·图形图像处理·　　文章编号：1000-3428(2022)03-0271-09　　文献标志码：A　　中图分类号：TP391

基于多尺度多粒度融合的行人重识别方法

符进武，范自柱，石林瑞，郭心悦，黄祎婧

（华东交通大学 理学院，南昌 330013）

摘　要：行人重识别是指利用计算机视觉技术在给定监控的图像中识别目标行人，受拍摄场景视角和姿势变化、遮挡等因素的影响，现有基于局部特征的行人重识别方法所提取的特征辨别力差，从而导致重识别精度较低。为有效地利用特征信息，提出一种多尺度多粒度融合的行人重识别方法MMF-Net。通过多个分支结构学习不同尺度和不同粒度的特征，并利用局部特征学习优化全局特征，以加强全局特征和局部特征的关联性。同时，在网络的低层引入语义监督模块以提取低层特征，并将其作为行人图像相似性度量的补充，实现低层特征和高层特征的优势互补。基于改进的池化层，通过结合最大池化和平均池化的特点获取具有强辨别力的特征。实验结果表明，MMF-Net方法在Market-1501数据集上的首位命中率和mAP分别为95.7%和89.1%，相比FPR、MGN、BDB等方法，其具有较优的鲁棒性。

关键词：行人重识别；特征学习；局部特征；低层特征；池化；多尺度多粒度融合

开放科学（资源服务）标志码（OSID）：

中文引用格式：符进武，范自柱，石林瑞，等. 基于多尺度多粒度融合的行人重识别方法[J]. 计算机工程，2022，48(3)：271-279.

英文引用格式：FU J W，FAN Z Z，SHI L R，et al. Person re-identification method based on multi-scale and multi-granularity fusion[J]. Computer Engineering，2022，48(3)：271-279.

Person Re-Identification Method Based on Multi-Scale and Multi-Granularity Fusion

FU Jinwu，FAN Zizhu，SHI Linrui，GUO Xinyue，HUANG Yijing

（School of Science，East China Jiaotong University，Nanchang 330013，China）

【Abstract】Person re-identification refers to the use of computer vision technology to recognize the target person in a given monitored image. Factors，such as the change of shooting scene angle，posture，and occlusion，among others，result in poor feature discrimination extraction with existing person re-identification methods based on local features，associating these methods with low re-identification accuracy. To make effective use of feature information，this paper proposes a person re-identification method，MMF-Net，based on multi-scale and multi-granularity fusion，whereby features with different scales and granularity are learned by multiple branch structures. Local feature learning is used to optimize global features，thereby enhancing the correlation between global features and local features. A semantic supervision module is also introduced into the lower layer of the network to extract low-level features，which are used as a supplement to person image similarity measurement，to benefit from the complementary advantages of low-level and high-level features. On the improved pooling layer，the features with strong discrimination are obtained by combining the features of maximum and average pooling. The experimental results show that the Rank-1 and mAP of MMF-Net method is 95.7% and 89.1%，respectively. Compared with FPR，MGN，BDB method，it has better robustness.

【Key words】person re-identification；feature learning；local feature；low-level feature；pooling；multi-scale and multi-granularity fusion

DOI：10. 19678/j. issn. 1000-3428. 0060189

基金项目：国家自然科学基金（61991401，61673097，61702117）；江西省自然科学基金重点项目（20192ACBL20010）。

作者简介：符进武（1998—），男，硕士研究生，主研方向为图像处理、模式识别；范自柱（通信作者），教授；石林瑞、郭心悦、黄祎婧，硕士研究生。

收稿日期：2020-12-04　　**修回日期：**2021-02-20　　**E-mail：**2654695508@qq.com

计算机工程 www.ecice06.com

0 概述

行人重识别是跨设备的图像检索技术，指在给定监控的行人图像中，检索该行人在其他摄像头中出现的全部图像。该技术广泛应用于智能视频监控、安保、刑侦等领域，是当前计算机视觉的研究热点[1]。受拍摄场景复杂性（如低分辨率、遮挡、视角和姿势变化、光照不同等）的影响，行人重识别任务面临诸多挑战。

基于深度学习的方法在行人重识别领域中占据了主导地位，相比传统手工提取方法[2-3]具有更优的性能。文献[4]提出以ResNet-50为主干网络、以行人的ID作为训练标签的IDE网络。文献[5]将行人的性别、头发、衣着等属性信息作为额外的标注，并结合ID损失和属性损失以增强网络的泛化能力。在深度度量学习方面，文献[6]将三元组损失引入到行人重识别中，通过结合度量学习和表征学习对网络进行训练，以改进行人重识别效果。文献[7]使用生成对抗网络（Generative Adversarial Network，GAN）模拟训练数据的分布来生成新图像，并将其与原数据集图像共同进行训练，以提升网络的鲁棒性、行人重识别跨域和泛化能力，从而解决数据集规模偏小的问题。在行人图像相似度度量方面，文献[8]利用k倒排编码方法对结果进行重排序，有效地提高行人重识别的准确率。

