

基于深层特征融合的行人重识别方法^{*}

熊 炜^{1,2}, 熊子婕¹, 杨菡椿¹, 童 磊¹, 刘 敏¹, 曾春艳¹

(1. 湖北工业大学电气与电子工程学院, 湖北 武汉 430068;

2. 美国南卡罗来纳大学计算机科学与工程系, 南卡 哥伦比亚 29201)

摘 要:针对现有基于深度学习的行人重识别方法对于行人姿态变化、部分遮挡等引起的行人判别特征信息缺失的问题,提出了一种深层特征融合的行人重识别方法。首先,利用卷积层和池化层多次提取网络深层特征,从空间维度提升网络性能,使用融合后的深层特征作为行人图像的全局特征属性;其次,为提高模型的泛化能力,在深层融合特征后加入一个批量归一化层,同时采用标签平滑损失函数和三元组损失函数对模型进行联合训练。实验结果表明,所提的深层特征融合方法具有很好的表达能力。在 Market1501、DukeMTMC-reID、CUHK03 和 MSMT17 4 个数据集上对所提方法进行了验证,其中在 Market1501 数据集上,Rank-1 值达到了 95.0%,mAP 达到了 85.6%。

关键词:行人重识别;深层特征融合;Se-resnet50;批量归一化;标签平滑损失;三元组损失

中图分类号:TP183

文献标志码:A

doi:10.3969/j.issn.1007-130X.2020.02.022

Pedestrian re-identification based on deep feature fusion

XIONG Wei^{1,2}, XIONG Zi-jie¹, YANG Di-chun¹, TONG Lei¹, LIU Min¹, ZENG Chun-yan¹

(1. School of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan 430068, China;

2. Department of Computer Science and Engineering, University of South Carolina, Columbia, SC 29201, USA)

Abstract: Aiming at the problem that the existing deep learning based pedestrian re-identification methods lack pedestrian discriminative feature information caused by pedestrian posture change and partial occlusion, a pedestrian re-identification network model based on deep feature fusion is proposed. Firstly, the convolution layer and the pooling layer are used to extract the deep features of the network many times to improve the network performance from the spatial dimension. Secondly, in order to improve the generalization ability of the model, a batch normalization layer is added after the deep fusion features. Finally, the label smoothing loss function and the triplet loss function are used to train the model. The experimental results show that the proposed deep feature fusion method has good expression ability. The proposed method is validated on four datasets: Market1501, Duke MTMC-reID, CUHK03 and MSMT17. On Market1501, Rank-1 and mAP reach 95.0% and 85.6%.

Key words: pedestrian re-identification; deep feature fusion; Se-resnet50; batch normalization; label smoothing loss; triplet loss

1 引言

行人重识别(Pedestrian Re-identification),是

利用计算机视觉技术检索出不同摄像头下相同行人的技术^[1]。在监控视频中,由于相机分辨率和拍摄角度的缘故,通常无法得到质量非常高的人脸图像,此时行人重识别就成为了一个非常重要的替代

^{*} 收稿日期:2019-07-13;修回日期:2019-08-29

基金项目:国家自然科学基金(61571182,61601177);湖北省自然科学基金(2019CFB530);国家留学基金(201808420418)

通信地址:430068 湖北省武汉市湖北工业大学电气与电子工程学院

Address: School of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan 430068, Hubei, P. R. China

技术。由于行人重识别在视频监控应用中的重要地位,以及越来越多的大规模行人重识别数据集的提出,行人重识别在计算机视觉中越来越受重视。然而由于行人姿态变化、环境光照影响、遮挡等问题,行人重识别是一个相当具有挑战性的问题。

目前行人重识别方法主要分为2类:基于特征表示的方法和基于度量学习的方法。

基于特征表示的方法主要提取具有鲁棒性的鉴别特征来表示行人。Wei等^[2]提出采用 deeper cut 算法估算行人的4个关键点,将输入图像分成3个部分,然后从全局图像和部件图像中学习描述子,以此来学习更加精细的局部特征。Su等^[3]为解决行人姿态变化以及视角差异的问题,提出采用姿势估计的方法对行人图像分块,并对不同块加权来增强行人细节特征。Liu等^[4]提出了一种基于注意力机制的深度神经网络 HPNet (HydraPlus Network),引入来自多层级的多尺度注意力特征,并将行人的全局和局部特征联合组成新的行人特征。Zhao等^[5]为解决人体错位的问题,提出一种人体区域对齐的方法,利用训练好的网络将人体分割成几个有区分力的区域,分别计算每个区域的特征向量。Yao等^[6]提出了部分损失网络 PL-Net (Part Loss Network),对现有大部分只提取全局特征且容易过拟合而不考虑局部特征的方法做了改进,该网络可自动检测人体部件,增加模型的判别性。

