

基于特征金字塔分支和非局部关注的行人重识别

孙明浩, 王洪元, 吴琳钰, 张 继, 周群颖

(常州大学计算机与人工智能学院/阿里云大数据学院/软件学院, 常州 213164)

摘 要: 关注全局轮廓和行人局部细节对现有行人重识别方法非常重要。为了能够提取这些更具代表性的特征, 提出一种基于特征金字塔分支和非局部关注模块的行人重识别网络方法来提取行人全局和局部表征特征。该方法首先引入一种轻量级别特征金字塔分支结构, 从不同的网络层中提取特征, 并且聚合成一个双向金字塔结构。其次为进一步提高行人重识别的精度, 使用非局部关注模块提取全局特征, 这样既能获取行人的全局信息, 又能注重行人的局部细节, 使两者最终融合的特征更具代表性。最后将不同层间的特征融合起来, 并使用联合损失函数策略对网络模型进行训练, 显著提高骨干网络的性能。通过在 MSMT17、Market1501、DukeMTMC-ReID 和 PersonX 四个公共行人重识别数据集上的大量实验, 证明所提出的基于特征金字塔分支和非局部关注的方法相较于目前一些先进的行人重识别方法, 具有一定的竞争力。

关键词: 行人重识别; 特征金字塔分支; 注意力机制; 非局部关注; 联合损失函数策略

中图分类号: TP391 **文献标志码:** A

Person Re-identification Based on Feature Pyramid Branch and Non-local Attention

SUN Minghao, WANG Hongyuan, WU Linyu, ZHANG Ji, ZHOU Qunying

(College of Computer Science and Artificial Intelligence/College of Aliyun and Big Data/College of Software, Changzhou University, Changzhou 213164, China)

Abstract: Paying attention to the global contour and the person local details is very important for the existing person re-identification methods. In order to extract these more representative features, a person re-identification network method based on the feature Pyramid branches and the non-local attention modules is proposed to extract the global and local characterization features of person. Firstly, this method introduces a lightweight feature Pyramid branch structure, extracts features from the different network layers, and aggregates them into a two-way Pyramid structure. Secondly, in order to further improve the accuracy of person re-identification, the non-local attention module is used to extract the global features, which can not only obtain the global information of person, but also pay attention to the local details of person, so that their final fusion features are more representative. Finally, the characteristics of different layers are fused, and the joint loss function strategy is used to train the network model to significantly improve the performance of the backbone network. Through a large number of experiments on the four public person re-identification datasets, MSMT17, Market1501, DukeMTMC-ReID and PersonX, it is proved that the proposed method based on the feature Pyramid branch and the non-local attention is competitive compared

基金项目: 国家自然科学基金(61976028); 江苏省研究生科研创新项目(KYCX22_3067)。

收稿日期: 2022-01-04; **修订日期:** 2022-06-07

with some advanced person re-identification methods.

Key words: person re-identification; feature Pyramid branch; attention mechanism; non-local attention; joint loss function strategy

引 言

行人重识别旨在解决跨摄像机检索匹配行人图像或视频的问题^[1],即给定一个目标行人的查询图片或视频,从图库集中精确地检索出相同的身份,并给出查询结果。随着深度卷积神经网络的快速发展,作为智能监督系统关键技术的行人重识别也取得了巨大的进步。过去的研究中,大多数行人重识别的研究是基于有监督的方式,即需要大量的有标签的数据样本。近年来,由于实际视频监控的需求不断增加,使用人工标注图像样本的方式在人力耗费和时间耗费上都存在很大问题^[2],已经不能满足现实的需要,使得深度无监督行人重识别的任务显得极为重要,受到越来越多的关注^[3]。根据是否使用额外的标签数据,无监督的行人重识别主要分为两类:无监督域自适应(Unsupervised domain adaptation, UDA)^[4-6]和完全无监督(Fully unsupervised, FU)^[7-9]行人重识别。在UDA中,有1个标记的源域和1个未标记的目标域。两个域的数据分布不同,可以用来训练在目标域上具有良好泛化能力的模型。由于只提供未标记的图像来训练深度模型,所以完全无监督的行人重识别更具挑战性。一般情况下,由于引入了外部源域,UDA的性能优于完全无监督的行人重识别。然而,UDA训练过程复杂,要求源域和目标域之间的差异并不显著。本文将主要关注完全无监督的行人重识别,为简单起见,将其称为无监督的行人重识别。

最近流行的无监督行人重识别方法^[7-10]主要采用聚类方法对未标记样本生成伪标签,使模型在半监督的方式下训练。采用伪标签生成和模型训练迭代的方法来训练精确的深度模型。尽管现有的方法很有效,但在这个过程中往往忽略了两个重要的因素:(1)由于摄像机的视角变化、光照变化、背景纷杂和障碍物遮挡等因素的影响,导致同一个行人的图像在不同摄像机下的差异很大,特征提取器提取出来的特征代表性不足;(2)对于图片来说,通常使用较大的卷积核来捕捉较远距离的像素之间的关系。然而,传统的卷积神经网络都只是在其时间或空间的很小邻域内进行捕捉,却很难捕获到更远位置的数据依赖关系。基于此,本文提出一种特征金字塔分支(Feature Pyramid branch, FPB)和非局部关注(Non-local attention)模块融合的方法。

