DOI:10.16136/j.joel.2022.09.0883

双金字塔结构引导的多粒度行人重识别方法

刘 粤[1],赵 迪[1],田紫欣[1],熊 炜[1,2,3*],许婷婷[1],李利荣[1,2]

(1.湖北工业大学 电气与电子工程学院,湖北 武汉 430068;2.襄阳湖北工业大学 产业研究院,湖北 襄阳 441003;3.美国南卡罗来纳大学 计算机科学与工程系,南卡罗来纳州 哥伦比亚 29201)

摘要:针对杂乱场景下难以有效地提取行人关键信息和局部遮挡时全局特征方法失效的问题,提出了一种双金字塔结构引导的多粒度行人重识别(person re-identification,ReID)方法。首先在ResNet50中嵌入注意力金字塔,引导网络由粗到细依次挖掘不同粒度的特征,使网络更倾向于关注复杂环境中行人的显著区域;其次通过结构不对称的双重注意力特征金字塔分支(double attention feature pyramid branch,DFP branch)提取多尺度的行人特征,丰富特征的多样性,同时双重注意力机制可使分支从浅层信息中捕获高细粒度的局部特征;最后将粒度较粗的全局特征与多层级细粒度的局部特征融合,两种金字塔相互作用,以此获得更多具有鉴别性的多粒度特征,改善行人遮挡问题。在多个数据集上进行了实验,结果表明,各项评价指标均高于目前大多数主流模型,其中在DukeMTMC-reID数据集上,Rank-1、*mAP*和平均逆负处罚(mean inverse negative penalty,*mINP*)分别达到了91.6%、81.9%、48.1%。

关键词:行人重识别;注意力金字塔;双重注意力特征金字塔分支(DFP branch);多粒度特征

中图分类号:TP183 **文献标识码**:A **文章编号**:1005-0086(2022)09-0959-09

Multi-granularity person re-identification method guided by double pyramid structure

LIU Yue[1], ZHAO Di[1], TIAN Zixin[1], XIONG Wei[1,2,3*], XU Tingting[1], LI Lirong[1,2]

(1. School of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan, Hubei 430068, China; 2. Xiangyang Industrial Research Institute, Hubei University of Technology, Xiangyang, Hubei 441003, China; 3. Department of Computer Science and Engineering, University of South Carolina, Columbia, South Carolina 29201, USA)

Abstract: Aiming at the problem that it is difficult to effectively extract the key information of pedestrians in the chaotic scene and the global feature method is invalid in the case of partial occlusion, a multi-granularity person re-identification (ReID) method guided by a double pyramid structure is proposed. First, the attention pyramid in is embedded ResNet50 to guide the network to dig out features of different granularities from coarse to fine, making the network more inclined to focus on the significant areas of pedestrians in complex environments; secondly, the branch of the double attention feature pyramid (DFP) with asymmetric structure is adopted. Multi-scale pedestrian features are extracted to enrich the diversity of features. At the same time, the dual attention mechanism allows branches to capture finer-grained local features from shallow information; finally, the coarser-grained global features are merged with multi-level and fine-grained local features, The two kinds of pyramids interact to retain more discriminative multi-granularity features to improve the pedestrian occlusion problem. Experiments on multiple data sets have shown that the evaluation indicators are higher than most current mainstream models. Among them, on the DukeMTMC-reID data set, Rank-1, *mAP* and mean inverse negative penalty (*mINP*)

* **E-mail**: xw@mail.hbut.edu.cn

收稿日期:2021-12-28 **修订日期**:2021-01-28

基金项目:国家自然科学基金(61571182,61601177)、湖北省自然科学基金(2019CFB530)、湖北省科技厅重大专项(2019ZYYD020)、襄阳湖北工业大学产业研究院科研项目(XYYJ2022C05)和国家留学基金(201808420418)资助项目

reached 91.6%, 81.9% and 48.1%, respectively.

Key words: person re-identification (Person ReID); attention pyramid; double attention feature pyramid branch (DFP branch); multi-granularity feature

1 引　言

行人重识别(person re-identification, ReID)旨在将不同地点和不同摄像机拍摄到的某个行人图像关联起来,以检索跨监控图像和设备视频中的特定行人。随着该技术被广泛应用于智能视频监控、大规模人员跟踪等领域,越来越多极具挑战性的难题也暴露而出,如行人图像受到背景杂乱、局部遮挡、姿态变化和尺度变化等因素影响[1]。因此,如何在复杂场景下的行人图像中提高特征识别能力是当前 ReID 研究的首要任务。

