计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.200600043

基于多尺度多粒度特征的行人重识别

王 栋[1] 周大可[1,2] 黄有达[1] 杨 欣[1]
1 南京航空航天大学自动化学院 南京 211100
2 江苏省物联网与控制技术重点实验室(南京航空航天大学) 南京 211100
(m15150690108@163.com)

摘 要 针对现有的基于卷积神经网络的行人重识别方法所提取的特征辨识力不足的问题,提出了一种基于多尺度多粒度特征的行人重识别方法。在训练阶段,该方法在卷积神经网络的不同尺度提取特征;然后对获得的多尺度特征图进行分块和池化,从而得到不同尺度的全局特征和局部特征的多粒度特征,使用不确定性权重调节 Softmax 损失和三元组损失来对特征向量进行监督训练。在推理阶段,对所获得的多尺度多粒度的特征进行融合,使用融合特征在图像库中进行相似度匹配。在 Market-1501 和 DukeMTMC-ReID 数据集上的实验表明,所提方法相比基准网络 ResNet-50 在 Rank-1 评价指标上分别提升了 4.3%和 3.6%,在 mAP 评价指标上分别提升了 6.2%和 6.6%。实验结果表明,所提方法能够增强提取特征的辨识力,提高行人重识别的性能。

关键词:机器视觉;卷积神经网络;行人重识别;多尺度特征;多粒度特征
中图法分类号 TP399

Multi-scale Multi-granularity Feature for Pedestrian Re-identification

WANG Dong[1],ZHOU Da-ke[1,2],HUANG You-da[1] and YANG Xin[1]
1 School of Automation Engineering,Nanjing University of Aeronautics and Astronautics,Nanjing 211100,China
2 Jiangsu Key Laboratory of Internet of Things and Control Technologies (Nanjing University of Aeronautics and Astronautics),Nanjing 211100,China

Abstract In order to address the problem of insufficient discriminative features for pedestrian re-identification extracted by existing convolutional neural network,a novel multi-scale multi-granularity feature learning for pedestrian re-identification method is proposed. In the training phase,the method extracts multi-scale features at different stages of the convolutional neural network,and then blocks and pools these feature maps to obtain multi-granularity features containing global and local features,uses uncertainty to weight Softmax loss and triples loss and to supervise training process on feature vectors. In the inference phase,the obtained multi-scale multi-granularity features are concatenated,and finally the concatenated features are used to perform similarity matching in the gallery. Experiments on the Market-1501 and DukeMTMC-ReID datasets show that the proposed method improves the Rank-1 evaluation index by 4.3% and 3.6%,respectively,compared with the benchmark network ResNet-50,and improves the mAP evaluation index respectively 6.2% and 6.6%. The results show that the proposed method can enhance the discrimination of extracted features and improve the performance of pedestrian re-identification.

Keywords Machine vision,Convolutional neural network,Pedestrian re-identification,Multi-scale features,Multi-granularity features

1 引言

行人重识别(person re-identification)也称行人再识别,被广泛认为是图像检索的一个子问题,是利用计算机视觉技术判断图像或者视频中是否存在特定行人的技术,即给定一幅监控行人图像,检索跨设备下的该行人图像[1]。随着公共安全意识的增强,行人重识别技术在视频监控、智能安防等领域有着重要应用。在实际应用中,行人重识别存在两个主要问题:1)类内差异大,即由于不同摄像机场景、视角等环境因素,造成不同摄像机所得到的同一行人的图像有很大差异;2)类间差异小,即由于背景环境杂乱、遮挡、分辨率低等问题造成的不同的人会有较大的外观相似性。

到稿日期:2020-06-05 返修日期:2020-09-18 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家自然科学基金(61573182)
This work was supported by the National Natural Science Foundation of China(61573182).
通信作者:周大可(dkzhou@nuaa.edu.cn)

解决上述两个问题的关键在于提取具有辨识力的特征。传统的行人重识别方法利用手工设计的图像特征，如KISSME[2]，XQDA[3]，ACF[4]等，然而手工设计的特征表示能力有限，无法满足复杂场景下行人重识别任务的要求。随着深度学习的发展，利用神经网络来学习复杂的特征，显著地提高了特征提取的性能，从而提高了行人重识别的性能。Zheng 等[5]提出了 IDE(ID-embedding)方法，该方法将每个行人当作分类问题的一个类别，以 ResNet-50 为骨干网络来训练分类网络，通过预测结果和行人 ID 类别标签来计算 ID 损失。IDE 网络已经成为了现有很多基于深度学习的行人重识别方法的基准网络。Hermans 等[6]使用三元组损失函数来训练模型，三元组损失是度量学习损失的一种，其训练目标是缩小同类样本之间的距离，扩大不同类样本之间的距离。现有的深度学习方法常常将 ID 损失和 Triplet 损失相结合来训练神经网络。Zhang 等[7]提出了一种基于特征融合的行人重识别方法，融合后的特征对行人图像具有更好的表述能力。本文采取了融合深度神经网络的不同尺度的特征图来增强行人特征图的表达能力。

