

多方向分区网络结构的行人再识别

唐一星 刘学亮 胡社教

合肥工业大学计算机与信息学院 合肥 230031 (elsa kf2016@163.com)

摘 要 将全局特征与局部特征相结合是提高行人再识别(re-identification)任务识别能力的重要解决方案。以往主要借助姿态估计等外部信息来定位有相应语义的区域,从而挖掘局部信息,这种方法大多是非端到端的,训练过程复杂且缺乏鲁棒性。针对该问题,文中提出了一种能有效挖掘局部信息并且能结合全局信息与局部信息进行端到端特征学习的方法,即多方向分区网络(Multi-orientation Partitioned Network, MOPN),该网络有3个分支,一个用于提取全局特征,两个用于提取局部特征。该算法不依靠外部信息,而是在不同的局部分支分别将图像按水平方向和竖直方向切分为若干横条纹和竖条纹,从而得到不同的局部特征表示。在 Market-1501、DukeMTMC-reID、CUHK03 和跨模态素描数据集 SketchRe-ID 上的综合实验表明,该算法的整体性能优于其他对比算法,具备有效性和鲁棒性。

关键词:行人再识别;深度学习;多分支网络;局部特征;全局特征

中图法分类号 TP391

Multi-orientation Partitioned Network for Person Re-identification

TANG Yi-xing, LIU Xue-liang and HU She-jiao

School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230031, China

Abstract Combining global features with local features is an important solution to improve discriminative performances in person re-identification (Re-ID) task. In the past, external information was used to locate regions with corresponding semantics, thus mining local information. Most of these methods are not end-to-end, so the training process is complex. To solve this problem, a multi-orientation partitioned network (MOPN) is proposed, which can effectively mine local information and combine global information with local information for end-to-end feature learning. The network has three branches; one for extracting global feature and two for mining local information. Without relying on external information, the algorithm divides pedestrians' images into horizontal and vertical stripes in different local branches respectively, so as to obtain different local feature representations. Plenty of experiments conducted on Market-1501, DukeMTMC-reID, CUHK03 and cross-modal dataset SketchRe-ID show that the proposed method has better overall performance than other comparison algorithms, and is effective and robust.

Keywords Person Re-identification, Deep learning, Multi-branch network, Local feature, Global feature

1 引言

行人再识别(Re-ID)被广泛认为是图像检索的一个子问题,旨在解决不重叠监控设备下行人图像的匹配问题,在罪犯定位识别、特定人物跟踪等方面具有重要作用[1]。由于监控视频图像的场景具有复杂性,对行人的身份识别的主要挑战来自行人的姿态变化(见图 1(a)、图 1(b))、物体遮挡(见图 1(c))、背景差异(见图 1(d)、图 1(e))、光照变化(见图 1(f)、图 1(g))、数据集中错误的行人检测(见图 1(h))等。随着深度学习的发展,卷积神经网络(CNN)也开始应用于行人再识别任务。目前,基于深度学习的方法已成为行人再识别任

务的常见方法。文献[2-3]最早将深度学习方法用于行人再识别任务。传统的手工设计提取特征的方法往往关注的是一些浅层特征,如颜色、形状等。与早期的手工设计提取视觉特征的方法^[4]相比,基于深度学习的行人再识别方法有着显著的优势,其对行人图像识别有更好的辨识度和鲁棒性。近年来,一些基于深度学习的行人再识别方法^[5-11]取得了突破性进展。

早期的基于深度学习模型的行人表示方法比较直观,其直接通过神经网络提取行人图像的全局特征,这种方法比较简单。如今行人数据愈发复杂,仅使用全局特征已无法满足现实应用。全局特征学习过程旨在捕获行人最明显的外观信息,如服饰颜色,以此区分不同行人。由于数据集规模有限、

到稿日期:2021-03-12 返修日期:2021-05-26

基金项目:科技部重点研发计划(2018AAA0102002);国家自然科学基金(61932009,61632007,61976076)

This work was supported by the National Key R&D Program of China(2018AAA0102002) and National Natural Science Foundation of China (61932009,61632007,61976076).

通信作者:刘学亮(Liuxueliang1982@gmail.com)

多样性不足等因素的制约,仅关注全局特征会漏掉一些不明显或出现次数较少的信息,还会造成空间信息的损失,也就是说无法充分学习到数据集中的行人特征,使得全局特征学习对学习有辨识度、区分度的特征的效果不好,因此仅有全局特征学习的行人再识别方法在类间差异小、类内差异大的数据集上表现不佳。为了缓解这一问题,人们做了很多研究,提取行人的局部特征逐渐被证明是一种有效的方法,在一定程度上解决了行人姿态变化、行人不对齐等问题。



