 个标签类别分别维护一个质心特征向量 $\boldsymbol{c}_k$，称之为每个类的原型(prototype)，并在训练迭代中遵循动量机制进行更新。基于所维护的原型，首先利用各类别原型计算分支输出特征 q_i 和原型的相似性分数 s_i^k，如式(2)所示：

$$s_i^k = \frac{\exp(q_i \cdot \boldsymbol{c}_k/\tau)}{\sum_{k=1}^{K}\exp(q_i \cdot \boldsymbol{c}_k/\tau)} \tag{2}$$

其中，τ 为设置的超参数。

其次，利用所得到的对应 i 类别的相似性分数 s_i^k 和分类概率 p_i 融合得到判别参数，若此参数大于设定阈值，则视为此特征距离 i 类别原型更接近，将标签更新为对应类别，否则不更新标签。利用原型对比损失函数来约束每一个样本，使其距离所属类别的中心更近，具体函数如式(3)所示，此时 y_i 为更新后的标签。

$$L_{pro}^i = -\log\frac{\exp(q_i \cdot c_{y_i}/\tau)}{\sum_{j=1}^{K}\exp(q_i \cdot c_j/\tau)} \tag{3}$$

1.1.3　标签引导对比损失

样本间的对比学习一直作为自监督学习的有效手段，PNL 方法在利用双分支输出间样本对比损失基础上加入标签引导，在训练过程中除了维护包含正负特征对的队列外，还同时维护样本所对应的标签 y_i 来辅助判别样本正负，避免误判。具体标签引导对比损失如式(4)所示：

$$L_{lgc}^i = -\frac{1}{|P(i)|}\cdot \log\frac{\sum_{k^+\in P(i)}\exp(q_i \cdot k^+/\tau)}{\sum_{k^+\in P(i)}\exp(q_i \cdot k^+/\tau) + \sum_{k^-\in N(i)}\exp(q_i \cdot k^-/\tau)} \tag{4}$$

其中，$P(i)$ 表示正特征集；$N(i)$ 表示负特征集；k^+ 表示正特征集中的特征；k^- 表示负特征集的特征。

最终由 3 个损失模块共同监督训练，总体损失函数，如式(5)所示：

$$L_i = \lambda_1 L_c^i + \lambda_2 L_{pro}^i + \lambda_3 L_{lgc}^i \tag{5}$$

考虑到训练初期标签噪声的影响，本文设置 $\lambda_1 = 0.995$，$\lambda_2 = \lambda_3 = 1$，在此联合函数的监督下，在数据集 LUperson-NL 上完成对主干网络的预训练。

1.2　特征融合跳接模块

残差跳接将残差块的输入直接加至对应的输出部分，使网络只学习恒等映射以外的残差部分，解决梯度爆炸和梯度消失的问题。本文为应对其伴生的信息冗余、特征利用有限等问题，引入基于注意力的特征融合模块并对其进行改良，来取代残差跳接的直接特征对应相加。改良后的多尺度注意力模块(Multi Scale-hard Channel Attention mechanism，MS-hCAM)结构如图 2(a)所示，基于此实现的注意力特征融合模块(hard Attentional Feature Fusion，hAFF)如图 2(b)所示，其所构成的残差块结构如图 2(c)所示。

在多尺度注意力模块中，输入特征经过两个分支，全局分支类似 SE(Squeeze Excitation)机制，首先通过全局平均池化获取全局特征，经过一次卷积降维将通道数降至 C/τ，其中 τ 为衰减参数，对于 ResNet50 本文将其设置为 16，之后经过激活函数并通过一次卷积还原通道数。另一侧的局部分支，相较于全局分支去掉全局平均池化操作，而保留细粒度的局部特征，其后的操作与全局分支相同。在完成两个分支的操作后，将全局分支 1×1 尺度特征通过广播至局部分支输出特征进行相加，并通过 Sigmoid 函数得到最终的注意力权重，将其赋予原输入特征。为降低参数量，整个过程中的卷积操作均采用逐点卷积而非使用大小不同的卷积核来适应不同尺度。为了降低 *ReLU* 以降低过度特征屏蔽带来的影响，本文以 hardswish[13] 激活函数将其取代，具体如公式(5)，其中以 $ReLU6(x+3)/6$ 来拟合 swish[14] 激活函数中运算成本最高的 sigmoid 部分，在低运算前提下有效提升模块效果。