上述方法只使用了全局特征，当目标存在遮挡的情况时，这些特征并不能提供良好的辨别能力。因此，研究者开始使用全局特征和局部特征相结合的方法来解决这个问题。一些研究者会借助身体的姿势信息[8-10]来获得局部特征。Wei 等[8]利用提取人体骨骼关键点的方法(Global-local-alignment，GLAD)把行人图片分为头部、上身和下身 3 个部分，之后将整张行人图片与 3 张局部图片一起输入到一个参数共享的卷积神经网络中，最终提取的特征融合了全局和局部的特征。但是基于身体姿势信息的方法有着明显的局限性，因为姿势检测器需要预先训练好。针对这个问题，Sun 等[11]提出了一个 PCB(Part-based Convolutional Baseline)分块模型，它将最后获取的特征图进行水平分割，得到了 6 个平均切分的小块，将这 6 个小块融合进行最后的预测。对于行人身体不对齐的问题，他们使用了 RPP(Refined Part Pooling)网络模型来调整每个分块。鉴于分块策略的简单有效性，本文也采取了分块策略，与 PCB 不同的是，本文的分块策略是在不同尺度上进行的，而且执行了一个由粗到细分块的过程，以不同的粒度结合全局信息和局部信息。

还有一些研究者基于注意力机制[12-14]和生成对抗网络[15-16]的方法来获取更具辨识力的特征。此外，一些预处理和后处理的方法被用于行人重识别网络。Zhong 等[17]提出了一种随机擦除的数据增强方法，增加了所提取特征的鲁棒性，这种数据增强的方法被广泛用于行人重识别网络中。Zhong 等[18]提出了一种重新排序的后处理方法，使得行人重识别性能得到很大的提高。

本文提出的基于多尺度多粒度特征的行人重识别方法以 ResNet-50 网络为基础，基于 FPN(Feature Pyramid Network)提取网络的不同尺度的特征，然后对不同尺度的特征图进行全局平均池化，将不同尺度的特征图分成多个区域，最后对多尺度多粒度的特征图使用不确定性调节的 Softmax 损失和三元组损失进行监督训练。在推理阶段，将所有的多尺度多粒度特征图进行融合，并以最终融合的特征作为匹配依据。本文将提出的方法在目前比较流行的大型数据集 Market-1501 和 DukeMTMC-ReID 上进行了有效性验证，实验结果表明，本文方法提高了行人重识别的性能。

2　基于多尺度多粒度特征的行人重识别

2.1　网络结构

本文方法的基本框架由 3 个部分组成，分别是 ResNet-50[19]骨干网络、FPN[20]多尺度特征融合模块以及多粒度处理模块，如图 1 所示。考虑到性能和速度的平衡，本文最终选择了 ResNet-50 作为本文方法的骨干网络。首先使用 ResNet-50 网络提取输入图片的不同尺度的特征图，然后借助 FPN 网络融合不同尺度的特征图，最后多粒度处理模块使用全局平均池化的方法对不同尺度的特征图进行池化，从而得到多尺度多粒度的特征向量。在训练阶段，对 8 个 256 维的向量分别计算 Softmax 损失，将 8 个 256 维向量拼接成 2048 维向量计算 triplet 损失；在测试阶段，将拼接的 2048 维向量作为行人表示和行人库中的行人特征计算欧氏距离，然后排序得到匹配结果。