基于度量学习的方法主要通过学习一个距离函数,使同一行人的距离小于不同行人的距离^[7]。它的目的是寻找一种新的距离度量,将原始的行人特征(如 HOG 特征^[8]和 SIFT 特征^[9])转换到一个新的度量空间中,在这个度量空间中,具有相同标签的样本更近,标签不同的样本具有较大的距离。基于深度学习的行人重识别方法中,这种距离度量学习算法被称为排序损失。其中最典型的排序损失是三元组损失 (Triplet Loss)^[10],给定一个目标样本,其中一个正样本与目标样本具有相同的标签,另一个负样本具有不同的标签,通过三元组损失调节使目标样本正距离小于目标样本负距离。除了三元组损失外,还有其他类型的度量学习损失,如柱状图损失 (Histogram Loss)^[11]和四元组损失 (Quadruplet Loss)^[12]。

本文提出了一种深层特征融合的行人重识别网络模型,从空间维度提升网络性能。首先,利用卷积层和池化层多次提取网络深层特征,使用融合后的深层特征作为行人图像的特征属性,融合后的

特征具有更好的细粒度特征表达能力。其次,为提高模型的泛化能力,在深层融合特征后加入一个批量归一化层,同时采用标签平滑损失 (Label Smoothing Loss) 函数和三元组损失函数对模型进行联合训练。与 2017 和 2018 年优秀的算法进行对比实验结果表明,本文所提方法在 Market1501、DukeMTMC-reID、CUHK03 和 MSMT17 4 个数据集上都取得了更高的行人重识别率。

2 基于深层特征融合的网络模型

2.1 网络结构

目前行人重识别在特征提取部分主要采用残差卷积网络 Resnet-50^[13]作为骨干网络 (Backbone),然后根据特征映射进行距离度量,计算图像相似度。考虑到从空间维度提升网络性能以及提高模型的泛化能力,本文采用 Se-resnet50^[14]作为骨干网络,利用卷积层和池化层多次提取网络深层特征并对融合特征进行批量归一化 BN (Batch Normalization) 处理。本文采用的行人重识别网络结构如图 1 所示,网络结构包括深层特征提取和融合特征批量归一化 2 个部分。

在深层特征提取部分,本文采用 Se-resnet50 网络提取图像的全局特征,其网络内部结构如图 2 所示,该网络是将 SENet (Squeeze-and-Excitation Networks) 嵌入 Resnet-50 中,SENet 分为压缩和激励 2 个部分。SENet 采用“特征重标定”方法进行特征通道间的融合,通过自己学习获取特征通道的权重并进行分配,提升有用特征通道的权重,同时削弱相关性小的特征通道的权重。在压缩部分,对 Resnet-50 输出的 C 个空间维度为 $H \times W$ 的特征映射 f_i ,通过全局池化压缩成 C 个空间维度为 1×1 的特征向量。在激励部分,通过如图 2 所示的瓶颈结构 (Bottleneck) 捕捉信道之间的内部关联,学习信道的注意力因子。最终将学习到的注意力因子分配到相应的信道上,输出 C 个大小为 $H \times W$ 的带有注意力标注的特征映射 f'_i 。

本文提出将 Se-resnet50 的 2 个中间层特征提取出来获得图像深层次的全局特征属性表示,提取的 2 个中间层特征维度分别为 512 和 1 024,对这 2 个不同尺度的中间层特征分别采用平均池化方法,获得 2 048 维的深层特征 f'_1, f'_2 ,最后将 2 个中间层特征与网络输出的 2 048 维的特征映射 f' 拼接为融合特征 $f'_{\text{fusion}} = [f', f'_1, f'_2]$,以提高模型对图像深层特征的描述能力。