随着深度学习的火热,诸多拥有高特征提取性能的网络孕育而生,其中基于深度学习卷积神经网络的 ReID 逐渐将传统的识别方法淘汰,并且获得了突破性的发展。在其发展的早期阶段,大多方法的目的是从行人的整体图像中获取显著信息,以此获取行人图像的全局特征表达。但是,如果存在行人部分身体被遮挡或检测错误等情况将会导致摄像机无法捕捉整体行人,且全局特征表达容易忽略一些关键的局部信息,影响全局特征的性能,导致识别精度下降,因此引入更为复杂的局部特征成为研究热点。伴随着技术的进一步发展,图片切块[2]、人体姿态关键点对齐[3]和姿态评估[4]成为提取局部特征的常用方法,文献[2]通过将行人图片分为若干等分,再由 AlignedReID 网络实现自上而下的动态对准;文献[3]利用姿态信息提取骨架关键点,将提取的 8 个特征在不同的尺度上使用融合操作,从而得到多个局部特征与一个全局特征;文献[4]使用 GLAD(global-local-alignment descriptor)特征描述子把行人图片分为头和上、下身,再将其与整图一起输入网络,以得到局部和全局特征。但将行人图片切分的方法,易导致人体各部分之间的上下文缺失,且现有的姿态估计方法也往往不能很好地泛化 ReID 真实场景,以至于无法可靠地估计人体部位。同时,为了更好地提升网络的抗干扰性,提取更具辨别性的特征,近些年来,越来越多的注意力机制[5,6]被用于 ReID 系统中,其通过卷积神经网络快速读取行人的各处信息,从而捕获极具特征的关键信息区域,并在关键区域加入相对更多的注意力。但当前注意力模型大多只学习全局的注意力信息,虽然能在复杂背景下关注行人,却往往无法准确地获取其中行人极具判别性的特征。

为解决上述问题,本文提出了一种可有效提取底层语义信息中具有高细粒度表征能力的局部特征,同时更关注多粒度全局特征的 ReID 网络模型。利用 APNet 注意力金字塔[7]引导网络从粗粒度到细粒度渐进式地学习注意力,从而以多尺度的方式关注重点信息,使网络更好关注杂乱背景下行人的显著特征,降低无关信息的干扰。并且通过双重注意力特征金字塔分支(double attention feature pyramid branch, DFP branch)来提取多样性的局部细粒度特征,有效弥补遮挡时全局信息失效的缺陷,并通过强制网络关注特征之间的关联性与剩余的区域来提升网络的泛化能力。

2 双金字塔结构引导的多粒度 ReID 模型

本文提出的双金字塔结构网络模型如图 1 所示,网络框架主要由嵌入了注意力金字塔的骨干网

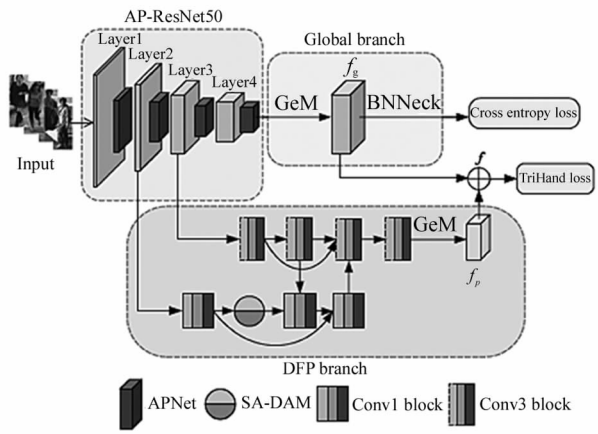

图 1　双金字塔结构引导的多粒度 ReID 网络框架

Fig. 1　Multi-granularity ReID network framework guided by the double pyramid structure

络(AP-ResNet50)、全局分支(global branch)和 DFP branch 3 个部分构成。

2.1　ReID 网络框架

本网络框架首先选用 ReID 中常见的 ResNet50 残差网络[8],以更公平地与主流模型的性能做比较,与此同时在其 4 个残差块(block)之后添加 APNet 注意力金字塔,构成 AP-ResNet50 作为本框架所使用的骨干网络(Backbone),以引导网络由粗至细依次挖掘更具判别性的多粒度特征。然后删除 AP-ResNet50 网络末尾的池化层与全连接层,并将网络最后一个 block 的步长由原始的 2 改为 1,增大输出特征的尺寸,显著改善了高空间的分辨率,从而使细粒度特征更加丰富,且不会增加额外的训练参数;再将增大尺寸后的 feature map 通过广义均值池化(generalized mean pooling,GeM)得到大小为 1×1×2048 的 global feature(f_g),同时 GeM 可自适应调节参数,使网络能够关注更多不同粒度的区域。为了让 cross entropy loss 更容易收敛,各个行人之间的特征差异性更加明显,因此将 f_g 通过 BNNeck 归一化处理得到 f_{gb},从而构成多粒度全局分支。并且正则化可平衡各个维度的特征,使得同一人的特征更加紧密,然后将 f_{gb} 送入 cross entropy loss 进行分类;另外提取 Layer2 和 Layer3 输出的浅层特征,将其输入 DFP branch,得到大小为 24×8×1024 局部特征,非对称的特征金字塔分支结构可以使特征的多样性更加丰富,同时双重注意力机制让分支获取更具鉴别性的细粒度特征。将局部特征 f'_p 通过 GeM 得到大小为 1×1×1024 的 f_p,并通过 1×1 卷积将 f_p 的 1024 通道数升为 2048 后,与全局特征 f_g 进行融合,结合两种金字塔结构的输出,以得到大小为 1×1×1024 的多粒度特征 f,用于 TriHard 的训练。