早期，研究人员仅关注行人图像的全局特征，却忽略了图像局部细节信息。局部特征具有更丰富的细节信息和更强的辨别性，结合全局和局部特征表示行人图像逐渐成为主流方向[9]，并取得了较优的效果。目前，基于局部特征的行人重识别方法主要包括基于人体姿态和骨架关键点等先验知识的方法、基于注意力机制的方法、仅简单地将图像划分成若干条带进行学习的方法。文献[10]通过骨架关键点提取网络来提取人体关键点，进而获取人体结构的感兴趣区域。文献[11]采用姿态估计模型估计人体关键点，同时通过仿射变换使得两张图像中的行人对齐。但这种方法需要额外引入姿态估计模型，并且不同源域数据集之间的偏差会影响姿态估计模型的鲁棒性。文献[12]通过构建自上而下的注意力机制网络增强空间像素特征的显著性。文献[13]结合通道和空间信息，避免注意力机制过度集中于前景。由于注意力机制本身的局限性，因此其提取的区域缺乏语义解释且不包括具有辨别力的特征。文献[14]提出基于部分的卷积基线（Part-based Convolutional Baseline，PCB），通过将图像均匀水平划分成条带来学习局部特征，同时设计一种部分精炼池化（Refined Part Pooling，RPP）策略使得分块内部具有一致性。PCB仅使用了局部特征，并没有结合全局特征学习，此外，通过引入RPP使得网络无法进行端到端训练。文献[15]提出多分支结构的多粒度网络（Multi-Granularity Network，MGN），每个分支可以学习全局特征和不同粒度的局部特征，因此融合这些特征能够较全面地表示行人图像。虽然MGN结合了全局和局部特征，但结构上仍然将两者分开进行单独学习，因此割裂了全局和局部特征的关联性。此外，MGN选择在主干网络中间进行分支，仅共享网络前面的权重参数，从而增加了整个网络的参数量且降低了计算效率。

高层特征包含丰富的语义信息，有助于模型对其进行辨别。目前大部分方法仅关注高层特征的使用，但是当不同类别的图像具有较高的语义相似度时（常见于行人重识别数据集），仅依靠高层特征的辨别能力无法进一步提升模型性能。低层特征包含更多的细节信息，用于度量细粒度相似性。由于低层特征具有过多的干扰信息而无法直接使用，一般将高层特征和低层特征相融合，但是直接融合带入的干扰信息可能会削弱高层语义信息的辨别力。

本文提出一种多尺度多粒度融合的行人重识别方法，利用不同特征的优势互补，弥补单一类别特征的不足。通过将提取的低层特征作为行人图像相似性度量的补充，在网络的高层中学习多尺度和多粒度的特征表示，并结合平均池化和最大池化的特性提取具有强辨别力的特征。

1 多尺度多粒度融合网络

1.1 MMF-Net网络结构

本文提出的多尺度多粒度融合（MMF-Net）网络结构如图1所示。MMF-Net的主干网络可以选择当前的经典网络，如ResNet、GoogLeNet和VGG。由于ResNet-50在行人重识别任务中表现出较好的效果，因此本文选择ResNet-50作为主干网络，并移除了后面的平均池化层和全连接层。此外，为丰富特征的粒度，取消了第5层中的下采样操作。因此，主干网络的下采样率为1/16。

1）语义监督模块

本文对ResNet-50第4层特征进行采样，引入语义监督模块（Semantic Supervision Module，SSM）后，特征图先经过池化层聚合成1 024维特征向量，然后使用全连接层整合到512维，作为该分支的特征表达。SSM提取的特征通过softmax损失和triplet损失进行约束。

网络不同层的特征包含的信息差别较大，如何选择引入SSM的位置对模型性能有着重要的影响。本文选择引入SSM的位置，主要考虑以下3个方面：（1）网络低层感受野较小，主要关注图像的空间细节，能够利用更多的细粒度特征信息，但也会带入杂乱背景和歧义语义的干扰信息；（2）网络高层感受野较大，更多关注语义特征，提取具有鉴别力的特征，却缺少了细节信息；（3）基于低层特征和高层特征的特性分析，本文选择在主干网络的次高层（即主干网络的第4层）引入SSM。次高层特征不仅具有较优的语义信息和细节信息，同时避免产生过多的干扰信息。针对模块的数量问题，过多或过早的约束可能导致模型出现过拟合现象，因此不能盲目增加

计算机工程 www.ecocon.com

SSM的数量。此外，SSM提取的特征不能与高层特征相融合，而是作为独立的特征表示，在测试时将其与高层特征拼接作为行人图像的最终表示，实现低层和高层特征的优势互补。

