

基于多尺度卷积特征融合的行人重识别

徐龙壮, 彭力*

江南大学物联网工程学院物联网应用技术教育部工程中心, 江苏 无锡 214122

摘要 针对现有的基于卷积神经网络的行人重识别方法对于遮挡和复杂背景引起的判别信息缺失问题, 提出了一种基于多尺度卷积特征融合的行人重识别算法。在训练阶段, 使用金字塔池化方法对卷积特征图进行分块和池化, 获得包含全局特征和多尺度局部特征的多个特征向量; 对每一个特征向量进行独立分类, 并在各个分类的最后内积层上归一化权重和特征, 以提升分类性能; 最后使用梯度下降法优化全部的分类损失。在识别阶段, 将池化后的多个特征向量融合成一个新向量, 使用新向量在库中进行相似性匹配。在 Market-1501、DukeMTMC-reID 数据库上对所提算法的有效性进行实验验证。结果表明, 本文模型提取的特征具有更好的识别效果, Rank-1 精度和平均准确率也优于大多数先进算法。

关键词 机器视觉; 卷积神经网络; 金字塔池化; 多尺度特征融合; 行人重识别; 权重与特征归一化

中图分类号 TP391.4

文献标识码 A

doi: 10.3788/LOP56.141504

Person Reidentification Based on Multiscale Convolutional Feature Fusion

Xu Longzhuang, Peng Li*

Engineering Research Center of Internet of Things Technology Applications of the Ministry of Education, School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China

Abstract Existing methods of person reidentification based on convolutional neural network lack discriminative information, due to occlusion and complex backgrounds. To solve these problems, a method based on multi-scale convolutional feature fusion is proposed herein. In the training phase, pyramid pooling is used to extract multiple eigenvectors containing global features and multi-scale local features for blocking and pooling of the convolutional feature map. Afterward, each feature vector is classified independently, and the weights and features on the last inner layer of each class are normalized to improve the classification performance. Finally, a gradient descent algorithm is applied to optimize the sum of losses for each classification. In the recognition phase, pooled multiple feature vectors are concatenated into a new vector for similarity matching. The efficiency of the proposed algorithm is verified on datasets Market-1501 and DukeMTMC-reID, in which the results indicate that features obtained by the proposed model are more discriminative and that the Rank-1 accuracy and average accuracy are both better than most state-of-the-art algorithms.

Key words machine vision; convolutional neural network; pyramid pooling; multi-scale feature fusion; person reidentification; weight and feature normalization

OCIS codes 150.0155; 100.4996; 100.4999

1 引言

行人重识别是一项跨摄像机检索行人的任务,

即给定一个感兴趣的行人进行查询, 其目标是从其他多个摄像机收集的数据图像库中检索出这个行人的所有图片。在此任务中, 行人在不同摄像机场景

收稿日期: 2019-01-04; 修回日期: 2019-02-03; 录用日期: 2019-02-26

基金项目: 国家自然科学基金(61873112)、教育部-中国移动科研基金(MCM20170204)、江苏省博士后科研资助计划(1601085C)

* E-mail: pengli@jiangnan.edu.cn

下的外观和所处环境会经历很多变化,比如人的姿势、环境光照和物体遮挡等,学习到更有稳健性的特征表示仍是行人重识别的一大挑战。行人重识别的精确度很大程度上取决于行人特征表示,对于行人的特征信息获取越全面,行人重识别的效果就越好。文献[1]中提出了一种基于卷积神经网络判别特征学习的模型,提升了对样本间距离关系的约束,可使网络得到判别性强的特征。文献[2]中采用增强聚合通道特征(ACF)算法检测行人,将直方图和纹理特征结合作为行人特征描述子,很好地提升了特征表示能力。近几年,由于深度学习学习到的特征表示相比于传统手工提取特征的方法具有更高的辨识能力,因此深度学习方法在行人重识别研究领域占据主导地位。文献[3]中提出了一个新的网络模型,同时学习特征和对应的相似性度量,以提升特征的鉴别能力。文献[4]中提出了一个多任务深度网络,使用排序任务和分类任务在不同的网络层进行优化,得到更具有识别力的特征表示。由于这些方法仅利用行人的全局特征,当检测目标存在关键部分信息缺失的情况时,这些特征并不能提供良好的辨别能力。