2.2　注意力金字塔 APNet

复杂场景往往会对行人识别造成大量干扰,使网络难以准确地关注行人的重点区域,因此本文采用了一种轻量化的注意力金字塔模块 APNet,通过“Split-Attend-Merge-Stack”4 步操作逐级实现,如图 2 所示。输入行人图像为 I,首先通过骨干网络的残差块提取特征 $F\in R^{H\times W\times C}$,$C$ 表示通道数,H 和 W 表示特征图的高与宽。再通过注意力金字塔 P_i 来提取特征 F 上多尺度的显著区域,其中 P_i 表示第 i 个金字塔层。不同的金字塔层可以提取不同粒度的判别信息,从而使粗粒度注意力引导接下来的细粒度注意力学习。以上可以由一个连续学习过程所表示:

$$F_i = P_i(F_{i-1}), \tag{1}$$

式中,F_i 表示 APNet 第 i 层的特征图。

在APNet结构框架中,进行拆分操作时以特征图 F_i 作为输入,将 F_i 拆分为 n 个特征张量 $F_{i,j}$ 作为输出,其中 $j=\{1,2,4,\cdots,n\}$,$n=2^i$,即被拆分(Split)数在金字塔层中呈指数增长,金字塔层数越高,特征粒度越细。由于本文在 APNet 框架中采用的是通道注意力模块(channel attention module,CAM),因此特征图 F_i 将被拆分为 $F_{ij}=R^{C\times\frac{H}{n}\times W}$ 以作为通道注意力的输入。在关注(Attend)与合并(Merge)操作中,通过一组子通道注意力模块 $M_{i,j}$ 来挖掘每个特征张量中具有判别性的信息。再将所有组子注意力模块与拆分操作的逆过程进行合并,而

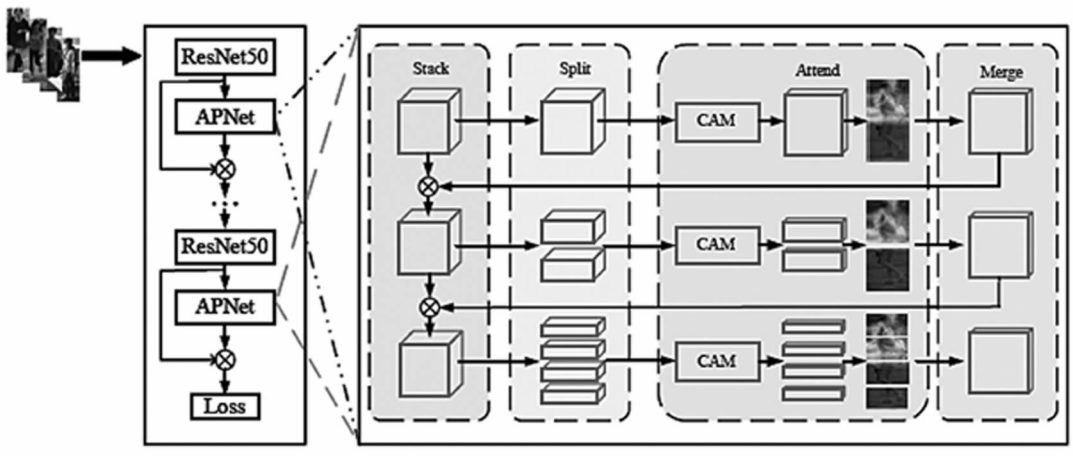

图 2　注意力金字塔结构图

Fig. 2　The structure diagram of the attention pyramid

不是将特征张量进行简单的融合，从而得到完整且尺度相同的注意力图 M_i，其表达式如下：

$$M_i = [M_{i,j}(F_{i,j})]_{j=1}^{n}, \tag{2}$$

式中，$[\cdots]_{j=1}^{n}$ 表示 n 个注意力图的专注度。将注意力由粗到细的进行堆叠，从而形成金字塔结构，以此引导网络逐步关注重要特征，从而完成最后的堆叠(Stack)操作。通过学习到的注意力 $M_{i,j}$ 来使本网络发现更多具有判别性的特征 F_i，表达式如下：

$$F_i = \alpha(A_i) * F_{i-1}, \tag{3}$$

式中，$*$ 表示逐元素相乘。对注意力图进行归一化处理后的特征将被重新加权，权值为 α。然后输入到金字塔的下一层以引导更细粒度的注意力学习。最后将 APNet 注意力金字塔应用于 ResNet50 网络的 4 个残差块之后，可促使本文的 AP-ResNet50 骨干网络通过渐进式细化操作来捕获更具辨别力的特征。APNet 无需额外的特征提取模块，取而代之的是拆分和堆叠操作。因此，该注意力金字塔通过较低的计算成本便可引导网络从粗到细依次挖掘不同粒度的显著特征，降低背景干扰，有效地提取更具鉴别性的特征。