图 1 行人再识别的挑战 Fig. 1 Challenges of Re-ID

近年来,一些基于局部特征的深度学习方法有着很好的 表现,其中一些方法[5-7]取得了接近人类甚至是优于人类的表 现。学习有显著区分能力的局部特征的重要前提是定位局部 区域。根据将身体部位分区的划分方式可以分为两组方法。 第一组是利用外部信息帮助的方法,如姿态估计的研究,需要 借助外部数据集和大计算量的姿态估计步骤。例如,PIE^[8] 算法将行人与标准姿态对齐,以此来减少姿态变化对行人再 识别性能的影响。GLAD[9]不是直接对齐行人,而是检测出 4 个人体姿态的关键点,然后通过构建的4个共享权重的卷积 子网分别提取全局特征和人体头部、上半身和下半身3个区 域的局部特征,直接利用4个分类损失函数来学习网络参数。 Spindle Net[10] 通过人体关键点提取网络提取出 14 个人体关 节点,产生7个身体局部区域,然后分阶段逐渐将身体相邻区 域融合,再与全局特征融合为一个256维的特征向量。另一 组不依靠来自外部语义信息的帮助,相比第一组,其不需要人 体部位的标注信息依然获得了很有竞争力的准确率。例如, PCB^[5]、MSCN^[6]和文献「11]中的方法是以水平分块为基础, 通过对网络进行合理精细的设计,达到了先进的水平。PCB 把行人图片均分为 6 块,并对每块独立做一个分类。PCB 网 络结构简洁,训练难度和经典的分类网络接近,在2018年其 是一种性能先进的方法,现在也是行人再识别任务中常见的 baseline。MSCN 是一种多尺度约束网络,它从多个尺度获取 详细信息,并独立训练每个尺度。

上述两组方法相比,基于姿态评估的方法可以定位身体 部位,但是将姿态评估模型迁移至行人再识别任务过程中存 在数据集适用偏差的问题,还会引入姿态估计关键点检测误 差,也会影响识别准确率。同时,基于姿态估计的方法大多不 是端到端的,这会增加训练难度。鉴于此,本文在不使用外部 信息的 PCB 算法的启发下,为解决局部区域划分和特征提取 的有效性问题,提出了一种多分支多方向分区的网络结构,该 网络有一个全局分支和两个局部分支,将全局特征和局部特 征相结合。以往的分区多为横向切分,而本文实验发现横向 切分与竖向切分相结合可以取得更好的效果,因此两个局部 分支分别将行人特征图等分为若干横条和竖条,横向切分与 竖向切分互为补充,能够有效挖掘局部信息,从而提高了行人 横向和竖向不对齐的容错率。该网络不依靠姿态估计等外部 信息的帮助,依然能提取出有显著区分性的局部特征,并与全 局特征结合进行端到端的训练。网络结构简洁,多分支协作, 使行人再识别算法的精度得到了提升。

2 行人再识别的相关工作

基于深度学习的行人再识别方法已成为主流。CNN 在ImageNet 分类任务中的优异表现表明分类网络能够提取出具有辨识度的图像特征。因此,在目标数据集上微调的分类网络被一些研究[12-13]作为 Re-ID 的特征提取器。例如,文献[12]提出了一种新的随机失活(dropout)方法,用多个数据集联合来训练分类模型。文献[13]结合手工设计的直方图特征和深度特征来对分类网络进行微调。分类网络可能会只关注行人身体的某一部位,如上身,因为行人上身往往包含了大部分的主要信息,所以忽视了其他部位,文献[14]对此进行了实验和说明。

除了分类网络,孪生网络(siamese network)和三元组网络(triplet network)等度量学习方法^[2-3,15]也是行人再识别任务中常用的方法。孪生网络以一对图像为输入,经过训练来判别两幅图像之间的相似程度。文献[2-3]最早将深度学习方法应用于行人再识别任务,它们都使用孪生网络,有着比当时手动设计提取特征的传统方法更优异的表现。文献[15]提出了一种通过三元组损失(triplet loss)训练来学习行人语义属性的网络。

近年来,一些工作从局部区域生成行人的深度特征表示,且都具有很好的表现。文献[11,16]根据 14 个人体关节点来分别提取行人不同身体区域的局部特征,然后融合所提取的身体局部区域特征。然而,这种提取深度局部区域特征的方法在一些情况下会受到限制,例如借助外部信息会引入姿态估计相关数据集和需要耗费很多计算量的估计器,以及两种数据集之间存在偏差。而且这些方法大多只聚焦于有特定语义的特定部位,不能涵盖所有需要判别的区域,同时这些方法大都不是端到端的学习过程,这会增加学习的难度。

因此,综合考虑上述问题,本文提出了一种端到端易于训练、横竖向切分互为补充共同协作,从而有效挖掘局部信息、结合全局信息与局部信息以及分类网络和度量学习的行人再识别模型。本文方法可以在不使用任何额外的外部信息的情况下,在几个主流的行人再识别数据集上达到较为理想的性能,并且在跨模态的素描数据集上较数月前提出的算法在精度方面有明显的提升,可帮助司法部门进行罪犯定位,因此在司法素描图像匹配及案件关键人物检索方向有着重要的应用价值,表明了本文模型的鲁棒性。

3 多方向分区网络结构

3.1 网络结构

多方向分区网络结构如图 2 所示,其网络的主干是 Res-Net50^[17]。ResNet50 是行人再识别任务的常用结构,在很多工作中^[5-7]都有着不错的表现。不同于 ResNet50 的是,本文方法在 res_conv4x 之后的部分接 3 个独立分支,即全局分支 (Global branch)、局部水平分支(Part_H branch)、局部竖直分支(Part_V branch),这 3 个独立的分支结构与原始的 Res-Net50 网络相似。

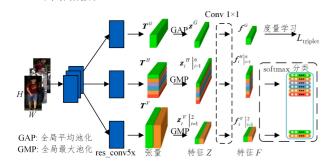


图 2 多方向分区网络结构

Fig. 2 Multi-orientation partitioned network architecture

表 1 列出了 3 个分支的结构设置。在全局分支上,行人图像数据经 res_conv1-5,得到特征图张量 T^c ,之后对 T^c 做 GAP(Global Average Pooling)操作,得到维度为 2048 的特征向量 z^c 。 z^c 经过正则化及 1×1 的卷积层进行降维,得到一个 256 维的特征向量 f^c ,这个分支学习的是全局特征表示而不需要任何分区的局部信息,因此将该分支记作全局分支。