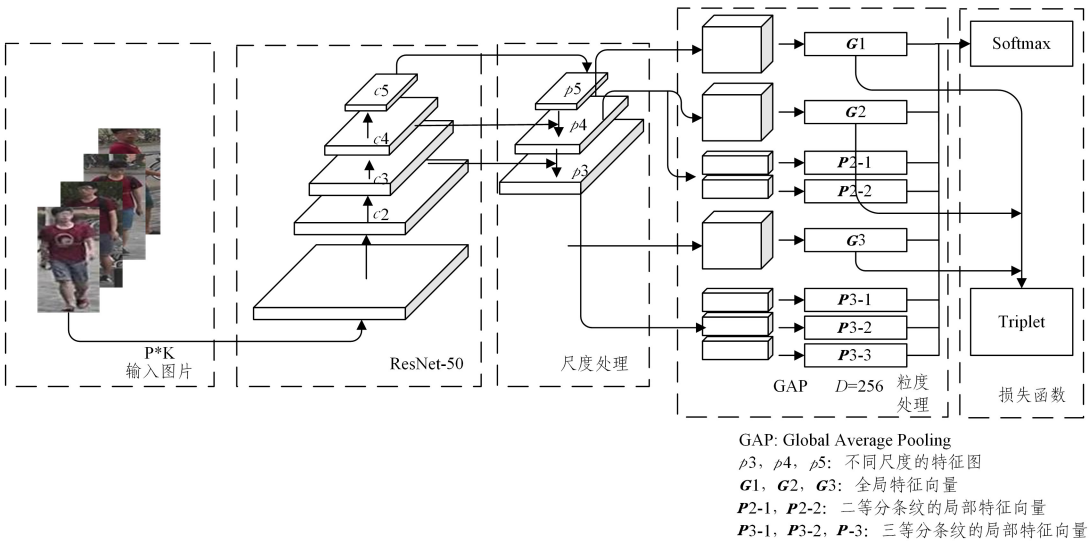


图 1　整体网络结构示意图

Fig. 1　Schematic diagram of the overall network structure

2.2 多尺度特征提取

本文提出的行人重识别方法的多尺度特征的实现依赖于不同网络层的特征图，多尺度的特征图包含浅层和深层特征。通过结合浅层和深层特征，可以使行人的特征描述更具有辨识力。由于卷积神经网络的性质，随着网络层数的增加，提取到的特征在不同尺度的特征图上所表示的信息并不相同。在浅层的特征图中，网络的注意力集中于背景区域，而在深层的特征图中，网络的注意力聚焦于目标区域，也就是说，深层次的特征图具有更好的抽象和语义表达能力。因此只利用深层网络提取到的特征图，会忽视行人的细节信息，从而影响行人重识别的性能。ResNet-50 有 4 个残差块，每经过一个残差块，特征图的尺寸就变为原来的一半，最终生成 4 个尺度的特征图 $c2$，$c3$，$c4$，$c5$。这里选取 $c3$，$c4$ 和 $c5$ 这 3 个尺度的特征图，舍弃掉 $c2$ 的原因在于，太过浅层的特征图包含了过多的背景信息，会干扰对行人目标的匹配。FPN 多尺度特征融合模块通过上采样和横向链接将不同尺度的特征图进行融合，得到了 $p5$，$p4$ 和 $p3$，如图 1 所示。

2.3 多粒度特征提取

提取行人特征最直接的方法就是提取行人的整个身体作为最后的匹配特征。全局特征能够捕获每个行人外观上最显著的特征来区分不同的行人。然而在复杂的大型监控场景中，仅使用全局特征作为行人重识别的特征很难得到满意的效果，因为在全局特征的学习过程中一些不显著的信息往往被忽略。为了解决上述问题，一些研究者通过提取局部特征来优化所提取的全局特征，从而提高了行人重识别的准确性。提取局部特征通常是基于身体部位的方法和分块策略。但是基于身体部位的方法有着明显的局限性，因为这些方法几乎只关注具有固定语义的身体部位，这些固定的部位在摄像机的角度变化时会发生显著变化，从而影响局部信息的可靠性。另外，基于身体部位的方法一般不能实现端到端的训练，增加了训练难度。分块策略是通过全局池化的方法将特征图分成几个固定大小的横条纹。PCB 方法使用了从上到下 6 个局部部位的信息来预测行人身份，由于检测精度的问题，图片之间行人存在错位、不对齐等状况，每个分块越小，这种影响就有可能被放大。如图 2 的 PCB 分块所示，如果分块更小，那么最上层的分块就仅包含背景部分。然而，如果同一行人的另一张图片紧贴图像边界，那么他的最上层分块包含行人头部。这种情况的存在会影响到最终的行人重识别性能。

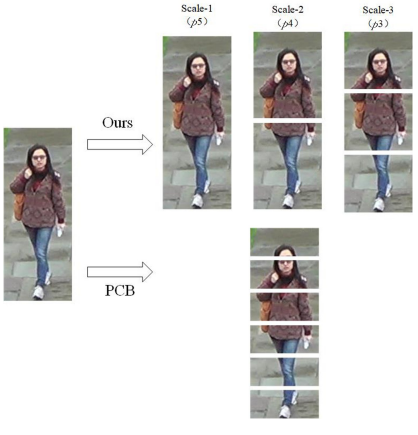


图 2 分块策略示意图

Fig. 2 Examples of image block

本文的方法也采取了分块策略。与 PCB 不同是，本文方法首先将特征图分成原始特征图、二等分图和三等分图，这样既包含了全局特征和局部特征的多粒度结合的全面信息，又使得每个分块的异常值比 PCB 分块少；其次，二等分图和三等分图之间有重叠的部位，而不是将全局特征简单地分开再融合(无重叠部分)；最后，本文方法对特征图的分块是在不同尺度上进行的。使用这种策略的原因是，越深层次的特征图具有越好的抽象和语义表达能力，而浅层的特征图具有更多的细节。本文分块策略与 PCB 分块策略的对比如图 2 所示。