2.3　双重注意力特征金字塔 DFP

通过全局特征进行识别的技术需要行人的整体信息，由于行人图像受物体遮挡和行人姿态变化等因素影响，会让此类技术因丢失具有判别性的人体局部信息而导致识别精度不高。为了解决该问题，本文通过对文献[9]进行改进，即在特征金字塔分支 FPB 内嵌入了双重注意力机制 SA-DAM，从而构成本文所提出了 DFP 双重注意力特征金字塔，可以充分利用行人不同尺度的空间信息，即使在重要信息丢失的情况下，仍能更准确地识别候选行人图像。

2.3.1　特征金字塔分支 FPB

为使网络更加关注行人的局部区域，本文使用 FPB 以在 ReID 的多层级系统中获取多样性的局部特征。首先提取 Layer2 和 Layer3 输出的特征。将 Layer2 的输出特征先通过 3 个 Conv1 block，将不同通道数的特征图统一转换为 256。Conv1 block 由一个标准的 1×1 卷积、BN 层和 ReLU 层组成。Layer3 的输出特征通过 4 个 Conv3 block，用于完成不同尺度的特征聚合，Conv3 block 除采用的是 3×3 的卷积之外，其余结构与 Conv1 block 一致，其中 Conv1 block 与 Conv3 block 的结构如图 3 所示。在该两条局部支路中，不同空间分辨率的特征图之间存在两种跨尺度连接。一种自顶而下的过程通过最近邻插值法的上采样方式将顶层的小特征图放大到与下一个特征图一样的大小。相反，一个自底向上的过程是通过 2×2 的最大池化实现的。同时，每一层还存在一个类似于 ResNet 中残差结构的下采样操作，使第一个 Conv block 与第三个 Conv block 相连，以增强特征之间的相关性与多样性，该残差结构如图 1 DFP branch 中的弯曲箭头所示。

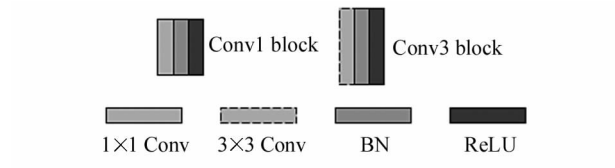

图 3　Conv1 block 和 Conv3 block 的结构图

Fig. 3　Structure diagram of Conv1 block and Conv3 block

因为 ReID 系统需要来自不同尺度的复杂信息，以便对相对较多的行人 ID 进行分类。所以本结构局部支路通过具有相对较大的卷积核和较小通道数的 Conv block 来提取信息，获得了更浅层的特征作为输入，该浅层特征可从图像中保留更多的局部细节信息，同时具有较低维度的特征可减少分支内学习的参数量，使本分支在仅增加较少参数量的代价，可获得更具多样性的特征，提高本系统的识别精度。

2.3.2　双重注意力机制 SA-DAM

Self-Attention 机制可利用特征之间的关联性来帮助模型关注更多相关特征，并且能充分减少网络对外部信息的依赖性。由于浅层特征包含更多的是局部底层粗略信息，为引导 FPB 分支能够从粗略信息中挖掘更显著的细粒度局部特征，本文将 Self-Attention 模块与 APNet 进行并行连接，构成本文所提出的 SA-DAM 双重注意力机制，如图 4 所示。本注意力机制还可以通过网络强制关注特征之间的关联性和剩余的区域，有效提升模型的泛化能力。

本文中 Self-Attention 机制基于位置注意力模块(PAM)来实现，输入特征 $X \in R^{H\times W\times C}$。将每个位置的特征 X 通过 1×1Conv 映射并重塑至两个较低维度的子空间上，从而得到 $\boldsymbol{Q}$(Query)$\in R^{\frac{C}{r}\times S}$ 和 $\boldsymbol{K}$(Key)$\in R^{\frac{C}{r}\times S}$。其中 $S=H\times W$，表示特征空间的大小，r 是用于调节子空间维度的超参数，本文将 r 设置为 8。则 X 的注意力表达式如下：

$$X_p = \boldsymbol{V}\sigma(\boldsymbol{A}) = \boldsymbol{V}\sigma(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{K}), \tag{4}$$

式中，$\sigma(\cdot)$ 表示 Softmax 函数，$\boldsymbol{V}$(Value)$\in R^{C\times S}$ 是特征 X 的另一个恒等映射。在忽略所有可学习参数后，可将位置亲和矩阵 $\boldsymbol{A}$ 视为格拉姆矩阵(Gram matrix)，以衡量 X 不同位置特征之间的关联性。λ 表示可学习参数，以调节注意力的影响。同时使特征

X 输入金字塔层数为 2 的 APNet 模块,将特征 X 进行多次拆分后以便于本注意力模块关注高细粒度特征,再通过合并与堆叠来引导网络渐进地关注显著信息,最后将其得到的特征与 Self-Attention 模块输出的特征使用 Concat 操作进行融合,以此获得更具判别性的细粒度特征。