1 相关工作

1.1 无监督的行人重识别

当前基于伪标签的无监督行人重识别基线方法通常包括3个阶段:内存字典初始化、伪标签生成和神经网络训练^[11]。以前的工作对部分或整个基线方法都有很大的改进,如:文献[7]将每个单独的样本作为一个聚类,然后逐渐将相似的样本分组到一个聚类中,生成伪标签;文献[12]预测由相似度计算和周期一致性组成的质量伪标签,然后将模型训练为一个多分类问题;文献[13]提出了一种新的自步对比学习框架,逐步创建更可靠的聚类,以完善包含源域和目标域数据集特征的混合记忆字典。本文一方面关注预先训练好的骨干网络的改进,观察到它能够提高性能,尤其是在大规模的数据集测试时,例如MSMT17;另一个方面,在神经网络训练阶段也引入不同的机制来提高模型的性能。使用非局部注意模块,以学习更多的全局特征和局部特征融合,聚集不同尺度的特征,提高特征的多样性。本文除了使用非局部关注模块的策略外,还使用标签平滑正则化损失^[14]减轻伪标签噪声带来的影响。

1.2 特征金字塔网络

特征金字塔网络(Feature Pyramid networks, FPN)是一种广泛应用的多尺度特征表示方法,用于检测不同尺度的目标。FPN^[15]提出了一种自上而下的途径来融合不同分辨率和尺度特征。此外,为了提高FPN的信息传播能力,提出了多种FPN变体,如PANet^[16]增加了自底向上的路径,Bi-FPN^[17]提出了一种新的跨尺度连接,或以非线性方式连接高级和低级特征^[18]。近年来,特征金字塔方法在语义分割^[19]和行人重识别^[20]等多个领域取得了巨大的成功。与基于FPN的方法在连接不同分辨率特征的检测任务中不同的是,这些方法首先通过额外的多尺度分支提取不同尺度的特征,并对其进行学习聚合。如Li等^[19]用多个卷积块提取不同尺度的特征,并融合它们来学习注意,而Zheng等^[20]将提取出的特征图分割成不同尺度的部分,学习每个部分的多分支局部特征。本文将FPN引入行人重识别的任务中,实现在全局和局部尺度上进行特征融合与匹配,实验表明FPN可以给模型的性能带来显著的提升。

1.3 注意力机制

最初的注意模块是针对自然语言处理任务提出的,通常称为多头注意。它通过关注上下文信息减少输入特征的歧义。最近,多头注意和Transformer架构也被证明在各种视觉任务中是有效的。另一种注意力机制,即位置注意模块也被广泛应用于行人重识别任务中。它在计算矩阵相乘时,每个像素只抽取特征图中对应十字位置的像素进行点乘,计算相似度。除了位置注意模块,还有一系列的注意模块,比如全局注意力机制,用来关注整个图像的全局关系。与自然语言处理任务中原有的多头注意模块不同,视觉任务中应用最广泛的注意模块即非局部关注神经网络结构。它的目的是编码不同位置的特征之间的相关性,从而输出更关注的特征。由于卷积核作用的感受野是局部的,经过很多层的累积以后,整个图像的不同区域才能关联起来。因此,它们只能在一个很小的范围内进行处理,对于一张图片没有办法将上下两个点之间的信息进行有效表达,而非局部关注模块可以解决卷积核只能处理局部之间信息的交互,无法进行长距离信息表达的问题。

2 本文方法

本文提出的特征金字塔分支和非局部注意模块融合的网络架构,可以从图像的局部和整体融合行人的特征。方法的整体框架如图1所示。整个框架主要包含两个部分:特征提取和神经网络训练。在特征提取阶段,本文使用的网络架构为嵌入特征金字塔的ResNet-50残差网络,用来提取图像的局部纹理细节。在ResNet-50残差网络的第1~4层之间加入非局部关注作为辅助模块,提取图像的全局特征。

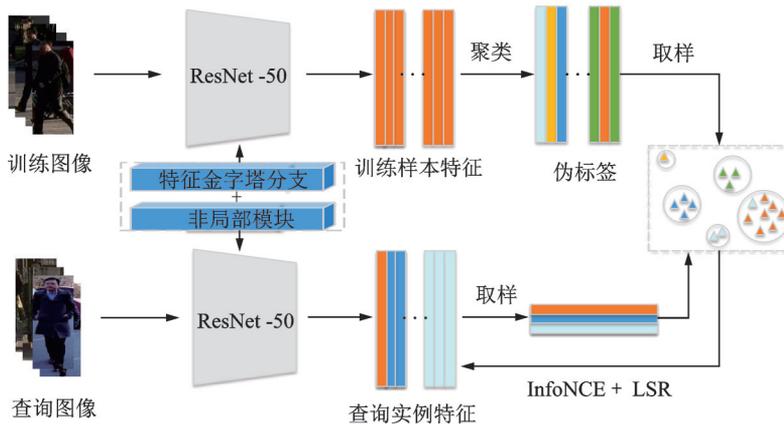


图1 整体框架流程

Fig.1 Overview of the framework

经过局部特征和全局特征的融合,从而得到行人的显著特征,最后使用聚类算法,将属于同一类的特征聚合到一起。由于前面特征提取比较具有代表性,因此聚类效果良好。在神经网络训练阶段,同样使用以上特征提取器提取 Query 特征,然后使用多分类噪声对比估计(Info noise-contrastive estimation, InfoNCE)损失和标签平滑正则化(Label smoothing regularization, LSR)损失进行联合优化,有效减少伪标签的噪声影响,提高行人重识别模型性能。基于一个具有 25.56×10^6 参数的骨干网络 ResNet50,本文提出的特征金字塔分支和非局部注意模块融合的网络架构仅通过引入少于 1.5×10^6 的额外参数就在所有基准测试中实现了最佳性能。