2.4 损失函数

为了使所提出的网络提取到具有辨识力的特征，本文采用 Softmax 损失函数[21]和三元组损失函数[6]联合训练的方法，这也是行人重识别网络中经常使用的方法。因为 Softmax 损失函数用于分类任务，而三元组损失函数用于度量学习任务，本文认为所提网络是用于多任务学习，所以使用多任务似然[22]来调节这两个损失函数。

Softmax 损失函数。对于基本的识别任务，本文将其视为一个多分类问题，即不同的人对应不同的类别，对于第 i 个学习的特征 $\boldsymbol{F}_i$，Softmax 损失可以表示为：

$$L_s=-\sum_{i=1}^{N}\log\frac{e^{\boldsymbol{w}_{y_i}^{\mathrm{T}}\boldsymbol{F}_i+b_{y_i}}}{\sum_{k=1}^{C}e^{\boldsymbol{w}_k^{\mathrm{T}}\boldsymbol{F}_i+b_k}} \tag{1}$$

其中，N 表示训练过程中的批次样本的大小，C 表示训练集中行人类别数。

本文方法使用的是批次硬三元组损失，是三元组损失的一个改进版本。批次硬三元组损失可以表示为：

$$L_t=-\sum_{i=1}^{P}\sum_{a=1}^{K}\Big[\alpha+\max_{p=1\cdots K}\|\boldsymbol{F}_a^{(i)}-\boldsymbol{F}_p^{(i)}\|_2-\min_{n=1\cdots k,j=1\cdots p,j\neq i}\|\boldsymbol{F}_a^{(i)}-\boldsymbol{F}_n^{(i)}\|_2\Big]_+ \tag{2}$$

其中，P 表示训练批次中行人的类别数；K 表示每个类别中图片的数量；α 表示正负样本的距离间隔大小，这里的正样本和负样本指与样本有相同或不同身份的行人。

多任务损失函数。Softmax 损失函数和三元组损失函数的结合可以表示为：

$$L_{\text{total}}(\lambda)=\lambda_s L_s+\lambda_t L_t \tag{3}$$

本文将通过最大化同方差不确定性的最大高斯似然来调节不同损失函数的权重。对于分类任务：

$$p(\boldsymbol{y}|\boldsymbol{f}^{\boldsymbol{W}}(\boldsymbol{x}))=\text{softmax}(\boldsymbol{f}^{\boldsymbol{W}}(\boldsymbol{x})) \tag{4}$$

对于回归任务：

$$p(\boldsymbol{y}|\boldsymbol{f}^{\boldsymbol{W}}(\boldsymbol{x}))=N(\boldsymbol{f}^{\boldsymbol{W}}(\boldsymbol{x}),\sigma^2) \tag{5}$$

Softmax 损失和 Triplet 损失结合的多任务损失函数为：

$$\begin{aligned}L_{\text{total}}(\boldsymbol{W},\sigma_s,\sigma_t)&=-\log p(y_1,y_2=c|\boldsymbol{f}^{\boldsymbol{w}}(\boldsymbol{x}))\\&=-\log N(y_1;\boldsymbol{f}^{\boldsymbol{w}}(\boldsymbol{x}),\sigma_t^2)\cdot\text{softmax}(y_2=c;\boldsymbol{f}^{\boldsymbol{w}}(\boldsymbol{x}),\sigma_s)\\&=\frac{1}{2\sigma_t^2}\|y_1-\boldsymbol{f}^{\boldsymbol{w}}(\boldsymbol{x})\|^2+\log\sigma_t-\log p(y_2=c|\boldsymbol{f}^{\boldsymbol{w}}(\boldsymbol{x}),\sigma_s)\\&\approx\frac{1}{2\sigma_t^2}L_t(\boldsymbol{W})+\frac{1}{\sigma_s^2}L_s(\boldsymbol{W})+\log\sigma_t+\log\sigma_s\end{aligned} \tag{6}$$

其中，$\boldsymbol{W}$ 是参数矩阵；σ_s，σ_t 是观测噪声；$L_t(\boldsymbol{W})=\|y_1-\boldsymbol{f}^{\boldsymbol{w}}(\boldsymbol{x})\|^2$ 是基于欧氏距离的三元组损失；$L_s(\boldsymbol{W})=-\log p(y_2=c|\boldsymbol{f}^{\boldsymbol{w}}(\boldsymbol{x}),\sigma_s)$ 是基于多分类任务的 Softmax 损失函数。优化的目的是同时寻找最优的 $\boldsymbol{W}$，σ_s，σ_t。在本文中，只有 3 个

全局特征使用三元组损失函数进行训练，而获得的每一个特征都使用 Softmax 损失函数进行多分类训练，如图 1 所示。

3　实验结果与分析

3.1　实验设置

3.1.1　数据集

实验是在 Market-1501 和 DukeMTMC-ReID 两个行人数据集上进行，这两个数据集都是大型数据集。在本文的实验中使用单查询的方式来评估实验结果。