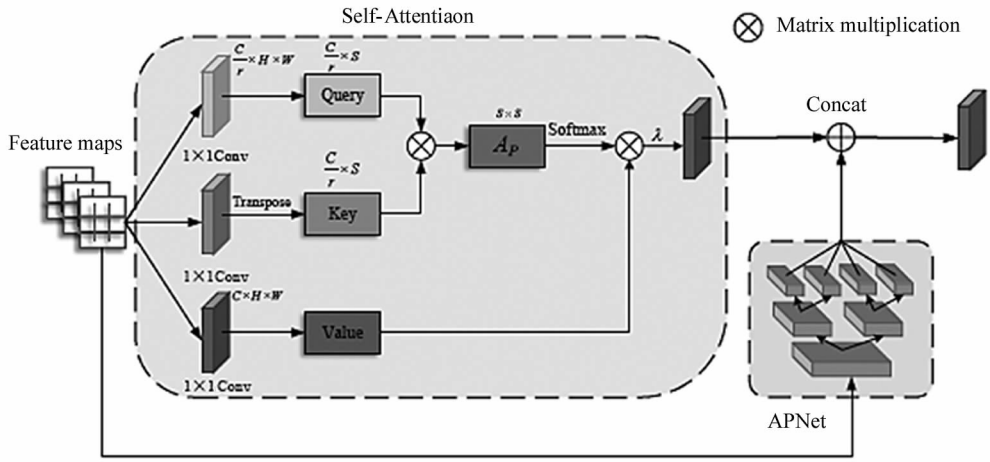

图 4　SA 双重注意力机制结构图

Fig. 4　Structure diagram of SA dual attention mechanism

2.4　损失函数

基于本网络模型,采用三元组难样本挖掘损失(TriHard loss)和交叉熵损失(cross entropy loss)进行联合训练。其中 TriHard loss 在每个训练的批次中任意读取出 N 个不同的行人,然后从每个不同的行人中再读取 E 张不一样的图像。再从每一批次中选取任意图像 a,对 a 选取一个最难正样本 b 和一个最难负样本 n,以此形成一个三元组样本[10],设置阈值为 β。则 TriHard loss 表达式如下:

$$L_{\text{th}}=\frac{1}{N\times E}\sum_{a\in\text{batch}}(\max_{b\in B}d_{a,b}-\min_{n\in D}d_{a,n}+\beta)_{+},\tag{5}$$

式中,B 表示和图像 a 为相同行人的图像集,D 为其余不同行人的图像集,本损失在训练过程中比三元组损失函数对正负样本在样本空间距离的调整更加精细并且具有更强的驱动力。

因交叉熵损失对行人标签的正确率要求很高,则本文通过标签平滑的方法(label smoothing)对行人标签平滑处理,来提升模型的泛化能力。其损失表达式如下:

$$L_{\text{ce\_ls}}=\sum_{\varphi=1}^{G}-q_{\varphi}\log(p_{\varphi}),q_{\varphi}=\begin{cases}\dfrac{\varepsilon}{G},m\neq\varphi\\[2ex]1-\dfrac{G-1}{G}\varepsilon,m=\varphi\end{cases},\tag{6}$$

式中,G 为所有行人的数量,ε 为设定的错误率,m 为行人的标签,q_{φ} 为真伪标签,p_{φ} 为本模型预测此行人属于标签 φ 行人的概率。

本模型使用两个损失函数进行联合训练,其总损失表达式如下:

$$L_{\text{total}}=L_{\text{th}}+L_{\text{ce\_ls}},\tag{7}$$

通过 L_{total} 可加快模型的收敛速度,同时更好地学习具有判别性的多粒度特征。

3　实验与分析

为充分验证本模型的性能并客观地进行对比,文本在 4 个主流数据集上进行了实验,所有实验均重复进行 3 次,取平均值作为最终实验结果,在保证实验结果准确性的同时降低偶然性。

3.1　数据集和评价指标

本文实验选用在清华大学所采集的 Market1501 数据集、在香港中文大学采集的 CUHK03 数据集、在杜克大学采集的 DukeMTMC-ReID 数据集和在北京大学采集的 MSMT17 数据集进行实验,如表 1 所示。

表 1　数据集

Tab. 1　Datasets

Dataset	Training Set		Test Set		Camera
	ID	Image	ID	Image	
Market1501	751	12 936	750	19 732	6
CUHK03	767	7 368	700	6 729	10
DukeMTMC-ReID	702	16 522	702	19 889	8
MSMT17	4 101	32 621	3 060	93 820	15

为更准确与全面地评测本模型性能,本文采用

ReID常用的两个评价指标Rank-1和mAP之外，还引入了文献[11]提出的平均逆负处罚(mean inverse negative penalty，$mINP$)评价指标。$mINP$具有衡量检索最难正确匹配项的能力。首先，使用负处罚(negative penalty，NP)查找最难的正确匹配，NP表达式如下：

$$NP_d = \frac{R_d^{\text{hard}} - |U_d|}{R_d^{\text{hard}}}, \tag{8}$$

式中，最难匹配项的击中位置为R_d^{hard}，查询第d个图片所有正确匹配项的数量为$|U_d|$。则$mINP$表达式如下：

$$mINP = \frac{1}{n}\sum_{d=1}^{n}(1 - NP_d) = \frac{1}{n}\sum_{d=1}^{n}\frac{|U_d|}{R_d^{hard}}, \tag{9}$$