为解决这个问题,研究者提出了一些专注于学习有判别力的部分特征表示,以增强行人重识别的稳健性。文献[5]中利用提取人体骨骼关键点的方法(GLAD)把行人图片分为头部、上身和下身三个部分,之后将整张行人图片与三个局部图片一起输入到一个参数共享的卷积神经网络(CNN)中,最后提取的特征融合了全局和局部的特征。但这种方法首先需要预先训练好骨骼关键点模型,其次,分块区域的参数共享使得局部有效信息不能被充分挖掘。针对这个问题,文献[6]中提出了一个基于部分的卷积基准模型(PCB),这种方法在特征提取层后的池化层对行

人特征图进行水平切分,平均切分为6个小块,并分别送入不共享权重的全连接层进行预测。由于身体部分不对齐的问题会使得小分块预测不准确,作者又使用一个精确部分池化(RPP)网络来精炼每一分块的特征,但这个RPP网络属于后续处理操作,整个模型不能使用端到端的方式进行训练,而且单一固定的局部尺寸划分也不能充分地提取出有效的局部信息。

本文提出了一种多尺度卷积特征融合的方法,可解决全局特征判别信息不足的问题。采用金字塔池化方法将特征图切分为多个空间区域,然后对每个空间区域特征输出都独立地学习分类。在每一个分类器之前归一化特征向量和权重,以提升各个分类器的分类性能,在预测阶段将多个特征向量进行融合,以判别行人身份。在数据集 Market1501^[7]、DukeMTMC-reID^[8]上对所提算法的有效性进行实验验证,结果表明,本文算法可以有效地提升行人的重识别精度。

2 基于多尺度特征融合的行人重识别

2.1 网络结构

如图1所示,本文的网络结构主要由骨干网络层、金字塔池化层、L2归一化层和全连接(FC)层构成,图中 D 表示特征向量的维数,L2表示使用L2归一化范数。骨干网络采用Resnet50网络,对该网络进行了以下修改:1)移去Resnet50网络最后的平均池化层(GAP)和全连接层;2)卷积(Conv)层conv4_1的步长由2设置为1,结果骨干网络提取出的特征图尺寸变为输入图片尺寸的1/16。在金字塔池化层,把输入的特征图深度复制为3份,各自采用不同尺度的平均池化方法,输出得到6个包含不同特征信息的2048维特征向量。之后采用 1×1 卷积核把特征向量维数降低为512。再经由归一化层对特征向量和权重向量作归一化处理,输出的6个

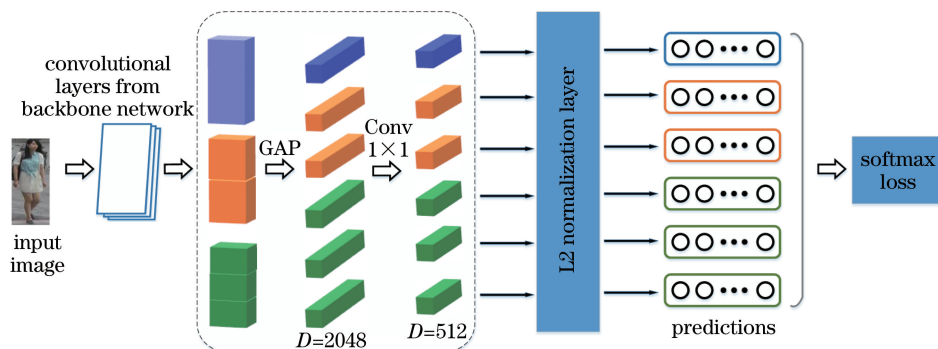


图1 网络结构

Fig. 1 Network structure

512 维向量分别送入不共享权重的全连接层,使用 softmax 损失函数和随机梯度下降方法进行优化。在识别阶段,查询图片使用串联 6 个 2048 维向量的新向量与图库中的新向量进行相似性匹配。

2.2 金字塔池化的多尺度特征

提取局部特征的常用思路主要有图像切块、注意力机制、利用骨架关键点定位以及姿态校正等。图像切块是一种不依赖辅助模型、简单而又有效的方法。本文使用空间金字塔池化对图片进行水平分割和池化,获得全局特征和多尺度局部特征的特征向量。全局特征可以对行人进行快速识别,同时考虑了整幅行人图像各部件之间的联系。局部特征可以很好地解决遮挡和视角变化大引起的关键信息缺失问题^[9],两者的有效融合可以增强特征的代表能力。