Market-1501 数据集中的行人图片由 6 台不同的摄像机捕获，包含 1 501 个不同身份的人。这些行人图片是使用 DPM 检测器从摄像机捕获的图片中裁剪得来，所以存在图像不对齐的问题。整个数据集被分为训练集和测试机两个部分，训练集包含 751 个人的 12 936 张图片，测试集包含 750 个人的 3 368 张查询图片和 19 732 张库图片。

DukeMTMC-ReID 数据集是从 DukeMTMC 数据集中分离出来的用于行人重识别的一个子集，这些行人图片由 8 台高分辨率的摄像机捕获得到，包含 1 404 个人的 36 411 张图片。整个数据集被分为训练集和测试集两个部分，训练集包含 702 个人的 16 522 张图片，测试集包含来自 702 个人的 2 228 张查询图片和 17 661 张库图片。对于行人重识别来说，这可能是目前最具挑战性的数据集，因为存在不同行人之间具有很高的相似性而相同的行人之间差异很大的情况。

3.1.2　评价指标

为了评估行人重识别模型的性能，本文使用了 CMC(Cumulative Matching Characteristics)的等级 1(Rank-1)和 mAP(mean Average Precision)这两个评价指标。

3.1.3　实验环境 10.3

将输入图片的尺寸调整为 384×128 像素。使用在 ImageNet 的图像分类数据集上预训练的 ResNet-50 模型参数来初始化本文的骨干网络。在训练阶段，本文使用了随机翻转、随机裁剪和随机擦除来对训练数据进行数据增强，然后对输入的图像进行归一化处理。每一训练批次包含 48 张训练图片，随机采样 P 个人，每个人随机采样 K 张图片。在本文的实验中，P 设置为 12，K 设置为 4。对于三元组损失，本文将间隔超参数设置为 1.2。本文使用 SGD 优化器来训练本文的模型，其中将 *momentum* 设置为 0.9，将 *weight_decay* 设置为 5×10^{-4}。对于学习率，本文对骨架网络和新增加的网络设置不同的学习率，骨架网络的学习率设置为新增网络的 1/10，将骨架网络的初始学习率设置为 0.01，然后每经过 40 个 epoch 后衰减为原来的 1/10，经过 120 个 epoch 完成训练。在测试阶段，本文对查询图片进行尺寸调整和归一化处理，使用欧氏距离来计算查询图像和库图像之间的相似性。本文实验的训练平台和推理平台均为英伟达 RTX2080，框架为 PyTorch。对于不同的数据集，本文都采用上面的实验设置。

3.2　实验结果

3.2.1　对比实验

表 1 列出了本文方法与其他方法在 Market-1501 和 DukeMTMC-ReID 两个行人数据集上的对比结果。与其他方法相比，本文方法取得了不错的效果。在 Market-1501 和 DukeMTMC-ReID 数据集上，本文的 *Rank*-1 得分分别为 95.1%和 86.9%，*mAP* 得分分别为 86.3%和 76.4%。与经典的分块技术 PCB 相比，本文方法在 *Rank*-1 上分别提升了 2.7%和 5.0%，在 *mAP* 上分别提升了 8.9%和 10.3%。与 PCB+RPP 相比，本文方法在 *Rank*-1 上分别提升了 1.3%和 3.6%，在 *mAP* 上分别提升了 4.7%和 7.2%。此外，从表 1 可以看出，现有的行人重识别方法在 Market-1501 和 DukeMTMC-ReID 两个行人数据集的性能几乎达到了饱和。但是与这些方法相比，本文方法在 Market-1501 和 DukeMTMC-ReID 数据集上实现了总体最好的性能。在 DukeMTMC-ReID 数据集上，OSNet 实现了最好的 *Rank*-1 性能 88.6%，优于本文方法的 1.7%，但是在 *mAP* 指标上，本文方法优于 OSNet 2.9%。

表 1　本文方法与其他方法在 Market-1501 和 DukeMTMC-ReID 数据集上的对比结果

Table 1　Comparison results of proposed method with other methods on Market-1501 and DukeMTMC-ReID datasets

(单位：%)

Method	Market-1501		DukeMTMC-ReID	
	Rank-1	*mAP*	*Rank*-1	*mAP*
BOW+kissme[23]	34.4	14.1	25.1	12.1
LOMO+XQDA[3]	43.8	22.2	30.7	17.0
IDE[5]	72.5	46.0	65.2	44.9
TriNet[6]	84.9	69.1	72.4	53.5
GLAD[8]	89.9	73.9	—	—
Aligned-ReID[24]	90.6	77.7	81.2	67.4
PCB[11]	92.4	77.4	81.9	66.1
PCB+RPP[11]	93.8	81.6	83.3	69.2
AANet[25]	93.9	83.4	—	—
IANet[26]	94.4	83.1	87.1	73.4
BagTricks[27]	94.5	85.9	86.4	76.4
OSNet[28]	94.8	84.9	88.6	73.5
Ours	95.1	86.3	86.9	76.4