3.2 实验设置

本实验平台为Ubuntu18.04操作系统，基于Pytorch框架实现。硬件配置为：GeForce RTX3060显卡，AMDR5-3600X处理器，16GB内存。在数据预处理阶段，设置输入图像大小为384×128，使用随机擦除、随机增补、随机裁剪以及水平翻转4种数据增强方法，以提升网络对遮挡行人图像的鲁棒性。本模型在训练前，注意力金字塔层数取值为2，训练批次为64，使用Adam优化器，学习率采用Warmup策略，初始学习率为3.5×10^{-5}，训练90个epoch。

3.3 实验内容

为验证APNet$i$不同金字塔层数的性能，则分别以AP-ResNet50骨干网络的不同金字塔层数在本模型上使用Market1501、DukeMTMC-ReID和CUHK03数据集完成实验，$i=\{1,2,3\}$，即APNet1表示APNet所使用的金字塔层数为1，实验结果如表2所示。由表2可知，虽然当金字塔层数为3时，有个别评价指标与层数为2时的持平或略高，但APNet2的整体性能最佳。因此相较于1、3层，当金字塔层数为2时，APNet注意力金字塔引导网络挖掘多粒度特征的能力更优，并之后实验都使用APNet2进行。

表2 不同金字塔层数的实验结果

Tab. 2 Experimental results of different pyramid levels

Method	Market1501			DukeMTMC-ReID			CUHK03		
	Rank-1	mAP	$mINP$	Rank-1	mAP	$mINP$	Rank-1	mAP	$mINP$
APNet1	95.2	88.2	65.7	90.3	80.8	47.2	79.3	76.2	65.4
APNet2	**96.1**	**89.5**	**66.9**	**91.6**	**81.9**	**48.1**	**80.5**	77.9	**66.5**
APNet3	95.8	**89.5**	66.7	91.1	81.6	47.8	80.2	**78.1**	66.4

针对本文提出的SA-DAM双重注意力机制，分别使用Self-Attention、APNet2和SA-DAM(Self-Attention + APNet2)3个模块作为本网络DFP Branch中的注意力机制部分，通过Market1501和DukeMTMC-ReID数据来验证SA-DAM的效果，实验数据如表3所示。在表3数据中，将Self-Attention和APNet2两种注意力机制按本文方法进行组合后形成的SA-DAM双重注意力机制整体达到了最优的评价指标，证明本文提出的SA-DAM作用于DFP branch中的注意力机制部分时，较Self-Attention和APNet2注意力机制能更充分捕获显著的细粒度特征。

表3 SA-DAM双重注意力的验证实验

Tab. 3 SA-DAM double attention verification experiment

Methods	Market1501			DukeMTMC-ReID		
	Rank-1	mAP	$mINP$	Rank-1	mAP	$mINP$
Self-Attention	**96.1**	89.1	66.6	91.3	81.1	47.2
APNet2	95.3	88.4	66.3	91.0	81.0	47.1
SA-DAM	**96.1**	**89.5**	**66.9**	**91.6**	**81.9**	**48.1**

为了进一步验证本网络中各个模块的有效性，本消融实验以ResNet50和Global Branch组成Baseline，再使用APNet2、FPB和SA-DAM 3个模块在DukeMTMC-ReID数据集上进行。本网络模型的消融实验数据见表4。通过实验可以发现，在Baseline上依次加入APNet2、FPB和SA-DAM模块后，其评价指标Rank-1、mAP与$mINP$都有较为明显的提升。由此证明本算法所使用的3个模块可

有效提升识别任务的准确率。

表 4　网络的消融实验

Tab. 4　Ablation experiment of the network

	Baseline	APNet2	FPB	SA-DAM	Rank-1	*mAP*	*mINP*
1	√	—	—	—	88.1	78.8	44.5
2	√	√	—	—	89.3	79.7	45.8
3	√	√	√	—	90.6	81.0	47.0
4	√	√	√	√	**91.6**	**81.9**	**48.1**

3.4　结果可视化

使用本文的方法,行人图像 Rank-10 检索结果可视化如图 5 所示,每行的第一列是被检索行人图像,后 10 列是以相似度排序的检索图,黑色加粗框体表示错误检索图,未加粗框体表示正确检索图。由图可发现,仅在第一行的被检索行人存在少部分遮挡的情况下,第 6 位检索出现错误,在第二行被检索行人无遮挡时,检索结果全部正确,表明本模型具有相当不错的识别性能。

图 5　Rank-10 检索结果示例图

Fig. 5　Example image of Rank-10 search results

本文还采用了 Grad-CAM 类激活热力图对 DFP branch 最后输出的特征图进行可视化,如图 6 所示。由图可发现,本分支更注重行人的腿、背包、手中物体和衣服的特定颜色等多样性的局部细节特征。主要是因为 DFP branch 使用了不对称的结构,会导致所提取的特征丰富多样,同时SA-DAM双重注意力机制引导分支捕获到了更显著的细粒度特征。