2.1 特征金字塔分支策略

特征金字塔分支架构由 BiFPN 改进而来,它最初用于目标检测任务,但是 FPB 与 BiFPN 存在本质的区别。首先,FPN 是通过聚合不同尺度的特征来解决不同大小的物体出现的问题,而 FPB 是将不同尺度的不同特征整合到最终匹配中。因此,FPB 将单个输出的特征聚合到平均池操作,而与 BiFPN 在每层都有多个输出不同。其次,本文使用比 BiFPN 更宽的滤波器实现内部节点和降采样连接。这是因为,行人重识别的任务需要来自不同尺度的复杂信息来对数量相对较大的身份进行分类。特征金字塔分支的关键结构是两层的 FPN。由于金字塔的所有层都像传统的特征图像金字塔一样使用共享分类器,因此本文在所有特征图中固定特征维度为 256。在特征金字塔分支中不同层间的特征之间存在两种跨尺度的联系。通过最近邻插值算法实现自顶向下的连接,以增加特征图的大小。另外,一个自底向上的连接是使用 2×2 卷积核通过最大池化的方式实现。每一层还存在两个向下采样操作,即从输入节点到输出节点的额外边。如图 2 中曲线箭头所示,类似于 ResNet 中的残差结构本文实现了向下采样。如图 2 所示,自上而下的过程就是把更抽象、语义更强的高层特征图进行上采样,横线连接的过程是将上采样的结果和原来的特征进行融合,横向连接的两层特征在空间尺寸上相同,这样做可以利用底层定位细节信息。然后通过按元素相加,将上采样映射与相应的原图映射(经过 1×1 卷积层用以减少通道维数)合并,此过程迭代进行,以此增强这些特征。最后,在每个合并的特征图上添加 3×3 卷积,生成最终的特征图,减少上采样的混叠效应。对于 FPB 提取出来的特征主要以局部细节为主,对于失去全局特征的图像使用非局部关注模块提取,实现局部和整体特征的融合,提升模型适应性和重识别精度。

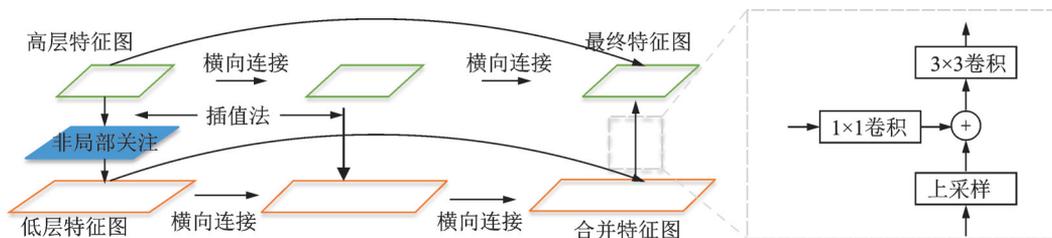


图 2 特征金字塔分支结构

Fig. 2 Feature Pyramid branch structure

2.2 非局部关注模块

计算机视觉中,目前传统的卷积神经网络都只是在其时间或空间的很小的邻域内进行捕捉,却很难捕获到更远位置的数据依赖关系。而非局部关注是一种全局注意力机制,用来关注整张图像的全局关系,因此本文提出的网络方法采用非局部关注模块对图像进行特征提取。具体来说,有

$$y_i = \frac{1}{C(x)} \sum_{v_j} f(x_i, x_j) g(x_j) \quad (1)$$

式中: i 表示某一特定位置; j 表示所有可能关联的位置; x 是输入信号,可以是图像、序列和视频,通常是特征图; y 是和 x 尺度一样的输出图; f 是配对计算函数,计算第 i 个位置和其他所有位置的相关性; g 是1个映射函数,将1个点映射成1个向量,可以看成是计算1个点的特征,目的是进行信息变换; $C(x)$ 是归一化函数,保证变换前后整体信息不变。

为了能让非局部关注操作作为一个组件,可以直接插入任意的神经网络中,将非局部关注模块设计成残差块的形式,让非局部关注操作去学 x 的残差。将非局部定义为

$$z_i = W_z y_i + x_i \tag{2}$$

式中: W_z 实际上是一个卷积操作,它的输出通道数与 x 一致; z_i 表示非局部关注经过残差连接的输出; $+x_i$ 表示残差连接。

图3表示一个非局部关注块。给定从大小为 $H \times W \times C$ 的 T 个特征图获得的特征向量 x ,经过非局部关注公式得到相应输出 y_i 。最终输出 $Z = W_z Y + X$,该输出为原始特征张量 x 加上通过卷积操作 W_z 的输出 y_i 。其中 $T \times H \times W \times 1024$ 表示1024个通道, \oplus 表示元素求和, \otimes 表示矩阵乘法,每一行执行softmax激活函数。最后经过 1×1 卷积恢复输出通道,保证输入输出完全相同。