表 1 本文网络结构的各分支设置

Table 1 Settings for each branch in MOPN

| 分支名 | 特征图尺寸 | 特征维度 | 特征名 |
|--------|---------------|------------------|----------------------|
| Global | 12×4 | 256 | f^G |
| Part_H | 24×8 | 256×6 | $f_i^H \mid_{i=1}^6$ |
| Part_V | 24×8 | $256\!\times\!2$ | $f_i^V \mid_{i=1}^2$ |

两个局部分支与全局分支有着相似的网络结构,输入数据经过 res_conv1-5,其中在 res_conv5_1 中没有进行下采样,这样可以保留更多的细节信息,有助于学习局部特征。然后,局部水平分支输出的特征图在水平方向上被均匀地分割成若干条,局部竖直分支输出的特征图在竖直方向上被均匀地分割为若干条,并对每一条做全局最大池化,得若干 2048 维特征向量 $\{z_i^H|_{i=1}^m, f_i^V|_{i=1}^m\}$,接着送入全连接层,最后进行softmax 分类。另外,通过实验发现,全局分支使用平均池化操作的表现优于最大池化,而两个局部分支使用最大池化操作的表现优于平均池化。出现该现象的原因可能是:全局分支使用平均池化可以更全面地学习到特征信息,而局部分支使用最大池化可以学习到更多的细节信息。因此,本文网络在全局分支使用平均池化操作,在两个局部分支使用最大池化操作。

在测试阶段,为了获得更强的识别能力,本文将降维后的特征连接起来作为最终的特征 $\{f^G,f_+^H|_{i=1}^m,f_-^V|_{i=1}^m\}$,将全局信息和局部信息结合起来,提高了学习特征的区分性。

3.2 损失函数

为了提高本文网络结构学习到的特征的辨别能力和区分能力,本文使用 softmax 损失用于分类学习,使用三元组损失 (triplet loss)用于度量学习。这两种损失函数在深度学习的行人再识别方法中应用广泛,许多已有研究[7.18]已经证明了在行人再识别任务中将上述两种损失函数联合训练的有效性。

对于基本的区分性的学习,本文将其看作多分类任务。对于第i个特征 f_i ,softmax 损失的计算式为:

$$L_{\text{softmax}} = -\sum_{i=1}^{N} \ln \frac{e^{\mathbf{W}_{y_i}^{\mathsf{T}} f_i + b_{y_i}}}{\sum_{C} e^{\mathbf{W}_{k}^{\mathsf{T}} f_i + b_{k}}}$$
 (1)

其中,N 为训练过程中 mini-batch 的大小,C 为训练数据集中类的数量,即行人 ID 数量。对局部特征 $\{f_i^H\mid_{i=1}^m, f_i^V\mid_{i=1}^n\}$ 使用 softmax 损失函数进行分类,其中 m 为横条纹数量,n 为竖条纹数量。

对降维后的全局特征 f° 进行三元组损失训练来提高排序性能。本文使用 batch-hard 三元组损失[19],这是在原有的 semi-hard 三元组损失的基础上进行改进的版本。该损失函数可表示为:

$$L_{\text{triplet}} = -\sum_{i=1}^{P} \sum_{a=1}^{K} \left[\alpha + \max_{p=1,\dots,K} \| f_a^{(i)} - f_p^{(i)} \|_2 - \min_{\substack{j=1,\dots,K\\j=1,\dots,P}} \| f_a^{(i)} - f_p^{(j)} \|_2 \right]_{+}$$
(2)

其中, $f_a^{(i)}$, $f_b^{(i)}$, $f_b^{(i)}$ 分别是 anchor、正(positive)样本、负(negative)样本提取的特征。正样本指和 anchor ID 相同的行人,负样本指和 anchor ID 不同的行人。候选三元组由最难的正、负样本对构建,即在每个 mini-batch 中选取 P个身份,每个身份选取 K 张图像,然后在其中选取最难正、负样本对。所谓最难正样本就是在这个 mini-batch 中与 anchor 距离最远的正样本,也就是与 anchor 是同一 ID 但差距很大且很不像的图像;最难负样本就是在这个 mini-batch 中与 anchor 距离最近的负样本,也就是与 anchor 不同标签但是差距很小且很像的图像。改进后的三元组损失增强了度量学习的鲁棒性,同时进一步提高了性能。

3.3 局部特征的水平和竖直分区

从图 3 可以比较直观地看出横向和竖向分区的作用。 图 3 中,绿、红线分别代表合理和不合理的切分方向。图3(a) 是同一行人的不同图像,可以明显看出这两张图中的行人不 对齐,这会给准确识别带来障碍。图 3(a)中左图的行人在图 偏下位置,而右图的行人主要处于图像上半部分,如果使用红 线进行横向切分,那么左图上端切出来的部分几乎都是背景, 而右图上端切出来的是行人的上半身。如果使用绿线进行竖 向切分,那么图 3(a)左右两张图切分后的结果包含的身体部 位大致相同,这样两图便更易匹配。同理,图 3(b)的左右两 张图也是相同行人的,可以看出绿线对应的横向切分可以让 两图的局部特征得到更好的匹配。行人再识别面临的挑战很 大一部分来源于行人不对齐,通过上述直观的描述可以看出, 竖向切分可以缓解行人竖向不对齐的问题,横向切分可以缓 解行人水平方向不对齐的问题。因此,对于不同的图像,其对 应的合理的切分方式不同,而本文模型通过多方向切分、横向 切分和纵向切分互为补充,共同学习到有区分性的行人局部 特征,有效提高了图片中行人不对齐的容错率,从而改善了模型的表现,增强了模型的鲁棒性。