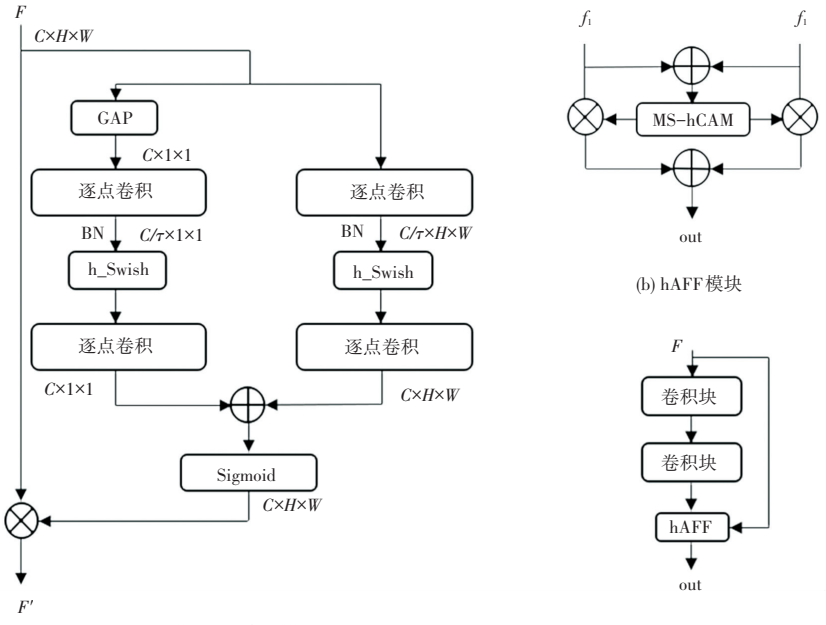

图 2 多尺度注意力模块、hAFF 模块以及 hAFF 残差块

Fig. 2 MS-hCAM module, hAFF module and hAFF residual block

$$hardswish(x) = x\frac{ReLU6(x+3)}{6} = \begin{cases} 0 & \text{if } x \leqslant -3 \\ x & \text{if } x \geqslant +3 \\ x * \dfrac{x+3}{6} & \text{otherwise} \end{cases}$$

注意力特征融合模块在上述多尺度注意力模块的基础上实现。对于特征 f_1 和 f_2，首先将两特征进行直接相加得到初始融合特征，作为多尺度注意力模块的输入，最终经过 Sigmoid 函数输出的权重本身作为 f_1 的权重，以 1 减去其对应元素所得的差作为 f_2 的权重，两特征分别逐元素与权重相乘后再次相加，得到最终的融合特征。应用于残差网络时，本文选择将跳接残差作为输入模块的 f_1，当前残差块的输出特征作为 f_2。

1.3 坐标注意力模块

为进一步强化所提取特征的判别性，本文在网络中嵌入坐标注意力机制，坐标注意力模块结构如图 3 所示。对于输入特征，采用类似坐标系结构的方式，分别进行 X 方向平均池化和 Y 方向平均池化，得到通道和尺寸维度分别为 $C \times H \times 1$ 和 $C \times 1 \times W$ 尺度的方向感知特征，此时将两特征进行空间维度级联和 1×1 卷积通道压缩，将通道数压缩至 C/τ，经过批归一化和非线性激活函数处理后再通过分割操作再次分为 $C \times H \times 1$ 和 $C \times 1 \times W$ 尺度的两个特征，将其分别输入 Sigmoid 函数获取 X 方向和 Y 方向权重，重新赋予输入特征得到最终的输出特征图。