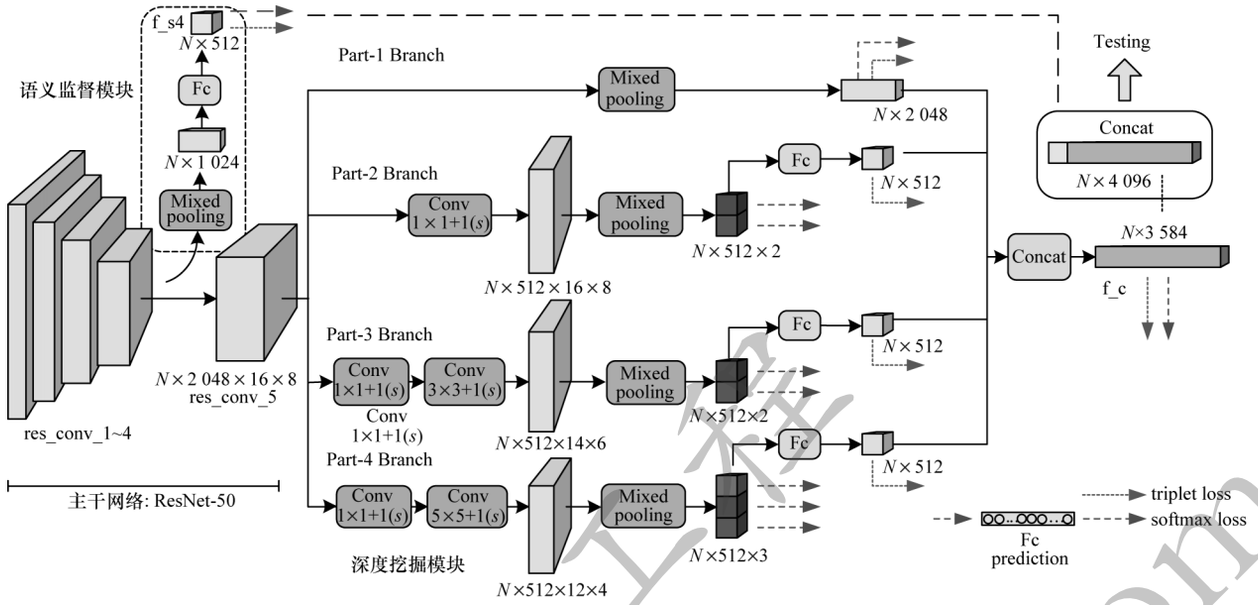

图1　多尺度多粒度融合网络结构

Fig.1　Structure of multi-scale and multi-granularity fusion network

2)深度挖掘模块

在主干网络中有4个分支，第1个分支不进行降维、划分条带等其他操作，称为原始分支。其他3个分支分别使用1×1、3×3、5×5卷积核进一步提取特征，同时将维度降至512维。为减少参数量且增加非线性，在使用3×3、5×5卷积核前先采用1×1卷积核进行降维操作。本文设置卷积的步长为1和填充为0，经卷积后形成了特征金字塔。MMF-Net网络中4个分支的参数设置如表1所示。

表1　MMF-Net网络中4个分支的参数设置

Table 1　Parameter settings of the four branch in MMF-Net network

分支	特征图大小/像素	维度	粒度
Part-1	16×8	2 048	1
Part-2	16×8	512	2
Part-3	14×6	512	2
Part-4	12×4	512	3

特征金字塔首先将池化后的特征图均匀地划分成若干水平条带，以提取局部特征用于学习分类；然后将特征图拉平成一个n×512维（n是划分水平条带的数目）特征向量，将其传入全连接层并整合到512维，作为该分支最后的特征表示；最后通过拼接的方式融合4个分支的特征，以得到一个包含更多鉴别信息的特征。

在该模块中，利用不同感受野的卷积层进一步提取特征，从而形成不同尺度和不同感受野的特征图。相对感受野越大的特征图，每个部分对应的原始图像区域则更大，因此将特征图划分为更多条带，不会因分区细小而丢失过多的语义信息。粒度越小（即条带越多）的分支能学习到更加精细的特征。此外，本文并没有割裂全局和局部特征学习，而是将特征图拉平后作为全局特征进行度量学习，通过局部特征学习使得全局特征更精细化。将两者相融合的设计加强了全局和局部之间的联系，使得全局特征能包含更多的细节信息，因此在测试阶段能够减少局部特征的使用。最后，每个分支学习了不同粒度和不同尺度的特征信息，通过融合各分支得到一个辨别能力更强的特征。

在测试阶段中，本文将SSM提取的低层特征和上述的融合特征（高层特征）串联，作为行人图像最后的特征表示。

3)混合池化

混合池化的结构如图2所示。

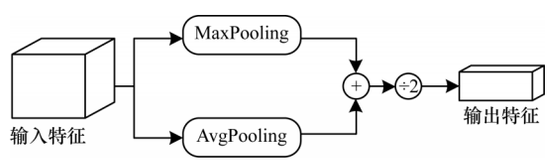

图2　混合池化结构

Fig.2　Structure of mixed pooling

网络高层的特征图尺寸较小，且包含的高级语义信息能够有效地识别模型，因此在网络最后的池化层一般使用平均池化来聚合特征图的全局信息，以保证信息的完整性。最大池化能够保留特征图中响应强烈的信息。因此本文结合两者的特性，使用平均池化和最大池化在保留特征图鉴别信息完整性的同时凸显其中最具有鉴别性的信息。本文将这种

计算机工程 www.ecice06.com

池化方式称为混合池化。将特征图F分别输入到平均池化层和最大池化层，最后将池化后的特征进行融合。混合池化计算如式(1)所示：

$$F_{\text{out}}=\frac{1}{2}\times(A_{\text{AvgPool}}(F)+M_{\text{MaxPool}}(F)) \tag{1}$$

1.2 损失函数

为了使模型能获得较好的表征学习能力，本文结合多分类任务和深度度量学习对模型所提取的特征进行约束。

针对模型的分类问题，通过一个全连接层作为分类器预测每一类的得分，再采用softmax损失函数进行优化。给定进行分类识别的特征为f，softmax损失函数如式(2)所示：

$$L_{\text{soft max}}=-\sum_{i=1}^{N}q_i\ln\frac{e^{\boldsymbol{W}_i^{\mathrm{T}}f}}{\sum_{k=1}^{N}e^{\boldsymbol{W}_k^{\mathrm{T}}f}}$$

$$q_i=0,y\neq i;\ q_i=1,y=i \tag{2}$$

其中：$\boldsymbol{W}_i$为全连接层对应i类别的权重向量；N为训练集的类别总数；y为输入图像的真实标签。由于测试集和训练集的行人ID是不同的，为了防止模型过度拟合训练集中的行人ID，提高模型的泛化能力，本文采用softmax损失函数的另一种形式Label smoothing[16]，q_i的表示如式(3)所示：

$$q_i=\begin{cases}\dfrac{\varepsilon}{N}, & y\neq i\\ 1-\dfrac{N-1}{N}\varepsilon, & y=i\end{cases} \tag{3}$$