图 6　DFP branch 输出特征的可视化热力图

Fig. 6　Visualized heat map of DFP branch output characteristics

3.5　实验对比

将本文所提出的模型与目前主流模型进行对比分析,以验证本算法模型的优异性能。其对比实验分别在上述 4 个数据集上进行。

表 5 为本模型在 Market1501 和 DukeMTMC-ReID 数据集上的对比数据。从该表中数据可发现,虽然 Market1501 数据集由于错误标注较多,以至于难以更进一步提高系统识别性能,但是在未使用重排序(Re-ranking,RK)测试技巧时,本模型对比本年度公布的 APNet-S 模型在 Rank-1 持平的情况下,*mAP* 依旧提升了 0.5%,*mINP* 指标相较于 ABD-Net 提升了 0.7%。在 DukeMTMC-ReID 上,本模型在未使用 Re-ranking 测试技巧时,相较于 APNet-S 模型,Rank-1 和 *mAP* 分别提升了 2.3%、3.1%;相较于 AGW 模型,*mINP* 提升了 2.4%。

表 6 为本模型在 CUHK03 和 MSMT17 数据集上与其他主流模型的对比结果。从该表中结果可发现,本模型在 CUHK03 数据集上 Rank-1 和 *mAP* 略低于 APNet-S,但使用 Re-ranking 测试技巧后,

Rank-1、*mAP* 和 *mINP* 分别达到了 84.9%、87.0%、85.3%。在更加接近真实环境下的大数据集MSMT17上，Rank-1提升了1.1%，*mAP*提升了0.3%，同时也比其余模型的评价指标高，由此证明本模型在大数据集上表现出了更佳的识别效果。

在真实场景下，摄像头所拍摄的行人图像之间难免会存在风格差异很大的情况，因此模型的泛化能力也尤为重要。为验证本模型的泛化能力，本文进行了跨域实验。该跨域实验数据及其对比其他模型性能的结果如表7所示。可见，本模型的 Rank-1 和 *mAP* 明显高于其中绝大部分模型，证明本模型的泛化性能更优。但当前指标与未跨域前的指标存在不小差距，说明仍难以应对风格多变的复杂场景，泛化能力还需进一步提升。

表5　Market1501 和 DukeMTMC-ReID 数据集与主流模型的性能对比

Tab. 5　Performance comparison between Market1501 and DukeMTMC-ReID data sets and mainstream models

Methods	Publication	Market1501			DukeMTMC-ReID		
		Rank-1	*mAP*	*mINP*	Rank-1	*mAP*	*mINP*
Bag-of-Tricks[12]	CVPRW'19	94.5	85.9	59.4	86.4	76.4	40.7
Pyramid[13]	CVPR'19	95.7	88.2	—	89.0	79.0	—
ABD-Net[14]	ICCV'19	95.6	88.3	66.2	89.0	78.6	42.1
FastReID[15]	Arxiv'20	95.4	88.2	64.8	90.3	80.3	46.5
AGW[11]	TPAMI'20	95.1	87.8	65.0	89.0	79.6	45.7
RGA-SC[6]	CVPR'20	**96.1**	88.4	—	—	—	—
APNet-S[7]	TIP'21	**96.1**	89.0	—	89.3	78.8	—
CBDB-Net[16]	TCSVT'21	94.4	85.0	—	87.7	74.3	—
Ours		**96.1**	**89.5**	**66.9**	**91.6**	**81.9**	**48.1**
Ours(RK)		96.0	**94.4**	**86.8**	**92.6**	**90.7**	**73.5**

表6　CUHK03 和 MSMT17 数据集与主流模型的性能对比

Tab. 6　Performance comparison between CUHK03 and MSMT17 datasets and mainstream models

Methods	Publication	CUHK03			MSMT17		
		Rank-1	*mAP*	*mINP*	Rank-1	*mAP*	*mINP*
Bag-of-Tricks[12]	CVPRW'19	58.0	56.6	43.8	63.4	45.1	12.4
FastReID[15]	Arxiv'20	88.2	74.9	64.8	81.8	58.4	13.9
AGW[11]	TPAMI'20	63.6	62.0	50.3	68.3	49.3	**14.7**
RGA-SC[6]	CVPR'20	79.6	74.5	—	80.3	57.5	—
APNet-S[7]	TIP'21	**80.9**	**78.1**	—	80.8	59.0	—
CBDB-Net[16]	TCSVT'21	75.4	72.8	—	—	—	—
Ours		80.5	77.9	**66.5**	**81.9**	**59.3**	14.5
Ours(RK)		**84.9**	**87.0**	**85.3**	—	—	—

表7　跨域实验结果对比

Tab. 7　Comparison of cross-domain experiment results

Methods	D→M		M→D		C→D		C→M	
	Rank-1	*mAP*	Rank-1	*mAP*	Rank-1	*mAP*	Rank-1	*mAP*
Bag-of-Tricks[12]	54.3	25.5	41.4	25.7	—	—	—	—
ATNet[17]	55.7	25.6	45.1	24.9	—	—	—	—
PUL[18]	45.5	20.5	30.0	16.4	23.0	12.0	41.9	18.0
HHL[19]	62.2	31.4	46.9	27.2	42.7	**23.4**	56.8	29.8
UCDA[20]	60.4	30.9	47.7	31.0	—	—	—	—
Ours	**64.79**	**33.4**	**56.5**	**35.3**	**43.2**	23.3	**63.4**	**36.6**