图 3 多方向分区局部区域(电子版为彩色)
Fig. 3 Multi-orientation partitioned local regions

3.4 重要参数设置

本文的多方向分区网络结构得益于细粒度的空间集成。 下文列出了几个重要参数的设置,如输入图像的大小(即[H,W])、张量T的空间大小(即[M,N])、分割出的横条纹数量 (即 m)和竖条纹数量(即 n),这些参数的设定对于本文模型 的表现有很重要的影响。 $\lceil M, N \rceil$ 是由给定固定大小输入的 骨干模型的空间下采样率决定的。一些基于深度学习的目标 检测方法,如 SSD^[20],其结果表明降低下采样率可以有效地 丰富特征的粒度,文献[7]也在 ResNet50 中去除最后的空间 下采样操作来增加张量 T的大小,从而成功地提高了行人检索 的准确性,且只增加了非常少的计算成本。因此,在本文网络 结构中的两个局部分支均去除了 res conv5 1 中的下采样操 作,以便学习到有区分性的局部特征。文献[5]中的实验结果 出现了将张量T划分成太多条状从而损害学到特征的辨别能 力的现象,根据实验结果,本文设置 m=6。因此,本文参考文 献[5,7],将一些重要参数具体设置为:输入图像大小调整为 384×128,即图像高宽比为 3:1;全局分支的张量 T 的尺寸为 12×4:两个局部分支的张量 T 的尺寸为 24×8;局部水平分支 分割出的横条纹数量 m 设置为 6;根据本文的实验结果,局部竖 直分支分割出的竖条纹数量 n 设置为 2,具体内容见实验部分。

4 实验

4.1 数据集

本文在 4 个行人再识别数据集上对本文模型进行评估,分别为 3 个比较常用的真实场景中可见光相机采集的数据集 Market- $1501^{[21]}$ 、DukeMTMC-reID^[22]、CUHK03^[3] 和一个素 描行人再识别数据集 SketchRe-ID^[23],后面会对该数据集进行详细介绍。

Market-1501 包含由 6 个相机摄制的 1 501 张行人的图像。训练集有 751 个行人,图像数为 12 936;底库(gallery)有 750 个行人,图像数为 19 732;查询集图像数为 3 368。

DukeMTMC-reID 由 8 个相机摄制,包含 1 404 个人, 16522 张训练集图像,2228 张查询图像,17 661 张底库图像。它是对行人再识别任务具有挑战性的一个数据集。该数据集中常出现的情况是,不同身份的行人图片之间高度相似,相同身份的行人图片之间有很大的差异。

CUHK03 由 10 个相机摄制,包含 1 467 个行人的 14 097 张图片。

以上3个常用数据集都是单模态的,在现实生活的一些场景中,单模态数据集的应用会受到限制,于是引入了多模态

行人数据,如司法场景中行人再识别方法也可以帮助寻找案 件的关键人物,但很多时候无法获得相应人物的查询照片,只 能由专业画家根据目击者的描述绘制出相应的素描图像,因 此文献「23]创建了跨模态行人再识别数据集 SketchRe-ID。 它包含200个行人,每个人都有一张素描画家绘制的素描图 像和两张在白天由两台不同摄像机捕捉到的照片,照片是经 手动裁剪原始图像(或视频帧)得到的,以确保每张照片都包 含一个特定的行人, SketchRe-ID 的一些例子如图 4 所示。该 数据集的构建模拟了现实司法场景中的素描照片的匹配过 程,即选取一些志愿者扮演目击者,与专业的素描画家交流。 作为目击者,这些志愿者看了照片一段时间后向素描画家描 述照片中行人(在司法场景中,通常为嫌疑人或是其他与案件 相关的重要人物)的外貌,画家可以根据目击者的描述反复修 改直到完成素描图像。总共有5位画家绘制素描图。图5给 出了从每位画家绘制的素描图中随机选出的3张图,本文将其 作为例子,从中可以明显看出每位画家都有不同的绘画风格。



图 4 SketchRe-ID 数据集图像示例 Fig. 4 Examples of images in SketchRe-ID

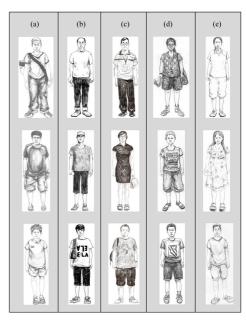


图 5 每种绘画风格示例 Fig. 5 Examples of each painting style

4.2 实验设置

为了能从行人图像中获取更详细的信息,本文参考文献 [5],将所有输入图片尺寸预处理为 384×128。本文使用在 ImageNet 上预训练的 ResNet-50 的权重来初始化网络主干。 在 res_conv4x 之后,网络中的不同分支都重复了相应层的预 训练权值。每个 mini-batch 是从训练集中随机抽取 P 个身 份,然后为这P个身份中的每个身份随机抽取K张图片,以 满足三元组损失的要求,这里设置 P=16 和 K=4 来训练本 文提出的模型(当实验数据集为 SketchRe-ID 时,K=3)。本 文选择 SGD 优化器,动量为 0.9。L2 正则化的权重衰减因子 设为 5×10⁻⁴。按 0.5 的概率翻转图像以扩增数据。至于学 习率的设定,本文先将学习率设置为 0.01,然后经过 40 个和 60个 epoch 的训练后将学习率分别衰减到10⁻³和10⁻⁴,整个 训练过程持续 100 个 epoch。本文实验使用 PyTorch 框架, 操作系统为 Ubuntu16.04,编程语言为 python3.6,CPU 型号 为 Intel(R) Xeon(R) CPU E5-2620,在 Market-1501 上一个 完整的训练过程用两个 NVIDIA GTX TITAN X(显存为 12 GB)并行大约需要 3h,本文在不同数据集上的实验都遵循 上述设置。