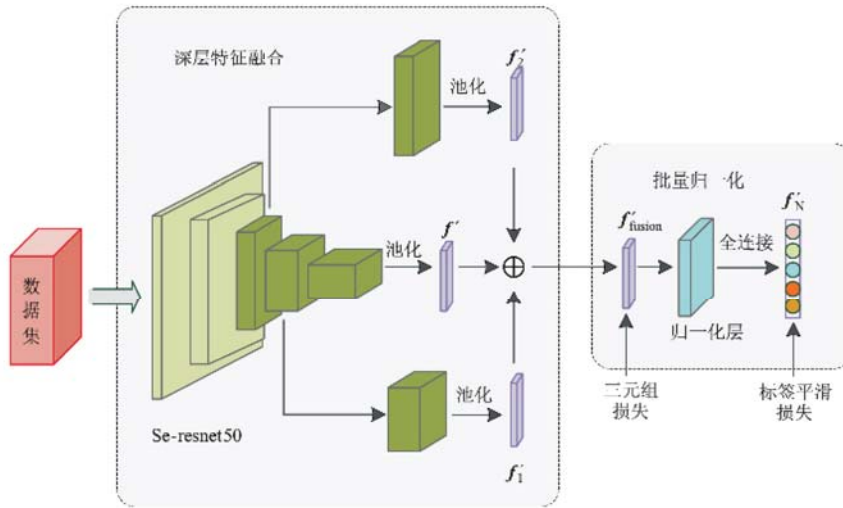


Figure 1 Structure of pedestrian re-identification network

图1 行人重识别网络结构

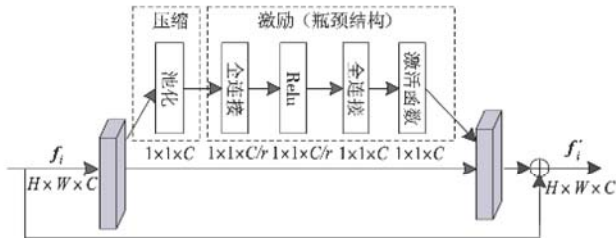


Figure 2 Structure of Se-resnet50

图2 Se-resnet50 结构

在融合特征批量归一化区域,为提高网络的泛化能力,在输出的 C 个融合特征(即批处理特征值)后面加入一个归一化层,首先计算批处理特征值的均值,然后计算其方差,对批处理特征值进行归一化处理,使其均值为 0,方差为 1。同时,在批量归一化层中引入了 2 个可学习参数 γ, β ,将 C 个融合特征的特征值作为样本对参数 γ, β 进行训练,对批处理特征值进行尺度变换和偏移。具体过程如图 3 所示,其中图 3a 为没经过任何处理的融合特征的 sigmoid 函数,深层融合特征分布在梯度平滑的区域时,网络在训练过程中将很难收敛。图 3b 对融合特征进行归一化处理后,将数据移到 sigmoid 函数中间区域,该区域具有最大梯度,能更有效地训练网络。为保证模型的非线性表达能力,图 3c 对归一化的深层融合特征进行尺度变换和偏移,使其更接近于真实分布。

2.2 损失函数

2.2.1 标签平滑损失函数

在行人重识别系统中,行人训练样本通常会存在少量的错误标签,这些错误标签可能会影响预测结果,标签平滑损失函数被用来防止模型在训练过

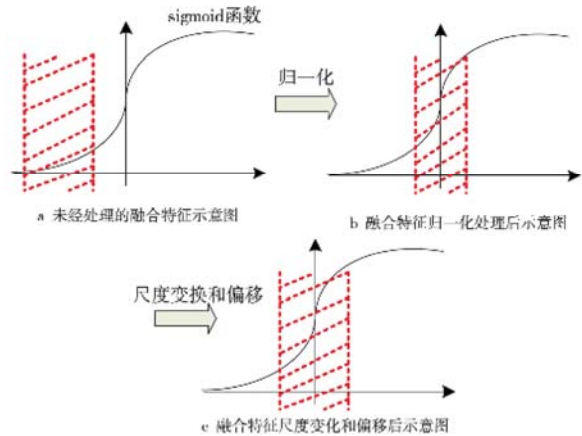


Figure 3 Batch normalization process of fusion feature

图3 融合特征批量归一化过程

程中过度依赖标签。

假设输入行人图像且其类别为 i 时, y_i 为图像中行人的标签,若 y_i 属于类别 i ,则其值为 1,否则为 0。本文对行人标签进行了平滑处理,在每次训练过程中,给标签设置一个错误率 ϵ ,则实际标注的行人标签 y'_i 如式(1)所示:

$$y'_i = \begin{cases} \epsilon, & y_i = 0 \\ 1 - \epsilon, & y_i = 1 \end{cases} \quad (1)$$