其中：ε为一个在[0,1]之间的小常量，用于降低模型在训练集上的置信度，本文设置ε为0.1。

在模型的深度度量学习过程中，本文基于triplet损失函数进行改进[6]。triplet损失函数如式(4)所示：

$$L_{\text{triplet}}=\sum_{i=1}^{P}\sum_{a=1}^{K}\left[m+\max_{p=1,2,\cdots,K}D\left(f_a^i,f_p^i\right)-\min_{\substack{j=1,2,\cdots,P\\ n=1,2,\cdots,K\\ j\neq i}}D\left(f_a^i,f_n^i\right)\right]_+ \tag{4}$$

其中：f_a^i、f_p^i和f_n^i分别为固定样本、正样本和负样本对应的特征；m为一个阈值超参数；$[\cdot]_+$为$\max(\cdot,0)$。在训练过程中每一个批次包含P个ID的行人，每个行人有K张不同的图像，即一个批次总共有$P\times K$幅图像。本文在批次中每幅图像(固定样本)找出其对应的最难正样本和负样本并计算triplet损失，通过优化难样本的损失，有效地提高模型的表征学习能力和泛化能力。

在MMF-Net网络中，为了使softmax损失函数和triplet损失函数相互作用，结合这两种损失对每个部分提取的特征进行约束。深度挖掘模块中3个分支需考虑到分块局部特征的不对齐问题，本文仅使用softmax损失对局部特征进行约束。然而，拉平特征图得到的特征向量保留了全局信息，可用triplet损失对其进行约束。

2 实验与结果分析

本节对MMF-Net在4个主流行人重识别数据集(Market-1501[17]、DukeMTMC-ReID[18]、CUHK03[19]、MSMT17[20])上进行实验分析。本文在4个数据集上将MMF-Net与当前最新的方法进行性能对比。此外，通过一系列(以Market-1501数据集为例)消融实验，评估MMF-Net各模块的有效性。

2.1 数据集

Market-1501数据集是通过6个摄像头拍摄完成的图像，使用可变形部件模型(DPM)检测器[21]检测行人边界框。Market-1501数据集总共包含1 501个行人的36 036张图像，分为训练集和测试集。其中训练集由750个行人的12 936张图像组成。测试集有751个行人，包含19 732张图库图像和3 368张查询图像。其中，图库中存在6 617张干扰图像，包括误检和标注不合格的图像。

DukeMTMC-ReID数据集是DukeMTMC数据集的行人重识别子集，是目前最大的行人重识别数据集之一，包含8个摄像头拍摄的36 411张图像，有1 404个有效的行人身份和408个干扰身份(只出现在一个摄像头中)。图像采用人工标注的方法裁剪出行人边界框。702个行人的16 522张图像用于训练，其余702个行人的17 661张图库图像和2 228张查询图像用于测试。

CUHK03数据集采集了包含1 467个行人的14 097张图像，其中每个行人拍摄于2个摄像机(1对摄像机)。数据集提供两种类型行人边界框，分别由手工标注和DPM检测。关于数据集的使用协议，本文使用新的训练/测试协议[8]。训练集包含767个行人身份，测试集(查询图像和图库图像)包含700个行人身份。

MSMT17数据集是目前公开的规模最大和更接近真实场景的行人重识别数据集。数据集使用15个摄像头在不同天气条件下采集不同时段(上午、中午、下午)的行人图像。行人边界框使用Faster RCNN[22]检测裁剪。数据集总共有4 101个行人的126 441张图像，采用1∶3的比例划分为训练集和测试集。最后训练集包含1 401个行人的32 621张图像，测试集包含2 700个行人的11 659张查询图像和82 161张图库图像。复杂的场景和庞大的数目使得MSMT17成为最具挑战性的行人重识别数据集。此外，由于该数据集发布时间尚短，因此在该数据集上验证的方法较少。

在实验中，本文使用累积匹配特性(CMC)曲线中的Rank-1精度和平均精度均值(mAP)作为评估模型性能的指标。Rank-1为搜索结果中第一张图像是正确结果的概率。mAP为所有查询图像在精度召

计算机工程 www.ecice.com

回曲线下方的面积（AP）的平均值，能更加全面地评估ReID算法的性能。此外，为了简化评估过程，本文所有的实验均采用单查询模式（即一个ID仅使用某摄像头中的一张图像作为查询图像），并且没有使用重新排序算法[8]。