4.3 评价标准

评估 Re-ID 算法时,累计匹配特性曲线(CMC)以及平均准确率均值(mAP)很常用。

CMC 通常以 Rank-k 命中率的形式出现,表示在查询 (query)集中的行人图像与底库(gallery)中相似度最高的 k 张图像中命中相应行人的概率。

当候选集存在多个 ground truth 时,学者提出使用 mAP 进行评估。mAP 为取多个查询图像平均正确率的均值来代表算法的精确度。该指标体现出查询图像在候选集中所有可以匹配的图像排在检索结果前面的程度。相比 CMC,mAP 是一个更全面的评价指标,其计算公式为:

$$mAP = \frac{1}{m} \sum_{i=1}^{m} AP_{i} \tag{3}$$

其中,m 为查询集中查询图像的数量; AP_i 为第 i 张查询图片的平均正确率。

4.4 在数据集 Market-1501 上的实验结果

Market-1501 数据集上的实验结果如表 2 所列, MOPN w/o Part V表示 MOPN 不带竖直局部分支 Part V的实验, MOPN(n=4)代表将竖直局部分支 Part V 切分成 4 个条纹 的实验。从实验结果中可以看出,基于局部的方法普遍比基 于全局的方法表现好,而本文使用的网络 MOPN 通过合理的 局部区域的分区方式以及全局信息与局部信息的有效结合取 得了优于对比实验的结果。实验结果显示,不带竖直局部分 支的 MOPN 与 MOPN 相比, Rank-1 降低了 0.84%, mAP 降 低了1.17%,这说明局部竖直分支对模型有正向提升的作 用。同时,将局部竖直分支划分为 4 块的 MOPN 与划分为 2 块的 MOPN 相比, Rank-1 降低了 0.37%, mAP 降低了 0.24%,可以看出对于特征图的分区并不是越多越好,划分出 太多的条状会使得粒度太细,这反而会损害所学习到特征的 区分能力,而划分得太粗也不能充分提取到有显著区分性的 局部特征。本文通过实验和理论分析设置 m 为 6,n 为 2,在 该设置下,模型可以取得很好的表现。而且根据文献[5]的实 验结果可以看出,水平分区为6块时,模型表现最好,而本文 的输入图像尺寸经预处理后统一调整为384×128,则图像的 高宽比为 3:1,这也印证了本文将 MOPN 模型的竖直局部分 支分区数量设置为 2 的合理性。MOPN 与只用了水平方向 切分的 PCB 算法相比, Rank-1 提高了 2.28%, mAP 提高了 7.72%, Rank-1和 mAP 都有显著的提高对, 说明全局分支和 竖直局部分支可以提高行人的识别能力。MOPN 与比 PCB 性能更好的 PCB+RPP 相比, Rank-1 和 mAP 分别提升了 0.78% 和 3.52%,可以看出,mAP 的提升比 Rank -1 的提升 更明显,这证明了度量学习对排序的正向提升作用。度量学 习本质上就是拉近同类样本的距离,推开不同类样本的距离, 因此对于 mAP 的提升更明显, MOPN w/o Part V 相比 PCB, Rank-1 和 mAP 分别有 1.44% 和 6.55% 的提升, mAP 的提升明显大于Rank-1的提升,这能更直接地看出度量学习 对排序能力的提升作用。GP-reid 是行人再识别任务常见的 baseline, MOPN 与基于全局特征的 GP-reid 相比, Rank-1 提 升了 2.38%, mAP 提升了3.92%。综上可以看出, 本文模型 在 Market-1501 数据集上有明显的优势。

表 2 在 Market-1501 数据集上的实验结果 Table 2 Experiment results on Market-1501

(单位.%)

| | 1# ml | | CMC | | | |
|-----------------------|---|---------------|-------|---------|--------|--|
| 模型 | | Rank-1 Rank-5 | | Rank-10 | - mAP | |
| 本文 | MOPN(本文方法) | 94.58 | 97.61 | 98.34 | 85. 12 | |
| | MOPN w/o Part_V | 93.74 | 97.45 | 98.11 | 83.95 | |
| | MOPN(n=4) | 94.21 | 97.54 | 98.24 | 84.88 | |
| 基于局部 特征的深度 学习方法 | $PCB+RPP^{[5]}$ | 93.8 | 97.5 | 98.5 | 81.6 | |
| | $PCB^{[5]}$ | 92.3 | 97.2 | 98.2 | 77.4 | |
| | $\mathrm{GLAD}^{\llbracket 9 bracket}$ | 89.9 | _ | _ | 73.9 | |
| | MultiScale ^[24] | 88.9 | _ | _ | 73.1 | |
| TVVA | PartLoss ^[14] | 88.2 | _ | _ | 69.3 | |
| | MultiLoss ^[25] | 83.9 | _ | _ | 64.4 | |
| 基于全局 特征的深度 学习方法 | GP -reid $^{[26]}$ | 92.2 | 97.9 | _ | 81.2 | |
| | $DML^{[27]}$ | 87.7 | _ | _ | 68.8 | |
| | Triplet Loss ^[19] | 84.92 | 94.21 | _ | 69.14 | |
| | $PAN^{[28]}$ | 82.81 | _ | _ | 63.35 | |
| | $SVDNet^{[29]}$ | 82.3 | 92.3 | 95.2 | 62.1 | |
| | $SOMAnet^{[30]}$ | 73.87 | 88.03 | 92.22 | 47.89 | |