设 q_i 是网络预测实际标注的行人标签 y'_i 属于类别 i 的概率,如式(2)所示:

$$q_i = \begin{cases} \frac{\epsilon}{M}, & y_i = 0 \\ 1 - \frac{M-1}{M}\epsilon, & y_i = 1 \end{cases} \quad (2)$$

其中, M 为数据集集中的行人个数。由交叉熵损失函数可得标签平滑损失如式(3)所示:

$$L_{id} = \sum_{i=1}^N -q_i \log(p_i) \quad (3)$$

其中, p_i 为网络预测标签 y_i 属于类别 i 的概率。

2.2.2 三元组损失函数

本文通过三元组损失训练模型使得任意目标样本与正样本之间的距离最小,与负样本之间的距离最大。三元组损失函数如式(4)所示:

$$L_{\text{tri}} = [d_p - d_n + \alpha]_+ \quad (4)$$

其中, d_p 为目标样本与正样本之间的欧氏距离, d_n 为目标样本与负样本之间的欧氏距离, α 为三元组损失的阈值,本文中阈值设为 0.3。

2.2.3 联合损失函数

由于本文采用标签平滑损失^[15]和三元组损失^[12]对模型进行联合训练,然而这 2 种损失计算的特征映射是不一致的。标签平滑损失需要构造多个超平面,将每个类的特征分布在不同的子空间中。在使用标签平滑损失函数优化模型时,余弦距离更适用于标签平滑损失函数。但是,三元组损失函数是对欧氏距离中的类内紧密性和类间可分离性进行了增强。如果使用这 2 个损失函数同时对融合特征映射进行优化,会存在目标不一致的问题,甚至会增加训练难度。

为联合标签平滑损失和三元组损失,本文将批量归一化层加入到融合特征映射 f'_{fusion} 和全连接层之间,融合特征映射 f'_{fusion} 进入批量归一化层后输出标准化融合特征映射 f'_N 。由于批量归一化层将所有特征属性全部分配给了同一个标签,故采用 f'_N 计算标签平滑损失,用 f'_{fusion} 计算三元组损失。最终,本文模型的损失函数为 $L_{\text{loss}} = L_{\text{id}} + L_{\text{tri}}$ 。

3 实验结果及分析

3.1 实验数据与评估标准

本文在 Market1501^[16]、CUHK03^[17]、DukeMTMC-reID^[18]和 MSMT17^[19]4 个数据集上对提出的行人重识别方法进行对比实验,各数据集的介绍如表 1 所示。

Table 1 Datasets introduction

表 1 数据集介绍

数据集	时间	行人数	图像数	相机数
CUHK03	2014	1 467	13 164	10
Market1501	2015	1 501	32 668	6
DukeMTMC-reID	2017	1 812	36 441	6
MSMT17	2018	4 101	126 441	15

Market1501 数据集^[16]采集于清华大学,包含 32 668 幅行人图像,一共有 1 501 个行人标签。数

据集中包含训练集以及测试集,训练集中有 751 个行人标签,平均每人约有 17.5 幅图像;测试集中有 750 个行人标签。

CUHK03 数据集^[17]采集于香港中文大学,包含 13 164 幅行人图像,一共有 1 467 个行人标签。该数据集是 Matlab 数据文件格式,由检测、标注和测试 3 个部分组成。

DukeMTMC-reID 数据集^[18]采集于杜克大学,包含 36 441 幅行人图像,一共有 1 812 个行人标签。数据集中包含训练集以及测试集,训练集中有 702 个行人标签,平均每人约有 23.5 幅图像;测试集包含 17 661 幅图像。