2.2　实验设置

本文使用ImageNet中预训练的ResNet-50的权重参数来初始化MMF-Net主干网络。将输入图像大小调整为256×128像素。在训练阶段，通过水平翻转、随机擦除和归一化增强数据，而在测试阶段仅使用归一化对数据进行处理。为满足三元组损失的训练要求，从训练集中随机挑选P个身份，每个身份由K张图像组成一个训练批次，因此训练批次大小为$P\times K$。本文设置P=8，K=4，批次大小为32。对于三元组损失的阈值参数，本文设置m=0.3。在训练过程中采用自适应优化器（Adam）优化模型参数，并使用权重衰减因子为0.000 5的L2正则化。基本学习率设置为3.5×10^{-4}。在训练过程中学习率在前10个周期内从3.5×10^{-6}线性增加到3.5×10^{-4}，并在第60个周期和第130个周期分别下降到3.5×10^{-5}和3.5×10^{-6}。模型训练总共持续240个周期。本文将测试阶段串接第4个阶段提取的特征f_s4和后面融合的特征f_c作为查询图像的特征表示。其中，f_s4较小的比重设置为α，本文设置α=0.8。本文在Pytorch框架上搭建模型，并使用一个NVIDIA TITAN Xp GPU进行加速训练。在Market-1501数据集上完整地训练一个模型大概需要6 h。在4个数据集上的所有实验都保持与上述相同的实验设置

2.3　不同方法对比

本文选择MMF-Net方法与主流方法SVDNet[23]、Part-aligned[24]、PCB+RPP[14]、MLFN[25]、HA-CNN[26]等进行对比。

在Market-1501数据集上不同方法的评价指标对比如表2所示。本文方法MMF-Net的Rank-1为95.7%，mAP为89.1%。MMF-Net的Rank-1指标与MGN和Pyramid相同，MMF-Net的mAP相较于性能最接近的方法DSAP提高了1.5个百分点，比MGN提高了2.2个百分点。

在DukeMTMC-ReID数据集上不同方法的评价指标对比如表3所示。本文方法MMF-Net的Rank-1精度和mAP分别为89.7%的和79.9%。MMF-Net方法的Rank-1相比MHN提高了0.6个百分点，MMF-Net方法的mAP相比Pyramid提高了0.9个百分点。由于Pyramid主要依靠更大的主干网络（ResNet-101）和使用了更多的特征（21个特征，本文方法只用了5个），因此其与MMF-Net方法的性能更接近。在同样以ResNet-50为主干网络的条件下，本文方法相比性能最优的FPR方法在Rank-1精度和mAP上分别提高了1.1和1.5个百分点。

表2　在Market-1501数据集上不同方法的评价指标对比

Table 2　Evaluation indexs comparison among different methods on Market-1501 dataset　%

方法	Rank-1	mAP
SVDNet方法	82.3	62.1
Part-aligned方法	91.7	79.6
PCB+RPP方法	93.8	81.6
MLFN方法	90.0	74.3
HA-CNN方法	91.2	75.7
MGN方法[15]	95.7	86.9
AANet方法[27]	93.9	83.4
CASN方法[28]	94.4	82.8
IANet方法[29]	94.4	83.1
CAMA方法[30]	94.7	84.5
DGNet方法[7]	94.8	86.0
DSAP方法[31]	95.7	87.6
Pyramid方法[32]	95.7	88.2
Auto-ReID方法[33]	94.5	85.1
OSNet方法[34]	94.8	84.9
DCDS方法[35]	94.8	85.8
MHN方法[36]	95.1	85.0
P^2-Net方法[37]	95.2	85.6
BDB方法[38]	95.3	86.7
FPR方法[39]	95.4	86.6
MMF-Net方法	95.7	89.1

表3　在DukeMTMC-ReID数据集上不同方法的评价指标对比

Table 3　Evaluation indexs comparison among different methods on DukeMTMC-ReID dataset　%

方法	Rank-1	mAP
SVDNet方法	76.7	56.8
Part-aligned方法	84.4	69.3
PCB+RPP方法	83.3	69.2
MLFN方法	81.2	62.8
HA-CNN方法	80.5	63.8
MGN方法[15]	88.7	78.4
AANet方法[27]	87.7	74.3
CASN方法[28]	87.7	73.7
IANet方法[29]	87.1	73.4
CAMA方法[30]	85.8	72.9
DGNet方法[7]	86.6	74.8
DSAP方法[31]	86.2	74.3
Pyramid方法[32]	89.0	79.0
OSNet方法[34]	88.6	73.5
DCDS方法[35]	87.5	75.5
MHN方法[36]	89.1	77.2
P^2-Net方法[37]	86.5	75.1
BDB方法[38]	89.0	76.0
FPR方法[39]	88.6	78.4
MMF-Net方法	89.7	79.9

计算机工程 www.ecice06.com

在CUHK03数据集上不同方法的评价指标对比如表4所示。因为数据集有手工标记和检测器检测两种标注类型,所以将数据集分成两种情况进行验证。在手工标记情况下,本文方法MMF-Net的Rank-1精度和mAP分别达到了78.9%和76.6%,其中相比P^2-Net的Rank-1精度提高了0.6个百分点,mAP提升了3.0个百分点。在检测情况下,MMF-Net具有76.4%的Rank-1精度和74.1%的mAP,相比FPR的Rank-1精度和mAP分别提升了0.3和1.8个百分点。所有方法在手工标记的情况下能取得更好的结果,这说明行人边界框标注的准确性对识别结果产生一定的影响,即更精准的行人边界框标注能提升行人重识别的准确度。