金字塔池化(SPP)是由 He 等^[10]2015 年提出的,主要是为了解决深度卷积网络中输入图片尺寸固定的问题。现有的 CNN 中对于结构已经确定的

网络需要输入一张固定大小的图片,对于希望检测各种大小图片的时候,往往要经过裁剪或缩放等操作,这样会降低识别检测的精度,于是作者提出空间金字塔池化方法使现有的网络结构可以适应任意大小尺寸的图片。具体操作为:把输入图片的特征图复制为 3 份,对每一份采用 1×1 、 2×2 、 4×4 的网格划分,再对每个网格中的特征块使用最大池化方法进行处理,得到 21 个固定的特征向量,然后把把这些向量平铺组成一个新的特征向量,并送入全连接层、softmax 层进行预测分类。本文在以上金字塔池化策略的基础上做出修改:1) 固定网络输入图片的尺寸,因而骨干网络提取出的特征图尺寸也固定,使用金字塔池化的目的是产生多个包含不同局部信息的特征向量;2) 使用平均池化并修改池化的内核和步长,将产生的 6 个单维固定通道数的特征向量送入不共享参数的全连接层,目的是更充分地提取各种特征的关键信息。

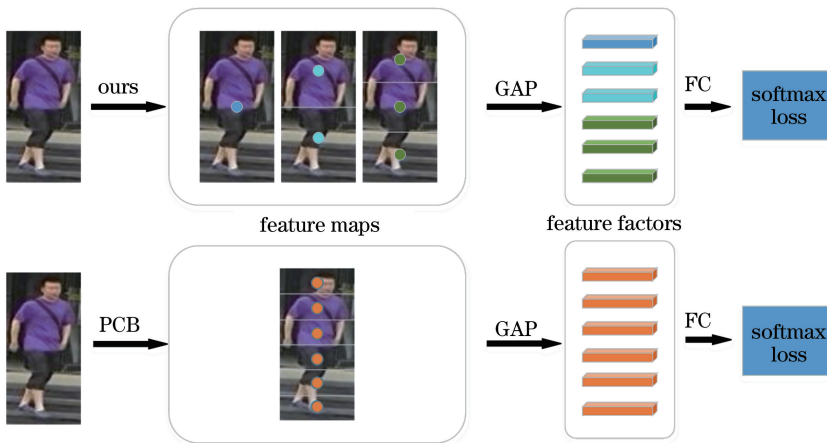


图 2 所提分块策略与 PCB 分块策略比较

Fig. 2 Comparison of our blocking strategy with PCB blocking strategy

在图片的分块策略上,PCB^[6]方法也是把图片分成 6 个固定尺寸的矩形块,然后进行独立分类,如图 2 所示。直观上看,PCB 方法只使用了从上到下 6 个局部部位信息来预测行人身份,由于图片之间行人存在错位、不对齐等状况,每个分块分得越小,异常值存在的状况越多,例如有的行人图片最上方的分块包含头部和背景,而另一行人图片最上方分块却只包含背景,这种情况的分类会降低总的识别精度。而本文使用金字塔池化的方法把特征图水平分成原特征图、二等分图和三等分图,这样既包含了全局特征和局部特征粗细结合的更为全面的信息,还使得每一分块上存在的异常值比 PCB 的分块要少。PCB 的高识别率主要来自后处理操作中 RPP

方法对于每一分块异常值的重新定位,但 RPP 不能使优化模型使用端到端的方式进行训练,而本文分块对于行人错位引起的异常值更具稳健性,故未使用 RPP 策略。

2.3 L2 归一化层

目前,深度卷积网络在分类问题中^[11]使用 softmax 函数,因为其计算简单,对于类间距离优化效果显著。但是,由于权重向量和特征向量的尺度不一致,在少数情况下仍会出现两类之间的距离小于某一类的类间距离,softmax 函数易发生误判。为了解决这个问题,使用 L2 归一化层对 softmax 的所有特征和权重向量做归一化处理,去除偏置向量,提升 softmax 的分类效果。传统的 softmax 损失函