MSMT17 数据集^[19]是目前最大且最接近真实场景的数据集,该数据集采集于校园内,通过 15 个摄像头采集 180 小时,最终得到 126 441 幅行人图像以及 4 101 个行人标签。数据集中包含训练集以及测试集,测试集中有 3 060 个行人标签,平均每人约有 30.6 幅图像;测试集中有 11 659 幅查询图像以及 82 161 幅行人图像。

实验中使用第 1 次命中匹配准确率(Rank-1)以及平均准确率 mAP(mean Average Precision)作为评价指标对模型进行评估。

3.2 参数设置与分析

3.2.1 参数设置

本文方法在 Pytorch 深度学习工具包上实施,实验平台是基于 64 位的 Ubuntu16.04 操作系统,显卡型号为 iGame GeForce RTX 2070,卷积神经网络选用的是 Se-resnet50,运用随机遮挡、裁剪和旋转对训练数据进行增强,输入行人图像大小为 224×224 。初始学习率设为 3.5×10^{-5} ,随着迭代次数逐渐增加到 3.5×10^{-4} ,然后分别在迭代到 40 和 70 次时降到 3.5×10^{-5} 和 3.5×10^{-6} ,迭代次数设为 120 次。训练时批次大小设为 32,优化器为 ADAM。采用标签平滑损失函数与三元组损失函数联合训练模型。

3.2.2 骨干网络的选择

为验证 Se-resnet50 的优越性,本文分别以 Resnet-50 和 Se-resnet50 为骨干网络在 Market1501 数据集上进行训练,训练过程中仅采用 softmax 损失对骨干网络进行优化,实验结果如表 2 所示。

由表 2 可知,Se-resnet50 网络相比于 Resnet50 网络,虽然在训练时间上增加了 16.11 min,但 Se-resnet50 在 Rank-1 上提升了 2%,在 mAP

Table 2 Experimental results based on Resnet-50 and Se-resnet50 backbone networks on Market1501 dataset

表 2 Market1501 数据集上基于 Resnet50 和 Se-resnet50 骨干网络的实验结果

Backbone	Rank-1	mAP	Time/min
Resnet-50	88.2	73.9	96.12
Se-resnet50	90.2	77.7	112.23

上提升了 3.8%，该结果表明本文采用的 Se-resnet50 网络可以更有效地提取行人特征。

3.2.3 深层特征融合模型

本文为提取行人深层特征属性，将 Se-resnet50 中间层特征分支提取出来先输入池化层然后融合，为验证深层特征属性的有效性，对只提取中间层特征 f'_1 与原始特征 f' 融合训练，提取 2 个中间层特征 f'_1, f'_2 与原始特征 f' 融合训练，提取 3 个中间层特征 f'_1, f'_2, f'_3 与原始特征 f' 融合训练以及原始特征单独训练进行实验，实验结果如表 3 所示。本次实验仅采用三元组损失函数训练模型，并未在网络结构中加批量归一化层。

Table 3 Experimental results of extraction of deep feature fusion on Market1501 dataset

表 3 Market1501 数据集上提取深层特征融合的实验结果

Se-resnet50	Rank-1	mAP	Time/min
f'	90.2	77.7	112.23
$f' + f'_1$	91.1	79.4	159.13
$f' + f'_1 + f'_2$	92.4	82.6	172.38
$f' + f'_1 + f'_2 + f'_3$	92.3	82.4	202.52

由表 3 可知，在 Market1501 数据集上，提取中间层特征与原始特征融合训练的方法相较于原始特征单独训练方法均有提升，最高可在 Rank-1 和 mAP 指标上提升 2.2% 和 4.7%，表明了深层特征提取对模型优化的有效性。

由实验结果可知，提取 2 个中间层特征与原始特征融合训练的方法在 Rank-1 与 mAP 指标上，高于提取 1 个中间层特征与原始特征融合训练的方法 1.3% 和 3.2%，高于提取 3 个中间层特征与原始特征融合训练的方法 0.1% 和 0.2%。虽然提取 2 个中间层特征与原始特征融合训练的方法与提取 3 个中间层特征与原始特征融合训练的方法结果相差不大，但在提取 2 个中间层特征与原始特征融合训练的方法的模型训练时间更短。与提取 3 个中间层特征的方法相比，虽然在 Rank-1 与 mAP 指标上相差不大，但模型训练时间减少了 30.14 min。综合行人重识别率与模型训练时间分