表4 在CUHK03数据集上不同方法的评价指标对比

Table 4 Evaluation indexs comparison among different methods on CUHK03 dataset %

方法	手工标记		检测器检测	
	Rank-1	mAP	Rank-1	mAP
SVDNet方法	40.9	37.8	41.5	37.3
PCB+RPP方法	—	—	63.7	57.5
MLFN方法	—	—	54.7	49.2
HA-CNN方法	—	—	44.4	41.0
MGN方法[15]	68.0	67.4	68.0	66.0
CASN方法[28]	73.7	68.0	71.5	64.4
CAMA方法[30]	70.1	66.5	66.6	64.2
Auto-ReID方法[33]	77.9	73.0	73.3	69.3
MHN方法[36]	77.2	72.4	71.7	65.4
P^2-Net方法[37]	78.3	73.6	74.9	68.9
FPR 方法[39]	—	—	76.1	72.3
MMF-Net方法	78.9	76.6	76.4	74.1

在MSMT17数据集上不同方法的评价指标对比如表5所示。由于该数据集发布时间尚短,因此在该数据集上验证的方法不多。实验结果表明,本文方法MMF-Net在该数据集上的Rank-1和mAP优于其他方法,Rank-1精度和mAP分别达到了80.6%和59.3%,相比OSNet[34]的Rank-1和mAP分别提升了1.9和6.4个百分点。

表5 在MSMT17数据集上不同方法的评价指标对比

Table 5 Evaluation indexs comparison among different methods on MSMT17 dataset %

方法	Rank-1	mAP
PDC方法[40]	58.0	29.7
GLAD方法[41]	61.4	34.0
IANet方法[29]	75.5	45.8
DGNet方法[7]	77.2	52.3
Auto-ReID方法[33]	78.2	52.5
OSNet方法[34]	78.7	52.9
MMF-Net方法	80.6	59.3

2.4 消融实验

本文以Market-1501数据集为例,在该数据集上设计一系列实验来评估每个模块的有效性。其中,使用softmax损失和triplet损失训练的全局特征(即原始分支P1)作为基线(baseline)模型。

2.4.1 语义监督模块分析

本文在基线和MMF-Net上,通过设置不同SSM数量和位置,以评估该模块的性能。SSM一方面旨在提前引入语义监督,加强主干网络的特征提取能力,另一方面该模块提取的特征可以作为高层特征的补充,与高层特征一起作为行人图像的最终特征表示。实验结果如表6所示,表中的s3、s4表示在主干网络的第3、4层中引入SSM,p表示深度挖掘模块中的分支(Part-*n*),w/wo表示with/without。在次高层(即主干网络第4层)中引入SSM能够有效提升基线和MMF-Net的性能(1%~2%),而在第3层引入SSM反而降低了网络性能,其原因可能是由于网络层次过低会包含过多的干扰信息。此外,过多的模块数量(Baseline+s3+s4,MMF-Net w/s3)并不能提升性能指标。实验结果验证了高层特征和低层特征互补的有效性。

表6 在Market-1501数据集上MMF-Net不同分支设置的评估结果

Table 6 Evaluation results of different branch settings of MMF-Net on Market-1501 dataset %

方法	Rank-1	mAP
Baseline(w/p1)	93.1	83.2
Baseline+s3	92.8	82.0
Baseline+s4	94.0	85.8
Baseline+s3+s4	93.0	84.6
Baseline+s4+p2	94.9	87.9
Baseline+s4+p2+p3	95.0	88.5
Baseline+s4+p2+p3+p4(MMF-Net)	95.7	89.1
Baseline+s4+p2+p3+p4+p5	94.5	87.3
MMF-Net w/s3	94.6	88.1
MMF-Net w/os4	94.7	87.4
MMF-Net	95.7	89.1

2.4.2 深度挖掘模块中分支数目分析

网络多分支的设置可以增加模型表达的多样性。每个分支可以进行不同的特征学习任务,合理地设置网络分支能够有效提升模型性能。本文在基线和SSM模型的基础上,逐步增加分支的数量,以评估分支数量对模型性能的影响。从表6可以看出,随着分支数目的增加,模型的性能逐步提升,当分支数目达到4时,模型的性能最佳。此外,本文额外增加Part-5分支,根据模块的构建思路采用感受野更大的卷积核(7×7)进行特征提取。然而,这样的设置反而降低了模型的性能,验证了MMF-Net设置4个分支的必要性和有效性。由于此时特征图尺寸较小并且包含的信息足够完整,因此使用大卷积核会造成信息损失,反而降低模型的表达能力。

计算机工程 www.ecice06.com

2.4.3　池化层选择

池化层主要用于特征的聚合和降维。主干网络最后的池化层所聚合的特征信息一般直接或间接用于最后的特征表示。本文在MMF-Net网络上以不同的池化方式进行实验，评估池化方式对模型性能的影响，实验结果如表7所示。

表7　在Market-1501数据集上不同池化方法的评估指标对比

Table 7　Evaluation indexs comparison among different pooling methods on Market-1501 dataset　%

方法	Rank-1	mAP
平均池化	94.8	88.3
最大池化	94.7	88.4
混合池化	95.7	89.1

从表7可以看出，平均池化和最大池化的精度基本相同，而结合两种池化方式的混合池化能取得更好的效果。平均池化关注特征图的全局性，考虑信息的完整性。最大池化关注特征图中响应强烈的部分信息，旨在保留最具辨别性的信息。混合池化将两种方式结合到一个模型中，能够保留更有效的鉴别信息，强化了模型特征表达能力。