经过以上设置,将该模块嵌入ResNet-50网络的Layer、Layer 2、Layer 3、Layer 4层之间,效果比较鲁棒。对于跨摄像头的图片,连续变化的图片会将同一行人在不同状态下的特征保留在时序信息中,非局部关注可以更加有效关注全局信息,这恰好解决了行人重识别存在的遮挡、光照、视角以及跨摄像头等问题。

2.3 联合损失策略

无监督的行人重识别中,由于没有行人身份的真实标签,并且伪标签在一定程度上随着模型的优化会发生变化,于是人们使用非参数分类损失,比如InfoNCE损失函数为

$$L_q = -\lg \frac{\exp(q \cdot c^+ / \tau)}{\sum_{i=0}^K \exp(q \cdot c_i / \tau)} \tag{3}$$

给定一个query实例 q ,使用InfoNCE损失函数将它与特征空间中的所有聚类特征 C 进行比较,其中 c^+ 为查询实例 q 的正聚类特征向量, τ 为温度因子。当 q 和它的正聚类特征 c^+ 相似而与其他聚类特征不同时,损失值较低。但是,这种实例级的InfoNCE损失计算出了每个聚类特征向量与每个实例特征向量之间的平均分类损失。因为伪标签不可避免的包含噪声,所以简单的平均总是引入噪声特征向量来表示聚类特征。然而对于在噪声标签条件下训练出来的网络往往存在过拟合问题。针对这个问题,本文在训练时加入标签平滑正则化损失函数,能有效缓解模型训练中的过拟合问题。该损失为

$$L_{LSR} = -(1 - \epsilon) \lg(p(y)) - \frac{\epsilon}{K} \sum_{k=1}^K \lg(p(k)) \tag{4}$$

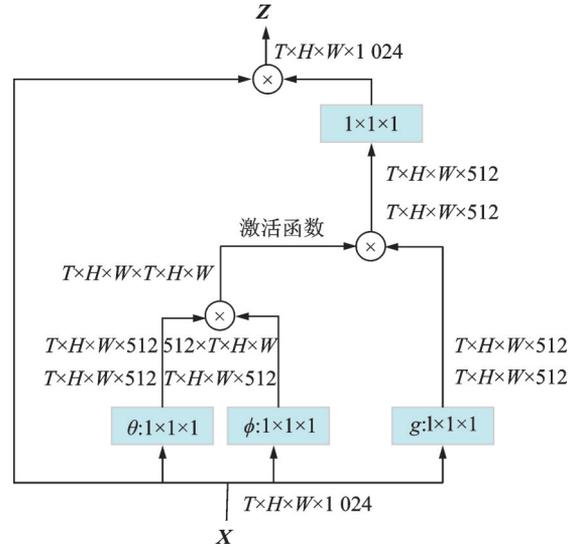


图3 非局部关注模块

Fig.3 Non-local attention module

式中: $k \in \{1, 2, \dots, K\}$ 为训练数据预先定义的类别; $p(k)$ 为样本被预测为 k 的概率; $p(y)$ 为样本预测为真实标签的概率; $\varepsilon \in [0, 1]$ 为一个超参数。由式(3,4)得到本文的联合损失函数为

$$L = \lambda L_q + (1 - \lambda) L_{\text{LSR}} \quad (5)$$

式中 λ 为一个超参数。

3 实验与结果

3.1 数据集和评价指标

本文实验使用的数据集是当前图片行人重识别领域主流的大规模数据集: Market1501^[21], DukeMTMC^[22], MSMT17^[23] 和 PersonX^[24]。4个数据集的训练集和查询集如表1所示,其中查询集由测试集抽取的部分样本组成。Market1501^[20]包含6个摄像头捕捉的1501个身份的32668张标记好的行人图片。DukeMTMC^[22]包含36411张图像,来自超过两个摄像头捕获的1404个身份,以及仅由1个摄像头捕获的408个身份作为干扰,共计1812个行人身份。MSMT17^[23]数据集包含了来自4101个身份的126441张照片,是使用15个摄像头(12个室外,3个室内)拍摄的1个月内4天内不同天气情况的数据集。PersonX^[24]数据集是基于游戏开发引擎Unity创建的1个3D数据合成系统,包含1266个行人身份。

表1 数据集信息

Table 1 Information of datasets

数据集	摄像头/个	图像/张	身份/个		图像/张	
			Train	Test	Train	Query
Market-1501	6	32 668	751	750	12 936	3 368
DukeMTMC-reID	8	36 441	702	702	16 522	2 228
MSMT17	15	126 441	1 041	3 060	32 621	11 659
PersonX	6	45 792	410	856	9 840	5 136

在本文的实验中,采用了累积匹配特性曲线(Cumulative matching characteristic, CMC)和平均查准率(Mean average precision, mAP)来评估提出方法的性能。CMC曲线是反映检索的准确率,用来判断行人重识别模型的排名能力;mAP反映真实排名结果,根据准确率和召回率曲线计算而得。