析可知，提取 2 个中间层特征进行融合训练是最有效的。

为验证批量归一化层与联合损失函数的有效性，本文以 Se-resnet50 为骨干网络以及提取 2 个中间层特征与原始特征融合的方法为基础模型，对是否加入批量归一化层在 Market1501 数据集上进行实验，实验结果如表 4 所示。

Table 4 Experimental results with or without batch normalization layer on Market1501 dataset

表 4 Market1501 数据集上有无批量归一化层的实验结果

Methods	Rank-1	mAP	Time/min
$f' + f'_1 + f'_2$	92.4	82.6	172.38
$f' + f'_1 + f'_2 + \text{BN}$	95.0	85.6	173.56

由表 4 可知，加入批量归一化层训练的方法比未加批量归一化层训练的方法在 Market1501 数据集上有大幅度提升，在 Rank-1 和 mAP 指标上分别提升了 2.6% 和 3.0%。虽然在模型训练时间上与未加批量归一化层训练的时间相近，增加了 1.17 min，但有效地提高了行人重识别率。

3.3 相关方法比较

为验证所提行人重识别方法的有效性，分别在 Market1501、DukeMTMC-reID、CUHK03 和 MSMT17 4 个数据集上实验，实验中未采用 Re-ranking 对行人重识别率进行提升。通过与 2017 和 2018 年优秀的算法进行对比，在 4 个数据集上本文方法的结果均有显著提升。图 4 是部分实验的查询结果图，第 1 列为行人查询图像，后 10 列为查询结果，其中黑框为错误结果，无框为正确结果。

在 Market1501 数据集上加入深层特征融合以及批标准化处理后，实验结果如表 5 所示，对比方法的评估结果分别摘自文献[20-23]。在该数据集上本文方法均优于其他的方法，且在 mAP 指标上有大幅提升，达到了 85.6%，Rank-1 达到了 95.0%。

Table 5 Performance comparison with mainstream methods on Market1501 dataset

表 5 Market1501 数据集上与主流方法的性能比较

Methods	Publication	Rank-1	Rank-5	Rank-10	mAP
SVDNet ^[23]	ICCV17	82.3	92.3	95.2	62.1
HA-CNN ^[20]	CVPR18	91.2	-	-	75.7
PCB ^[21]	ECCV18	92.4	97.0	97.9	77.3
Manes ^[22]	ECCV18	93.1	-	-	81.1
Ours	-	95.0	98.5	99.0	85.6