注:加粗字体为每列的最优值

4.5 在数据集 DukeMTMC-reID 上的实验结果

在数据集 DukeMTMC-reID 上的实验结果如表 3 所列,可以看出 MOPN 相比其他方法取得了最好的实验结果。

表 3 在 DukeMTMC-reID 数据集上的实验结果 Table 3 Experiment results on DukeMTMC-reID

(单位:%)

| 模型 | Rank-1 | mAP |
|-------------------------|--------|--------|
| MOPN(本文方法) | 88.04 | 76. 22 |
| MOPN w/o Part_V | 86.56 | 74.61 |
| MOPN(n=4) | 87.88 | 75.96 |
| GP-reid ^[26] | 85.2 | 72.8 |
| $\mathrm{Mancs}^{[31]}$ | 84.9 | 71.8 |
| $PCB+RPP^{[5]}$ | 83.3 | 69.2 |
| $PCB^{[5]}$ | 81.7 | 66.1 |
| $GSRW^{[32]}$ | 80.7 | 66.4 |
| $SVDNet^{[29]}$ | 76.7 | 56.8 |
| $PAN^{[28]}$ | 71.59 | 51.51 |
| | | |

注:加粗字体为每列的最优值

没有局部竖直分支的 MOPN 与 MOPN 相比,Rank-1 降低了 1.73%,mAP 降低了 1.61%,这表明竖直局部分支 part-

V 对于网络的表现有着明显的提升效果。另外,将局部竖直 分支划分为 4 块的 MOPN 与划分为 2 块的 MOPN 相比, Rank-1 降低了 0.16%, mAP 降低了 0.26%, 可以看出, 增加 竖条纹数并没有使模型表现更好,其表现反而略逊于2条竖 条纹的模型。MOPN 与只有横向划分成局部区域的 PCB 相 比, Rank-1 和 mAP 分别有了 6,34% 和 10,12% 的提升,可以 看出本文模型的优势明显,同时也说明全局分支和竖直局部 分支给网络带来了整体提升。MOPN w/o Part-V 相比 PCB, Rank-1 和 mAP 分别提升了 4.86%和 8.51%,可以看出度量 学习对模型表现的提升十分明显。另外,本文方法与比 PCB 性能更好的 PCB+RPP 相比, Rank-1 提高了 4.74%, mAP 提高了7.02%,可以看出,较 PCB 而言,本文方法在 Duke-MTMC-reID 上提升十分明显。DukeMTMC-reID 数据集规 模较大,且该数据集中经常出现不同身份的行人图片高度相 似和相同身份的行人图片之间差异很大的情况,因此 Duke-MTMC-reID 是对行人再识别任务很有挑战性的数据集, MOPN 在该数据集上的良好表现证明了本文方法的鲁棒性。

4.6 在数据集 CUHK03 上的实验结果

在数据集 CUHK03 上的实验结果如表 4 所列,可以看出 MOPN 较其他方法取得了最好的实验结果。

表 4 在 CUHK03 数据集上的实验结果 Table 4 Experiment results on CUHK03

(单位:%)

| 模型 | Rank-1 | mAP |
|---------------------------|--------|-------|
| MOPN(本文方法) | 67.71 | 66.85 |
| MOPN w/o Part_V | 66.00 | 65.97 |
| MOPN(n=4) | 67.42 | 66.49 |
| $GOG^{[33]}$ | 67.3 | _ |
| $MB-DML^{[34]}$ | 65.04 | _ |
| PersonNet ^[35] | 64.80 | _ |
| $PCB+RPP^{[5]}$ | 63.7 | 57.5 |
| $DNS^{[36]}$ | 62.55 | _ |
| Ensemble ^[37] | 62.10 | _ |
| GatedSCNN ^[38] | 61.80 | 51,25 |
| $PCB^{[5]}$ | 61.3 | 54.2 |
| SVDNet ^[29] | 41.5 | 37.3 |
| $PAN^{[28]}$ | 36.29 | 34.00 |

注:加粗字体为每列的最优值

没有局部竖直分支的 MOPN 与 MOPN 相比, Rank-1 降低了 1.71%, mAP 降低了 0.88%, 这说明局部竖直分支对网络有着正向提升的作用。将局部竖直分支划分为 4 块的 MOPN 与划分为 2 块的 MOPN 相比, Rank-1 降低了 0.29%, mAP 降低了 0.45%, 结果接近, 可以看出过多的分区并不能有效提高模型的表现, 反而会增加计算成本和模型复杂度。MOPN 与只有横向划分出局部区域的 PCB 相比, Rank-1 和 mAP 分别有着 6.41%和 12.65%的提高, 说明全局分支和竖直局部分支对模型有正向提升的作用。同时, MOPN 与性能优于 PCB的 PCB+RPP 相比, Rank-1 提高了 4.01%, mAP 提高了 9.35%, 优势十分明显。