3.2 实验设置

本文采用ResNet-50^[25]作为特征提取的骨干网络,使用ImageNet^[26]预先训练的参数初始化模型。在Layer 4之后,去掉所有子模块层,加入全局平均池化层,然后加入批处理归一化层^[27]和 L_2 归一化层,生成2048维特征。在测试过程中,利用全局平均池化层的特征来计算距离。在每个轮次的开始,使用DBSCAN^[28]进行聚类,生成伪标签。对于Market 1501、DukeMTMC-reID和PersonX数据集,输入图像大小调整为256像素 \times 128像素,对于MSMT17,大小调整为224像素 \times 224像素。对于训练图像,执行随机水平翻转、像素填充、随机裁剪和随机擦除^[29]。每个小批次包含256个图像,包含16个伪人身份(每个人16个实例图像)。采用Adam优化器对权重衰减 $5e-4$ 的行人重识别模型进行训练。初始学习率设置为 $3.5e-4$,在总共50个Epoch中,每20个Epoch将其降低到以前值的1/10。与文献[13]的聚类方法一样,本文使用DBSCAN和Jaccard距离^[30]对 k 个最近邻进行聚类,其中 $k = 30$ 。 λ 取值为0.5,用于平衡两部分损失函数对模型训练的影响。对于DBSCAN,两个样本之间的最大距离设置为0.4,一个核心点的最小近

邻数设置为4。整个实验基于Ubuntu16.04、Cuda10和Cudnn7.6的环境和Python 3.6、Pytorch 1.3.1和torchvision0.4.2深度学习框架进行,实验的硬件配置包括:4块GPU GTX2080Ti(显存为11 GB)。

3.3 实验结果和对比

为了验证所提出的方法的有效性,本文在4个大规模数据集上进行实验并与一些最新的方法进行了比较。如表2、3所示,与本文比较的方法有BUC(Bottom-up clusterin)^[7]、SSL(Softened similarity learning)^[8]、HCT(Hierarchical clustering with hard-batch triplet loss)^[9]、SPCL(Self-paced contrastive learning)^[13]、MMT(Mutual mean-teaching)^[6]、UGA(Unsupervised graph association)^[31]以及Cluster contrast基线(Baseline)等方法。在MSMT17数据集上,在基线方法加入特征金字塔分支和非局部关注方法以后,mAP提高了6.8%,Top-1提高了6.0%。在Market1501、PersonX和DukeMTMC-reID三个数据集上,加入该方法以后,mAP分别提高了3.9%、4.6%、2.5%,Top-1提高了1.7%、1.5%、0.8%。超过MMT、SPCL、CCL等先进方法。因此,本文所提出的基于特征金字塔分支和非局部关注的行人重识别方法在综合性能上优于所对比的先进的行人重识别方法。

表2 在Market1501、DukeMTMC-reID数据集上与先进方法的比较

Table 2 Comparison with state of the art on Market1501 and DukeMTMC-reID datasets

Method	Market 1501				DukeMTMC-reID			
	Source	mAP	Top-1	Top-5	Source	mAP	Top-1	Top-5
BUC ^[7]	—	38.3	66.2	79.6	—	27.5	47.4	62.6
SSL ^[8]	—	37.8	71.7	83.8	—	28.6	52.5	63.5
MMCL ^[12]	—	45.5	80.3	89.4	—	—	—	—
MMCL ^[12]	Duke	60.4	84.4	92.8	Market	51.4	72.4	82.9
HCT ^[9]	—	56.4	80.0	91.6	—	50.7	69.6	83.4
CycAs ^[32]	—	64.8	84.8	—	—	60.1	77.9	—
AD-Cluster ^{++[33]}	Duke	68.3	86.7	94.4	Market	54.1	72.6	82.5
UGA ^[31]	—	70.3	87.2	—	—	53.3	75.0	—
MMT ^[6]	MSMT17	75.6	89.3	95.8	Market	65.1	78.9	88.8
SPCL ^[13]	—	73.1	88.1	95.1	—	—	—	—
SPCL ^[13]	MSMT17	77.5	89.7	96.1	Market	68.8	82.9	90.1
CCL ^[11]	—	82.6	93.0	97.0	—	72.8	85.7	92.0
Ours	—	86.5	94.7	98.0	—	75.3	86.5	92.9

表3 在MSMT17、PersonX数据集上与先进方法的比较

Table 3 Comparison with state of the art on MSMT17 and PersonX datasets

Method	MSMT17				PersonX			
	Source	mAP	Top-1	Top-5	Source	mAP	Top-1	Top-5
ECN ^[5]	Duke	10.2	30.2	41.5	—	—	—	—
TAUDL ^[34]	—	12.5	28.4	—	—	—	—	—
UTAL ^[35]	—	13.1	31.4	—	—	—	—	—
SPCL ^[13]	—	19.1	42.3	55.6	—	72.3	88.1	96.6
UGA ^[31]	—	21.7	49.5	—	—	—	—	—
MMT ^[6]	Market	24.0	50.1	63.5	MSMT17	78.9	90.6	96.8
CycAS ^[32]	—	26.7	50.1	—	—	—	—	—
SPCL ^[13]	Market	26.8	53.7	65.0	Market	78.5	91.1	97.8
CCL ^[11]	—	33.3	63.3	73.7	—	84.8	94.5	98.4
Ours	—	40.1	69.3	79.7	—	89.4	96.0	99.0