数公式为

$$L_{\text{softmax}} = -\frac{1}{N} \sum_{i=1}^N \ln \frac{\exp(\mathbf{W}_{y_i}^T \mathbf{x}_i + \mathbf{b}_{y_i})}{\sum_{j=1}^C \exp(\mathbf{W}_j^T \mathbf{x}_i + \mathbf{b}_j)}, \quad (1)$$

式中: N 为训练样本的数目; C 为分类的数目; \mathbf{x}_i 为第 i 个样本的特征; y_i 为对应于 \mathbf{x}_i 的标签; \mathbf{W} 和 \mathbf{b} 为 softmax 损失之前的最后内积层的权重矩阵和偏置向量。 \mathbf{W}_j 是 \mathbf{W} 的第 j 行, 对应于第 j 类, 在测试阶段, 单个样本分类表达式为

$$\text{Class}(\mathbf{x}) = i = \arg \max_i (\mathbf{W}_i^T \mathbf{x} + \mathbf{b}_i). \quad (2)$$

从(2)式得到 $(\mathbf{W}_i^T \mathbf{x} + \mathbf{b}_i) - (\mathbf{W}_j^T \mathbf{x} + \mathbf{b}_j) \geq 0, \forall j \in [1, C]$, 测试样本的真值类表达式值比其他所有非真值类表达式值大。如图 3 所示, 假设有一个二分类问题, \mathbf{x} 属于类别 2, \mathbf{W}_1 和 \mathbf{W}_2 分别为类别 1 和类别 2 的权重向量, θ_1 与 θ_2 分别表示 \mathbf{W}_1 和 \mathbf{W}_2 与 \mathbf{x} 之间的角度, 若不考虑 \mathbf{b} 的影响, 则有 $\mathbf{W}_2^T \mathbf{x} > \mathbf{W}_1^T \mathbf{x}$,

即 $\|\mathbf{W}_2\| \|\mathbf{x}\| \cos \theta_2 > \|\mathbf{W}_1\| \|\mathbf{x}\| \cos \theta_1$, 在图 3(a) 中, 原始的 softmax loss 由于未对权重向量和特征向量进行处理, 故而导致 $\|\mathbf{W}_2\| \|\mathbf{x}\| \cos \theta_2 < \|\mathbf{W}_1\| \|\mathbf{x}\| \cos \theta_1$, \mathbf{x} 被误判为类别 1。在图 3(b) 中, 使用 L2 归一化层对 softmax 之前的内积层的参数做归一化处理, 即

$$\mathbf{W}_j^* = \frac{\mathbf{W}_j}{\|\mathbf{W}_j\|}, \mathbf{x}^* = \frac{\mathbf{x}}{\|\mathbf{x}\|}, \mathbf{b} = 0, \quad (3)$$

式中 $j \in [1, C]$, $\|\cdot\|$ 为向量的 L2 范数。此时 $\|\mathbf{W}_1^*\| = \|\mathbf{W}_2^*\| = 1, \|\mathbf{x}^*\| = 1$, 二分类的判别条件是比较余弦角度大小。在图 3 中可以看到, $\theta_2 < \theta_1$, 则有 $\cos \theta_2 > \cos \theta_1$, \mathbf{x} 被判别为类别 2。对权重向量和特征向量进行归一化处理, 可以减少损失函数优化参数的数量, 只考虑各类权重向量与特征向量之间的角度, 间接提升了 softmax 分类器的分类性能。

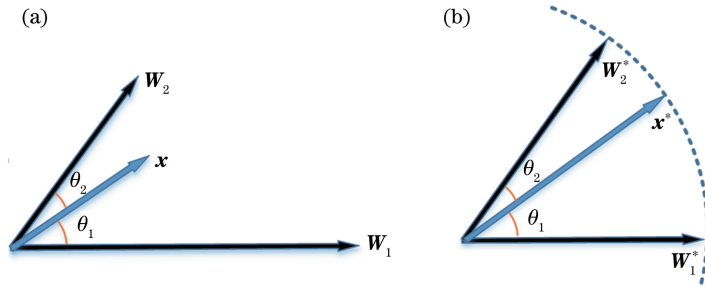


图 3 二分类的几何示意图。(a)原始的 softmax loss; (b)归一化权重和特征后的 softmax loss

Fig. 3 Geometric diagram of the second classification. (a) Original softmax loss; (b) softmax loss with normalized weights and features