4　结　论

本文提出了一种双金字塔结构引导的多粒度 ReID 方法，该方法通过 AP-ResNet50 金字塔骨干网络来挖掘不同粒度的显著特征，并使用了 DFP branch 提取不同尺度的多样性特征，可让本模型更多地关注局部信息，同时使网络强制性地注意特征之间的关联性来提高模型的泛化能力。双金字塔结构形成互补，从而引导网络更加关注杂乱背景前的行人显著区域和具有判别性的行人局部细节特征，提升模型在真实场景中的识别精度。大量的实验数据证明，本方法相比其他主流方法，各项评价指标均有较为明显的提升，但泛化能力仍有所欠缺，因此未来的研究重点将是提升跨域场景下的识别性能。

参考文献：

[1] GONG X，YAO Z，LI X，et al. LAG-Net：multi-granularity network for person re-identification via local attention system[J]. IEEE Transactions on Multimedia，2021，24：217-229.

[2] ZHANG X，LUO H，FAN X，et al. AlignedReID：surpassing human-level performance in person re-identification[EB/OL]. (2018-01-31) (2021-12-28). https://arxiv. org/abs/1711.08184.

[3] ZHAO H，TIAN M，SUN S，et al. Spindle Net：person re-identification with human body region guided feature decomposition and fusion[C]//IEEE Conference on Computer Vision and Pattern Recognition，July 21-26，2017，Honolulu，HI，USA. New York：IEEE，2017：907-915.

[4] WEI L，ZHANG S，YAO H，et al. GLAD：global-local-alignment descriptor for scalable person reidentification[J]. IEEE Transactions on Multimedia，2019，21(4)：986-999.

[5] CHEN G，LIN C，REN L，et al. Self-critical attention learning for person re-identification[C]//IEEE/CVF International Conference on Computer Vision，October 27-November 02，2019，Seoul，Korea (South). New York：IEEE，2019：9636-9645.

[6] ZHANG Z，LAN C，ZENG W，et al. Relation-aware global attention for person re-identification[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition，June 13-19，2020，Seattle，WA，USA. New York：IEEE，2020：3183-3192.

[7] CHEN G，GU T，LU J，et al. Person re-identification via attention pyramid[J]. IEEE Transactions on Image Processing，2021，30：7663-7676.

[8] HE K，ZHANG X，REN S，et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition，June 27-30，2016，Las Vegas，NV，USA. New York：IEEE，2016：770-778.

[9] ZHANG S，YIN Z，WU X，et al. FPB：feature pyramid branch for person re-identification[EB/OL]. (2021-08-04)[2021-12-28]. https://arxiv. org/abs/2108.01901.

[10] HERMANS A，BEYER L，LEIBE B. In defense of the triplet loss for person re-identification[EB/OL]. (2017-03-22)[2021-12-28]. https://arxiv. org/abs/1703.07737.

[11] 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.

[12] LUO H，GU Y，LIAO X，et al. Bag of tricks and a strong baseline for deep person re-identification[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops，June 15-20，2019，Long Beach，CA，USA. New York：IEEE，2019：1487-1495.

[13] ZHENG F，DENG C，SUN X，et al. Pyramidal person re-identification via multi-loss dynamic training[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition，June 15-20，2019，Long Beach，CA，USA. New York：IEEE，2019：8514-8522.

[14] CHEN T，DING S，XIE J，et al. Abd-net：attentive but diverse person re-identification[C]//IEEE/CVF International Conference on Computer Vision，October 27-November 02，2019，Seoul，Korea (South). New York：IEEE，2019：8351-8361.

[15] HE L，LIAO X，LIU W，et al. FastReID：a pytorch toolbox for real-world person re-identification[EB/OL]. (2020-06-04) [2021-12-28]. https://arxiv. ogr/abs/2006.02631.

[16] TAN H，LIU X，BIAN Y，et al. Incomplete descriptor mining with elastic loss for person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology，2021，32(1)：160-171.

[17] LIU J，ZHA Z J，CHEN D，et al. Adaptive transfer network for cross-domain person re-identification[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition，June 15-20，2019，Long Beach，CA，USA. New York：IEEE，2019：7202-7211.

[18] FAN H，ZHENG L，YAN C，et al. Unsupervised person re-identification：clustering and fine-tuning[J]. ACM Transactions on Multimedia Computing，Communications，and Applications (TOMM)，2018，14(4)：1-18.

[19] ZHONG Z，ZHENG L，LI S，et al. Generalizing a person retrieval model hetero-and homogeneously[C]//European Conference on Computer Vision，September 8-14，2018，Munich，German. Berlin：Springer，2018：172-188.

[20] QI L，WANG L，HUO J，et al. A novel unsupervised camera-aware domain adaptation framework for person reidentification[C]//IEEE/CVF International Conference on Computer Vision，October 27-November 02，2019，Seoul，Korea (South). New York：IEEE，2019：8080-8089.

作者简介：

熊　炜　(1976—)，男，博士，副教授，硕士生导师，主要从事数字图像处理和计算机视觉方面的研究.