3.2.2　消融实验

为了展现本文方法如何提高行人重识别的性能，本文设计了一系列消融实验，展示了在 Market-1501 和 DukeMTMC-ReID 上，多尺度、多粒度以及权重损失对行人重识别性能的影响，实验结果如表 2 和表 3 所列。

表 2　Market-1501 数据集上消融实验的结果

Table 2　Ablation experiment results on Market-1501 dataset

Method	Multi-scale	Multi-granularity	Weight loss	*Rank*-1/%	*mAP*/%
ResNet-50	—	—	—	90.8	80.1
	√	—	—	92.5	82.5
	—	√	—	94.3	85.4
	—	—	√	91.9	80.7
Ours	√	√	√	95.1	86.3

注："√"表示添加了相应的模块，"—"表示没有使用相应的模块

表 3　DukeMTMC-ReID 数据集上消融实验的结果

Table 3　Ablation experiment results on DukeMTMC-ReID dataset

Method	Multi-scale	Multi-granularity	Weight loss	*Rank*-1/%	*mAP*/%
ResNet-50	—	—	—	83.3	69.8
	√	—	—	85.1	72.7
	—	√	—	86.3	75.7
	—	—	√	84.2	70.8
Ours	√	√	√	86.9	76.4

注："√"表示添加了相应的模块，"—"表示没有使用相应的模块

由表 2 和表 3 的消融实验结果可知，与骨架网络 ResNet-50 相比，多尺度、多粒度以及权重损失对行人重识别的性能

都有所提升。在 Market-1501 数据集上，这 3 个模块的 *Rank*-1 和 *mAP* 分别提升了 1.7%/2.4%，3.5%/5.3% 和 1.1%/0.6%；在 DukeMTMC-ReID 数据集上，这 3 个模块的 *Rank*-1 和 *mAP* 分别提升了 1.8%/2.9%，3.0%/5.9% 和 0.9%/1.0%。使用权重损失虽然能在一定程度上提高行人重识别性能，但是效果不明显。最终本文将不同的模块结合在一起，与基准方法相比，其在 Market-1501 和 DukeMTMC-ReID 数据集上的 *Rank*-1 评价指标上分别提升了 4.3% 和 3.6%，在 *mAP* 评价指标上分别提升了 6.2% 和 6.6%。

3.2.3 结构和速度分析

为了进一步探索不同的网络结构对行人重识别性能的影响，本文在 Market-1501 数据集上对不同结构进行了实验，结果如表 4 所列。首先对不同的骨干基准网络 ResNet-34，ResNet-50 以及 ResNet-101 进行对比，可以看到，ResNet-50 实现了最好的性能，*Rank*-1/*mAP* 为 90.8%/80.1%。在本文的实验中，ResNet-101 的性能出现了下降(−4.1%/−5.5%)，可能存在两方面的原因：1)更深的网络结构增加了训练的难度；2)行人重识别的数据集的数据量相对较少，在网络结构更加复杂的情况下，网络模型可能会出现过拟合。然后在 ResNet-50 网络的基础上，对不同的尺度选择进行了相关实验，Scalar-2 表示使用了 $p5$，$p4$ 这 2 个尺度的特征，Scalar-4 表示使用了 $p5$，$p4$，$p3$，$p2$ 这 4 个尺度的特征。从实验结果可以看出，当使用 4 个尺度的特征时，性能出现了下降(−0.9%/−0.6%)，我们猜测是因为过低层级的特征包含了过多的背景信息，从而在特征融合时影响了特征对行人的表达能力。最终，我们选取了 $p5$，$p4$，$p3$(Scalar-3)这 3 个尺度的特征。本文对不同条纹的划分也进行了实验，Stripe1-2-4 表示在 $p5$，$p4$，$p3$ 尺度上分别做 1 等分、2 等分和 4 等分，其他同理。Stripe1-2-4 和 Stripe1-3-4 的实验结果表明，当深层的特征图使用较小的分块时会得到更好的实验结果，因为网络层数越深，语义表达能力越强，网络的注意区域越集中，而较多的分块可能会破坏语义信息。最终，本文选择了 Stripe1-2-3 分块结构。

表 4 Market-1501 数据集上不同结构的实验结果

Table 4 Experimental results of different structures on Market-1501 dataset

(单位：%)

architecture	*Rank*-1	*mAP*
ResNet-34	86.5	73.4
ResNet-50	90.8	80.1
ResNet-101	86.7	74.6
Scalar-2	91.5	81.3
Scalar-4	91.6	81.9
Scalar-3	92.5	82.5
Stripe1-2-4	94.6	85.8
Stripe1-3-4	94.2	85.6
Stripe1-2-3	94.7	85.8
Ours	95.1	86.3