3 实验分析

3.1 数据集和评估标准

为了验证本文所提算法的有效性, 选择常用的行人重识别数据集 Market-1501^[7] 和 DukeMTMC-reID^[8] 评估该方法, 并和几种最新的行人重识别算法进行比较。

Market-1501 包含从 6 个摄像机视图中收集的 1501 个行人的 32668 张标记图像。这些图像是由目标检测算法——应变部件模型 (DPM)^[12] 检测得出。数据集分为两部分: 包含用来训练的 751 个行人的 12936 张图像和用于测试的包含 750 个行人的 19732 张图像。在训练集中平均每个行人有 17.2 张图像。在测试中, 使用由 750 个行人组成的 3368 张样本用作查询对象, 使用单查询方式评估检索数据库中的匹配行人。

DukeMTMC-reID 是一个新发布的大规模行人

重识别数据集, 包括从 8 个摄像机采集的 36411 个标记图像, 共包含 1404 个不同行人。与 Market-1501 类似, DukeMTMC-reID 也把数据集平均分成两部分, 使用包含 702 个行人的 16522 张图片用于训练, 余下 702 个行人用于测试, 抽去 2228 张图片用作查询图像, 使用单查询方式在由 17661 张图片组成的数据库中匹配相似行人。对于这两个数据集, 都使用 rank-1 精度和平均精确率 mAP 这两个指标作为评估标准。

3.2 实验细节

选用残差网络 Resnet50 作为提取特征的骨干网络, 权重的初始化使用在数据集 ImageNet 上预训练好的参数, 以便减少训练时间。调整输入行人图片的尺寸为 384 pixel × 192 pixel, 并且对 Resnet50 网络的 conv4_1 层的步长设置为 1, 去除之后的平均池化层和全连接层, 输入图片经过骨干网络得到的特征图大小为 24 pixel × 12 pixel ×

2048 pixel。特征图经过金字塔池化之后输出 6 个 2048 维的特征向量。

在训练过程中,设定骨干网络的初始学习率为 0.01,其他网络层的学习率为 0.1,在 40 代之后学习率呈指数衰减,衰减系数为 0.01。模型总共训练 60 代,使用随机梯度下降(SGD)法对每一个 mini-batch 进行更新参数,mini-batch 的大小设置为 16,动量为 0.9。对于池化后的每一个特征向量独立进行分类,模型的总损失为每个分类损失之和。在验证时,把输入图片经过卷积池化后得到的多个特征向量串联起来,获得一个 12288 长度的新特征向量,查询图像使用新特征向量与数据库中的每张图片的新特征向量进行欧式距离计算,按照距离从大到小进行排列。

实验平台是基于 64 位的 Ubuntu16.04 操作系统和 NVIDIA GTX 1070 GPU,采用的深度学习框架为 Pytorch,python 版本为 3.6,训练时间约为 240 min。

3.3 实验结果分析

选用 CamStyle^[14] 和 PCB 中使用的基准模型作为本文算法评估的基准模型,该模型精度比 IDE^[15] 高很多,重新复现这个基准并对参数进行微调,所得结果与报道的结果稍有不同,在 Market-1501 数据集上 rank-1 精度为 86.88%,平均精确率为 69.59%。

为了测试本文算法的有效性,分别把多尺度金字塔池化层和 L2 归一化层与基准模型结合,观察所提算法对基准网络的提升效果,结果如表 1 所示。在 Market-1501 上,基准网络使用金字塔池化(SPP)方法,rank-1 精度和 mAP 分别提升了 2.5% 和 1.8%,使用 L2 归一化权重和特征,rank-1 精度和 mAP 分别提升了 1.8% 和 2.8%,在 DukeMTMC-reID 数据集上两种方案的 rank-1 精度和 mAP 提升更大,说明本文应用的两种方法对于行人重识别的检测效果都有较好的提升。将这两种方法相结合,最终在 Market-1501 数据集上取得 rank-1 精度为 91.36%,mAP 为 75.46%;若使用 Zhong 等^[16] 提出的重排列(Re-ranking)算法,则 rank-1 精度和 mAP 分别提升至 93.29% 和 88.81%。在 DukeMTMC-reID 数据集上,rank-1 精度为 82.05%,mAP 为 65.97%,使用 Re-ranking 算法后,rank-1 精度和 mAP 分别提升至 86.49% 和 81.93%。