4.7 在数据集 SketchRe-ID 上的实验结果

表 5 列出了不同绘画风格的素描图像的数量。鉴于素描图像绘画风格变换不被认为是素描图像行人再识别问题的一部分,因此从每种绘画风格中随机抽取 3/4 进行训练,将剩下

的 1/4 用于测试,以此来消除不同绘画风格对实验结果的 影响。

表 5 每种绘画风格素描图的数量

Table 5 Numbers of sketches of each painting style

| 绘画风格 | 数量 | 训练集 | 测试集 |
|------|-----|-----|-----|
| (a) | 46 | 35 | 11 |
| (b) | 20 | 15 | 5 |
| (c) | 79 | 59 | 20 |
| (d) | 33 | 25 | 8 |
| (e) | 22 | 16 | 6 |
| 总和 | 200 | 150 | 50 |
| | | | |

在 SketchRe-ID 数据集上的实验结果如表 6 所列。Triplet SN 是设计用于手绘物体草图识别的网络,先从行人照片提取边缘图片,然后将提取的行人边缘图片和行人素描图片送入 Triplet SN。GN Siamese 是有两个 GoogleNet^[39]分支的网络,该模型可学习相同语义空间的两个不同域。AFL Net提出了一种跨域深度对抗特征学习模型来联合学习行人身份特征和域不变特征。DIFL Net 的基础结构是一个三元组分类网络,通过融合深度神经网络的中级特征和高级特征来提升网络表现。

表 6 在 SketchRe-ID 数据集上的实验结果

Table 6 Experiment results on dataset SketchRe-ID

(单位:%)

| 模型 | CMC | | | |
|----------------------------|--------|--------|---------|---------|
| 侠型 | Rank-1 | Rank-5 | Rank-10 | Rank-20 |
| MOPN | 53.0 | 68.8 | 80.0 | 94. 2 |
| MOPN w/o Part_V | 50.2 | 66.5 | 77.0 | 92.4 |
| MOPN(n=4) | 52.6 | 68.0 | 80.4 | 94.0 |
| DIFL Net ^[40] | 49.0 | 70.4 | 80.2 | 92.0 |
| AFL Net ^[23] | 34.0 | 56.3 | 72.5 | 84.7 |
| GNSiamese ^[41] | 28.9 | 54.0 | 62.4 | 78.2 |
| Triplet SN ^[42] | 9.0 | 26.8 | 42.2 | 65.2 |

注:加粗字体为每列的最优值

从表 6 可以看出,本文方法与对抗特征学习网络 AFL Net 相比, Rank-1 提高了 19%, 另外 AFL Net 是先在 Market-1501 上预训练去除跨域对抗特征学习模块的模型部分, 然后在 SketchRe-ID 上微调,而本文方法是端到端的, 易于训练。本文方法与几个月前提出的域不变特征学习网络 DIFL Net 相比, Rank-1 提高了 4%。实验结果表明,本文模型在跨模态素描数据集 SketchRe-ID 上同样表现优异, 说明了本文模型可以学习到行人图片的域不变深层高级特征,在跨模态数据集上依然优异的表现也证明了本文方法的鲁棒性与泛化能力。

结束语 本文提出了一种新的多分支深度网络——多方向分区网络,用于在行人再识别任务中学习有区分性的特征表示。本文模型中的每个分支按照不同的切分方式学习全局或局部区域的特征表示。该方法直接学习水平分割和竖直分割的特征条纹上的局部特征,完全是端到端的学习,不引人区域建议和姿态估计等局部定位操作,不依靠外部信息。实验结果表明,本文方法在主流的行人再识别数据集 Market-1501,DukeMTMC-reID和 CUHK03 上均取得了优于对比方法的表现,在智能安防等领域有着广泛的应用前景;同时,本文方法在跨模态素描行人再识别数据集 SketchRe-ID 上也有着优异的表现,这在司法素描图像匹配及案件关键人物检索

方面有着重要的应用价值。在以上 4 个数据集上的表现证明 了本文模型的鲁棒性以及泛化能力。下一步将会就多方向分 区网络与注意力机制的结合以及进一步提高在跨模态数据集 上的正确率进行研究。

参考文献

- [1] LI H. TANG M. LIN J W., et al. Cross-modality Person Re-identification Framework Based on Improved Hard Triplet Loss
 [J]. Computer Science, 2020, 47(10):180-186.
- [2] YI D,LEI Z,LIAO S,et al. Deep metric learning for person reidentification [C] // 2014 22nd International Conference on Pattern Recognition. Piscataway: IEEE Press, 2014:34-39.
- [3] LI W,ZHAO R,XIAO T,et al. Deepreid: Deep filter pairing neural network for person re-identification [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2014:152-159.
- [4] ZHENG L, YANG Y, HAUPTMANN A G. Person re-identification: Past, present and future [J]. arXiv:1610.02984,2016.
- [5] 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. Berlin: Springer Press, 2018; 480-496.
- [6] LISS,LIUX L,ZHAO Y, et al. Person re-identification based on multi-scale constraint network [J]. Pattern Recognition Letters, 2020, 138, 403-409.
- [7] 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: ACM Press, 2018:274-282.
- [8] 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.
- [9] 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.
- [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 the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway; IEEE Press, 2017;1077-1085.
- [11] CHEN L Y, LI W J. Multishape part network architecture for person re-identification [J]. Journal of Image and Graphics, 2019,24(11):1932-1941.
- [12] XIAO T, LI H, OUYANG W, et al. Learning deep feature representations with domain guided dropout for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 1249-1258.
- [13] WU S,CHEN Y C,LI X,et al. An enhanced deep feature representation for person re-identification [C] // 2016 IEEE Winter