本文就多尺度多粒度带来的参数量和计算量的提升对速度的影响与基准网络 ResNet-50 进行了对比，实验结果如表 5 所列，其中 Time 表示提取一张图片所花费的时间。可以看出，本文方法所带来的参数量和计算量的增加很小，对单张图片的特征提取时间增加了 0.28 ms(9.46%)。参数量和计算量的提升几乎都是 FPN 特征融合模块带来的，因为分块模块的全局平局池化不会引起参数量的变化。

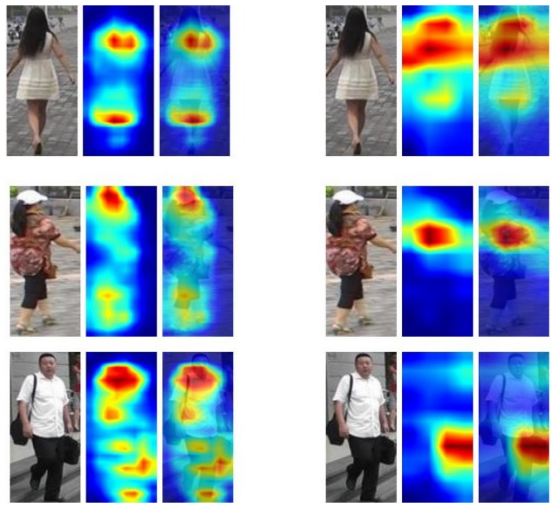
表 5 参数量、计算量和时间的对比

Table 5 Comparison of parameter, calculation and time

Method	GPU	*Params*/M	*FLOPs*/G	*Time*/ms
ResNet-50	RTX2080	24.03	6.18	2.96
Ours	RTX2080	24.42	6.33	3.24

3.2.4 可视化结果

为了验证本文所提网络提取特征的有效性，对网络最后提取的特征图进行可视化处理。激活的特征图的计算方式为：沿通道对特征图的每个像素的绝对值求和，然后进行空间 L2 归一化处理。图 3 给出了本文方法与基准网络的激活特征图的对比。可以看出，基准方法的特征图将更多焦点聚集在某一块区域(如头发、手提包)，这增加了网络模型过拟合的风险。相反，本文方法的特征图不仅将焦点聚集于行人区域，使行人从背景中分离出来，还将焦点聚集在一些局部区域。这样的特征图能更好地区分不同的行人。



(a)本文方法 (b)基准方法

图 3 可视化的激活特征图

Fig. 3 Vualized activation feature map

此外，本文还给出了使用本文方法得到的一些查询图片的 Rank-10 结果，如图 4 所示，右边的深色边框代表错误匹配。

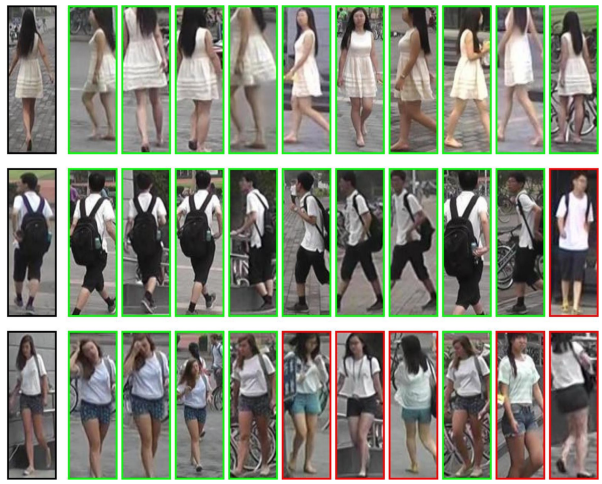


图 4 一些查询图的 Rank-10 结果

Fig. 4 Rank-10 results of some query graphs

从图 4 可以看出，错误匹配图片的穿着与查询图片的穿着非常相似，例如，第 3 行错误匹配和查询图片有着相同的着装组合：T-恤、短裤和背包。这说明本文方法不仅将注意力集中在行人整体，而且注意到了局部的重要信息，这与前面的特征图的结果是一致的，与本文设计的基于多尺度多粒度特征的行人重识别网络的动机也是一致的。

结束语　为了提高行人重识别方法在背景杂乱和行人不对齐的场景下提取特征的辨识力，本文提出了一种多尺度多粒度的行人重识别方法。该方法通过结合多尺度、多粒度的行人信息，获得了全局和局部的特征信息。实验表明，所提方法能改善行人重识别的性能。本文方法的多粒度特征的获取方式较为简单，只是做了等分分割。下一步将探索多粒度特征更好的获取方式，以优化行人的对齐问题。

参 考 文 献

[1] LUO H，JIANG W，FAN X，et al. A Survey on Deep Learning Based Person Re-identification[J]. Acta Automatica Sinic，2019，45(11)：2032-2049.