此外，本文还进行了部分可视化实验。行人特征响应热力图对比如图3所示（彩色效果见《计算机工程》官网HTML版）。从图3可以看出，MMF-Net特征图的响应区域更活跃，并且关注具有鉴别信息的局部区域，例如衣服图案和纹理。在不同数据集上本文方法的前5个检索结果如图4所示，图中第1行和第2行分别是基线和本文方法的检索结果，虚线框表示错误的检索结果。当不同类别的行人图像有较高的相似度时，本文所提的MMF-Net仍能够有效地将其识别出，具有较优的鲁棒性。

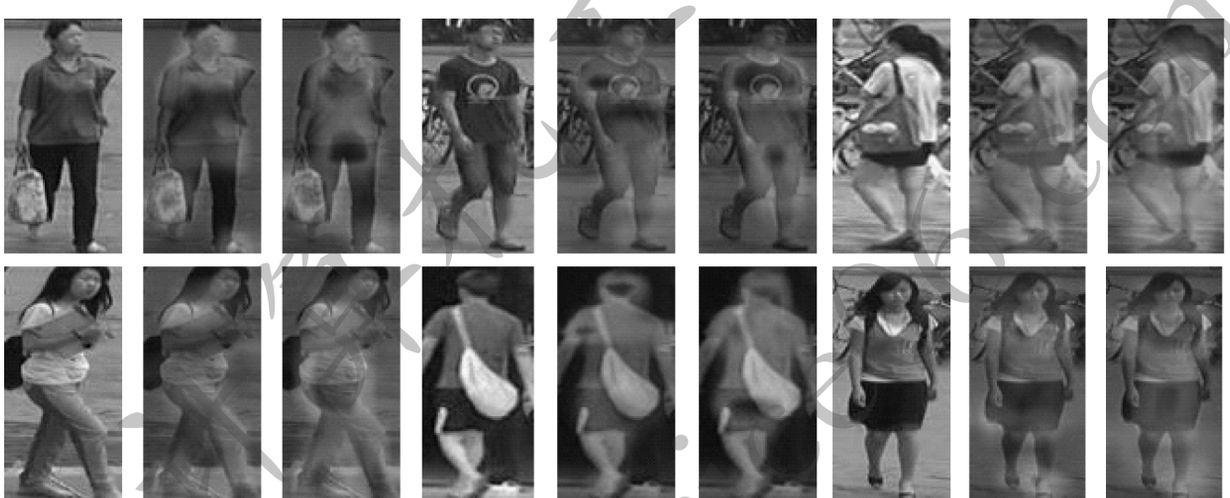

图3　行人特征响应热力图对比

Fig.3　Thermal map comparison of person features response



图4　本文方法特征识别结果

Fig.4　Feature recognization results of the proposed method

3 结束语

针对局部特征学习存在的局限性和低层特征利用率低的问题，本文提出多尺度多粒度融合的行人重识别方法。通过学习不同尺度和粒度的特征表示，并融合全局和局部特征学习，以加强全局和局部特征的关联性。同时将低层特征作为行人图像表示的一部分，实现低层和高层特征的优势互补。此外，在改进的池化层上结合平均池化和最大池化的特性，获取具有强鉴别力的特征。实验结果表明，本文方法在Market-1501数据集上的Rank-1和mAP分别为95.7%和89.1%，相比FPR、MGN等方法能够有效提升行人重识别性能。下一步将通过引入注意力机制，并将其与分块局部特征相结合，以提取鲁棒性更优的行人特征。

参考文献

[1] 李幼蛟，卓力，张菁，等. 行人再识别技术综述[J]. 自动化学报，2018，44(9)：1554-1568.
LI Y J，ZHUO L，ZHANG J，et al. Overview of person re-identification technology[J]. Acta Automatica Sinica，2018，44(9)：1554-1568. (in Chinese)

[2] LIAO S，HU Y，ZHU X，et al. Person re-identification by local maximal occurrence representation and metric learning [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2015：2197-2206.

[3] 姜国权，肖禛禛，霍占强. 基于RGB与灰度信息融合的行人再识别方法[J]. 计算机工程，2021，47(4)：226-233，240.
JIANG G Q，XIAO Z Z，HUO Z Q. Person re-identification based on the fusion of RGB and grayscale information[J]. Computer Engineering，2021，47(4)：226-233，240. (in Chinese)

[4] ZHENG L，YANG Y，HAUPTMANN A G. Person re-identification：past，present and future[EB/OL]. [2020-11-02]. https：//arxiv. org/pdf/1610. 02984. pdf.

[5] LIN Y，ZHENG L，ZHENG Z，et al. Improving person re-identification by attribute and identity learning[J]. Pattern Recognition，2019，95：151-161.

[6] HERMANS A，BEYER L，LEIBE B. In defense of the triplet loss for person re-identification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2017：1-15.

[7] ZHENG Z，YANG X，YU Z，et al. Joint discriminative and generative learning for person re-identification [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2019：2138-2147.

[8] ZHONG Z，ZHENG L，CAO D，et al. Re-ranking person re-identification with k-reciprocal encoding[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2017：1318-1327.

[9] 罗浩，姜伟，范星，等. 基于深度学习的行人重识别研究进展[J]. 自动化学报，2019，45(11)：2032-2049.
LUO H，JIANG W，FAN X，et al. Research progress of person re-identification based on deep learning[J]. Acta Automatica Sinica，2019，45(11)：2032-2049. (in Chinese)

[10] ZHAO H，TIAN M，SUN S，et al. Spindle Net：person re-identification with human body region guided feature decomposition and fusion [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2017：1077-1085.