为了验证本文分块策略的合理性,复现了未使用 RPP 方法 PCB 模型的结果,并且改变金字塔池化策略,由之前分割的 6 个水平分块变成 7 个和 9

表 1 两种方法对基准模型的提升效果
Table 1 Effects of two methods for improving baseline model

Model	Market-1501		DukeMTMC-reID	
	rank-1	mAP	rank-1	mAP
Baseline	86.88	69.59	72.62	54.93
Baseline+SPP	89.31	71.41	80.83	64.99
Baseline+L2_softmax	88.72	72.42	76.93	60.18
Ours	91.36	75.46	82.05	65.97
Ours+Re-ranking	93.29	88.81	86.49	81.93

个水平分块,分块细化,实验结果如表 2 所示。相比 PCB,本文模型的 rank-1 精度和 mAP 都比 PCB 的要好,因为虽然分割的小块数相同,但本文算法结合了全局特征和多尺度的局部特征的信息,而 PCB 只使用了等分的局部特征。其次,七分块与九分块模型的整体效果并没有比六分块模型的效果好,主要是因为随着分割块数增多,行人错位和姿态改变会使得小的分割块更容易受到异常值的干扰,对于总体识别效果没有助益。图 4 为其中几类算法的识别效果,R5 表示算法计算出的与探寻行人(query)匹配度最高的前 5 张图片,从图中可看出,本文六分块方案识别效果最佳。

表 2 不同分块模型性能比较

Table 2 Performance comparison of different block models

Model	Market-1501		DukeMTMC-reID	
	rank-1	mAP	rank-1	mAP
Baseline	86.88	69.59	72.62	54.93
PCB(out RPP)	88.48	69.89	80.21	64.27
Baseline+SPP (1+2+3)	89.31	71.41	80.83	64.99
Baseline+SPP (1+2+4)	89.41	70.68	79.35	62.55
Baseline+SPP (1+2+6)	88.09	68.75	80.05	64.89

3.4 与目前主流算法的比较

为了验证本文所提算法的优越性,将本文算法与目前主流的 7 种行人重识别算法进行对比,这几类算法分别是精准姿态嵌入(PSE)^[17]、全局-局部特征描述子(GLAD)^[5]、奇异向量分解(SVDNet)^[18]、多尺度特征(MultiScale)^[19]、多层特征融合(MLFN)^[20]、注意力机制(HA-CNN)^[21] 和深度金字塔特征学习(DPFL)^[19],都是基于卷积神经网络提取特征的算法,各类算法的 rank-1 精度和 mAP 如表 3 所示。由表中数据可以看出,本文算法在两个数据集上的 rank-1 精度和 mAP 都超过了对比的卷积算法,这表明本文算法的性能较其他算法有明显的提升。



图4 算法识别效果图。(a) Baseline;(b) PCB (out RPP);(c) baseline+SPP (1+2+6);(d) ours

Fig. 4 Results of recognition algorithm. (a) Baseline; (b) PCB (out RPP); (c) baseline+SPP (1+2+6); (d) ours

表3 本文算法与主流算法的性能比较

Table 3 Performance comparison of our algorithm with mainstream algorithms

Method	Market-1501		DukeMTMC-reID	
	rank-1	mAP	rank-1	mAP
PSE	87.7	69.0	79.8	62.0
GLAD	89.9	73.9	—	—
SVDNet	82.3	62.1	76.7	56.8
MultiScale	88.9	73.1	79.2	60.6
MLFN	90.0	74.3	81.0	62.8
HA-CNN	91.2	75.7	80.5	63.8
DPFL	88.9	72.6	79.2	60.6
Ours	91.4	75.5	82.1	66.0
Ours+Re-ranking	93.3	88.9	86.5	81.9

4 结 论

本文所提的多尺度卷积特征融合方法具有如下特征:1) 把使用金字塔池化后得到的全局特征和多尺度局部特征向量独立分类,以弥补全局特征判别信息不足的缺陷;2) 归一化每个分类器的权重与特征,使得损失函数只优化权重与特征之间的角度,提升了分类性能;3) 融合多个不同特征向量预测行人身份。结果表明,所提方法取得了很好的行人重识别效果。下一步的研究工作是获取更加精细化的局部特征与全局特征融合,提升行人重识别性能。