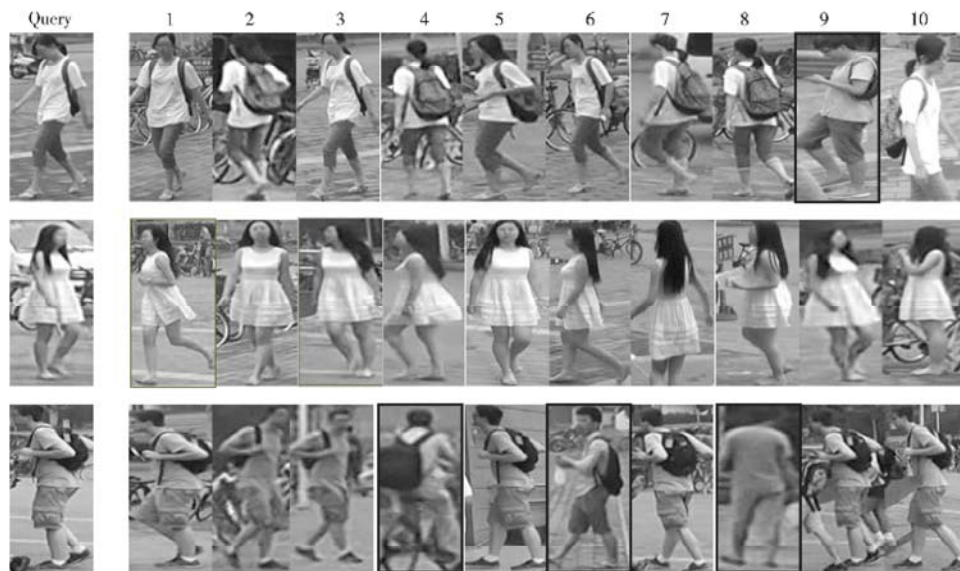


Figure 4 Query results of pedestrian re-identification

图 4 行人重识别的查询结果图

在 DukeMTMC-reID 数据集上加入深层特征融合以及批标准化处理后,实验结果如表 6 所示,对比方法的评估结果分别摘自文献[20—23]。在该数据集上本文方法在 $Rank-1$ 以及 mAP 2 个评价指标上均取得了大幅提升,分别提升了 2.2% 和 5.3%,达到了 87.1% 和 77.1%。

Table 6 Performance comparison with mainstream methods on DukeMTMC-reID dataset
表 6 DukeMTMC-reID 数据集上与主流方法的性能比较

Methods	Publication	$Rank-1$	$Rank-5$	$Rank-10$	mAP
SVDNet ^[23]	ICCV17	76.7	86.4	89.9	56.8
HA-CNN ^[20]	CVPR18	80.5	—	—	63.8
PCB ^[21]	ECCV18	83.3	—	—	69.2
Manes ^[22]	ECCV18	84.9	—	—	71.8
Ours	—	87.1	94.7	96.8	77.1

在 CUHK03 数据集上加入深层特征融合以及批标准化处理后,实验结果如表 7 所示,对比方法的评估结果分别摘自文献[20,21,23]。在该数据集上本文方法均优于其他的方法,且在 mAP 指标上有大幅提升,达到了 61.1%, $Rank-1$ 达到了 62.0%。

Table 7 Performance comparison with mainstream methods on CUHK03 dataset
表 7 CUHK03 数据集上与主流方法的性能比较

Methods	Publication	$Rank-1$	$Rank-5$	$Rank-10$	mAP
SVDNet ^[23]	ICCV17	40.9	—	—	37.8
HA-CNN ^[20]	CVPR18	41.7	—	—	38.6
PCB ^[21]	ECCV18	61.3	—	—	54.2
Ours	—	62.0	81.9	89.3	61.1

针对 2018 年新提出的数据集,在 MSMT17 数据集上加入深层特征融合以及批标准化处理后,实验结果如表 8 所示,与其对比方法的实验均来自于文献[19]。本文方法对该数据集的行人重识别率有显著提升,在 $Rank-1$ 以及 mAP 2 个评价指标上均取得了大幅提升,分别提升了 13.5% 和 17.8%,达到了 74.9% 和 51.8%。

Table 8 Performance comparison with mainstream methods on MSMT17 dataset
表 8 MSMT17 数据集上与主流方法的性能比较

Methods	Publication	$Rank-1$	$Rank-5$	$Rank-10$	mAP
GoogleNet ^[19]	CVPR18	47.6	65.0	—	23.0
PDC ^[19]	CVPR18	58.0	73.6	—	29.7
GLAD ^[19]	CVPR18	61.4	76.8	—	34.0
Ours	—	74.9	86.5	90.2	51.8

4 结束语

本文提出了一种深层特征融合的行人重识别方法,从空间维度提升网络性能。该方法利用卷积层和池化层多次提取 Se-resnet50 网络深层行人特征,使用融合后的深层特征作为行人图像的特征属性。在深层融合特征后加入一个批量归一化层,同时采用标签平滑损失函数和三元组损失函数对模型进行联合训练,提高模型的泛化能力。实验结果表明,本文所提行人重识别方法在 Market1501、DukeMTMC-reID、CUHK03 和 MSMT17 4 个数据集上都取得了更高的行人重识别率。在未来工作中将进一步研究更为有效的特征提取方法,提高

行人重识别性能。

参考文献:

- [1] Song Wan-ru, Zhao Qing-qing, Chen Chang-hong, et al. Survey on pedestrian re-identification research[J]. CAAI Transactions on Intelligent Systems, 2017, 12(6): 770-780. (in Chinese)
- [2] Wei L H, Zhang S L, Yao H T, et al. GLAD: Global-local-alignment descriptor for pedestrian retrieval [C]//Proc of the 2017 ACM International Conference on Multimedia(MM), 2017: 420-428.
- [3] Su C, Li J, Zhang S, et al. Pose-driven deep convolutional model for person re-identification [C]//Proc of the 2017 IEEE International Conference on Computer Vision(ICCV), 2017: 3980-3989.
- [4] Liu X, Zhao H, Tian M, et al. Hydraplus-net: Attentive deep features for pedestrian analysis [C]//Proc of the 2017 IEEE International Conference on Computer Vision(ICCV), 2017: 350-359.
- [5] Zhao L M, Li X, Zhuang Y T, et al. Deeply-learned part-aligned representations for person re-identification [C]//Proc of the 2017 IEEE International Conference on Computer Vision(ICCV), 2017: 3239-3248.
- [6] Yao H, Zhang S, Hong R, et al. Deep representation learning with part loss for person re-identification [J]. IEEE Transactions on Image Processing, 2019, 28(6): 2860-2871.
- [7] Shen Yuan-yuan, Yan Yan, Wang Han-zi. Recent advances on supervised distance metric learning algorithms [J]. Acta Automatica Sinica, 2014, 40(12): 2673-2686. (in Chinese)
- [8] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]//Proc of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2005: 886-893.
- [9] Lowe D. Object recognition from local scale-invariant features [C]//Proc of the 7th IEEE International Conference on Computer Vision(ICCV), 1999, 1150-1157.
- [10] Zhang S Z, Zhang Q, Wei X, et al. Person re-identification with triplet focal loss [J]. IEEE Access, 2018, 6(1): 78092-78099.
- [11] Ustinova E, Lempitsky V. Learning deep embeddings with histogram loss [C]//Proc of the 2016 Annual Conference on Neural Information Processing Systems (NIPS), 2016: 4177-4185.
- [12] Chen W H, Chen X T, Zhang J G, et al. Beyond triplet loss: A deep quadruplet network for person re-identification [C]//Proc of the 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2017: 1320-1329.
- [13] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]//Proc of the 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016: 770-778.
- [14] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//Proc of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2018: 7132-7141.
- [15] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision [C]//Proc of the 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016: 2818-2826.
- [16] Zheng L, Shen L Y, Tian L, et al. Scalable person re-identification: A benchmark [C]//Proc of the 2015 IEEE International Conference on Computer Vision, 2015: 1116-1124.
- [17] Li W, Zhao R, Xiao T, et al. Deepreid: Deep filter pairing neural network for person re-identification [C]//Proc of the 2014 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2014: 152-159.
- [18] Zheng Z D, Zheng L, Yang Y. Unlabeled samples generated by gan improve the person re-identification baseline in vitro [C]//Proc of the 2017 IEEE International Conference on Computer Vision(ICCV), 2017: 3774-3782.
- [19] Wei L H, Zhang S L, Gao W, et al. Person transfer gan to bridge domain gap for person re-identification [C]//Proc of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2018: 79-88.
- [20] Li W, Zhu X T, Gong S G. Harmonious attention network for person re-identification [C]//Proc of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2018: 2285-2294.
- [21] Sun Y F, Zheng L, Yang Y, et al. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline) [C]//Proc of the 2018 European Conference on Computer Vision(ECCV), 2018: 501-518.
- [22] Wang C, Zhang Q, Huang C, et al. Mncs: A multi-task attentional network with curriculum sampling for person re-identification [C]//Proc of the 2018 European Conference on Computer Vision(ECCV), 2018: 384-400.
- [23] Sun Y F, Zheng L, Deng W J, et al. Svdnet for pedestrian retrieval [C]//Proc of the 2017 IEEE International Conference on Computer Vision(ICCV), 2017: 3820-3828.

附中文参考文献:

- [1] 宋婉茹, 赵晴晴, 陈昌红, 等. 行人重识别研究综述 [J]. 智能系统学报, 2017, 12(6): 770-780.
- [7] 沈媛媛, 严严, 王茜子. 有监督的距离度量学习算法研究进展 [J]. 自动化学报, 2014, 40(12): 2673-2686.

作者简介:



熊炜(1976-), 男, 湖北宜昌人, 博士, 副教授, 研究方向为数字图像处理和计算机视觉。E-mail: xw@mail.hbut.edu.cn

XIONG Wei, born in 1976, PhD, associate professor, his research interests include digital image processing, and computer vision.