- Conference on Applications of Computer Vision. Piscataway: IEEE Press, 2016:1-8.
- [14] 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.
- [15] SU C, ZHANG S, XING J, et al. Deep attributes driven multicamera person re-identification [C] // European Conference on Computer Vision, Berlin; Springer Press, 2016:475-491.
- [16] SU C,LI J,ZHANG S,et al. Pose-driven deep convolutional model for person re-identification[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017; 3960-3969.
- [17] HE K,ZHANG X,REN S,et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016;770-778.
- [18] SONG G, LENG B, LIU Y, et al. Region-based quality estimation network for large-scale person re-identification [C] // Proceedings of Association for the Advancement of Artificial Intelligence, Menlo Park: AAAI, 2018: 7347-7354.
- [19] HERMANS A, BEYER L, LEIBE B. In defense of the triplet loss for person re-identification [J]. arXiv:1703.07737,2017.
- [20] LIU W.ANGUELOV D.ERHAN D.et al. Ssd:Single shot multibox detector[C] // European Conference on Computer Vision. Berlin:Springer Press, 2016; 21-37.
- [21] ZHENG L, SHEN L, TIAN L, et al. Scalable person re-identification: A benchmark [C] // Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2015:1116-1124.
- [22] ZHENG Z, ZHENG L, YANG Y. Unlabeled samples generated by gan improve the person re-identification baseline in vitro [C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017; 3754-3762.
- [23] PANG L, WANG Y, SONG Y Z, et al. Cross-domain adversarial feature learning for sketch re-identification [C] // Proceedings of the 26th ACM International Conference on Multimedia. New York; ACM, 2018; 609-617.
- [24] CHEN Y, ZHU X, GONG S. Person re-identification by deep learning multi-scale representations [C] // Proceedings of the IEEE International Conference on Computer Vision Workshops. Piscataway: IEEE Press, 2017; 2590-2600.
- [25] LI W.ZHU X.GONG S. Person re-identification by deep joint learning of multi-loss classification [C] // Proceedings of the 26th International Joint Conference on Artificial Intelligence. San Francisco: Morgan Kaufmann, 2017: 2194-2200.
- [26] SI J,ZHANG H,LI C G, et al. Dual attention matching network for context-aware feature sequence based person re-identification [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018; 5363-5372.
- [27] ZHANG Y,XIANG T,HOSPEDALES T M,et al. Deep mutual learning[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 4320-4328.

- [28] ZHENG Z,ZHENG L,YANG Y. Pedestrian alignment network for large-scale person re-identification [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 29 (10): 3037-3045.
- [29] SUN Y, ZHENG L, DENG W, et al. Sydnet for pedestrian retrieval [C] // Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017; 3800-3808.
- [30] BARBOSA I B, CRISTANI M, CAPUTO B, et al. Looking beyond appearances: Synthetic training data for deep cnns in reidentification[J]. Computer Vision and Image Understanding, 2018,167;50-62.
- [31] WANG C, ZHANG Q, HUANG C, et al. Mancs; A multi-task attentional network with curriculum sampling for person reidentification[C]//Proceedings of the European Conference on Computer Vision. Berlin; Springer Press, 2018; 365-381.
- [32] SHEN Y,LI H,XIAO T,et al. Deep group-shuffling random walk for person re-identification[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE Press, 2018: 2265-2274.
- [33] MATSUKAWA T, OKABE T, SUZUKI E, et al. Hierarchical gaussian descriptor for person re-identification [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016:1363-1372.
- [34] USTINOVA E,GANIN Y,LEMPITSKY V. Multi-region bilinear convolutional neural networks for person re-identification [C] // 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance. Piscataway: IEEE Press, 2017; 1-6.
- [35] WU L.SHEN C.HENGEL A. Personnet: Person re-identification with deep convolutional neural networks [J]. Computer Vision and Image Understanding, 2018, 167: 63-73.
- [36] ZHANG L,XIANG T,GONG S. Learning a discriminative null space for person re-identification [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016:1239-1248.

- [37] PAISITKRIANGKRAI S, SHEN C, VAN DEN HENGEL A. Learning to rank in person re-identification with metric en-sembles[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 1846-1855.
- [38] VARIOR R R, HALOI M, WANG G. Gated siamese convolutional neural network architecture for human re-identification [C] // European Conference on Computer Vision, Berlin; Springer Press, 2016; 791-808.
- [39] SZEGEDY C,LIU W,JIA Y,et al. Going deeper with convolutions[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 1-9
- [40] GUI S, ZHU Y, QIN X, et al. Learning Multi-level Domain Invariant Features for Sketch Re-identification[J]. Neurocomputing, 2020, 403, 294-303.
- [41] SANGKLOY P, BURNELL N, HAM C, et al. The sketchy database: learning to retrieve badly drawn bunnies [J]. ACM Transactions on Graphics, 2016, 35(4):1-12.
- [42] YU Q,LIU F,SONG Y Z,et al. Sketch me that shoe[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. Piscataway: IEEE Press, 2016:799-807.



TANG Yi-xing, born in 1995, postgraduate. Her main research interests include computer vision and so on.



LIU Xue-liang, born in 1981, Ph.D, professor, is a member of China Computer Federation. His main research interests include computer vision and multimedia information retrieval.