参 考 文 献

- [1] Chen B, Zha Y F, Li Y Q, *et al.* Person re-identification based on convolutional neural network discriminative feature learning [J]. *Acta Optica Sinica*, 2018, 38(7): 0720001.
陈兵, 查宇飞, 李运强, 等. 基于卷积神经网络判别特征学习的行人重识别[J]. *光学学报*, 2018, 38(7): 0720001.
- [2] Huang X Y, Xu J L, Guo G, *et al.* Real-time

pedestrian reidentification based on enhanced aggregated channel features [J]. *Laser & Optoelectronics Progress*, 2017, 54(9): 091001.

黄新宇, 许娇龙, 郭纲, 等. 基于增强聚合通道特征的实时行人重识别[J]. *激光与光电子学进展*, 2017, 54(9): 091001.

- [3] Ahmed E, Jones M, Marks T K. An improved deep learning architecture for person re-identification[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 3908-3916.
- [4] Chen W H, Chen X T, Zhang J G, *et al.* A multi-task deep network for person re-identification[C]// Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix. New York: AAAI Press, 2016: 3988-3994.
- [5] 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, October 23-27, 2017, Mountain View, California. New York: ACM, 2017: 420-428.
- [6] Sun Y F, Zheng L, Yang Y, *et al.* Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline) [M]// Ferrari V, Hebert M, Sminchisescu C, *et al.* Lecture notes in computer science. Cham: Springer, 2018, 11208: 501-518.
- [7] Zheng L, Shen L Y, Tian L, *et al.* Scalable person re-identification: a benchmark [C]// 2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 1116-1124.
- [8] Zheng Z D, Zheng L, Yang Y. Unlabeled samples generated by GAN improve the person re-

- identification baseline *in vitro* [C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 3774-3782.
- [9] Zhu X B, Che J. Personre-identification algorithm based on feature fusion and subspace learning [J]. Laser & Optoelectronics Progress, 2019, 56 (2): 021503.
朱小波, 车进. 基于特征融合与子空间学习的行人重识别算法[J]. 激光与光电子学进展, 2019, 56(2): 021503.
- [10] He K M, Zhang X Y, Ren S Q, *et al.* Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37 (9): 1904-1916.
- [11] Wu M, Zha Y F, Zhang Y Q, *et al.* Visual tracking algorithm based on classification-validation model[J]. Acta Optica Sinica, 2018, 38(5): 0515003.
吴敏, 查宇飞, 张园强, 等. 基于分类-验证模型的视觉跟踪算法研究[J]. 光学学报, 2018, 38(5): 0515003.
- [12] Felzenszwalb F, Girshick B, McAllester D, *et al.* Object detection with discriminatively trained part-based models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32 (9): 1627-1645.
- [13] Deng J, Dong W, Socher R, *et al.* ImageNet: a large-scale hierarchical image database [C] // 2009 IEEE Conference on Computer Vision and Pattern Recognition, June 20-25, 2009, Miami, FL, USA. New York: IEEE, 2009: 248-255.
- [14] Zhong Z, Zheng L, Zheng Z D, *et al.* Camera style adaptation for person re-identification [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 5157-5166.
- [15] Zheng Z D, Zheng L, Yang Y. A discriminatively learned CNN embedding for person reidentification [J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2018, 14 (1): 13.
- [16] Zhong Z, Zheng L, Cao D L, *et al.* Re-ranking person re-identification with k-reciprocal encoding[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 3652-3661.
- [17] Sarfraz M S, Schumann A, Eberle A, *et al.* A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 420-429.
- [18] Sun Y F, Zheng L, Deng W J, *et al.* SVDNet for pedestrian retrieval [C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 3820-3828.
- [19] Chen Y B, Zhu X T, Gong S G. Person re-identification by deep learning multi-scale representations [C] // 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 2590-2600.
- [20] Chang X B, Hospedales T M, Xiang T. Multi-level factorisation net for person re-identification [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 2109-2118.
- [21] Li W, Zhu X T, Gong S G. Harmonious attention network for person re-identification[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 2285-2294.