[2] MARTIN K，HIRZER M，WOHLHART P，et al. Large Scale Metric Learning from Equivalence Constraints[C]// IEEE Conference on Computer Vision and Pattern Recognition. Providence：IEEE Press，2012：2288-2295.

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

[4] HUAN X Y，XU J L，GUO G，et al. Real-time pedestrian re-recognition based on enhanced aggregate channel features[J]. Progress in Laser and Optoelectronics，2017(9)：119-127.

[5] ZHENG L，YANG Y，HAUPTMANN A G. Person re-identification：Past，present and future [J]. arXiv：1610. 02984.

[6] HERMANS A，BEYER L，LEIBE B. In Defense of the Triplet Loss for Person Re-Identification[J]. arXiv：1703. 07737.

[7] ZHANG G N，WANG J B，ZHANG Y F，et al. Pedestrian recognition method based on feature fusion[J]. Computer Engineering and Applications，2017(12)：190-194，245.

[8] 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：ACM Press，2017：420-428.

[9] SU C，LI J，ZHANG S，et al. Pose-driven Deep Convolutional Model for Person Re-identification[C]// IEEE Conference on Computer Vision and Pattern Recognition. Hawaii：IEEE Press，2017：3980-3989.

[10] 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. Hawaii：IEEE Press，2017：1077-1085.

[11] SUN Y，ZHENG L，YANG Y，et al. Beyond Part Models：Person Retrieval with Refined Part Pooling (and a Strong Convolutional Baseline)[C]// European Conference on Computer Vision. Munich：IEEE Press，2018：480-496.

[12] LIU Z Y，WAN P P. Feature extraction method for pedestrian re-recognition based on attention mechanism[J]. Journal of Computer Applications，2020，40(3)：672-676.

[13] LI W，ZHU X，GONG S. Harmonious Attention Network for Person Re-Identification[C]// IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City：IEEE Press，2018：2285-2294.

[14] SI J，ZHANG H，LI C G，et al. Dual Attention Matching Network for Context-Aware Feature Sequence based Person Re-Identification[C]// IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City：IEEE Press，2018：5363-5372.

[15] ZHONG Z，ZHENG L，ZHENG Z D，et al. CamStyle：A Novel Data Augmentation Method for Person Re-Identification[J]. IEEE Transactions on Image Processing，2019，28(3)：1176-1190.

[16] ZHANG S L，CAO X. Improved Person Re-Identification Algorithm on Camstyle [J]. CEA，2020，56(15)：124-131.

[17] ZHONG Z，ZHENG L，KANG G，et al. Random erasing data augmentation [J]. arXiv：1708. 04896.

[18] ZHONG Z，ZHENG L，CAO D，et al. Re-ranking Person Re-identification with k-reciprocal Encoding[C]// IEEE Conference on Computer Vision and Pattern Recognition. Hawaii：IEEE Press，2017：3652-3661.

[19] HE K，ZHANG X，REN S，et al. Deep Residual Learning for Image Recognition [C]// IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas：IEEE Press，2016：770-778.

[20] LIN T Y，DOLLAR P，GIRSHICK R，et al. Feature pyramid networks for object detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. Honolulu：IEEE Press，2017：2117-2125.

[21] CHRISTIAN S，VINCENT V，SERGER I，et al. Rethinking the inception architecture for computer vision[C]// IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas：IEEE Press，2016：2818-2826.

[22] CIPOLLA R，GAL Y，KENDALL A. Multi-task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics [C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City：IEEE Press，2018：7482-7491.

[23] ZHENG L，SHEN L Y，TIAN L，et al. Scalable Person Re-identification：A Benchmark[C]// The IEEE International Conference on Computer Vision. Santiago：IEEE Press，2015：1116-1124.

[24] ZHANG X，LUO H，FAN X，et al. Alignedreid：Surpassing human-level performance in person reidentification [J]. arXiv：1711. 08184.

[25] TAY C, ROY S, YAP K. AANet: Attribute Attention Network for Person Re-Identifications[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE Press, 2019: 7127-7136.

[26] HOU R B, MA B P, CHANG H, et al. Interaction-and-Aggregation Network for Person Re-identification [C]// The IEEE Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE Press, 2019: 9317-9326.

[27] LUO H, JIANG W, GU Y, et al. Bag of Tricks and A Strong Baseline for Deep Person Re-identification [C]// The IEEE Conference on Computer Vision and Pattern Recognition Workshops. Long Beach: IEEE Press, 2019: 4321-4329.

[28] ZHOU K, YANG Y, CAVALLARO A. et al. Omni-Scale Feature Learning for Person Re-Identification [C] // 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE Press, 2019: 3701-3711.

WANG Dong, born in 1996, postgraduate. His main research interests include target detection, pedestrian re-identification and target tracking.

ZHOU Da-ke, born in 1974, Ph.D, associate professor. His main research interests include digital image processing, computer vision and machine learning.