[11] ZHENG L，HUANG Y，LU H，et al. Pose-invariant embedding for deep person re-identification[J]. IEEE Transactions on Image Processing，2019，28(9)：4500-4509.

[12] 刘紫燕，万佩佩. 基于注意力机制的行人重识别特征提取方法[J]. 计算机应用，2020，40(3)：672-676.
LIU Z Y，WAN P P. Person re-identification feature extraction method based on attention mechanism [J]. Journal of Computer Applications，2020，40(3)：672-676. (in Chinese)

[13] CHEN T，DING S，XIE J，et al. Abd-Net：attentive but diverse person re-identification[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. ，USA：IEEE Press，2019：8351-8361.

[14] 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. Berlin，Germany：Springer，2018：480-496.

[15] 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. New York，USA：ACM Press，2018：274-282.

[16] SZEGEDY C，VANHOUCKE V，IOFFE S，et al. Rethinking the inception architecture for computer vision [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2016：2818-2826.

[17] ZHENG L，SHEN L，TIAN L，et al. Scalable person re-identification：a benchmark [C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. ，USA：IEEE Press，2015：1116-1124.

[18] RISTANI E，SOLERA F，ZOU R，et al. Performance measures and a data set for multi-target，multi-camera tracking [C]//Proceedings of European Conference on Computer Vision. Berlin，Germany：Springer，2016：17-35.

[19] LI W，ZHAO R，XIAO T，et al. Deepreid：deep filter pairing neural network for person re-identification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2014：152-159.

[20] WEI L，ZHANG S，GAO W，et al. Person transfer GAN to bridge domain gap for person re-identification [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2018：79-88.

计算机工程 www.ecice06.com

[21] FELZENSZWALB P，MCALLESTER D，RAMANAN D. A discriminatively trained，multiscale，deformable part model[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2008：1-8.

[22] REN S，HE K，GIRSHICK R，et al. Faster R-CNN：towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2017，39：1137-1149.

[23] SUN Y，ZHENG L，DENG W，et al. SVDNet for pedestrian retrieval[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. ，USA：IEEE Press，2017：3800-3808.

[24] SUH Y，WANG J，TANG S，et al. Part-aligned bilinear representations for person re-identification[C]//Proceedings of European Conference on Computer Vision. Berlin，Germany：Springer，2018：402-419.

[25] CHANG X，HOSPEDALES T M，XIANG T. Multi-level factorisation net for person re-identification [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2018：2109-2118.

[26] LI W，ZHU X，GONG S. Harmonious attention network for person re-identification [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2018：2285-2294.

[27] TAY C P，ROY S，YAP K H. AANet：attribute attention network for person re-identifications[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2019：7134-7143.

[28] ZHENG M，KARANAM S，WU Z，et al. Re-identification with consistent attentive siamese networks[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2019：5735-5744.

[29] HOU R，MA B，CHANG H，et al. Interaction-and-aggregation network for person re-identification[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2019：9317-9326.

[30] YANG W，HUANG H，ZHANG Z，et al. Towards rich feature discovery with class activation maps augmentation for person re-identification [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2019：1389-1398.

[31] ZHANG Z，LAN C，ZENG W，et al. Densely semantically aligned person re-identification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2019：667-676.

[32] ZHENG F，DENG C，SUN X，et al. Pyramidal person re-identification via multi-loss dynamic training [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2019：8514-8522.

[33] QUAN R，DONG X，WU Y，et al. Auto-reid：searching for a part-aware convNet for person re-identification[C]// Proceedings of IEEE International Conference on Computer Vision. Washington D. C. ，USA：IEEE Press，2019：3750-3759.

[34] ZHOU K，YANG Y，CAVALLARO A，et al. Omni-scale feature learning for person re-identification [C]// Proceedings of IEEE International Conference on Computer Vision. Washington D. C. ，USA：IEEE Press，2019：3702-3712.

[35] ALEMU L T，PELILLO M，SHAH M. Deep constrained dominant sets for person re-identification[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. ，USA：IEEE Press，2019：9855-9864.

[36] CHEN B，DENG W，HU J. Mixed high-order attention network for person re-identification[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. ，USA：IEEE Press，2019：371-381.

[37] GUO J，YUAN Y，HUANG L，et al. Beyond human parts：dual part-aligned representations for person re-identification [C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. ，USA：IEEE Press，2019：3642-3651.

[38] DAI Z，CHEN M，GU X，et al. Batch dropblock network for person re-identification and beyond[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. ，USA：IEEE Press，2019：3691-3701.

[39] HE L，WANG Y，LIU W，et al. Foreground-aware pyramid reconstruction for alignment-free occluded person re-identification [C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. ，USA：IEEE Press，2019：8450-8459.

[40] SU C，LI J，ZHANG S，et al. Pose-driven deep convolutional model for person re-identification [C]// Proceedings of IEEE International Conference on Computer Vision. Washington D. C. ，USA：IEEE Press，2017：3960-3969.

[41] WEI L，ZHANG S，YAO H，et al. GLAD：global-local-alignment descriptor for pedestrian retrieval [C]// Proceedings of the 25th ACM International Conference on Multimedia. New York，USA：ACM Press，2017：420-428.

编辑　薛晋栋

www.ecice06.com