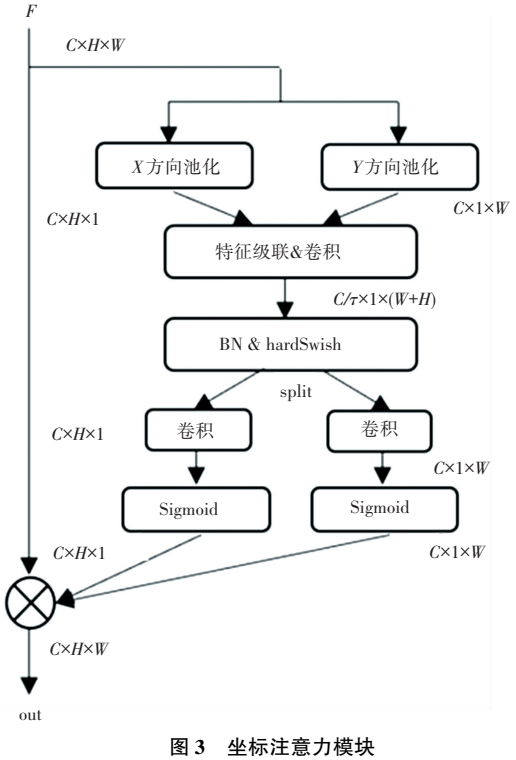

图 3 坐标注意力模块

Fig. 3 Coordinate attention module

相较于 CBAM 以 7×7 卷积的方式获取局部特

征，坐标注意力将 X、Y 方向池化后的特征进行级联后再进行卷积降维，建立感受野更大的长范围依赖。在特征分割操作后，重新通过卷积升维和 Sigmoid 生成两个方向的权重，对原特征图进行类似坐标方式的权重赋予。

卷积核尺度均为 1×1，保证了模块的低复杂度。本文选择在网络结构中加入坐标注意力强化所提取特征的表达能力。实验证明，在本文方法基础上，由于在各层间添加注意力，一定程度破坏了网络初始结构以及增加网络非线性，导致预训练效果降低，因此在网络最后一层输出后嵌入坐标注意力效果优于逐层嵌入坐标注意力。本文将主干网络最后一层卷积步长设置为 1，适当增大特征粒度，在特征输出和全局平均池化层间嵌入 CA 模块。

1.4　损失函数

本文网络模型采用融合交叉熵分类损失、三元组损失以及中心损失的联合损失函数对模型进行监督训练，主干网络输出特征图经过全局平均池化得到的特征向量用于三元组损失和中心损失，对此特征进行批归一化操作后，经过全连接层及 softmax 分类器获得输出，将其用于交叉熵分类损失监督。其中分类损失函数如式(6)所示：

$$L_{ce} = -\frac{1}{N}\sum_{i=1}^{N} q_i \log\left(p_i\right) \tag{6}$$

其中，N 表示样本类别数量；p_i 表示此样本属于对应类别的概率；q_i 由样本标签和 i 类别判定为 1 或 0，此处引入标签平滑来应对标签噪声。

三元组损失作为距离度量的有效损失函数，本文遵循 $P \times K$ 采样模式，在每个批次抽取 P 个身份的 K 张行人图片，从中抽取正负样本对，如式(7) 所示：