3.3.1 特征金字塔分支策略的有效性

在4个数据集上进行的消融研究如图4所示。FPB表示只使用特征金字塔分支策略,Non-local表示只使用非局部关注模块,FPB+non-local表示同时使用特征金字塔分支策略和非局部关注模块。从图4可以看出,两种方法的组合可以共同提高整个模型的鲁棒性,并且特征金字塔分支策略的效果要优于非局部关注模块的效果。具体的 Top-1 数值如表4所示。

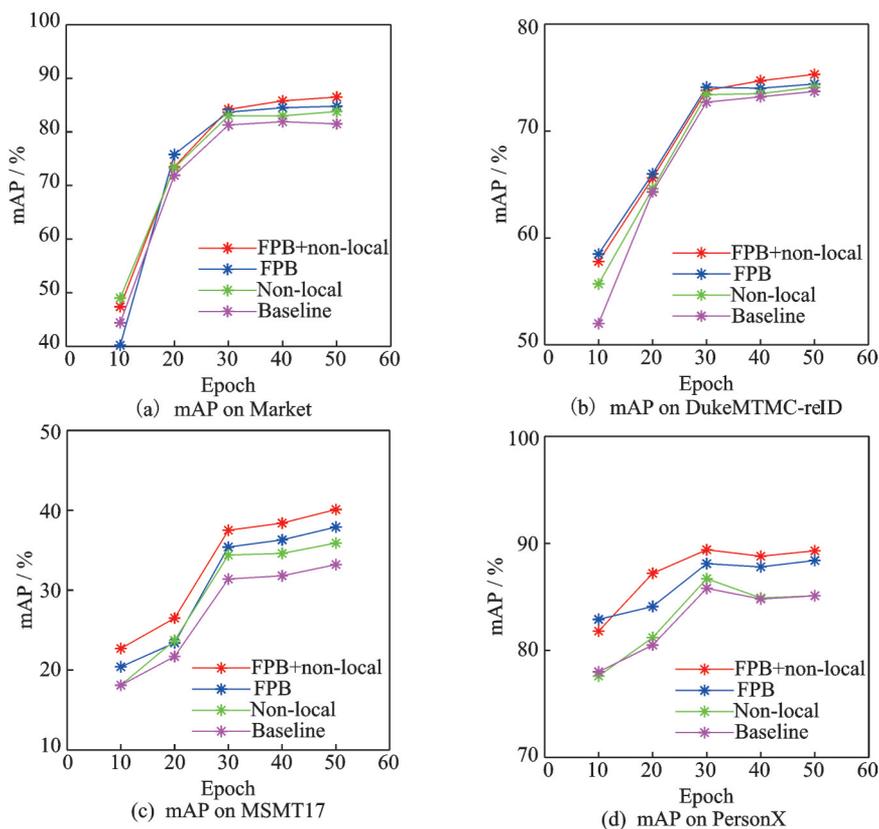


图4 在4个数据集上的消融实验

Fig.4 Ablation studies on four datasets

表4 在4个数据集上的消融实验

Table 4 Ablation studies on four datasets

%

Method	MSMT17		Market1501		PersonX		DukeMTMC-reID	
	mAP	Top-1	mAP	Top-1	mAP	Top-1	mAP	Top-1
CCL ^[11]	33.3	63.3	82.6	93.0	84.8	94.5	72.8	85.7
FPB	37.9	68.3	84.8	93.8	88.4	95.7	74.4	86.0
Non-local	35.9	65.7	83.8	93.4	86.7	95.1	74.1	86.0
FPB+non-local	40.1	69.3	86.5	94.7	89.4	96.0	75.3	86.5

3.3.2 非局部关注模块的有效性

从表4可见,使用非局部关注模块时,在Market1501数据集上的mAP、Top-1值分别为83.8%、93.4%,比CCL提高了1.2%、0.4%;在DukeMTMC-reID数据集上的mAP、Top-1值分别为74.1%、

86.0%,比CCL提高了1.3%、0.3%;在MSMT17数据集上的mAP、Top-1值分别为35.9%、65.7%,比CCL提高了2.6%、2.4%;在PersonX数据集上的mAP、Top-1值分别为86.7%、95.1%,比CCL提高了1.9%、0.6%。由于非局部关注是一种全局注意力机制,能够很好地关注整张图像的全局关系。由此可见,非局部关注模块能够辅助特征金字塔分支策略,进一步证明了其有效性。

3.4 对模型效率的影响

特征金字塔分支和非局部模块对模型效率的影响如表5所示。从表5可以看出,本文所提出的特征金字塔分支和非局部关注模块融合方法对模型影响是不同的。在Market-1501和DukeMTMC-reID数据集上模型训练时长提高,而在MSMT17和PersonX数据集上模型训练时长却减少了,模型效率实现了提高。

表5 特征金字塔分支和非局部模块对模型效率的影响

Table 5 Influence of feature Pyramid branches and nonlocal modules on model efficiency

Method	Time/(h:min:s)			
	Market 1501	DukeMTMC-reID	MSMT17	PersonX
Baseline	1:58:06	2:01:27	10:15:02	3:02:05
Non-local	2:01:46	2:13:11	5:30:49	2:12:54
FPB	1:59:34	2:06:21	5:23:22	2:18:10
FPB+non-local	3:27:03	3:35:31	5:33:16	2:16:57

3.5 模型参数量对比

与先进Re-ID方法的参数量对比如图5所示。从图5中可以看出,本文所提出的特征金字塔分支和非局部关注模块融合方法跟目前主流的Re-ID方法相比,参数量最少($<28 \times 10^6$)并且性能依然良好。其他具有接近性能的先进方法通常需要超过 32×10^6 个参数的模型,甚至比参数量更多的模型AsNet和Pyramid性能更强。

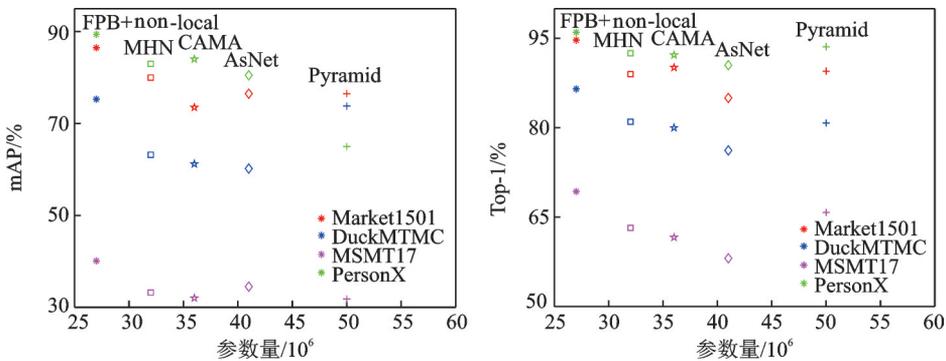


图5 与先进Re-ID方法的参数量对比

Fig.5 Parameter comparison with advanced Re-ID method

4 结束语

本文对目前主流的特征提取器ResNet-50网络进行了改进,对聚类的损失函数进行联合损失处理,显著提高了行人重识别的精度。本文提出的特征金字塔分支策略能够提取不同尺度的特征并聚合,同时结合使用非局部关注模块作为辅助模块提取全局特征,显著提高了模型的性能。本文在MSMT17、Market1501、PersonX和DukeMTMC-reID大型数据集上进行了实验,证明本文方法具有突出的优越性,消融实验结果也证明了网络中各个模块的有效性。虽然本文方法能够有效地提取更具代表性的特征,但是在聚

类过程中仍然存在一些具有噪声的伪标签,如何降低噪声标签的影响还需要进一步探索和研究。

参考文献:

- [1] 张云鹏,王洪元,张继,等.近邻中心迭代策略的单标注视频行人重识别[J].软件学报,2020,32(12):4025-4035.
ZHANG Yunpeng, WANG Hongyuan, ZHANG Ji, et al. One-shot video-based person re-identification based on neighborhood center iteration strategy[J]. *Journal of Software*, 2021, 32(12): 4025-4035.
- [2] 张文文,王洪元,万建武,等.基于稀疏学习的行人重识别算法[J].数据采集与处理,2018,33(5):855-864.
ZHANG Wenwen, WAN Hongyuan, WANG Jianwu, et al. A sparsity-learning-based person re-identification algorithm[J]. *Journal of Data Acquisition and Processing*, 2018, 33(5): 855-864.
- [3] WANG Hongyuan, WU Linyu, CHEN Fuhua, et al. Common-covariance based person re-identification model[J]. *Pattern Recognition Letters*, 2021, 146: 77-82.
- [4] WEI Longhui, ZHANG Shiliang, GAO Wen, 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*. [S.l.]: IEEE Computer Society, 2018: 79-88.
- [5] ZHONG Zhong, ZHENG Liang, LUO Zhiming, et al. Invariance matters: Exemplar memory for domain adaptive person re-identification[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.]: IEEE Computer Society, 2019: 598-607.
- [6] GE Yixiao, CHEN Dapeng, LI Hongsheng. Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification[EB/OL].(2020-01-30)[2022-01-01].<https://arxiv.org/pdf/2001.01526.pdf>.
- [7] LIN Yutian, DONG Xuanyi, ZHENG Liang, et al. A bottom-up clustering approach to unsupervised person re-identification [C]// *Proceedings of the AAAI Conference on Artificial Intelligence*. [S.l.]: AAAI, 2019: 8738-8745.
- [8] LIN Yutian, XIE Lingxi, WU Yu, et al. Unsupervised person re-identification via softened similarity learning[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.]: IEEE Computer Society, 2020: 3390-3399.
- [9] ZENG Kaiwei, NING Munan, WANG Yao, et al. Hierarchical clustering with hard-batch triplet loss for person re-identification[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.]: IEEE Computer Society, 2020: 13657-13665.
- [10] FAN Hehe, ZHENG Liang, YAN Yi, 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.
- [11] DAI Zuo Zhuo, WANG Guangyuan, YUAN Weihao, et al. Cluster contrast for unsupervised person re-identification[EB/OL].(2021-11-19)[2022-01-01].<http://128.84.21.203/pdf/2103.11568>.
- [12] WANG Dongkai, ZHANG Shiliang. Unsupervised person re-identification via multi-label classification[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.]: IEEE Computer Society, 2020: 10981-10990.
- [13] GE Yixiao, ZHU Feng, CHEN Dapeng, et al. Self-paced contrastive learning with hybrid memory for domain adaptive object Re-ID[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 11309-11321.
- [14] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]// *Proceedings of the IEEE Conference On Computer Vision And Pattern Recognition*. [S.l.]: IEEE, 2016: 2818-2826.
- [15] LIN Tsung-Yi, DOLLÁR P, GIRSHICK R, et al. Feature Pyramid networks for object detection[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.]: IEEE Computer Society, 2017: 2117-2125.
- [16] LIU Shu, QI Lu, QIN Haifeng, et al. Path aggregation network for instance segmentation[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.]: IEEE Computer Society, 2018: 8759-8768.
- [17] TAN Mingxing, PANG Ruoming, LE Q V. Efficientdet: Scalable and efficient object detection[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.]: IEEE Computer Society, 2020: 10781-10790.
- [18] KONG Tao, SUN Fuchun, TAN Chuanqi, et al. Deep feature pyramid reconfiguration for object detection[C]// *Proceedings of the European Conference on Computer Vision (ECCV)*. Cham: Springer, 2018: 169-185.
- [19] LI Hanchao, XIONG Pengfei, AN Jie, et al. Pyramid attention network for semantic segmentation[EB/OL].(2018-11-25)[2022-01-01].<https://arxiv.org/pdf/1805.10180.pdf>.
- [20] ZHENG Feng, DENG Chen, SUN Xing, et al. Pyramidal person re-identification via multi-loss dynamic training[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.]: IEEE, 2019: 8514-8522.