$$L_{triplet} = \left[d_p - d_n + \alpha\right]_+ \tag{7}$$

其中，d_p 表示与锚点特征为同一类别的正样本对的欧氏距离；d_n 表示负样本对的距离；α 表示边际参数，本文设置为 0.3，距离度量方式为欧氏距离。

为了解决三元组损失在随机抽样到正负样本对间距均较大时损失仍较小的问题，引入中心损失辅以监督，对类内间距加以约束，如式(8)所示：

$$L_c = \frac{1}{2}\sum_{j=1}^{s} \left\|f_i - c_{y_j}\right\|_2^2 \tag{8}$$

其中，s 表示一个迷你批次的大小；f_i 表示样本特征；c_{y_j} 则表示所维护的对应类别的中心特征。

总体损失函数如式(9)所示：

$$L = \lambda_1 L_{ce} + \lambda_2 L_{triplet} + \lambda_3 L_c \tag{9}$$

其中，本文设置 $\lambda_1 = \lambda_2 = 1$，$\lambda_3 = 0.005$，模型在上述联合损失监督下完成训练。

2　实验结果与分析

2.1　数据集和评价指标

本文在行人重识别领域最常用的两个公开数据集 Market-1501 和 DukeMTMC-reID 上完成了全部实验。Market-1501 数据集含有通过 5 个高分辨率一个低分辨率摄像头采集到的1 501个行人的图像，其中训练集占 751 个身份的 12 936 张图像，测试集占 750 个身份以及包含查询图像和待查询图像的 19 732张图像；DukeMTMC-reID 则是由 8 个高分辨率摄像头采集，含1 404个行人身份的36 411张图像，其中训练集占 702 个行人身份的16 522张图像，测试集为剩余 702 个身份的19 889张图像。上述两个数据集都有效囊括了不同背景、不同光照、不同清晰度、部分遮挡等现实因素。

在评价指标方面，传统行人重识别领域长期采用 CMC 曲线（$Rank-1$，$Rank-5$，$Rank-10$ 等）作为评价指标，但该指标不能有效表达同一类别存在多样本命中的情况，因此，后来引入了全类平均精度 mAP 来对模型性能进行更准确的评判。本文选择 mAP 和 $Rank-1$ 两项指标评价模型性能。

2.2　实验细节

本文实验环境为 Python3.6 以及 Pytorch1.2 版本，操作系统为 Ubuntu20.04，硬件环境为一块 NVIDIA GTX 1080Ti GPU 和一块 NVIDIA Tesla4 GPU。训练和测试时模型输入的图片尺寸为 256×128 像素，采用随机擦除和随机反转作为数据增强策略以防止过拟合。Batch-Size 设置为 64，选用 Adam 优化器，共训练 240 个 epochs，训练时对学习率采取 WarmUp 预热策略，预热后的基础学习率为 3.5e-4，在第 60 个 epoch 下降至 3.5e-5，在第 110 个 epoch 下降至 3.5e-6。

2.3　消融实验

本文在数据集 Market-1501 和 DukeMTMC-reID 上分别进行了对 PNL 预训练模块、注意力特征融合模块以及坐标注意模块的消融实验，并在数据集 Market-1501 上测试了不同坐标注意嵌入位置的效果。3 个模块的消融实验效果见表 1，表中 B 表示骨干网络，PRE、CA 和 hAFF 分别代表预训练、坐标注意力和注意力特征融合 3 个模块，从表 1 数据可以看出，本文采用原始视频预训练方法，骨干网络性能有较好提升，在 Market-1501 和 DukeMTMC-reID

上 *mAP* 分别提升了 3.4 和 4.1 个百分比，*Rank* - 1 也有着相应的提升。而注意力特征融合模块以及坐标注意力模块，在 ImageNet 预训练和原始视频预训练条件下均对网络性能有一定提升，在 ImageNet 预训练条件下，CA 模块在两个数据集上 *mAP* 分别提升了 0.7%以及 0.5%，*Rank* - 1 则在两个数据集上分别提升了 0.5%和 0.8%，hAFF 模块则使 *mAP* 分别提升了 0.2%和 0.4%，*Rank* - 1 分别提升了 0.4%和0.7%，两个模块共同作用时则可为 *mAP* 分别带来 1.5%和 2.1%的提升，*Rank* - 1 则分别提升了 0.9%和 2.7%；在引入预训练方法对网络初始权重进行激活的前提下，CA 模块在两个数据集上 *mAP* 分别提升 0.3%和 0.2%，*Rank* - 1 均提升 0.2%，hAFF 模块则使 *mAP* 分别提升 0.9%和 1.8%，*Rank* - 1 分别提升0.7%和 1.7%，两个模块共同作用时达到最好效果，*mAP* 分别为 91.5%和 83.5%，*Rank* - 1 分别为 96.3%和 92.3%。实验结果证明在绝大多数情况下，3 个模块及其分别组合均能带来较好的性能提升。

表 1 各模块消融实验结果

Table 1 The ablation experiment results of the modules

网络结构	Market-1501		DukeMTMC-reID	
	mAP	*Rank* - 1	*mAP*	*Rank* - 1
B	86.4	94.1	76.5	87.0
B+CA	87.1	94.6	77.0	87.8
B+hAFF	86.6	94.5	76.9	87.7
B+hAFF+CA	87.9	95.0	78.6	89.7
B+PRE	89.8	95.2	80.6	89.8
B+PRE+CA	90.1	95.4	80.8	90.0
B+PRE+hAFF	91.0	96.1	82.6	91.7
B+PRE+hAFF+CA	91.5	96.3	83.5	92.3

除了消融实验外，本文还在 Market-1501 数据集上对比了两种预训练情况下不同 CA 模块嵌入方式的影响，结果见表 2。由表 2 数据可以看出，在未引入预训练方法来激活网络初始权重时，在每个残差块间均嵌入 CA 模块效果优于在输出层添加 CA 模块，而在引入预训练方法后，嵌入过多 CA 模块一定程度破坏了预训练的结果，效果反而略逊于将其添加至输出层。因此最终本文将 CA 模块添加至骨干网络最后一个残差块和池化层之间。

2.4 网络总体性能对比

将本文方法的最优结果与其他方法的效果在数据集 Market-1501 和 DukeMTMC-reID 上进行对比，并附上对本文方法进行重排序(re-ranking)的最终结果，见表 3。与经典的 PCB+RPP 方法相比，本文方法在数据集 Market-1501 上 *mAP* 和 *Rank* - 1 上分别提升了 4.6%和 2.5%，在 DukeMTMC-reID 上分别提升了 13.9% 和 9.3%。BoT 方法是基于 ResNet50 并添加了许多训练技巧的基线方法，本文在训练技巧上也做了适当参考，与其相比本文方法在 Market-1501 上 *mAP* 和 *Rank* - 1 分别提升了 5.6%和 1.8%，在 DukeMTMC-reID 上 *mAP* 和 *Rank* - 1 分别提升了 6.7% 和 6.2%。其他方法如 ABD (Attentive But Diverse)、MGN 以及 SCR (Spatial-Channel Re-identification)算法均采用多分支结构来同时考虑全局和局部特征，本文方法在仅考虑全局特征的前提下，与上述多分支方法以及 LDS (Learning to Disentangle Scenes)、FlipReID 等先进方法相比仍可取得有竞争力的成果。

表 2 不同 CA 嵌入方式对网络性能的影响

Table 2 The influence of different CA embedding methods on the network performance

网络结构	Market-1501	
	mAP	*Rank* - 1
B	86.4	94.1
B+allCA	87.6	94.8
B+lastCA	87.1	94.6
B+PRE	89.8	95.2
B+PRE+allCA	90.0	95.3
B+PRE+lastCA	90.1	95.4

表 3 网络总体性能对比

Table 3 Comparison of the state-of-the-arts

方法	Market-1501		DukeMTMC-reID	
	mAP	*Rank* - 1	*mAP*	*Rank* - 1
PCB[4]	77.4	92.3	66.1	81.7
PCB+RPP[4]	86.9	93.8	69.2	83.3
BoT[6]	85.9	94.5	76.4	86.4
MGN[5]	86.9	95.7	78.4	88.7
BDB[15]	86.7	95.3	78.6	89.0
ABD[16]	88.3	95.6	78.5	89.0
SCR[17]	89.0	95.7	81.4	91.1
FlipReID[18]	89.6	95.5	81.5	90.9
LDS[19]	90.4	95.8	82.5	91.5
本文方法	**91.5**	**96.3**	**83.5**	**92.3**
本文方法+re-Ranking	**95.4**	**96.8**	**92.0**	**94.0**