- [21] ZHENG Liang, SHEN Liyue, TIAN Liu, et al. Scalable person re-identification: A benchmark[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Computer Society, 2015: 1116-1124.
- [22] RISTANI E, SOLERA F, ZOU R, et al. Performance measures and a data set for multi-target, multi-camera tracking[C]//Proceedings of 14th European Conference on Computer Vision, ECCV 2016. [S.l.]: Springer, 2016, 9914: 17-35.
- [23] WEI Longhui, ZHANG Shiliang, GAO Wen, 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. [S.l.]: IEEE Computer Society, 2018: 79-88.
- [24] SUN Xiaoxiao, ZHENG Liang. Dissecting person re-identification from the viewpoint of viewpoint[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE Computer Society, 2019: 608-617.
- [25] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE Computer Society, 2016: 770-778.
- [26] DENG Jia, WEI Dong, SOCHER Richard, et al. Imagenet: A large-scale hierarchical image database[C]//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE Computer Society, 2009: 248-255..
- [27] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//Proceedings of International Conference on Machine Learning. [S.l.]: PMLR, 2015: 448-456.
- [28] ESTER M, KRIEGEL H, SANDER J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[C]// Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. Menlo Park: AAAI Press, 1996: 226-231.
- [29] ZHONG Zhun, ZHENG Liang, KANG Guoliang, et al. Random erasing data augmentation[C]//Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2020: 13001-13008.
- [30] ZHONG Zhun, ZHENG Liang, CAO Donglin, et al. Re-ranking person re-identification with k-reciprocal encoding[C]//Proceedings of the IEEE Conference on Computer Vision And Pattern Recognition. [S.l.]: IEEE Computer Society, 2017: 1318-1327.
- [31] WU Jinlin, YANG Yang, LIU Hao, et al. Unsupervised graph association for person re-identification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. [S.l.]: IEEE Computer Society, 2019: 8321-8330.
- [32] WANG Zhongdao, ZHANG Jingwei, ZHENG Liang, et al. CycAs: Self-supervised cycle association for learning re-identifiable descriptions[C]//Proceedings of Computer Vision-ECCV 2020: 16th European Conference. Glasgow, UK: Springer International Publishing, 2020: 72-88.
- [33] ZHAI Yunpeng, LU Shijian, YE Qixiang, et al. Ad-cluster: Augmented discriminative clustering for domain adaptive person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE Computer Society, 2020: 9021-9030.
- [34] LI Minxian, ZHU Xiantian X, GONG Shangang. Unsupervised person re-identification by deep learning tracklet association [C]//Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer, 2018: 737-753.
- [35] LI Minxian, ZHU Xiantian X, GONG Shangang. Unsupervised tracklet person re-identification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 42(7): 1770-1782.

作者简介:



孙明浩(1996-),男,硕士研究生,研究方向:计算机视觉和行人重识别,E-mail: 873054203@qq.com。



王洪元(1960-),通信作者,男,教授,研究方向:图像处理、人工智能和模式识别。



吴琳钰(1996-),女,硕士研究生,研究方向:计算机视觉和视频行人重识别。



张继(1981-),男,副教授,研究方向:计算机视觉和行人检测。



周群颖(1999-),女,硕士研究生,研究方向:计算机视觉和目标检测。