3　结束语

本文设计了一种基于残差网络的行人重识别方法，采用基于带噪声标签原始视频序列数据集进行训练的方法对网络模型进行初始权重激活，优化网络的收敛过程；引入注意力特征融合模块取代残差网络中直接相加的残差跳接，提升网络的特征提取能力，降低信息冗余；在网络中融入坐标注意力模块，进一步强化所提取特征的判别性，整个网络不采用多分支结构，仅考虑全局特征。在行人重识别领域最常用的两个数据集上的所有实验结果表明，本文所设计的方法可提供有竞争力的精度。下一步的工作将在此基础上针对网络的泛化能力进行优化，以提升跨域情况下的重识别性能。

参考文献

[1] KOESTINGER M, HIRZER M, WOHLHART P, et al. Large scale metric learning from equivalence constraints [C]//2012 IEEE conference on computer vision and pattern recognition. IEEE, 2012: 2288-2295.

[2] 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. 2014: 152-159.

[3] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification [C]//Proceedings of the IEEE International Conference on Computer Vision, 2015: 1026-1034.

[4] 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.

[5] WANG G, YUAN Y, CHEN X, et al. Learning discriminative features with multiple granularities for person re-identification [C]//Proceedings of the 26th ACM international conference on Multimedia. 2018: 274-282.

[6] LUO Hao, GU Youzhi, LIAO Xingyu, et al. Bag of tricks and a strong baseline for deep person re-identification [J]. arXiv preprint arXiv:1903.07071, 2019.

[7] LI RX Y E O. Pedestrian re-identification combining random erasing and residual attention network[J]. Computer Engineering, 2022, 58(3): 215-221.

[8] FU D, CHEN D, BAO J, et al. Unsupervised pre-training for person re-identification [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 14750-14759.

[9] YANG Z, JIN X, ZHENG K, et al. Unleashing potential of unsupervised pre-training with intra-identity regularization for person re-identification [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 14298-14307.

[10] FU D, CHEN D, YANG H, et al. Large-scale pre-training for person re-identification with noisy labels[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 2476-2486.

[11] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13713-13722.

[12] DAI Y, GIESEKE F, OEHMCKE S, et al. Attentional feature fusion[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021: 3560-3569.

[13] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3 [C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 1314-1324.

[14] RAMACHANDRAN P, ZOPH B, LE Q V. Searching for activation functions[J]. arXiv preprint arXiv:1710.05941, 2017.

[15] DAI Z, CHEN M, GU X, et al. Batch dropblock network for person re-identification and beyond [C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 3691-3701.

[16] CHEN T, DING S, XIE J, et al. Abd-net: Attentive but diverse person re-identification [C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 8351-8361.

[17] CHEN H, LAGADEC B, BREMOND F. Learning discriminative and generalizable representations by spatial-channel partition for person re-identification [C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2020: 2483-2492.

[18] NI X, RAHTU E. Flipreid: closing the gap between training and inference in person re-identification [C]//2021 9th European Workshop on Visual Information Processing (EUVIP). IEEE, 2021: 1-6.

[19] ZANG X, LI G, GAO W, et al. Learning to disentangle scenes for person re-identification [J]. Image and Vision Computing, 2021, 116: 104330.

(上接第 94 页)

[28] GAO J, HU W, CHEN Y. Client: cross-variable linear integrated enhanced transformer for multivariate long-term time series forecasting[J]. arXiv preprint arXiv:2305.18838, 2023.

[29] PENG Y, ARORA S, HIGUCHI Y, et al. A study on the integration of pretrained ssl, asr, lm and slu models for spoken language understanding [C]//Proceedings of 2022 IEEE Spoken Language Technology Workshop (SLT). IEEE, 2023: 406-413.

[30] DIMITRIADIS T, GNEITING T, JORDAN A I, et al. Evaluating probabilistic classifiers: The triptych [J]. arXiv preprint arXiv: 2301.10803, 2023.